UNIVERSITAT DE BARCELONA

De la genómica a la modelización: estudios poblacionales a nivel individual en peces litorales del Mediterráneo Occidental

Héctor Torrado Mateo

DE LA GENÓMICA A LA MODELIZACIÓN: ESTUDIOS POBLACIONALES A NIVEL INDIVIDUAL EN PECES LITORALES DEL MEDITERRÁNEO OCCIDENTAL

Héctor
Torrado Mateo

Facultat de Biologia. Universitat de Barcelona.

Programa de Doctorat en Biodiversitat

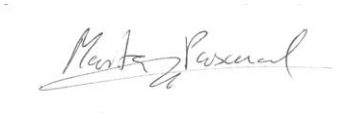
De la genómica a la modelización: estudios poblacionales a nivel individual en peces litorales del Mediterráneo Occidental

Memòria presentada per Héctor Torrado Mateo per optar al grau de doctor per la Universitat de Barcelona

Héctor Torrado Mateo

Departament d'Ecologia Marina, Centre d'Estudis Avançats de Blanes. Consejo Superior de Investigaciones Científicas (CEAB-CSIC)
Laboratori de Genètica Evolutiva, Departament de Genètica, Microbiologia i Estadística i IRBio. Universitat de Barcelona

Novembre 2020



Director de tesi:
Dr. Enrique Macpherson Mayol
Centre d'Estudis Avançats de Blanes

Directora/Tutora de tesi:
Dra. Marta Pascual Berniola
Universitat de Barcelona

Torrado, H (2020). De la genómica a la modelización: estudios poblacionales a nivel individual en peces litorales del Mediterráneo Occidental. Universitat de Barcelona.

De pell ferotgement dentada,
i esquelet indulgent.
D'escates d'argent
i esquelet sever.
De perfil lunar.
De cara de color curiositat.
D'espionatge i altres camuflatges.
D'eixams de sincronia coreogràfica.
De fluorescència caçadora.
De «fusiformitat»
amb la física de l'aigua.
I, sense cap dubte,
els simbionts
de la nostra idea d'oceà.

Peixos
Els nostres veïns submarins:
Bestiari contemporani de la Mediterrània
Laia Fontana i Bria

Agradecimientos

Quisiera empezar por agradecerte a ti, lector o lectora, el interesarte en el tema que ha sido mi vida durante los últimos cuatro años. Espero que, como yo, encuentres en esta tesis información que te ayude a comprender un poco mejor al mar y a sus habitantes y, sobre todo, que te lleve a hacerte más preguntas sobre ellos.

Larga sería la tarea de nombrar a todas las personas que han ayudado a que esta tesis llegue a realizarse, tanto a nivel profesional como personal, por lo que recurriré a ciertas generalizaciones que espero se me disculpen.

Debo empezar, indudablemente, por mis directores, Marta y Mac. Nada de esto hubiera sido posible sin ellos. Han conseguido, gracias a sus consejos y ayuda, que este pequeño proyecto de científico que acudió a ellos hace unos años haya dado pasos de gigante hacia ser un verdadero investigador. He aprendido muchísimo de ellos a todos los niveles y siempre han estado ahí cuando los he necesitado, no puedo más que dar gracias por haber tenido la oportunidad de haber hecho esta tesis con ellos.

Tampoco hubiera posible llevar a cabo esta tesis sin Núria, que además de venir a los muestreos, se ha encargado de la lectura de los otolitos que se han empleado en esta tesis y que constituyen la base de la mayoría de los artículos que la componen, además de enseñarme a mí a hacerlo. Muchas gracias por el enorme trabajo que has realizado.

También quisiera dar las gracias a Carlos, que siempre ha estado ahí para todo durante estos años, siendo prácticamente el tercer director de esta tesis.

Gracias también al resto de grupo del PopCOmics, las reuniones del proyecto siempre han estado llenas de ideas interesantes. Especialmente, gracias a Xavi, Iosune, Creu y Francesc, de los que también he aprendido mucho estos años. Tampoco puedo pasar sin agradecer a Teresa y a Rohan la ayuda siempre que la he necesitado y todo lo que me han enseñado.

Gracias a Paula y a Anna, mis “hermanas académicas” por todo lo que me han ayudado, enseñado y lo que hemos pasado juntos estos años. No podría haber compartido directores y doctorado con mejores compañeras y amigas.

Tampoco podría olvidarme de toda la gente que me ha acogido y enseñado en las estancias que he realizado estos años. Gracias a Rafaella Franch y Luca Bargelloni y al resto de la gente de la Università degli studi di Padova, por sus lecciones sobre el 2b-RAD. También a Baptiste Mourre y Joaquín Tintoré, y a todo el grupo del SOCIB, por haberme enseñado tanto sobre oceanografía. Finalmente, gracias a Eric Treml y al resto del Marine Spatial Ecology and Conservation Lab, que me enseñaron todo lo que sé sobre redes de conectividad.

Quisiera dar las gracias además a mis antiguos compañeros/as del laboratorio de biología marina de la Universidad de Valencia, y de Bioblau, con y de los que tanto aprendí. En especial a Romana y Ferran, que tanto me enseñaron en su día y que me han seguido ayudando y enseñando incluso después de irme. Sin ellos/as no hubiera conseguido llegar ni a empezar esta tesis.

Gracias también a mis compañeros y compañeras del CEAB y de la UB por haberme acompañado estos años. Especialmente a los primeros, con los que he compartido tantísimos buenos momentos a diario.

Por último, pero no menos importante, a mi familia, en especial a mis padres y hermano, y a mis amigos/as, tanto en Blanes como en Valencia o Calpe De las cosas que he de agradecerles a cada uno de ellos y ellas podría escribir mil páginas, por lo que me limitaré a decir muchas gracias por haber estado ahí siempre.

A tota la gent que he anomenat ací directa i indirectament, de nou, moltes gràcies per tot. Aquesta tesi és també vostra.

Abstract

Connectivity and local adaptation are two contrasting evolutionary forces influencing population structure. Due to their complex life cycle, marine fish species tend to be structured in metapopulations, connected almost exclusively by movements in the larval phase. In this thesis, we study the population genomic structure of several Mediterranean littoral fishes and the potential factors affecting their distribution. To asses this goal, we combine a set of different methodologies including population genomics, otolith reading, oceanographical dispersion models and graph theory.

We show a negative effect of temperature on pelagic larval duration but no effect on settlement size and thus raising temperatures would reduce dispersal capabilities of fishes. With an individual-based oceanographic dispersal model, we observed an effect of both hatching date and pelagic larval duration in the dispersal distances and orientations, but variable among species. Furthermore, we found a clear effect of the oceanographic fronts in dispersal capabilities of our species, allowing us to identify three hydrodynamic units in the Western Mediterranean delimited by these fronts.

We found different genomic structuring between sympatric species of *Symphodus* despite their similar early life traits. Nonetheless, we identified in both species candidate regions for local adaptation by combining outlier analysis with environmental and phenotypic association analyses. We provided tools and guidelines for laboratory and bioinformatics analyses to optimise studies using 2b-RAD sequencing on different non-model organisms with different genome sizes. Additionally, we found in three different localities clear trends of selective mortality for hatch date and lower for growth rate and pelagic larval duration in a common littoral fish. We confirmed these results with a phenotype-genotype association study, finding loci related with these traits, suggesting a genetic basis of differential mortality between settlers and survivors. Finally, we defined the main clusters and the main nodes of connectivity in three fish species in the Western Mediterranean. With this information, we evaluated the protection state of the areas with high importance for connectivity maintenance, finding a small proportion of them protected. All together, these results provide new valuable information about connectivity, population structure and adaptation in littoral fishes of the Mediterranean Sea.

Resumen

La conectividad y la adaptación local son dos fuerzas evolutivas contrastantes que influyen en la estructura poblacional. Debido a su complejo ciclo de vida, las especies de peces marinos tienden a estructurarse en metapoblaciones, conectadas casi exclusivamente por movimientos en la fase larvaria. En esta tesis estudiamos la estructura poblacional genómica de varios peces del litoral mediterráneo y los factores potenciales que afectan su distribución. Con este objetivo, combinamos un conjunto de diferentes metodologías que incluyen genómica de poblaciones, lectura de otolitos, modelos de dispersión oceanográfica y teoría de grafos.

Mostramos un efecto negativo de la temperatura sobre la duración de la fase pelágica larval, pero ningún efecto sobre el tamaño del asentamiento, por lo que el aumento de las temperaturas reduciría la capacidad de dispersión de los peces. Con un modelo de dispersión oceanográfica individual, observamos un efecto tanto de la fecha de eclosión como de la duración de la fase pelágica en las distancias y orientaciones de dispersión, pero variable entre especies. Además, encontramos un claro efecto de los frentes oceanográficos en las capacidades de dispersión de nuestras especies, lo que nos permite identificar la presencia de tres unidades hidrodinámicas delimitadas por estos frentes en el Mediterráneo Occidental.

Encontramos una estructura genómica diferente entre las especies simpátricas de *Symphodus*, a pesar de sus rasgos larvales similares. No obstante, en ambas especies identificamos regiones candidatas para la adaptación local mediante la combinación de análisis de valores atípicos con análisis de asociación ambientales y fenotípicos. Proporcionamos herramientas y pautas para análisis de laboratorio y bioinformáticos para optimizar los estudios utilizando la secuenciación 2b-RAD en diferentes organismos no modelo con diferentes tamaños de genoma. Además, encontramos en tres localidades diferentes tendencias claras de mortalidad selectiva para la fecha de nacimiento y, en menor medida, para la tasa de crecimiento y la duración de las larvas pelágicas en un pez litoral común. Confirmamos estos resultados con un estudio de asociación fenotipo-genotipo, encontrando loci relacionados con estos rasgos, lo que sugiere una base genética para esta mortalidad diferenciada entre colonos y supervivientes. Finalmente, definimos los principales grupos y los principales nodos de conectividad en tres especies de peces en el Mediterráneo Occidental. Con esta información evaluamos el estado de protección de las áreas de alta importancia para el mantenimiento de la conectividad presentes en la zona, encontrando una pequeña proporción de ellas protegidas. En conjunto, estos resultados proporcionan nueva y valiosa información sobre la conectividad, la estructura poblacional y la adaptación en los peces litorales del mar Mediterráneo.

Resum

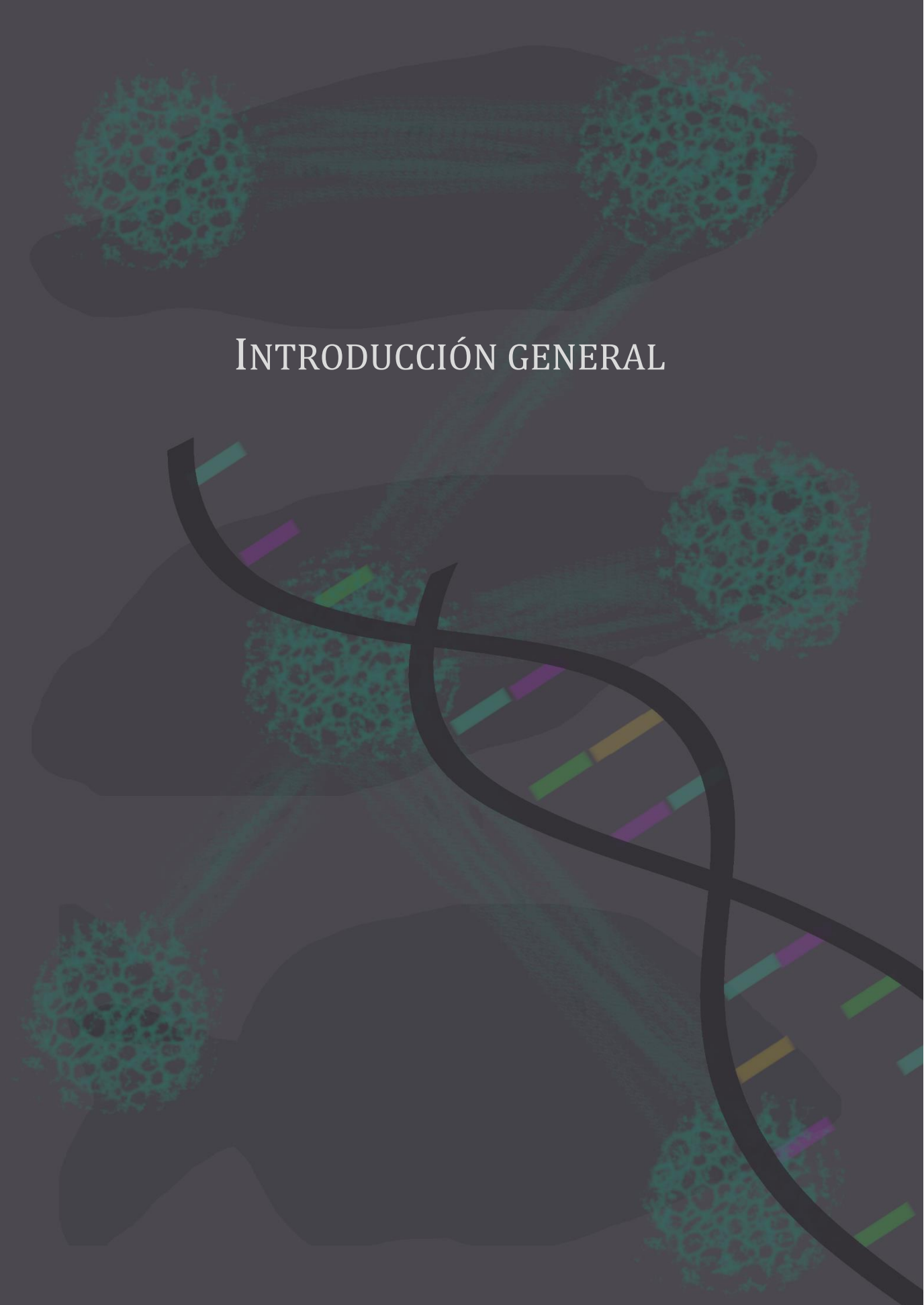
La connectivitat i l'adaptació local són dues forces evolutives contrastants que influeixen en l'estructura poblacional. A causa del seu cicle de vida complex, les espècies de peixos marins tendeixen a estructurar-se en metapoblacions, connectades quasi exclusivament pels moviments en la fase larvària. En aquesta tesi, estudiem l'estructura poblacional genòmica de diversos peixos litorals mediterranis i els possibles factors que afecten la seva distribució. Per avaluar aquest objectiu, combinem un conjunt de metodologies diferents, inclosa la genòmica de la població, la lectura d'otòlits, els models de dispersió oceanogràfica i la teoria de grafs.

Mostrem un efecte negatiu de la temperatura sobre la duració de la fase pelàgica larval, però cap efecte sobre la mida de l'assentament. Per tant, l'augment de les temperatures reduiria la capacitat de dispersió dels peixos. Amb un model de dispersió oceanogràfica individual, vam observar un efecte tant de la data d'eclosió com de la duració de la fase pelàgica larval en les distàncies i orientacions de dispersió, però variable entre les espècies. A més a més, vam trobat un clar efecte dels fronts oceanogràfics en la capacitat de dispersió de les nostres espècies, que ens permet identificar la presència de tres unitats hidrodinàmiques al Mediterrani occidental, delimitades per aquests fronts.

Trobem estructuracions genòmiques diferents entre espècies simpàtriques de *Symphodus,* malgrat dels seus trets larvaris similars. No obstant això, vam identificar en ambdues espècies regions candidates a l'adaptació local, combinant anàlisis de valors atípics amb anàlisis d'associacions fenotípiques i ambientals. Vam proporcionar eines i directrius per a anàlisis de laboratori i bioinformàtica per optimitzar estudis mitjançant la tècnica de seqüenciació 2b-RAD en diferents organismes no models amb diferents mides de genoma. A més, hem trobat en tres localitats diferents tendències clares de mortalitat selectiva relacionada en la data d'eclosió i, de forma més feble, relacionada en la taxa de creixement i la durada larvària pelàgica d'un peix litoral comú. Vam confirmar aquests resultats amb un estudi d'associació fenotip-genotip, trobant loci relacionats amb aquests trets, suggerint per tant una base genètica d'aquesta mortalitat diferenciada entre colons i supervivents. Finalment, hem definit els principals grups i els principals nodes de connectivitat de tres espècies de peixos al Mediterrani occidental. Amb aquesta informació, hem avaluat l'estat de protecció de les àrees amb gran importància per al manteniment de la connectivitat, trobant una xicoteta proporció protegida. En conjunt, aquests resultats proporcionen una nova i valuosa informació sobre la connectivitat, estructura poblacional i adaptació en peixos litorals del mar Mediterrani.

Índice

Introducción general



Introducción general

El mar Mediterráneo. Circulación en la cuenca Occidental.

El Mar Mediterráneo, a pesar de su reducido tamaño, es uno de los puntos calientes de biodiversidad marina de este planeta (Macpherson, 2002; Myers et al., 2000). Se trata de una cuenca semicerrada situada entre Eurasia y África. Únicamente se encuentra comunicado con el Océano Atlántico por el Estrecho de Gibraltar, el cual representa, junto al Nilo y los escasos ríos caudalosos de Europa, el principal aporte natural de agua a la cuenca. A través de este estrecho fluyen sendas corrientes de agua, una superficial desde el Atlántico y otra de aguas profundas desde el Mediterráneo hacia el océano.

El patrón de circulación y la topografía a lo largo de las costas sur y este de la Península Ibérica originan las tres discontinuidades oceanográficas principales del área (Figura 1): el Frente Almería-Orán, el Canal de Ibiza y el Frente Balear. El primero (Almería-Orán) es un frente de densidad a gran escala ubicado unos 400 km al este del Estrecho de Gibraltar y formado por la convergencia de la corriente superficial principal de agua entrante atlántica y las aguas que han recirculado en el Mar Mediterráneo (Tintore et al., 1988). La intensidad de dicho frente varía con los cambios del giro oriental del mar de Alborán (Renault et al., 2012). Al este de este frente, la corriente principal lleva el agua del Atlántico hacia levante, a lo largo de la costa norte africana. Parte de esta agua continúa hacia la cuenca oriental del Mediterráneo, atravesando el Canal de Sicilia, mientras que la otra parte fluye hacia el norte a través del Mar Tirreno, hasta el Mar Ligur y de retorno hasta la península Ibérica (Astraldi et al., 1999; Millot and Taupier-Letage, 2005).

La circulación en la cuenca occidental es principalmente ciclónica, con la corriente del norte fluyendo hacia el suroeste a lo largo de las costas francesa y española y bifurcándose al llegar al sur del mar Balear (Garcia Lafuente et al., 1995; Salat, 1996). Cerca de este punto se encuentra el Canal de Ibiza, de 80 km de ancho y 800 m de profundidad. Dicho canal corresponde al paso entre Ibiza y la Península

Ibérica, que cruza entre la cordillera balear y el cabo de La Nao. Se trata de un punto de estrangulamiento clave, con importantes intercambios entre aguas de reciente origen atlántico, por el lado sur, y aguas de características mediterráneas, más salinas, en el lado norte (Heslop et al., 2012; Pinot et al., 2002). Finalmente, la parte del flujo que se dirige hacia el noreste a lo largo del borde de la plataforma norte de las Baleares forma un frente de densidad bien definido, el Frente Balear, presente en los 200 m superiores (Figura 1).

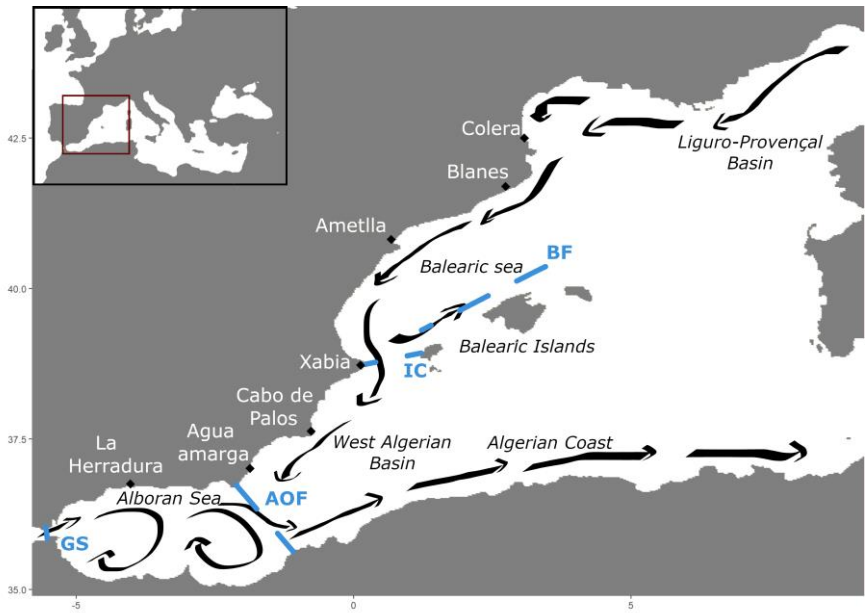


Figura 1: Patrones de circulación del Mediterráneo Occidental para el Agua Modificada Atlántica (MAW) y el Agua Intermedia Invernal (WIW). BF: Frente Balear, IC: Canal de Ibiza, AOF: Frente Almería-Orán. Editado de Millot, 1999.

El comportamiento dinámico de estas discontinuidades oceanográficas genera una variabilidad intra e interanual significativa que afecta los patrones de dispersión larval (Sabates et al., 2003; Sabatés and Olivar, 1996; Sponaugle and Cowen, 1996). Estos patrones de circulación, unidos a la amplitud latitudinal de la zona, crean un gradiente ambiental a lo largo de la costa de la península Ibérica entre el norte, más frío y productivo, y el sur, más cálido pero menos productivo. De este gradiente hay que excluir el mar de Alborán, que debido a la gran influencia Atlántica tiene

características de temperatura que se asemejan más al norte del Mediterráneo que a las aguas al otro lado del frente de Almería-Orán.

Los rasgos de la vida temprana y el uso de otolitos.

Parte de esta vulnerabilidad viene dada por la variabilidad en todo un conjunto de rasgos de esta fase temprana del desarrollo que pueden influir tanto en la supervivencia como en la capacidad dispersiva de las larvas. Los principales de estos rasgos son la duración de vida pelágica (PLD), el tamaño de asentamiento, la fecha de nacimiento y la tasa de crecimiento larval. Un crecimiento más rápido a menudo está relacionado con un tamaño más grande en el asentamiento y potencialmente un mejor éxito posterior al asentamiento (Gagliano et al., 2007; Grorud-Colvert and Sponaugle, 2011), aunque también con PLD más cortos (McLeod et al., 2015; O'Connor et al., 2007). Y es que la cantidad de tiempo que pasan las larvas en el medio pelágico (PLD) sigue siendo un indicador notablemente bueno de su capacidad de dispersión (Selkoe and Toonen, 2011), aunque no infalible ya que las diferentes especies varían también en otros parámetros importantes, como pueden ser la época de reproducción o el comportamiento de natación y las estrategias de asentamiento (Faillettaz et al., 2018b; Nanninga and Manica, 2018). En general, podemos decir que los PLD más cortos, debido al menor tiempo disponible para la dispersión, pueden dar lugar a poblaciones más aisladas (Pascual et al., 2017), haciéndolas vulnerables a fluctuaciones y condiciones de estrés puntuales y a las extinciones locales. Al modificar el PLD, las temperaturas cambiantes del océano podrían influir en la conectividad de la población a escala regional (Álvarez-Romero et al., 2018; Kleypas et al., 2016). Para obtener estos datos, una herramienta muy útil es la lectura de otolitos, pequeños huesos situados en el interior del sistema auditivo de los vertebrados. En el caso de los peces óseos, estos otolitos comienzan a formarse durante la vida larvaria. El desarrollo de los otolitos produce en ellos la formación de anillos de crecimiento diarios y con patrones diferenciados entre la fase larval y la juvenil (Figura 2). Gracias a ellos, es posible retrocalcular los rasgos de vida larvaria anteriormente mencionados, a partir de la

captura de juveniles y análisis de sus otolitos (Raventos and Macpherson, 2001; Wilson and McCormick, 1999), obteniendo así una visión completa de toda la fase larval.

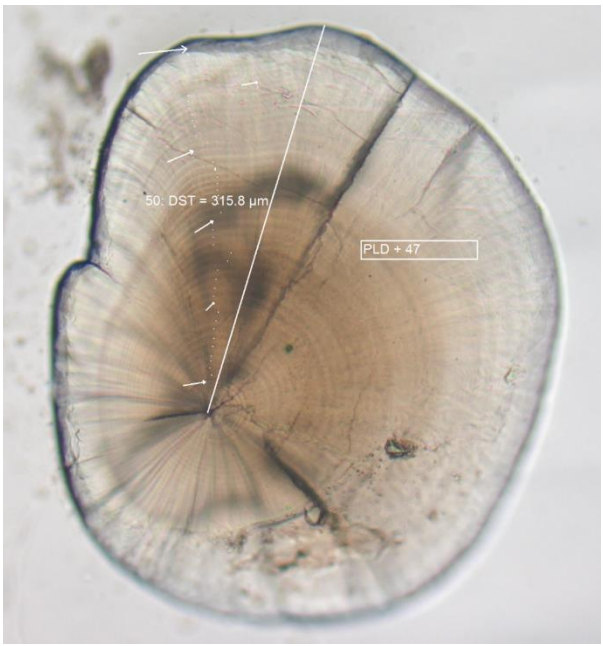


Figura 2: Otolito de *Symphodus tinca*, con marcas de incremento diario (anillos). La flecha central indica el anillo de asentamiento. El número de anillos antes del anillo de asentamiento permiten calcular la duración de vida larvaria (PLD). Los puntos y flechas blancos sirven para calcular la edad del individuo (en el rectángulo). La línea continua mide el tamaño del otolito en µm. Fotografía de: Núria Raventós.

Reclutamiento, conectividad y estructura poblacional en peces

La mayoría de peces óseos poseen un ciclo de vida complejo, que alternan dos fases diferenciadas a lo largo de la vida del animal. Lo más común en las especies bentónicas o demersales (y lo que ocurre en todas las estudiadas en esta tesis) es contar con una fase adulta de escasa movilidad que, al reproducirse (e independientemente de los cuidados parentales que pueda haber sobre los huevos), da lugar a una fase larvaria de vida pelágica, que es en la que se produce la mayor dispersión del individuo. Cuando la larva está suficientemente desarrollada, desciende hacia hábitats bentónicos (Figura 3).

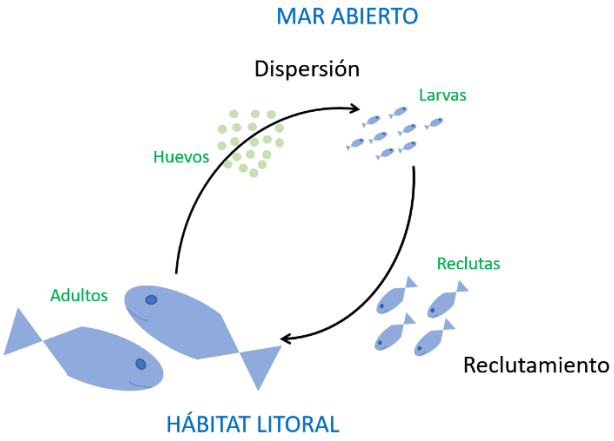


Figura 3: Ciclo vital de los peces bentónicos

Una vez dichas larvas llegan a las costas en hábitats favorables se produce su asentamiento y metamorfosis en juveniles. Dicho asentamiento es crucial para el mantenimiento de las poblaciones de la especie. Además, dada la escasa movilidad que suelen mostrar los adultos, la llegada de larvas de otras localidades es el mecanismo principal que mantiene la conectividad (entendiendo ésta no sólo como la llegada de individuos a la población, sino como la supervivencia de éstos hasta la fase reproductora). Sin embargo, también el autoreclutamiento de larvas en la misma zona de nacimiento es un hecho común en muchas especies (Almany et al., 2007; Carreras-Carbonell et al., 2007; Schunter et al., 2014; Swearer et al., 2002), lo que limita aún más el intercambio de individuos y resalta la importancia de aquellas larvas que sí se dispersan. Así, la estructuración poblacional está muy influenciada por el nivel de conectividad entre localidades, que viene condicionada por el potencial de dispersión larvario y las barreras al flujo génico, ambas con fuerte influencia de las condiciones oceanográficas a las que se ven sometidas en el pélagos, pero también determinada por la adaptación local. Como consecuencia de todo esto, estas especies suelen estructurarse en metapoblaciones con subpoblaciones de adultos distribuidas en parches de hábitats bentónicos, conectadas casi exclusivamente por los movimientos que ocurren en la fase larval (Cowen and Sponaugle, 2009).

Modelización oceanográfica.

La conectividad poblacional es difícilmente medible de manera empírica debido a la variabilidad de las escalas temporales y espaciales en las que ocurre y a los efectos de las diferentes características de las especies (Kool and Nichol, 2015). Por ello, se han empleado un considerable número de técnicas para cuantificarla, entre las que se incluyen el análisis químico de estructuras como los otolitos (Almany et al., 2007; Di Franco et al., 2012), análisis de parentesco genético (Berumen et al., 2012; Planes et al., 2009), de estructuración genómica (Carreras-Carbonell et al., 2007; Pascual et al., 2016), modelos biofísicos de dispersión (Andrello et al., 2013; Kough and Paris, 2015; Treml et al., 2012), o una combinación de ellas (Crandall et al., 2014; Schunter et al., 2011).

En lo que a los modelos de dispersión se refiere, el pequeño tamaño de estas larvas hace que el desplazamiento del que son capaces esté limitado, por lo que sus movimientos a gran escala están principalmente dirigidos por las corrientes oceanográficas predominantes en la zona. Así, para predecir el movimiento de las larvas pelágicas es necesario conocer sus movimientos junto al de estas corrientes. Para ello, es muy útil el uso de modelos de dispersión larvaria basados en simulaciones Lagrangianas (Andrello et al., 2017; Calò et al., 2018; Rossi et al., 2014; Schunter et al., 2011; Treml et al., 2012).

Dependiendo de los objetivos del estudio y de la información disponible, estos modelos se han abordado desde varios enfoques. En primer lugar, liberando partículas de áreas conocidas como fuentes potenciales y siguiendo las corrientes hasta sus áreas de asentamiento. Esta metodología se ha utilizado ampliamente, por ejemplo, para el estudio la dispersión de propágulos y el impacto potencial del cambio climático (Andrello et al., 2015) o la variabilidad espacial y temporal de la dispersión larvaria (Barbut et al., 2019; Di Franco et al., 2012; Ospina-Alvarez et al., 2015). Alternativamente, otro método consiste en que el modelo haga retroceder las partículas desde las áreas de asentamiento conocidas (ejecutando el modelo oceanográfico hacia atrás en el tiempo) para encontrar su origen potencial o áreas de

origen. Esto resulta especialmente útil cuando el estudio se centra en áreas de muestreo concretas, y se ha utilizado, por ejemplo, en el modelado de dispersión de larvas a nivel de población para evaluar el efecto de las estructuras oceanográficas (Holliday et al., 2012), para la identificación del área de eclosión potencial (Calò et al., 2018; Christensen et al., 2007; Fraker et al., 2015), para predicciones de reclutamiento en gestión pesquera (Allain et al., 2007) y para reconstruir la historia ambiental asociada a la tasa de crecimiento larvario (Payne et al., 2013; Ross et al., 2012). Se ha de tener en cuenta, además, que las capacidades natatorias de estas larvas aumentan con el desarrollo, y a partir de cierta edad son suficientes para nadar activamente (Faillettaz et al., 2018a; Staaterman et al., 2012). El desarrollo de esta mayor capacidad de natación, unido al desarrollo de capacidades sensoriales que les permiten detectar lugares de asentamiento propicios (Bottesch et al., 2016; Faillettaz et al., 2015; Leis et al., 2014; Mouritsen et al., 2013; O'Connor and Muheim, 2017; Paris et al., 2013) hace que las larvas puedan, en estados avanzados del desarrollo, influir de manera determinante en su desplazamiento debido a sus movimientos, por lo que añadir estimaciones de estos movimientos larvarios es crucial para obtener resultados fiables.

Por último, estas simulaciones oceanográficas pueden combinarse con parámetros de la vida larvaria y adulta de las especies para dar lugar a modelos biofísicos que simulen su dispersión. Estos modelos han demostrado ser muy útiles, ya que proporcionan una buena comprensión tanto de la dispersión como de la conectividad de las larvas y no requieren ni de un muestreo intenso o extenso ni de equipos especializados para tener éxito (Roberts et al., 2020). En concreto, este método permite producir matrices de conectividad que, al ser analizada como redes ecológicas, brindan una forma intuitiva de visualizar la estructura de la conectividad y sus propiedades emergentes, como la resiliencia de la red o la agrupación de los nodos dentro de esta (Cumming et al., 2010; Treml and Halpin, 2012; Urban et al., 2009). Además, también permiten identificar los nodos más importantes de la red de conectividad, como fuentes, sumideros y nodos muy centrales que funcionen como lugares importantes para la conexión entre muchas localidades (e.g. Andrello et al., 2013; Treml et al., 2012). Esta información es de gran importancia para una correcta

gestión espacial de las especies y, muy especialmente, para el diseño de redes de áreas marinas protegidas. Pues para asegurar su permanencia y utilidad a largo plazo, es necesaria una red bien conectada que las haga más resistentes a las perturbaciones estocásticas locales, en lugar de poblaciones aisladas autopersistentes.

La conectividad en el Mediterráneo Occidental ha sido estudiada en varias especies, tanto desde el punto de vista genético como oceanográfico (Andrello et al., 2013; Carreras et al., 2020; Faillettaz et al., 2018b; Galarza et al., 2009; Ospina-Alvarez et al., 2015; Pascual et al., 2017; Schunter et al., 2014, 2011), mostrando una amplia variabilidad que hace pensar en un importante efecto de la variabilidad en la biología de las especies sobre su dispersión y conectividad. Por ejemplo, las condiciones ambientales pueden imponer presiones selectivas diferentes sobre los rasgos fenotípicos debido a su variación espacial, por lo que su estudio combinado puede ayudar a conocer la capacidad adaptativa de la especie (Grummer et al., 2019). El uso de información ambiental junto con la estructura genética de las especies puede ser esencial para comprender cómo se adaptan a sus hábitats (Ahrens et al., 2018). Algunos estudios han investigado la relación entre la adaptación local y las condiciones ambientales en varias especies marinas siendo los valores de temperatura, salinidad y productividad factores importantes que afectan los diferentes procesos biológicos y la diferenciación genética a nivel poblacional (Benestan et al., 2016; Carreras et al., 2020; Hasan et al., 2017; Xuereb et al., 2018). Por ejemplo, los gradientes térmicos observados en el mar pueden explicar los patrones de distribución de numerosas especies, sus procesos de diversificación y la posible influencia de los cambios climáticos (Hansen et al., 2012).

Y es que estas fases tempranas del desarrollo son las más vulnerables a los cambios ambientales. En parte, esto es debido a que la vida en fase adulta de los peces bentónicos marinos es relativamente simple en comparación con la variabilidad ambiental a los que se ven sometidos y los complejos cambios de desarrollo que atraviesan los individuos en su fase planctónica pelágica (Leis, 2006). Este ciclo de vida dual hace que los procesos de la vida temprana sean especialmente fundamentales para la ecología de los peces. Aunque fugaz, la fase pelágica presenta una serie de

cuellos de botella críticos para las larvas que influyen en sus transiciones de desarrollo, la distancia que viajan, cuándo y dónde se establecen, y hasta en ciertos procesos de las primeras etapas posteriores al asentamiento (Cowen and Sponaugle, 2009; Kendall et al., 2016; Shima and Swearer, 2016).

Genómica de poblaciones y adaptación.

A pesar de la utilidad de los otolitos para obtener los fenotipos larvales, por sí solos son insuficientes para la comprensión de los procesos que están atravesando las especies, cómo se adaptan a ellos y cómo se estructuran sus poblaciones. Para estas preguntas hemos de recurrir a otras herramientas, como las genéticas. En las últimas décadas, los estudios de genética y genómica poblacional han sido fundamentales para identificar los patrones de estructuración genética y los procesos que los determinan en un gran número de especies (Carreras et al., 2020; Galarza et al., 2009; Manuzzi et al., 2019; Schweizer et al., 2016).

Las nuevas herramientas genómicas basadas en la secuenciación masiva de subconjuntos genómicos dirigidos por enzimas de restricción (RAD-seq) abren la posibilidad de trabajar con especies no modelo, pero ecológicamente relevantes, ya que la secuenciación de todo el genoma para estudios poblacionales todavía no es factible en estas especies. Dentro de estas técnicas, dos de las más empleadas son el genotipado por secuenciación (GBS) (Elshire et al., 2011) y el 2b-RAD (RAD-seq con enzimas de tipo 2b) (Wang et al., 2012). Ambos proporcionan un enfoque rentable para los estudios genómicos de poblaciones en especies no modelo, con una alta densidad de marcadores que pueden permitir identificar firmas genómicas no neutrales relacionadas con los procesos de adaptación, con potencial de anotación (Carreras et al., 2020; Stapley et al., 2010).

Una de las ventajas más interesantes del enfoque genómico en comparación con los marcadores tradicionales es la posibilidad de identificar marcadores genéticos relacionados con características ambientales o fenotipos específicos. Clásicamente, la evaluación de las señales genómicas de adaptación en organismos no modelo se realiza

de dos formas diferentes, análisis de valores atípicos ("Outlier Analysis", OA) y análisis de asociación genotipo-ambiente ("genotype-Environment Association Analysis", EAA) (Ahrens et al., 2018). Por un lado, los OA se basan en la búsqueda de marcadores genómicos que muestran valores elevados de diferenciación genética (FST) entre las poblaciones estudiadas (Carreras et al., 2017). Por otro lado, los EAA utilizan datos genómicos poblacionales combinados con datos ambientales, que se han utilizado con éxito para descubrir polimorfismos involucrados en la adaptación a las condiciones ambientales (Benestan et al., 2016; Carreras et al., 2020). No obstante, la información basada en individuos puede usarse para refinar aún más esas asociaciones. Ésta se emplea en los estudios de asociación del genoma ("Genome-Wide Association Studies", GWAS), que pueden identificar señales de adaptación tanto a las condiciones ambientales (EAA) como a los rasgos fenotípicos, a través del análisis de asociación fenotípica (PAA). Esta metodología ha demostrado ser una herramienta poderosa para estudiar características fenotípicas interesantes y la adaptación tanto en especies modelo como humanos, plantas y peces cautivos (Babbucci et al., 2016; Gonzalez-Pena et al., 2016; Zhong et al., 2017). Si bien el uso de información individual está extendido en las especies modelo, la aplicación en poblaciones silvestres es poco común y sólo se ha utilizado en unos pocos estudios combinados con información de datos ambientales (Schweizer et al., 2016). A día de hoy, hay una falta de estudios basados en datos individuales en poblaciones marinas naturales, lo que impide un conocimiento preciso de las relaciones entre los datos genómicos, el medio ambiente y los fenotipos de la vida temprana para determinar la estructura de la población en gradientes ambientales.

Algunas características larvarias, como la longitud de la duración de la vida larvaria (PLD), la tasa de crecimiento larvario, el tamaño de eclosión o el tamaño de asentamiento pueden afectar a las capacidades o a la mortalidad de éstos (Garrido et al., 2015; Hamilton et al., 2008; Johnson et al., 2015; Maldonado and Young, 1999; Raventos and Macpherson, 2005). No obstante, estos efectos son muy variables entre especies e incluso poblaciones (Searcy and Sponaugle, 2001), por lo que son necesarios estudios específicos de cada especie y que tengan en cuenta cierta variabilidad espacial

y ambiental. Además, algunos estudios han encontrado una base genética para estos rasgos bajo mortalidad selectiva (Lenfant and Planes, 2002; Planes and Romans, 2004; Vigliola et al., 2007). Esto hace que los rasgos encontrados bajo mortalidad selectiva sean aún más interesantes, ya que esta presencia de una base genética a su variabilidad los hace potenciales candidatos a selección. Por lo tanto, el estudio de rasgos larvarios individuales inferidos de otolitos, asociados a datos genómicos también individuales, podría proporcionar información muy valiosa sobre la adaptación y la evolución local de las especies.

Especies de estudio

En esta tesis, nos centramos en nueve especies de peces abundantes en los ecosistemas litorales del Mediterráneo occidental, incluyendo varias especies funcionalmente importantes (Prado et al., 2007; Sala et al., 2012). Estas especies abarcan un amplio espectro de la variabilidad que se puede encontrar en estos hábitats, incluyendo varios niveles tróficos (desde depredadores hasta herbívoros), diferentes comportamientos de reproducción y asentamiento. Todos ellos son principalmente sedentarios como adultos. A continuación se detallan algunas de sus características principales (Whitehead et al., 1984).

Sparidae

Diplodus puntazzo (Walbaum, 1792) (Figura 4)

Figura 4: Fotografía de *D. puntazzo*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico, con un tamaño adulto de hasta 60cm. Se alimenta tanto de algas como de invertebrados. De puesta pelágica, su reproducción tiene lugar alrededor de agosto-octubre. Se caracteriza por su boca en forma de pico y la presencia de varias líneas transversales y una mancha en la base de la aleta caudal oscuras. Tiene un PLD de entre 13 y 21 días. Se distribuye desde Cabo Verde hasta la bahía de Vizcaya, incluido el Mediterráneo.

Diplodus sargus (Linnaeus, 1758) (Figura 5)

Figura 5: Fotografía de *D. sargus*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico con un tamaño adulto de hasta 45cm. Se alimenta de algas y pequeños invertebrados como juvenil, como adulto principalmente de pequeños y medianos invertebrados. De puesta pelágica, su reproducción tiene lugar alrededor marzo-mayo. Se caracteriza por la presencia de varias líneas transversales y una mancha en la base de la aleta caudal y el rostro de color claro. Tiene un PLD de entre 14 y 32 días. Alcanza la madurez sexual a los 2 años. Se distribuye desde Cabo Verde hasta la bahía de Vizcaya, incluido el Mediterráneo

Diplodus vulgaris (Geoffroy Saint-Hilaire, 1817) (Figura 6)

Figura 6: Fotografía de *D. vulgaris*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico con un tamaño adulto de hasta 45cm. Se alimenta principalmente de pequeños invertebrados marinos. De puesta pelágica, reproducción tiene lugar alrededor octubre-enero. Se caracteriza por la presencia de dos líneas transversales oscuras, una cerca de la inserción de la aleta pectoral y otra en la base de la aleta caudal. Tiene un PLD de entre 22 y 43 días. Alcanza la madurez sexual en 2 años. Se distribuye desde Sudáfrica hasta la Bretaña, incluido el Mediterráneo.

Oblada melanura (Linnaeus, 1758) (Figura 7)

Figura 7: Fotografía de *O. melanura*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico con un tamaño adulto de hasta 30cm. Se alimenta principalmente de algas y pequeños invertebrados marinos. De puesta pelágica, su reproducción tiene lugar alrededor mayo-julio. Se caracteriza por la presencia de una

mancha oscura en la base de la aleta caudal. Tiene un PLD de entre 12 y 20 días. Se distribuye desde Angola hasta la bahía de Vizcaya, incluyendo al mar Mediterráneo.

Sarpa salpa (Linnaeus, 1758) (Figura 8)

Figura 8: Fotografía de *S. salpa*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico con un tamaño adulto de hasta 46cm. De juvenil mayoritariamente carnívoro (pequeños invertebrados), de adulto casi exclusivamente herbívoro. Muy gregario y de puesta pelágica, su reproducción tiene lugar alrededor septiembre-octubre. Se caracteriza por la presencia de unas líneas longitudinales amarillentas que le recorren todo el cuerpo. Tiene un PLD de entre 12 y 24 días. Se distribuye desde Sudáfrica hasta la Bahía de Vizcaya, incluyendo al mar Mediterráneo.

Pomacentridae

Chromis chromis (Linnaeus, 1758) (Figura 9)

Figura 9: Fotografía de *C. chromis*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico, con un tamaño adulto de hasta 15cm. Se alimenta de pequeños animales. De puesta pelágica, su reproducción tiene lugar alrededor junio-julio. Se caracteriza por el color marrón oscuro y una aleta caudal muy bifurcada. Tiene un PLD de entre 13 y 21 días. Se distribuye desde Angola hasta Portugal, incluyendo al mar Mediterráneo.

Labridae

Coris julis (Linnaeus, 1758) (Figura 10)

Figura 10: Fotografía de *C. julis*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico, con un tamaño adulto de hasta 25cm. Se alimenta principalmente de pequeños invertebrados marinos. De puesta pelágica, su reproducción tiene lugar alrededor mayo-julio. Se caracteriza por una coloración viva y con un marcado dimorfismo sexual, con hembras (y juveniles) marronáceas y machos de colores verdes y anaranjados. Tiene un PLD de entre 19 y 42 días. Alcanza la madurez sexual en 1 año. Se ha de destacar que sus juveniles son capaces de alargar su fase pelágica, retrasando su asentamiento unos días después de su metamorfosis, una flexibilidad de la que carecen el resto de especies de estudio. Se distribuye desde Noruega hasta Gabón, incluyendo al mar Mediterráneo.

Symphodus ocellatus (Linnaeus, 1758) (Figura 10)

Figura 11: Fotografía de *S. ocellatus*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico, con un tamaño adulto de hasta 12cm. Se alimenta principalmente de pequeños invertebrados. De puesta bentónica, construye de nidos y presenta cuidado parental de los huevos por parte del macho. Su reproducción tiene lugar alrededor junio-septiembre. Se caracteriza por una mancha ocelada en la parte superior del opérculo. Tiene un PLD de entre 7 y 14 días. Vive hasta 5 años, alcanzando la madurez sexual en 1-2 años. Su distribución está limitada al mar Mediterráneo.

Symphodus tinca. (Linnaeus, 1758) (Figura 11)

Figura 12: Fotografía de *S. tinca*. Imagen obtenida de Froese and Pauly, 2019.

Bentopelágico, con un tamaño adulto de hasta 35cm. Se alimenta principalmente de pequeños y medianos invertebrados marinos. De puesta bentónica, construye de nidos y presenta cuidado parental de los huevos por parte del macho. Su

reproducción tiene lugar alrededor mayo-junio. Se caracteriza por la presencia de bandas longitudinales en los laterales y una mancha en la base de la cola. Existe cierto dimorfismo sexual en el color, con hembras grisáceas/marronáceas y machos amarillento/verdosos con pequeños puntos de color azulado y rojizo. Tiene un PLD de entre 8 y 18 días. Puede vivir más de 7 años, alcanzando la madurez sexual en 2-3 años. Su distribución está prácticamente limitada al mar Mediterráneo, con menor presencia en el Atlántico Ibérico y Norteafricano.

Objetivos

El objetivo general de esta tesis es analizar la conectividad y los patrones de diferenciación poblacional a nivel genómico y mediante modelización en diferentes peces litorales del Mediterráneo occidental. Asimismo, estudiar la variabilidad de los rasgos biológicos sobre los procesos de dispersión y adaptación. Esto se concreta en los siguientes objetivos específicos:

1. Estudiar el efecto de la variabilidad ambiental sobre las características larvales durante el pre-asentamiento de los peces litorales del Mediterráneo Occidental.

2. Determinar el efecto de los frentes oceanográficos en la dispersión larval de nueve especies de peces litorales mediante simulaciones oceanográficas de dispersión de partículas Lagrangianas, utilizando datos larvarios individualizados.

3. Estudiar la estructuración genómica poblacional y la relación fenotípica con las condiciones ambientales en dos especies simpátricas y congenéricas: *Symphodus ocellatus* y *S. tinca*.

4. Evaluar y optimizar el protocolo de construcción de librerías genómicas 2b-RAD en estudios con especies no modelo.

5. Estudiar la presencia de mortalidad selectiva entre ejemplares recien asentados y juveniles supervivientes de *Diplodus puntazzo* y su posible base genética.

6. Evaluar la conectividad en tres especies (*S. ocellatus*, *S. tinca* y *D. puntazzo*) en el Mediterráneo occidental mediante modelos biofísicos validados por datos genómicos. Localizar las principales zonas fuente, sumidero, y nodos de alta conexión de las redes de conectividad de estas especies y evaluar su estado de protección.

Informe de los directores

Informe de los directores sobre la coautoría y factor de impacto de las publicaciones derivadas de la tesis presentada por Héctor Torrado Mateo titulada "De la genómica a la modelización: estudios poblacionales a nivel individual en peces litorales del Mediterráneo Occidental".

Héctor Torrado, Carlos Carreras, Núria Raventós, Enrique Macpherson, Marta Pascual (2020) Individual-based population genomics reveal different drivers of adaptation in sympatric fish. Scientific Reports 10:12683. https://doi.org/10.1038/s41598-020-69160-2

Factor de impacto: (2019 JCR Science Edition) 3.998 Q1 MULTIDISCIPLINARY SCIENCES (17/71)

Participación del candidato: Diseño del estudio, realización de la mayor parte del trabajo de laboratorio. Análisis bioinformático y tratamiento estadístico de datos. Escritura del manuscrito, producción de tablas y figuras.

Anna Barbanti*, **Héctor Torrado***, Enrique Macpherson, Luca Bargelloni, Rafaella Franch, Carlos Carreras, Marta Pascual (2020) Helping decision making for a reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. Molecular Ecology Resources (2020) 20:795– 806. https://doi.org/10.1111/1755-0998.13144

*la misma contribución en el artículo

Factor de impacto: (2019 JCR Science Edition) 6.286 D1 ECOLOGY (13/168)

Participación del candidato: Diseño del estudio, realización del trabajo de laboratorio, análisis de datos, y simulaciones bioinformáticas de *Diplodus puntazzo*.

Escritura del manuscrito, producción de tablas y figuras. El artículo también forma parte de la tesis de Anna Barbanti que realizó los análisis con *Caretta caretta*.

Núria Raventós, **Héctor Torrado**, Rohan Arthur, Teresa Alcoverro, Enrique Macpherson (submitted) Temperature reduces fish dispersal as larvae grow faster to their settlement size.

Manuscrito en segunda revisión

Participación del candidato: Análisis estadístico de los datos. Producción de tablas y figuras

Héctor Torrado, Baptiste Mourre, Núria Raventós, Carlos Carreras, Joaquín Tintoré, Marta Pascual, Enrique Macpherson (submitted) Impact of individual early life traits in larval dispersal: a multispecies approach using backtracking models.

Manuscrito en segunda revisión

Participación del candidato: Elaboración de los modelos de dispersión. Simulaciones de las dispersiones larvarias. Análisis de los datos y resultados. Escritura del manuscrito, producción de tablas y figuras.

Héctor Torrado, Núria Raventós, Carlos Carreras, Enrique Macpherson, Marta Pascual. Born late, die late: Genomic basis for early-life selective mortality in *Diplodus puntazzo*.

Manuscrito en preparación

Participación del candidato: Elaboración de librerías genómicas. Análisis bioinformáticos, estadísticos y funcionales de los datos obtenidos. Escritura del manuscrito, producción de tablas y figuras.

Héctor Torrado, Eric Treml, Carlos Carreras, Enrique Macpherson, Marta Pascual. Matching genomics with biophysical-based networks to evaluate connectivity

Manuscrito en preparación

Participación del candidato: Planteamiento de las hipótesis y análisis de datos moleculares y de dispersión. Elaboración de los modelos. Escritura del manuscrito, producción de tablas y figuras.

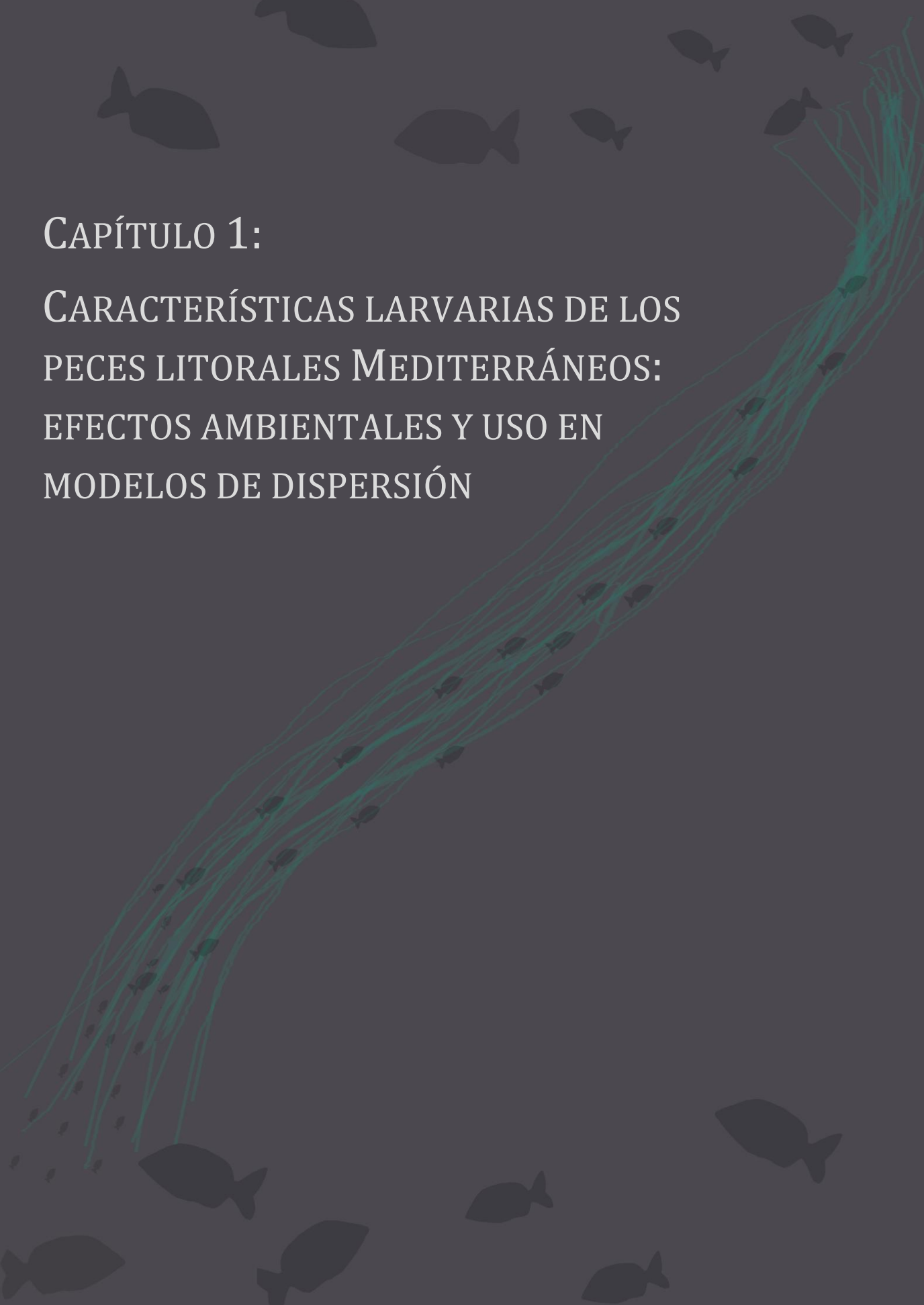
Barcelona, 11 noviembre 2020

Directores:

Dra. Marta Pascual Berniola Dr. Enrique Macpherson Mayol

Capítulo 1:

Características larvarias de los peces litorales Mediterráneos: efectos ambientales y uso en modelos de dispersión



Capítulo 1.1

Temperature reduces fish dispersal as larvae grow faster to their settlement size

Núria Raventós[1], Héctor Torrado [1,2], Rohan Arthur[3,1], Teresa Alcoverro[1] & Enrique Macpherson[1]

[1]Centre d'Estudis Avançats de Blanes, CEAB-CSIC, Blanes, Girona, Spain
[2]Department of Genetics, Microbiology and Statistics and IRBIO, University of Barcelona, Barcelona, Spain
[3]Nature Conservation Foundation, Amritha, 1311, 12th Cross, Vijayanagar 1st stage, Mysore 570017, India

Abstract

1. As species struggle to cope with rising ocean temperatures, temperate marine assemblages are facing major reorganisation. Many benthic species have a brief but critical period dispersing through the plankton, when they are particularly susceptible to variations in temperature. Impacts of rising temperatures can thus ripple through the population with community-wide consequences. However, responses are highly species-specific, making it difficult to discern assemblage-wide patterns in the life histories of different fish species.

2. Here we evaluate the responses to temperature in the early life histories of several fish species using otolith reconstructive techniques. We also assess the consequences of future warming scenarios to this assemblage.

3. We sampled recent settlers of 9 common species across a temperature gradient in the Mediterranean Sea and obtained environmental data for each individual. Using otolith microstructure we measured early life traits including pelagic larval duration (PLD), growth rate, settlement size, hatching and

settlement dates. We used a GLM framework to examine how environmental variables influenced early life history parameters.

4. We show that increasing temperature results in considerable reduction in the dispersal potential of temperate fish. We find a nearly universal, assemblage-wide decline in pelagic larval duration (PLD) of between 10-25%. This was because, with increasing temperature, larvae grew quicker to their settlement size. Settlement size itself was less affected by temperature and appears to be an ontogenetically fixed process.

5. These results suggest that evolutionary hard wiring might place strong limits to species adaptive capacities. Given current estimates of ocean warming, there could be an assemblage-wide reduction in larval dispersal of up to 50 km across the Mediterranean, reducing connectivity and potentially isolating populations as waters warm.

Key words: Early life traits, pelagic larval duration, productivity, temperature gradients, climate change

Introduction

Many marine organisms have complex life histories, that include a fleeting pelagic larval phase followed by a longer, more sedentary adulthood. This bipartite lifecycle makes early life processes especially pivotal, as young larvae navigate large scale oceanographic processes in the open ocean. The amount of time larvae spend in their pelagic form determines the length of time they are subject to movement by currents, influencing how far they can travel (Kinlan, Gaines, & Lester, 2005; Shanks, 2009). Like other species with a bipartite lifecycle, the pelagic phase of most marine fish presents a series of critical bottlenecks for fish larvae, influencing their developmental transitions, the distance they travel, when, and where they settle, and a host of post-settlement processes thereafter (Cowen & Sponaugle, 2009; Kendall, Poti,

& Karnauskas, 2016; Leis, 2006; Shima & Swearer, 2016). How species navigate these bottlenecks can determine adult densities, connectivity between populations, and the structuring of the entire adult community (Selkoe & Toonen, 2011; Shima & Swearer, 2016). However, despite the importance of this period, the pelagic phase of marine fishes remains enigmatic since following larvae from hatching to settlement in the water column is extremely complex. The recognition that the otoliths (ear stones) of many fish record daily growth increments and have clear settlement marks (Pannella, 1971; Wilson & McCormick, 1997), has made possible reconstructive ageing of fish to the day, giving us a window into the blackbox of fish planktonic stages. By counting and measuring daily rings of recently settled larvae it is possible to retrospectively examine early growth rate, size at hatching and settlement, back-calculated hatching date, and other early-life traits (Wilson & McCormick, 1997).

The length of the pelagic period can be influenced by a host of environmental variables including water temperature (Álvarez-Noriega et al., 2020; McLeod et al., 2015; O'Connor et al., 2007) that act on organismal development and mediate ontogenetic transitions. As pelagic waters warm with anthropogenic climate change, it can trigger subtle shifts in larval life histories with major implications for fish communities and the ecosystems they inhabit (Munday et al., 2009). Hatching success, larval growth and survival, metamorphosis and settlement are critical milestones that respond to cues in the pelagic environment such as sea surface temperature and productivity (Bergenius, McCormick, Meekan, & Robertson, 2005; McLeod et al., 2015; Robitzch, Lozano-Cortés, Kandler, Salas, & Berumen, 2016). These environmental parameters are changing alarmingly in temperate waters due to climate change (Adloff et al., 2015). While models vary, sea-surface temperatures (SST) are on a steady upward trajectory; in the Mediterranean for instance, SST is predicted to rise between +1.7ºC and 3ºC by the end of the century (2070-2099 vs. 1961-1990). In addition, projections indicate an overall increase of phytoplankton biomass, of about 13-28% over the next century (Moullec et al., 2019). It is still unclear how fish larval processes will respond to these changes at local or continental scales.

Most studies focus on three related developmental markers in fish larvae – growth rate, settlement size and pelagic larval duration (PLD) – which together influence how far larvae can travel and their post-settlement success. Although species also have their own swimming behaviour and settlement strategies(Faillettaz, Paris, & Irisson, 2018), the amount of time larvae spend in the water (PLD) is still a remarkably good indicator of their dispersal ability (Selkoe & Toonen, 2011), together with oceanographic fronts and currents (Pascual, Rives, Schunter, & Macpherson, 2017; White et al., 2010). Shorter PLDs can result in more isolated populations (Pascual et al., 2017), making them vulnerable to location-specific stresses and local extinctions. By modifying PLD, changing ocean temperatures could influence population connectivity at a regional scale (Kleypas et al., 2016). PLD itself is an emergent trait that results from the covarying relationship between larval growth and size at settlement. Faster growth is often linked to a larger size at settlement and potentially better post-settlement success and shorter PLDs (Gagliano, McCormick, & Meekan, 2007; Grorud-Colvert & Sponaugle, 2011; McLeod et al., 2015; O'Connor et al., 2007). However, the size at which larvae settle to the benthos is a complex trade-off between the uncertainties and physiological costs of continuing to grow in the plankton, and the risks of permanently settling in an unsuitable environment. Settlement size could influence post-settlement survival because of predation risk or resource availability, and could have long-term impacts on post-settlement growth and fitness (Searcy & Sponaugle, 2001). How larval growth and size at settlement respond to environmental changes will together influence PLD outcomes.

The emerging consensus is that fish larvae may have species-specific optima in their responses to increasing temperatures or changing productivity with no single directional response across the assemblage (McLeod et al., 2015). For instance, at high temperatures and low food availability, coral reef damselfish show an increase in PLD and a decrease in settlement size, associated with higher mortality (McLeod et al., 2015, 2013), suggesting that there is an optimum thermal range for these species beyond which they suffer significant fitness costs. Other species show an initial increase followed by a decrease in PLD across temperature gradients (e.g. (O'Connor

et al., 2007; Robitzch et al., 2016), and references cited therein). Apart from an overall increase in temperature, climate change is also disrupting the normal patterns of seasonal flux in environmental variables (Behrenfeld et al., 2016). This could signify differential impacts for species that spawn in different seasons. Additionally, each species is likely to have its own limit to the flexibility of its phenotypic response to shifting conditions either because of metabolic costs or metamorphic/developmental constraints. Studies typically examine one or two species within the assemblage (Cowen & Sponaugle, 2009; Munday et al., 2009; O'Connor et al., 2007), making it difficult to assess how the larger community is likely to respond to a changing climate. Species within an assemblage may each have unique responses based on the life history traits of its constituent species. In marine systems, genetic estimates of dispersal distances provide some evidence that species higher in the food web disperse more widely than species in lower trophic levels (Kinlan et al., 2005). Similarly, the spatial and temporal flexibility in spawning times or settlement behaviour may vary considerably between species, influencing their dispersal capabilities (Luiz et al., 2013).

An immediate question then, is how differential responses to climate change affect the dispersal and overall connectivity of fish assemblages. The early life history traits of temperate fish can respond to increasing temperatures either by changing the timing of critical life history processes (modifying phenology), or by altering their development milestones (modifying ontogenesis). Both these processes can influence the dispersal ability of species. Delaying or advancing hatching time could subject young larvae to slightly different oceanographic patterns, which could either facilitate or hinder larval flow. Alternatively, larvae could respond to increasing temperatures by modifying their growth or the size at which they settle; depending on the degree and direction of these responses, the duration of larval dispersal could either increase, decrease or remain relatively unchanged. Given how little we know of larval biology for most fish species, it is difficult to predict with any degree of certainty, how species will respond.

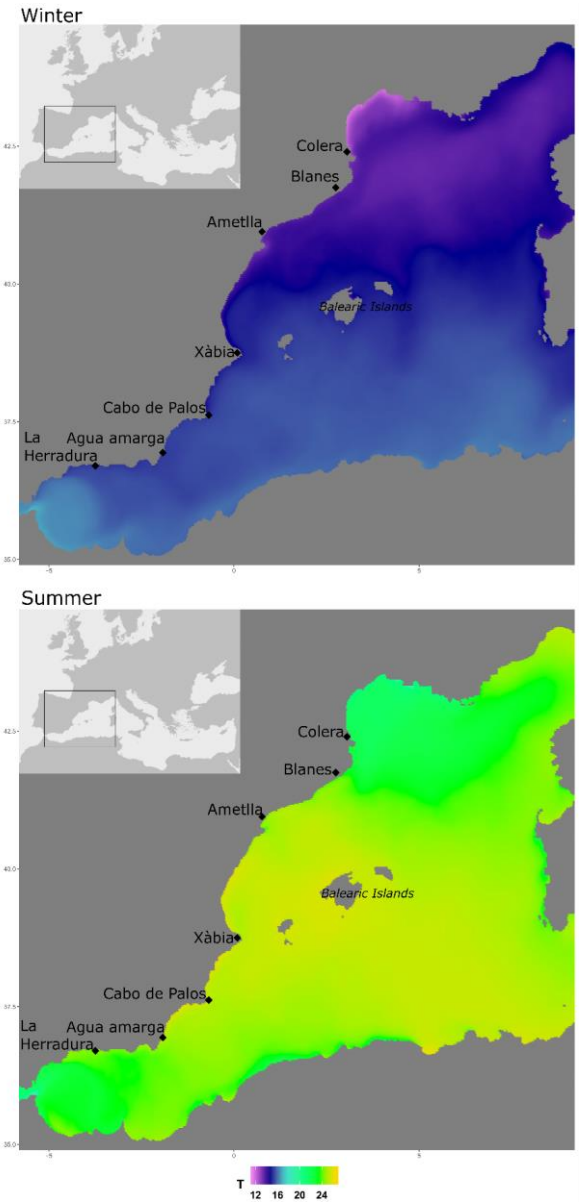


Figure 1: Sampling locations along a gradient of sea surface temperatures (SSTs) along the Mediterranean coast of Spain. Dots indicate sampling sites along the gradient where fish settlers were collected. Average annual differences between sites was approximately 4°C. Temperatures represent average SST for winter (above) and summer (below) based on SOCIB's WMOP model data (see methods).

In this study, we explored how temperature influences the time of hatching, larval growth, settlement size and PLD of an assemblage of marine fish in the Mediterranean. In addition, we explored the potential consequences of these patterns

for dispersal under different climate change scenarios. *A priori*, we expect that each species in the assemblage will respond differently to increasing temperature, by modifying either their phenology (time of hatching) or their ontogenesis (larval growth, size at settlement or PLD). How each individual species balances these responses will influence their dispersal potential under different climate change scenarios - with long-term assemblage-wide consequences as population ranges contract or expand relative to each other. In order to address this question we used a gradient of sea surface temperature regimes across the Mediterranean Sea (Figure 1) as a proxy for climate change. This makes for an ideal space-for-time substitution, with the simplifying but useful assumption that spatial and temporal variation are equivalent. While less precise in determining the exact magnitude of change, it can be a powerful means of evaluating qualitative trends or to generate more accurate hypotheses (Elmendorf et al., 2015; Pickett, 1989). We assessed early life traits of 9 common co-occurring fish (Table 1) across the gradient using reconstructive otolith microstructure techniques. Increases in temperatures could trigger species-specific responses with assemblage-wide consequences as oceans warm.

Methods

Site Selection and study design

We used a space-for-time substitution across a gradient of sea surface temperatures as a proxy of future climate change projected to affect the western Mediterranean Basin. This approach was used since manipulative experiments are often unfeasible during larval phases and to avoid the many problems of acclimation, scale and representativeness associated with studies that rely solely on laboratory techniques(Wernberg, Smale, & Thomsen, 2012). We selected 7 locations along the Iberian coasts of the Mediterranean Sea, which differ in their seasonal temperature regimes (Di Biagio et al., 2019), representing a gradient of approximately 4°C from the coldest sites (Colera and La Herradura in the north and south respectively) to the warmest (Agua amarga in south) during the summer. Across these sites, at the time of

hatching, larvae can experience temperatures of between 12ºC to 26ºC Sea Surface Temperature (SST). The sites spanned a range of 42°40'N, 3°11'E to La Herradura 36°44'N, 3°44'W and although the gradient had a latitudinal component, La Herradura in the south is influenced by Atlantic currents making it colder than northern locations in some seasons (Figure 1). At currently projected rates (between 0.24 ºC to 0.51 ºC decadal increase in SST) this gradient translates to the overall temperature changes the Mediterranean is likely to experience over the next century (Pachauri et al., 2014). Daily sea surface temperature (ºC) was obtained from the SOCIB system and WMOP model (Juza et al., 2016; Tintoré et al., 2013). In addition monthly mean values of ChLA concentration ($mg.m^{-3}$) were extracted for the study region from the Giovanni online data system, developed and maintained by NASA GES DISC (Acker & Leptoukh, 2007).

Measuring early life history milestones using otolith microstructure

Otoliths (ear bones) of most temperate fish leave distinctive daily growth rings from the moment of hatching, and lay a clear settlement band the day they transition from a pelagic to a benthic stage (Raventos & Macpherson, 2001; Wilson & McCormick, 1997) These daily structures are most clearly seen in young settlers. To test how early life history traits of fishes responded to the gradient of temperature, we collected settlers and small juveniles of nine representative species on SCUBA or by snorkelling. Overall, 1498 individuals were sampled belonging to the following species: *Chromis chromis* (n=146), *Coris julis* (n=150), *Diplodus puntazzo* (n=206), *Diplodus sargus* (n=175), *Diplodus vulgaris (n*=174*), Oblada melanura* (n=216), *Sarpa salpa* (n=111), *Symphodus ocellatus* (n=222) and *Symphodus tinca* (n=98) (Table 1). Together, these species form a representative assemblage of shallow benthic Mediterranean fish including several functionally important species; *Sarpa salpa*, is the most important herbivore in the Mediterranean Sea, while many of the *Diplodus spp*. are key meso-predators, responsible for regulating sea urchin populations (Prado, Tomas, Alcoverro, & Romero, 2007; Sala et al., 2012). They span several trophic levels (from predators to herbivores), different settlement seasons (cold water settlers and warm water settlers), and vary considerably in several other life history traits (Table 1).

Importantly, they are found ubiquitously across the gradient and all produce a clear settlement band in their otoliths (Raventos & Macpherson, 2001). Individuals were caught with hand nets by expert taxonomists (E.M and N.R.) and preserved in 100% ethanol. Individuals were collected over several years (2014, 2015 and 2017) in order to adequately sample all sites and species (since some species overlapped in their settlement times). To avoid interannual variation, we restricted our sampling to a single species across all sites within a single year. The settlement period for each species and locality is relatively short, usually 1-2 months, with the exception of *Diplodus vulgaris* that can settle over 3-4 months (see Table 1, García-Rubies & Macpherson, 1995). Collection and field procedures followed Spanish Law for Animal Experimentation (Royal Executive Order, 53/2013), in accordance with European Union directive 2010/63/UE.

Otolith extraction and processing was performed at CEAB's Otolith Research Lab (www.ceab.csic.es/en/otolith-research-lab-2). Otoliths (lapilli or sagitta) were extracted and mounted on microscope slides using thermoplastic glue (Crystalbond 509). To expose daily increments, the otoliths were mounted and polished using two different grained sandpapers (3 μm and 1 μm Imperial lapping film, 3M) to obtain a thin section exposing the nucleus and all the daily increments within the same plane. Reading of otolith increments was performed using a high-powered microscope with transmitted light (AxioPlan, Zeiss) connected to a ProgRes C10 camera (Jenoptic) and an image analysis system (Raventos & Macpherson, 2001). Otolith settlement marks in these species were always very clear (Raventos & Macpherson, 2001; Wilson & McCormick, 1997). We analysed the otoliths from the centre to the edge along the longest radius when possible. If not, counts were made along other axes in areas where increments were clearly visible. The length of the pelagic larval duration (PLD) for each individual was determined by counting the number of daily rings visible between the core and the settlement mark, and the age was determined by counting the total number of increments from core to margin in order to determine day of birth. Hatching day was calculated by subtracting the day of sampling from the total number of daily increments.

Table 1: Life history traits of sampled fish species. Early life history traits (based on sampled individuals): Hatching period (date range), Season at the time of hatching, PLD range (in days), Settlement size (in μm) and Days after settlement. Other species life history traits from secondary data (FishBase, Froese & Pauly, 2016) and personal observations: Otolith settlement size, Extended larval period, Laying behaviour, Trophic level, Adult length (cm) and Schooling behaviour (Small shoals < 10 individuals, large shoals 10s to 100s of individuals)

Species	N	Hatch dates (min-max)	Season	PLD range	Otolith settlement size (min-max)	Days after settlement	Extended larval period	Egg laying behaviour	Trophic level	Adult length	Schooling behaviour
Chromis chromis	146	02/06/2015-06/07/2015	Warm	13-21	89.0-153.0	10±5	No	Nester	3.38	13	Large school
Diplodus puntazzo	206	22/08/2015-22/10/2015	Warm	16-29	68.8-165.2	19±16	No	Pelagic spawning	3.07	30	Small school
Diplodus sargus	175	01/03/2015-15/05/2015	Cold	14-32	52.9-106.1	25±16	No	Pelagic spawning	3.24	22	Large school
Diplodus vulgaris	174	19/10/2014-31/01/2015	Cold	22-43	64.8-125.7	59±21	No	Pelagic spawning	3.34	22	Large school
Oblada melanura	216	22/05/2015-09/07/2015	Warm	dic-20	60.6-109.6	11±5	No	Pelagic spawning	3.38	20	Large school
Sarpa salpa	72	30/09/2017-27/10/2017	Warm	dic-24	46.9-88.8	18±8	No	Pelagic spawning	2	30	Large school
Coris julis	150	08/05/2015-01/07/2015	Warm	19-42	90.3-158.4	11±5	Yes	Pelagic spawning	3.24	20	Small school
Symphodus ocellatus	177	02/06/2014-12/09/2014	Warm	jul-14	31.6-71.5	98±16	No	Nester	3.34	12	Small school
Symphodus tinca	97	18/05/2015-22/06/2015	Cold	ago-18	29.0-48.0	42±13	No	Nester	3.23	25	Small school

We also recorded the size-at-hatching (radius from core to the first visible increment), size-at-settlement (radius from core to the settlement mark along the longest axis of the otolith), and growth rate before and after settlement (pre- and post-settlement growth, respectively). For pre-settlement growth, size-at-settlement was divided by PLD. To minimize errors, we repeated all measurements three times, and used the averaged values.

Environmental conditions during pelagic stage

For every fish sampled, we obtained values of SST and productivity by averaging environmental conditions for the area at which the individual was collected (ca. 100 km around each locality), from the time of hatching to the time of settlement (corresponding to the PLD value). This makes the assumption that the temperature at the settlement location was similar to hatching locations. For most species in the assemblage that do not have large dispersal ranges, this is a reasonable assumption. SST values were obtained from the SOCIB system (Balearic Islands Coastal Observing and Forecasting System, Tintoré et al., 2013) and productivity was estimated as Chlorophyll A concentration ($mg.m^{-3}$) extracted from the Giovanni online data system (Acker & Leptoukh, 2007).

Potential climate scenarios

The Mediterranean Sea has been identified as a hotspot of future climate change (Diffenbaugh & Giorgi, 2012), in terms of changes in mean seasonal climate features and extreme events (Paxian et al., 2015).

In this study, we projected the reduction in maximum linear dispersal capabilities of fish larvae under three different Representative Concentration Pathway (RCPs) which are projected scenarios for 2100 assuming different greenhouse gas concentrations. To determine how species responses could influence dispersal distance, we assumed an average dispersal capacity of 12 $km.day^{-1}$ based on Schunter et al. (2011). This dispersal was based on numerical simulations of particle dispersion using the Mediterraean Forecasting System (MFS) model (Tonani et al., 2009). It must

be noted that this represents the hypothetical maximum linear dispersal capacity a larva can travel. Maximum dispersal distance does not translate directly into the distance routinely travelled by particles since most will stay much closer to home because of eddies, non linear currents, physical and oceanographic barriers (Rossi, Ser-Giacomi, Lõpez, & Hernández-García, 2014). We modelled maximum linear dispersal capacity for RCP 2.6, RCP 4.5 and the RCP 8.5, ranging from the lowest to the highest greenhouse gas concentration scenarios, respectively (Meinshausen et al., 2011). These scenarios predict a variation in West Mediterranean mean sea surface temperature from +0.94 ºC (in RCP 2.6) to +3.55 ºC (in RCP 8.5). We calculated the mean difference between present and 2100 scenarios across the West Mediterranean (between 5.8ºW and 9ºE) based on the maximum, mean and minimum temperatures, and projected the effects (time reduction) on pelagic larval duration (PLD) for each species, in days, and translated to kilometers based on Schunter et al., (2011).

Statistical analysis

We used a GLM framework to examine how environmental variables influenced early life history parameters. To test how environmental conditions affected ontogenesis and phenology, we first used assemblage-level models for all species together to examine how i. Sea surface temperature (SST), productivity and season influenced pre-settlement growth, size of settlement and PLD (ontogenesis) and ii. average temperature influenced hatching periods (phenology). In addition, to explore in more detail how temperature influenced each species, we used iii. individual species models for pre-settlement growth, size at settlement and PLD with SST. Finally, iv. we built individual species models to project how dispersal would change under different climate change scenarios. The details of the models are as follows:

I. assemblage-wide responses of early life history traits to environmental gradients: We used a series of assemblage-level generalized linear mixed models (LMM) including as fixed factors: SST (calculated as the average temperature the individual was exposed over their PLD), season (spring/summer or autumn/winter, based on the hatching period of each

species) the interaction between SST and season, and productivity (ChlA, the average productivity each individual was exposed to over their PLD). In addition, sampling locality (7 levels) and species identity (9 levels) were included as random factors to take into account spatial and phylogenetic variability in our models. The final model took the form:

$$ELT = \beta_1 \text{SST} + \beta_2 \text{Season} + \beta_3 \text{ChlA} + \beta_4 \text{SST} * \text{Season} + \beta_5 \text{SST} * \text{ChlA} + \beta_6 \text{Season} * \text{ChlA} + \beta_7 \text{SST} * \text{Season} * \text{ChlA} + b_1 \text{Species} + b_2 \text{Locality}$$

Where ELT is each Early Life Trait (pre-settlement growth, size of settlement or PLD), β represents the fixed effects' coefficients and b represents the random effects coefficients. In order to make the coefficients of each variable comparable for the discussion, we standardized each early life trait by dividing each individual value by the mean value of the trait across all individuals of the same species (Table 2).

II. To evaluate possible changes to the start of hatching periods (a proxy of reproduction time) across the temperature gradient, we used an LMM with year-long mean temperature of the sampled year at each sampled locality (Tloc) as a fixed factor, with species identity and locality as random factors. The model took the form:

$$Hatching\ day \sim \beta_1\ Tloc + b_1 Species + b_2 Locality$$

III. Early life traits and SST at the species level were tested with a series of species-specific LMMs where locality was a random factor and SST a fixed factor, with the form:

$$ELT \sim \beta_1\ SST + b_1 Locality$$

IV. To estimate the reduction of PLD under different scenarios of increasing temperature in the Mediterranean Sea we performed additional linear models between PLD (without log transformation) and SST, using these coefficients to predict possible changes in PLD due to variation in temperature conditions

under the climate change scenarios RCP 2.6, RCP 4.5 and RCP 8.0 for each species.

Due to the large differences in variance of dependent variables between species, variables were standardized dividing individual values of each species by species' mean values. Pre-settlement growth and PLD were then transformed with natural logarithms. After transformation, the model residuals were normally distributed and homoscedastic for all variables. The LMM analyses were conducted in R, version 3.4.4 (R Core Team, 2018) using the lmer function from package lme4 version 1.1-17 (Bates, Mächler, Bolker, & Walker, 2015), and evaluated with lmer and ranova functions form lmerTest package version 3.0-1 (Kuznetsova, Brockhoff, & Christensen, 2017) and MuMIn package version 1.42.1 (Barton, 2018). Figures were built with ggplot2 (Wickham, 2016) and viridis (Garnier, 2018) R packages and/or GIMP 2.8.22 software (www.gimp.org, The GIMP team). We obtained projected reductions in pelagic larval duration (PLD) from species-specific linear models across the gradient of sampled temperatures. Present and projected climate change scenarios for the Mediterranean in the twenty-first century were obtained from Bio-ORACLE (Assis et al., 2018).

Results

Pre-settlement growth (model i and iii): Across the assemblage, pre-settlement growth showed a general trend of increase along the gradient of temperature (Table 2a). Almost every species in the assemblage responded to increased temperatures with higher pre-settlement growth (Table 3, Figure 2a). In contrast, the trend along the productivity gradient was much weaker (coefficient -0.08 *versus* 1.25, ANOVA mean squares 0.240 *versus* 0.362) (Table 2a, Table S1). Temperature response differed markedly between species hatching at different times of the year, with cold season hatchers (*Diplodus vulgaris* and *D. sargus*) responding far more than warm season hatchers (Table 3). This difference was confirmed by the interaction term in the model (T*season, Table 2a) and in the correlation between the slope of growth and the minimum hatching temperature for each species (Figure 3a). The species with the

weakest response, the wrasse *Coris julis*, was the only one in the assemblage with an extended larval period (Table 1), which allowed larvae to delay metamorphosis.

Settlement-size (model i and iii): In the assemblage-wide model, settlement size responded weakly to temperature and productivity (Table 2b). The species-specific model showed that this signal was not consistent across the assemblage since, unlike the rest of the assemblage, winter species showed a mild response in their settlement size (Figure 2b, Figure 3b, Table 3). Settlement size was clearly an attribute linked to species identity with some species showing settlement sizes that were twice as large as smaller-sized individuals (Table 1).

Tabla 1: ANOVA tables for the influence of environmental variables and species life history traits on pre-settlement growth rate (a), settlement size (b) and pelagic larval duration (PLD) (c). Locality and species identity are included as random factors (indicated with parentheses). Bold p-values indicate significance at α=0.05. ChlA: Chlorophyl A, SST: Sea Surface Temperature, Length: Mean adult length, Trophic: Trophic level, Schooling: Juvenile schooling behaviour, Lay: Egg laying behaviour, Tloc: mean locality temperature over the sampled year. For full Linear mixed effects model results see Table S1.

a

Larval Growth Rate	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
SST	0,362	0,362	1	1202,6	37,7	**< 0.001**
Season	0,379	0,379	1	936,5	39,5	**< 0.001**
ChlA	0,240	0,240	1	1231,4	25,0	**< 0.001**
SST:Season	0,485	0,485	1	1370,6	50,5	**< 0.001**
SST:ChlA	0,181	0,181	1	1199,5	18,9	**< 0.001**
Season:ChlA	0,165	0,165	1	1270,7	17,2	**< 0.001**
SST:Season:ChlA	0,101	0,101	1	1261,8	10,5	**0,001**

b

Settlement size	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
SST	0,016	0,016	1	1184,2	1,3	0,255
Season	0,248	0,248	1	1283,1	20,4	**< 0.001**
ChlA	0,168	0,168	1	1225,7	13,8	**< 0.001**
SST:Season	0,222	0,222	1	1336,2	18,3	**< 0.001**
SST:ChlA	0,178	0,178	1	1209,2	14,6	**< 0.001**
Season:ChlA	0,053	0,053	1	1264,8	4,4	**0,036**
SST:Season:ChlA	0,025	0,025	1	1267,5	2,1	0,151

c

PLD	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
SST	0,513	0,513	1	1230,2	51,2	**< 0.001**
Season	0,014	0,014	1	923,5	1,4	0,241
ChlA	0,011	0,011	1	1257,0	1,1	0,298
SST:Season	0,045	0,045	1	1386,4	4,5	**0,034**
SST:ChlA	0,002	0,002	1	1227,6	0,2	0,693
Season:ChlA	0,034	0,034	1	1295,8	3,4	0,065
SST:Season:ChlA	0,028	0,028	1	1287,8	2,8	0,096

Pelagic larval duration (PLD, model i and iii): In the assemblage-wide model, temperature and its interaction with hatching season were the only predictors with a significant effect on PLD (Table 2c). Moreover, in the species-specific analysis every species showed PLD declining across the temperature gradient (Table 3, Figure 2c). While decreases across the ca 4ºC gradient was modest in some cases (10% for *Chromis chromis*) it reduced by a quarter for some (24% for *Symphodus tinca,* Figure 2c). PLD was not affected by productivity or any other interaction term (Table 2c).

Tabla 2: Coefficients (slopes) of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05. For full model results see Table S2.

Specie	Larval Growth Rate		Settlement size		PLD	
	Mean Sq	Estimate±sd	Mean Sq	Estimate±sd	Mean Sq	Estimate±sd
Chromis chromis	**0.145**	**0.033±0.007**	0.002	-0.004±0.007	**0.16**	**-0.033±0.007**
Coris julis	0.007	0.011±0.008	**0.031**	**-0.023±0.011**	**0.072**	**-0.035±0.015**
Diplodus puntazzo	**0.892**	**0.057±0.005**	0.053	0.017±0.008	**0.272**	**-0.042±0.007**
Diplodus sargus	**1.444**	**0.075±0.007**	**0.127**	**0.022±0.006**	**0.759**	**-0.051±0.005**
Diplodus vulgaris	**1.97**	**0.083±0.005**	**0.519**	**0.044±0.006**	**0.421**	**-0.040±0.005**
Oblada melanura	**0.104**	**0.036±0.009**	0.046	-0.022±0.011	**0.795**	**-0.058±0.006**
Sarpa salpa	0.001	0.005±0.013	0.002	0.003±0.010	**0.048**	**-0.024±0.010**
Symphodus ocellatus	**0.142**	**0.022±0.007**	0.01	-0.006±0.006	**0.236**	**-0.028±0.007**
Symphodus tinca	**0.343**	**0.043±0.006**	0	-0.002±0.010	**0.339**	**-0.051±0.008**

Hatching day (model ii): Hatching day remained relatively constant across the temperature gradient (Table 4). There was distinct seasonality in hatching with cold-hatching species emerging between October and May and warm-hatching species emerging between April and September (Table 1). This general phenology did not vary across the gradient of temperature.

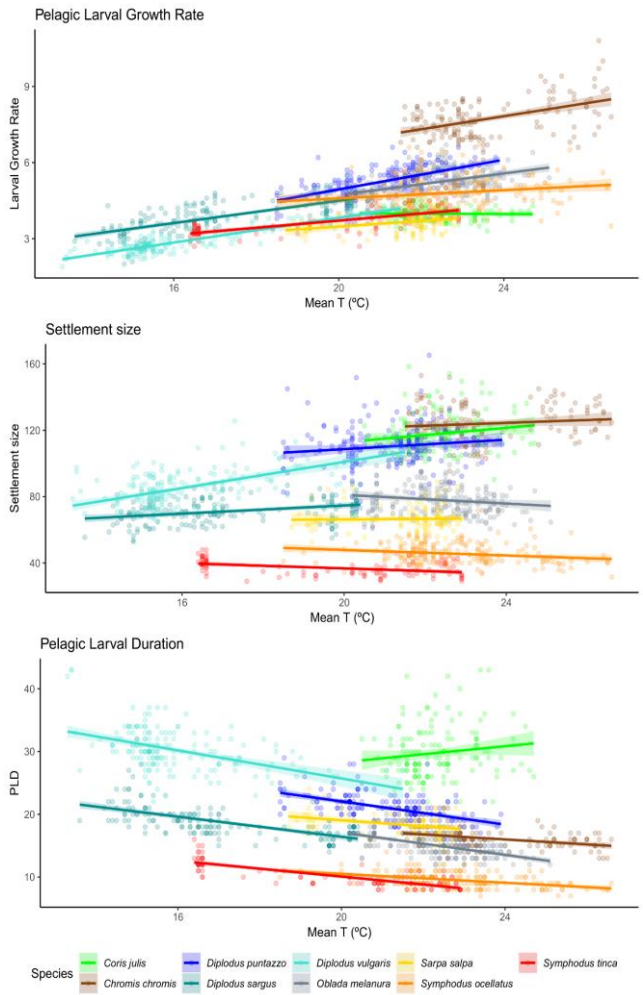


Figure 2: Influence of sea surface temperature on fish early life history traits. a. Pre-settlement growth rate (µm.day-1), b. Settlement size (µm) and c. Pelagic larval duration (PLD, in days). Lines represent best fit trend lines and 95% confidence bands for each specie. Temperatures (x axis) represent mean temperatures (°C) experienced by each sampled individual over its pelagic larval duration.

Future climate scenarios (model iv): In our predictions for the West Mediterranean (Table S2, Figure 4) the increase in SST at the end of the century under the most optimistic scenario (RCP 2.6) was <1ºC mean temperature (1.01ºC to 1.61ºC). The more realistic RCP 8.5 scenario projects a very different future, with increases of 3.55ºC of mean temperature, potentially rising to almost 5ºC. This implies that, by the turn of the century, temperatures in the West Mediterranean approach 32.5ºC, with a mean of 22.7ºC.

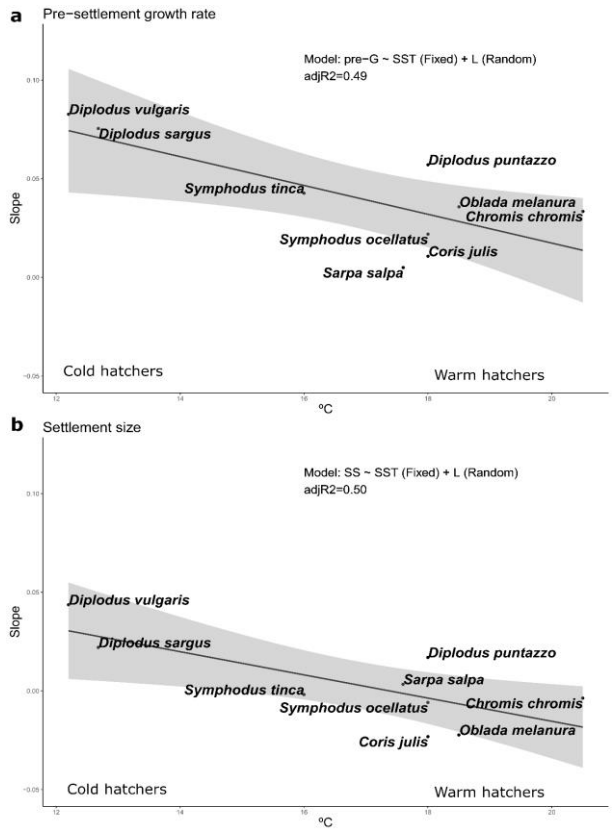


Figure 3: Influence of hatching season on pre-settlement growth rates and settlement size for each species. The x axis represents minimum hatching temperature for each species (°C). The y axis is the slope of a linear regression of pre-settlement growth (a) and settlement size (b) across the gradient of temperatures for all individuals of each species. Pre-G: Pre-settlement Growth Rate, SST: Sea Surface Temperature, SS: Settlement Size, L: Sampling locality.

Tabla 3: Linear mixed effects model coefficients and ANOVA table for the influence of temperature on hatching day. Locality and species identity are included as random factors (indicated with parentheses). Bold p-values indicate significance at α=0.05. T loc: mean locality temperature over the sampled year.

Model: Hatching day = Intercept + β*Tloc + b_1*Species + b_2*Locality

Fixed effects	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.796	0.129	9.648	6.19	**< 0.001**	
T loc	0.01	0.006	9.726	1.614	0.138	
	R^2 Fixed effects		0.019			
Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
(Specie)	4	1704.9	-3401.7	4.406	1	**0.036**
(Locality)	4	1644.9	-3281.7	124.407	1	**< 0.001**
R^2 complete model		0.148				

ANOVA table

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
T loc	0.015	0.015	1	9.726	2.605	0.139

Our results suggest that within the next century there would be a maximum reduction in linear dispersal distance of up to 50 km. The largest reductions would be experienced by the four species of sea breams (*Diplodus* spp. and *Oblada melanura*), whereas the two species of wrasses (*Symphodus* spp.), *Sarpa salpa* and *Chromis chromis* would likely experience smaller reductions (ca. 30 Km) (Figure 4). The wrasse *Coris julis* showed a different trend because of its capacity to delay metamorphosis.

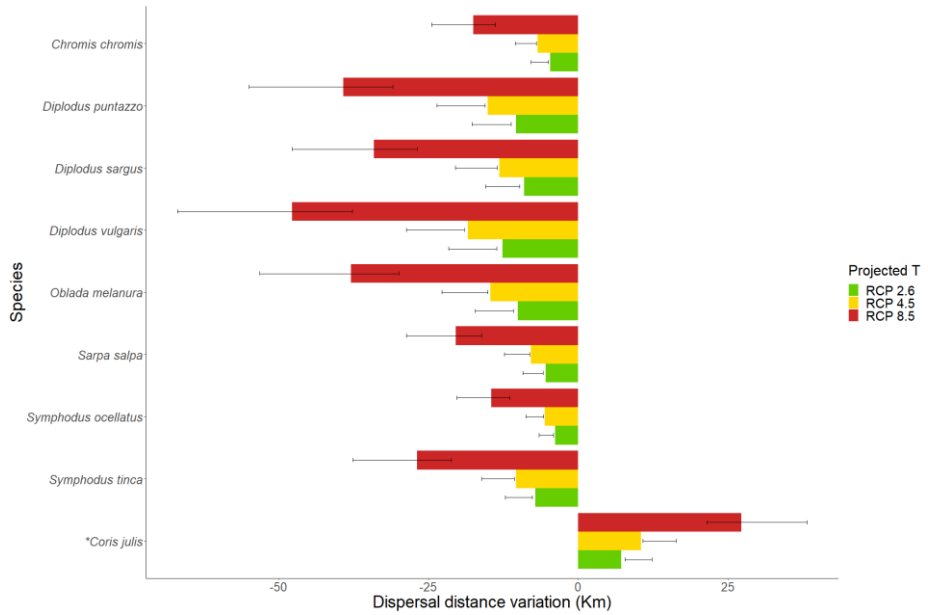


Figura 4: Projected reduction in maximal linear dispersal capabilities of fish larvae under a different warming scenarios (RCP 2.6, 4.5 and 8.5) for 2100. Columns mark changes in mean temperature, error bars mark changes in minimum and maximum temperatures. Reductions in pelagic larval duration (PLD) obtained from species-specific linear models across the gradient of sampled temperatures. We assume an average dispersal capacity of 12 km.day-1 based on (Schunter et al., 2011). Mediterranean present and projected scenario for the twenty-first century are obtained from Bio-ORACLE (Assis et al., 2018). *Coris julis is the only species with extended larval period and without significant correlation with temperature.

Discussion

Temperature modified several key processes in the early life history of fish larvae leaving others relatively unaffected. This differential set of responses acted together to influence the capacity of larvae to disperse in the plankton between hatching and settlement. Nearly all the species we sampled showed PLD declining across the temperature gradient (Table 2c, Figure 2c). Otolith microstructure helped

unpack the mechanisms of declining PLD, showing that it is the result of two contrasting responses to temperature. While larval growth responded strongly to increasing temperature (Table 2a and 3, Figure 2a), settlement size responded only weakly (Table 2b and 3, Figure 2b). As a result, larvae have considerably less time to disperse, as they grow quicker to a largely invariant settlement size. In this, cold-water hatchers showed the greatest change (between 25-30% of their maximum PLD), but this was mostly likely because they encompassed a larger temperature range than warm hatchers and had longer breeding seasons. Cold-season hatchers like *Diplodus vulgaris* and *D. sargus* did show mild increases in settler size at higher temperatures, but this response did not offset the overall decline in PLD (Figure 3b). Temperature was the strongest driver of larval growth and PLD confirming trends previously reported for other species (Sponaugle, Grorud-Colvert, & Pinkard, 2006). Productivity played a minor role in most early traits, particularly growth. While productivity has been shown to determine growth rates in many early life stages (e.g. Goldstein & Sponaugle, 2020), our results indicate that growth declined with productivity, which is likely a statistical artefact rather than an ecological reality. In addition, neither productivity nor its interaction with temperature significantly affected PLD (Table 2c).

Not every process in larval life history is pliable, or can change rapidly enough to keep up with changing temperatures. For instance, hatching day remained relatively constant across the temperature gradient (Table 4). There was distinct seasonality in hatching with cold-hatching species emerging between October and May and warm-hatching species emerging between April and September. This general phenology did not vary across the gradient of temperature. Similarly, as reported above, while pre-settlement growth showed a considerable assemblage-wide variation in response to increasing temperature, settlement size responded only weakly to temperature, even if there was a large species-specific range in settlement size (Table 3, Figure 2b). The patterns we report arise from the covariance between life history traits (larval growth and size at settlement) resulting in an overall decline in PLD with increasing temperatures. It is potentially true of all metamorphic processes that rely so heavily on the interaction between physiology and environment, that the multiple development

traits involved in ontogenesis covary closely, with some responding much more to environment than others. The declining PLD we report is really an emergent response based on these underlying relationships, but it has large population-level implications. Physiological flexibility in growth could occur fairly rapidly, but the size at which larvae settle may be a less pliant trigger that sets in motion a sequence of complex, ontogenetically unvarying development transitions essential to successful settlement (Leis, 2006). These include benthos seeking behaviours and diet shifts, among others, that may themselves be heavily size-dependent developmental milestones (Hare, Walsh, & Wuenschel, 2006). In fact, given how strong a bottleneck the first few weeks are in the post-pelagic life of fish, size at settlement may be a strongly selected trait linked to post-settlement survival processes. Settling at smaller sizes may come with significant fitness costs over the short term, linked to resource competition and predation, costs that could flow well into adulthood, even if settlers survive the first few weeks. While larger settler size can confer better survival in some instances (Caie & Shima, 2019; Meekan et al., 2006), it may be also associated with increased predation risk for shoaling settlers, where larger individuals may be much more conspicuous (the oddity effect, Rodgers, Downing, & Morrell, 2015). Of course, every species is likely to converge to very different optimal settlement sizes based on how they balance their relative costs and benefits.

The Mediterranean Sea is a hotspot in scenarios of global warming. where climate may strongly and rapidly impact hydrodynamics and marine ecosystems (Giorgi & Lionello, 2008). Models predict a rapid warming of surface waters (ca. 3-4ºC), although the warming is not homogeneous, with the Western Mediterranean regions warmer on average than the rest of the sea (Adloff et al., 2015). The most optimistic IPCC climate scenarios for the region indicate that reductions will be slight, but the more realistic business as usual scenarios (RCP 8.5) do not bode well for potential fish dispersal (Figure 4). Our projections suggest that within the next century, winter-hatching predators in the *Diplodus* genus could see a reduction in dispersal distance of up to 50 km from their natal areas (Figure 4). Clearly there will be species-specific differences in the assemblage with dynamically changing winners and losers as the

ocean warms. The wrasse *Coris julis* bucks this trend because of a unique capacity few species like it have to delay metamorphosis until they find a suitable settlement habitat (Laurel, Basilio, Danley, Ryer, & Spencer, 2015). *Coris julis* may be the only species able to buffer and even potentially benefit from future temperature rise (see Figure 4). Without this ability though, there was an across-the-board decline in PLD signifying that reducing dispersal ability with temperature was an assembly-wide phenomenon that transcended species life history. Higher temperatures have also been implicated in increased retention by improving larval swimming abilities (Leis, 2006), thereby decreasing passive transport through the plankton. To add to this, some models also predict changes in currents around the Balearic region with the penetration of Atlantic surface water toward the North adding to already known discontinuities which could further influence gene flow between populations (Adloff et al., 2015; and see Pascual et al., 2017 and references cited therein). Between declining dispersal capabilities, larval behaviour and oceanographic changes, climate change could result in strong changes in connectivity leaving populations subject to all the dangers of isolation and a potential contraction of ranges as waters warm (Selkoe & Toonen, 2011). This is particularly critical, because the species in this assemblage include a large range of trophic groups from piscivores to herbivores. Many play key roles in the ecosystems they inhabit. For instance, *Sarpa salpa* is the only native fish herbivore of macrophyte systems in the Mediterranean, and *Diplodus* species are important predatory agents of top-down control, regulating sea urchin populations from overgrazing macroalgal forests (Sala et al., 2012). In parallel, their populations also have to deal with the expanding ranges of exotic competitors currently entering the Mediterranean (Occhipinti-Ambrogi, 2007), and native communities may soon see their roles shrink along with their fragmenting populations. The ecological consequences of these community redistributions is still an open question, but it is likely to be large.

Using field gradients rather than controlled experimental studies is a handy means of exploring larval responses to environmental change, but it makes important simplifying assumptions. For one, we assume that the temperatures larvae were exposed to were similar to where settlers were collected from. Our temperature

estimates represent conditions within a 100 km radius of the settlement site, which is within the dispersal range of a majority of sampled species. For a few long-ranging species however (like *Diplodus* sps), this may introduce error in our estimates. We make the assumption that PLD translates to mean dispersal distance, although studies are equivocal. While there is often a clear relationship between PLD and maximum dispersal (Shanks, 2009), some studies indicate that larval behaviour and oceanographic currents may interact to weaken this relationship (D'Aloia et al., 2015; Shanks, 2009). In addition, average distances may not reflect maximum dispersal distance (Kinlan et al., 2005). A more detailed study of source and sink movement patterns will help more precisely determine early developmental temperatures. To be realistic, such studies need to integrate oceanographic patterns and physical features along the coast to understand how our predicted values of dispersal capacity translate to dispersal on the ground. What remains clear however is that across the assemblage, Mediterranean fish larvae are likely to be subject to these physical and oceanographic forces for far fewer days, reducing the overall capacity for dispersal. Another artefact of sampling a gradient is that locations may experience the same temperature conditions at slightly different times of the year. To reduce this effect, we concentrated our sampling to as small a window as possible (typically two weeks across the entire gradient) to ensure that temporal differences did not influence these patterns. However, while we sampled individuals within a week or two of them transitioning out of their pelagic phase, we cannot account for the considerable mortality that occurs within this critical period (Macpherson et al., 1997). At least some of the invariance in settler size we documented across species could be influenced by post-recruitment mortality acting on either smaller or larger sizes (Shima et al., 2020). Regardless of the mechanism, the universality of this signature points to how critical settlement size appears to be. Finally, it is important to highlight that although species show differential phenotypic responses across the gradient, the individuals we sampled may differ in their underlying genetic structure; our study cannot cleanly separate genetic variation from responses to environmental conditions. Species may well be able to respond to future climate change with a suite of genetic adaptations that this study was not designed to explore.

The galloping pace of climate change subjects species to unprecedented environmental flux. The relative ability of species to move, acclimate or adapt in the face of change can have assemblage-wide consequences as species scramble to respond, each within their own capacities. Not every species will have the flexibility and not every life history process can be altered at rates sufficient to offset changing conditions. Nor will every phenotypic response be adaptive (Grummer et al., 2019). Our results show that as species struggle to cope with a rapidly changing climate, evolutionary hard wiring may place strong limits on species adaptive capacity, with consequences that may flow on to juvenile survival, population connectivity and the composition of the entire assemblage.

Acknowledgements

The Spanish Ministry of Science and Innovation funded this research (projects UMBRAL; CTM2017-86695-C3-3-R; 'PopCOmics' CTM2017-88080 (MCIU, AEI/FEDER, UE) and project BBVA (PR17-ECO-0012)). The Spanish National Research Council supported R. Arthur's visitorship with the Memorandum of Understanding between CSIC and NCF. HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities (FPU15/02390). We thank Jordi Boada and Jordi Pages for their useful comments on statistical design.

Author contributions

E.M., R.A. and T.A. designed research. E.M., N.R. and H.T. did the sampling. N.R. did the otolith readings. H.T. conducted analyses with input from all authors. All authors contributed to conceptual development. All authors contributed equally to writing the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

Acker, J. G., & Leptoukh, G. (2007). Online analysis enhances use of NASA Earth Science Data. *Eos*, *88*(2). doi: 10.1029/2007EO020003

Adloff, F., Somot, S., Sevault, F., Jordà, G., Aznar, R., Déqué, M., … Gomis, D. (2015). Mediterranean Sea response to climate change in an ensemble of twenty first century scenarios. *Climate Dynamics*, *45*(9–10), 2775–2802. doi: 10.1007/s00382-015-2507-3

Álvarez-Noriega, M., Burgess, S. C., Byers, J. E., Pringle, J. M., Wares, J. P., & Marshall, D. J. (2020). Global biogeography of marine dispersal potential. *Nature Ecology and Evolution*, *4*(9), 1196–1203. doi: 10.1038/s41559-020-1238-y

Assis, J., Tyberghein, L., Bosch, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2018). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, *27*(3), 277–284. doi: 10.1111/geb.12693

Barton, K. (2018). *MuMIn: Multi-Model Inference. R package version 1.42.1. https://CRAN.R-project.org/package=MuMIn*. Retrieved from https://cran.r-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Behrenfeld, M. J., O'Malley, R. T., Boss, E. S., Westberry, T. K., Graff, J. R., Halsey, K. H., … Brown, M. B. (2016). Revaluating ocean warming impacts on global phytoplankton. *Nature Climate Change*, *6*(3), 323–330. doi: 10.1038/nclimate2838

Bergenius, M. A. J., McCormick, M. I., Meekan, M. G., & Robertson, D. R. (2005). Environmental influences on larval duration, growth and magnitude of settlement of a coral reef fish. *Marine Biology*, *147*(2), 291–300. doi: 10.1007/s00227-005-1575-z

Caie, P., & Shima, J. S. (2019). Patterns of selective predation change with ontogeny but not density in a marine fish. *Oecologia*, *189*(1), 123–132. doi: 10.1007/s00442-018-4303-3

Cowen, R. K., & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, *1*, 443–466. doi: 10.1146/annurev.marine.010908.163757

D'Aloia, C. C., Bogdanowicz, S. M., Francis, R. K., Majoris, J. E., Harrison, R. G., & Buston, P. M. (2015). Patterns, causes, and consequences of marine larval dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(45), 13940–13945. doi: 10.1073/pnas.1513754112

Di Biagio, V., Cossarini, G., Salon, S., Lazzari, P., Querin, S., Sannino, G., & Solidoro, C. (2019). Temporal scales of variability in the Mediterranean Sea ecosystem: Insight from a coupled model. *Journal of Marine Systems*, *197*, 103176.

Diffenbaugh, N. S., & Giorgi, F. (2012). Climate change hotspots in the CMIP5 global climate model ensemble. *Climatic Change*, *114*(3–4), 813–822. doi: 10.1007/s10584-012-0570-x

Elmendorf, S. C., Henry, G. H. R., Hollister, R. D., Fosaa, A. M., Gould, W. A., Hermanutz,

L., … Walker, M. D. (2015). Experiment, monitoring, and gradient methods used to infer climate change effects on plant communities yield consistent patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(30), E4156. doi: 10.1073/pnas.1511529112

Faillettaz, R., Paris, C. B., & Irisson, J. O. (2018). Larval fish swimming behavior alters dispersal patterns from marine protected areas in the North-Western Mediterranean Sea. *Frontiers in Marine Science*, *5*(MAR), 1–12. doi: 10.3389/fmars.2018.00097

Gagliano, M., McCormick, M. I., & Meekan, M. G. (2007). Survival against the odds: ontogenetic changes in selective pressure mediate growth-mortality trade-offs in a marine fish. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1618), 1575–1582.

García-Rubies, A., & Macpherson, E. (1995). Substrate use and temporal pattern of recruitment in juvenile fishes of the Mediterranean littoral. *Marine Biology*, *124*(1), 35–42. doi: 10.1007/BF00349144

Garnier, S. (2018). *viridis: Default Color Maps from "matplotlib" https://cran.r-project.org/package=viridis*. Retrieved from https://cran.r-project.org/package=viridis

Giorgi, F., & Lionello, P. (2008). Climate change projections for the Mediterranean region. *Global and Planetary Change*, *63*(2–3), 90–104. doi: 10.1016/j.gloplacha.2007.09.005

Goldstein, E. D., & Sponaugle, S. (2020). Juvenile reef fish growth and survival related to subregional patterns of primary production. *Marine Biology*, *167*(2). doi: 10.1007/s00227-019-3627-9

Grorud-Colvert, K., & Sponaugle, S. (2011). Variability in water temperature affects trait-mediated survival of a newly settled coral reef fish. *Oecologia*, *165*(3), 675–686. doi: 10.1007/s00442-010-1748-4

Grummer, J. A., Beheregaray, L. B., Bernatchez, L., Hand, B. K., Luikart, G., Narum, S. R., & Taylor, E. B. (2019). Aquatic Landscape Genomics and Environmental Effects on Genetic Variation. *Trends in Ecology & Evolution*, *34*(7), 641–654. doi: 10.1016/j.tree.2019.02.013

Hare, J. A., Walsh, H. J., & Wuenschel, M. J. (2006). Sinking rates of late-stage fish larvae: Implications for larval ingress into estuarine nursery habitats. *Journal of Experimental Marine Biology and Ecology*, *330*(2), 493–504. doi: 10.1016/j.jembe.2005.09.011

Juza, M., Mourre, B., Renault, L., Gómara, S., Sebastian, K., López, S. L., … others. (2016). Operational SOCIB forecasting system and multi-platform validation in the Western Mediterranean. *J. Oper. Oceanogr*, *9*, 9231.

Kendall, M. S., Poti, M., & Karnauskas, K. B. (2016). Climate change and larval transport in the ocean: Fractional effects from physical and physiological factors. *Global Change Biology*, *22*(4), 1532–1547. doi: 10.1111/gcb.13159

Kinlan, B. P., Gaines, S. D., & Lester, S. E. (2005). Propagule dispersal and the scales of marine community process. *Diversity and Distributions*, *11*(2), 139–148. doi: 10.1111/j.1366-9516.2005.00158.x

Kleypas, J. A., Thompson, D. M., Castruccio, F. S., Curchitser, E. N., Pinsky, M., & Watson, J. R. (2016). Larval connectivity across temperature gradients and its potential

effect on heat tolerance in coral populations. *Global Change Biology*, *22*(11), 3539–3549. doi: 10.1111/gcb.13347

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

Laurel, B. J., Basilio, A. J., Danley, C., Ryer, C. H., & Spencer, M. (2015). Substrate preference and delayed settlement in northern rock sole larvae *Lepidopsetta polyxystra* . *Marine Ecology Progress Series*, *519*, 183–193. doi: 10.3354/meps11090

Leis, J. M. (2006). Are Larvae of Demersal Fishes Plankton or Nekton? *Advances in Marine Biology*, *51*, 57–141. doi: 10.1016/S0065-2881(06)51002-8

Luiz, O. J., Allen, A. P., Robertson, D. R., Floeter, S. R., Kulbicki, M., Vigliola, L., … Madin, J. S. (2013). Adult and larval traits as determinants of geographic range size among tropical reef fishes. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(41), 16498–16502. doi: 10.1073/pnas.1304074110

Macpherson, E., Biagi, F., Francour, P., García-Rubies, A., Harmelin, J., Harmelin-Vivien, M., … Tunesi, L. (1997). Mortality of juvenile fishes of the genus *Diplodus* in protected and unprotected areas in the western Mediterranean Sea. *Marine Ecology Progress Series*, *160*, 135–147. doi: 10.3354/meps160135

McLeod, I. M., McCormick, M. I., Munday, P. L., Clark, T. D., Wenger, A. S., Brooker, R. M., … Jones, G. P. (2015). Latitudinal variation in larval development of coral reef fishes: Implications of a warming ocean. *Marine Ecology Progress Series*, *521*, 129–141. doi: 10.3354/meps11136

McLeod, I. M., Rummer, J. L., Clark, T. D., Jones, G. P., McCormick, M. I., Wenger, A. S., & Munday, P. L. (2013). Climate change and the performance of larval coral reef fishes: The interaction between temperature and food availability. *Conservation Physiology*, *1*(1), 1–12. doi: 10.1093/conphys/cot024

Meekan, M. G., Vigliola, L., Hansen, A., Doherty, P. J., Halford, A., & Carleton, J. H. (2006). Bigger is better: Size-selective mortality throughout the life history of a fast-growing clupeid, Spratelloides gracilis. *Marine Ecology Progress Series*, *317*(Anderson 1988), 237–244. doi: 10.3354/meps317237

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J., … van Vuuren, D. P. P. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1), 213–241. doi: 10.1007/s10584-011-0156-z

Moullec, F., Barrier, N., Drira, S., Guilhaumon, F., Marsaleix, P., Somot, S., … Shin, Y. J. (2019). An end-to-end model reveals losers and winners in a warming Mediterranean Sea. *Frontiers in Marine Science*, *6*(JUN), 1–19. doi: 10.3389/fmars.2019.00345

Munday, P. L., Leis, J. M., Lough, J. M., Paris, C. B., Kingsford, M. J., Berumen, M. L., & Lambrechts, J. (2009). Climate change and coral reef connectivity. *Coral Reefs*, *28*(2), 379–395. doi: 10.1007/s00338-008-0461-9

O'Connor, M. I., Bruno, J. F., Gaines, S. D., Halpern, B. S., Lester, S. E., Kinlan, B. P., & Weiss, J. M. (2007). Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(4), 1266–1271. doi:

10.1073/pnas.0603422104

Occhipinti-Ambrogi, A. (2007). Global change and marine communities: Alien species and climate change. *Marine Pollution Bulletin*, *55*(7–9), 342–352. doi: 10.1016/j.marpolbul.2006.11.014

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., … others. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Ipcc.

Pannella, G. (1971). Fish otoliths: daily growth layers and periodical patterns. *Science*, *173*(4002), 1124–1127.

Pascual, M., Rives, B., Schunter, C., & Macpherson, E. (2017). Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PLoS ONE*, *12*(5), 1–20. doi: 10.1371/journal.pone.0176419

Paxian, A., Hertig, E., Seubert, S., Vogt, G., Jacobeit, J., & Paeth, H. (2015). Present-day and future mediterranean precipitation extremes assessed by different statistical approaches. *Climate Dynamics*, *44*(3–4), 845–860. doi: 10.1007/s00382-014-2428-6

Pickett, S. T. A. (1989). Space-for-time substitution as an alternative to long-term studies. In *Long-term studies in ecology* (pp. 110–135). Springer.

Prado, P., Tomas, F., Alcoverro, T., & Romero, J. (2007). Extensive direct measurements of *Posidonia oceanica* defoliation confirm the importance of herbivory in temperate seagrass meadows. *Marine Ecology Progress Series*, *340*, 63–71. doi: 10.3354/meps340063

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/

Raventos, N., & Macpherson, E. (2001). Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Marine Biology*, *138*(6), 1115–1120. doi: 10.1007/s002270000535

Robitzch, V. S. N., Lozano-Cortés, D., Kandler, N. M., Salas, E., & Berumen, M. L. (2016). Productivity and sea surface temperature are correlated with the pelagic larval duration of damselfishes in the Red Sea. *Marine Pollution Bulletin*, *105*(2), 566–574. doi: 10.1016/j.marpolbul.2015.11.045

Rodgers, G. M., Downing, B., & Morrell, L. J. (2015). Prey body size mediates the predation risk associated with being "odd." *Behavioral Ecology*, *26*(1), 242–246. doi: 10.1093/beheco/aru185

Rossi, V., Ser-Giacomi, E., Lõpez, C., & Hernández-García, E. (2014). Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves. *Geophysical Research Letters*, *41*(8), 2883–2891. doi: 10.1002/2014GL059540

Sala, E., Ballesteros, E., Dendrinos, P., Di Franco, A., Ferretti, F., Foley, D., … Zabala, M. (2012). The structure of mediterranean rocky reef ecosystems across environmental and human gradients, and conservation implications. *PLoS ONE*, *7*(2). doi: 10.1371/journal.pone.0032742

Schunter, C., Carreras-Carbonell, J., Macpherson, E., Tintoré, J., Vidal-Vijande, E., Pascual, A., … Pascual, M. (2011). Matching genetics with oceanography:

Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, *20*(24), 5167–5181. doi: 10.1111/j.1365-294X.2011.05355.x

Searcy, S. P., & Sponaugle, S. (2001). Selective mortality during the larval-juvenile transition in two coral reef fishes. *Ecology*, *82*(9), 2452–2470. doi: 10.1890/0012-9658(2001)082[2452:SMDTLJ]2.0.CO;2

Selkoe, K. A., & Toonen, R. J. (2011). Marine connectivity: A new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, *436*, 291–305. doi: 10.3354/meps09238

Shanks, A. L. (2009). Pelagic Larval Duration and Dispersal Distance Revisited. *The Biological Bulletin*, *216*(3), 373–385. doi: 10.1086/bblv216n3p373

Shima, J. S., Osenberg, C. W., Alonzo, S. H., Noonburg, E. G., Mitterwallner, P., & Swearer, S. E. (2020). Reproductive phenology across the lunar cycle: parental decisions, offspring responses, and consequences for reef fish. *Ecology*, e03086.

Shima, J. S., & Swearer, S. E. (2016). Evidence and population consequences of shared larval dispersal histories in a marine fish. *Ecology*, *97*(1), 25–31. doi: 10.1890/14-2298.1

Sponaugle, S., Grorud-Colvert, K., & Pinkard, D. (2006). Temperature-mediated variation in early life history traits and recruitment success of the coral reef fish Thalassoma bifasciatum in the Florida Keys. *Marine Ecology Progress Series*, *308*, 1–15. doi: 10.3354/meps308001

The GIMP team. (n.d.). *GNU Image Manipulation Program (GIMP)*. Retrieved from www.gimp.org

Tintoré, J., Vizoso, G., Casas, B., Heslop, E., Pascual, A., Orfila, A., … Manriquez, M. (2013). SOCIB: The Balearic Islands Coastal Ocean Observing and Forecasting System Responding to Science, Technology and Society Needs. *Marine Technology Society Journal*, *47*(1), 101–117. doi: 10.4031/MTSJ.47.1.10

Tonani, M., Pinardi, N., Fratianni, C., Pistoia, J., Dobricic, S., Pensieri, S., … Nittis, K. (2009). Mediterranean Forecasting System: Forecast and analysis assessment through skill scores. *Ocean Science*, *5*(4), 649–660. doi: 10.5194/os-5-649-2009

Wernberg, T., Smale, D. A., & Thomsen, M. S. (2012). A decade of climate change experiments on marine organisms: procedures, patterns and problems. *Global Change Biology*, *18*(5), 1491–1498.

White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., & Toonen, R. J. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1688), 1685–1694. doi: 10.1098/rspb.2009.2214

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Retrieved from http://ggplot2.org

Wilson, D. T., & McCormick, M. I. (1997). Spatial and temporal validation of settlement-marks in the otoliths of tropical reef fishes. *Marine Ecology Progress Series*, *153*(1–3), 259–271. doi: 10.3354/meps153259

Supplementary material

Table S1: Detailed results of the linear mixed effects model of the influence of environmental variables and species life history traits on pre-settlement growth rate (a), settlement size (b) and pelagic larval duration (PLD) (c).

a

Larval Growth Rate	Fixed effects	Estimate	Std. Error	df	t value	Pr (>|t|)	Stand. Estimate
	Intercept	-1.01	0.1	334.89	-9.96	**< 0.001**	**-1.01**
	SST	0.06	0.01	1466.18	10.76	**< 0.001**	**1.25**
	Season	1.16	0.19	936.52	6.28	**< 0.001**	**1.16**
	ChlA	-0.27	0.18	1460.2	-1.51	0.131	-0.08
	SST: Season	-0.06	0.01	1370.62	-7.11	**< 0.001**	**-1.31**
	SST:ChlA	0.02	0.01	1459.26	1.5	0.134	0.1
	Season: ChlA	-2.39	0.58	1270.73	-4.14	**< 0.001**	**-0.72**
	SST:Season:ChlA	0.09	0.03	1261.79	3.24	**0.001**	**0.56**
		R2 Fixed effects	0.344				
	Random effects	npar	logLik	AIC	LRT	Df	Pr (>Chisq)
	(Species)	10	1228.7	-2437.3	120.4	1	**< 0.001**
	(Locality)	10	1278.1	-2536.1	21.6	1	**< 0.001**
		R2 complete model	0.513				

b

Settlement size	Fixed effects	Estimate	Std. Error	df	t value	Pr (>|t|)	Stand. Estimate
	Intercept	0.71	0.11	1428	6.7	**< 0.001**	**0.71**
	SST	0.02	0.01	1472	2.42	**0.016**	**0.31**
	Season	0.9	0.2	1283	4.51	**< 0.001**	**0.90**
	ChlA	-0.52	0.2	1468	-2.59	**0.01**	**-0.15**
	SST: Season	-0.04	0.01	1336	-4.27	**< 0.001**	**-0.87**
	SST:ChlA	0.04	0.01	1467	2.97	**0.003**	**0.23**
	Season: ChlA	-1.31	0.62	1265	-2.09	**0.036**	**-0.39**
	SST:Season:ChlA	0.04	0.03	1268	1.44	0.151	0.27
		R2 Fixed effects	0.064				
	Random effects	npar	logLik	AIC	LRT	Df	Pr (>Chisq)
	(Species)	10	1125.6	-2231.3	0	1	1
	(Locality)	10	1103.4	-2186.9	44.4	1	**< 0.001**
		R2 complete model	0.128				

Table S1(cont): Detailed results of the linear mixed effects model of the influence of environmental variables and species life history traits on pre-settlement growth rate (a), settlement size (b) and pelagic larval duration (PLD) (c).

c

	Fixed effects	Estimate	Std. Error	df	t value	Pr (>|t|)	Stand. Estimate
PLD	Intercept	0.69	0.1	314.5	6.65	**< 0.001**	**0.69**
	SST	-0.04	0.01	1464	-7.62	**< 0.001**	**-0.91**
	Season	-0.22	0.19	923.5	-1.17	0.241	-0.22
	ChlA	-0.23	0.18	1463	-1.29	0.199	-0.07
	SST: Season	0.02	0.01	1386	2.12	**0.034**	**0.4**
	SST:ChlA	0.02	0.01	1462	1.59	0.111	0.11
	Season: ChlA	1.09	0.59	1296	1.85	0.065	0.32
	SST:Season:ChlA	-0.05	0.03	1288	-1.66	0.097	-0.29
			R2 Fixed effects	0.249			
	Random effects	npar	logLik	AIC	LRT	Df	Pr (>Chisq)
	(Species)	10	1202.1	-2384.2	109.2	1	**< 0.001**
	(Locality)	10	1233	-2446.1	47.3	1	**< 0.001**
	R2 complete model	0.44					

Table S2: Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Coris julis

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.243	0.177	103.865	-1.373	0.173	
	SST	0.011	0.008	107.709	1.359	0.177	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	182.750	-359.510	15.535	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.007	0.007	1.000	107.709	1.846	0.177
PLD	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.788	0.328	138.657	2.405	**0.018**	
	SST	-0.035	0.015	143.222	-2.439	**0.016**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	85.295	-164.590	38.974	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.072	0.072	1.000	143.222	5.949	**0.016**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.516	0.242	137.017	2.136	**0.035**	
	SST	-0.023	0.011	141.810	-2.158	**0.033**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	133.420	-260.830	32.397	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.031	0.031	1.000	141.810	4.656	**0.033**

Table S2 (cont): Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Chromis chromis

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.793	0.171	60.692	-4.637	**< 0.001**	
	SST	0.033	0.007	64.077	4.635	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	138.740	-271.480	13.805	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.145	0.145	1.000	64.077	21.484	**< 0.001**
PLD	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.779	0.156	29.660	5.006	**< 0.001**	
	SST	-0.033	0.007	30.716	-5.053	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	149.830	-293.660	3.995	1.000	**0.046**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.160	0.160	1.000	30.716	25.528	**< 0.001**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.082	0.176	107.490	0.466	0.642	
	SST	-0.004	0.007	119.164	-0.496	0.621	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	131.700	-257.410	35.792	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.002	0.002	1.000	119.164	0.246	0.621

Table S2 (cont): Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Diplodus puntazzo

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-1.212	0.109	14.327	-11.090	**< 0.001**	
	SST	0.057	0.005	14.346	11.060	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	204.820	-403.630	0.000	1.000	0.988
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.892	0.892	1.000	14.346	122.256	**< 0.001**
PLD	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.874	0.143	52.939	6.108	**< 0.001**	
	SST	-0.042	0.007	54.036	-6.164	**< 0.001**	
	Random effects	npar	logLik	AIC		LRT	Df Pr(>Chisq)
	(Locality)	3.000	200.090	-394.180	6.632	1.000	**0.010**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.272	0.272	1.000	54.036	37.991	**< 0.001**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.365	0.178	30.827	-2.050	**0.049**	
	SST	0.017	0.008	31.143	2.014	0.053	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	143.030	-280.050	2.149	1.000	0.143
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.053	0.053	1.000	31.143	4.056	0.053

Table S2 (cont): Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Diplodus sargus

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-1.318	0.121	82.452	-10.900	**< 0.001**	
	SST	0.075	0.007	114.266	11.080	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	112.730	-219.460	28.677	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	1.444	1.444	1.000	114.266	122.847	**< 0.001**
PLD	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.879	0.094	48.058	9.301	**< 0.001**	
	SST	-0.051	0.005	56.423	-9.536	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	152.160	-298.320	12.207	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.759	0.759	1.000	56.423	90.941	**< 0.001**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.388	0.101	77.646	-3.827	**< 0.001**	
	SST	0.022	0.006	100.208	3.846	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	139.190	-272.380	30.820	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.127	0.127	1.000	100.208	14.790	**< 0.001**

Table S2 (cont): Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Diplodus vulgaris

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-1.333	0.079	128.265	-16.900	**< 0.001**	
	SST	0.083	0.005	140.553	16.930	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	172.860	-339.720	5.689	1.000	**0.017**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	1.970	1.970	1.000	140.553	286.530	**< 0.001**
PLD		Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.626	0.086	149.891	7.252	**< 0.001**	
	SST	-0.040	0.005	170.490	-7.517	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	154.380	-302.760	24.094	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.421	0.421	1.000	170.490	56.502	**< 0.001**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.708	0.098	149.156	-7.227	**< 0.001**	
	SST	0.044	0.006	167.800	7.278	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	134.460	-262.920	17.871	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.519	0.519	1.000	167.800	52.974	**< 0.001**

Table S2 (cont): Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Oblada melanura

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.805	0.202	87.972	-3.994	**< 0.001**	
	SST	0.036	0.009	90.997	3.988	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	207.780	-409.550	30.656	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.104	0.104	1.000	90.997	15.907	**< 0.001**
PLD	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	1.304	0.130	214.000	10.000	**< 0.001**	
	SST	-0.058	0.006	214.000	-10.060	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	209.440	-412.890	0.000	1.000	1.000
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.795	0.795	1.000	214.000	101.120	**< 0.001**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.493	0.255	49.812	1.934	0.059	.
	SST	-0.022	0.011	50.723	-1.964	0.055	.
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	152.040	-298.080	14.798	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.046	0.046	1.000	50.723	3.856	0.055

Table S2 (cont): Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Sarpa salpa

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.117	0.286	28.414	-0.409	0.685	
	SST	0.005	0.013	28.722	0.374	0.711	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	83.373	-160.750	5.673	1.000	**0.017**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.001	0.001	1.000	28.722	0.140	0.711
PLD	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.515	0.223	19.289	2.312	**0.032**	
	SST	-0.024	0.010	19.064	-2.337	**0.031**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	95.338	-184.680	2.455	1.000	0.117
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.048	0.048	1.000	19.064	5.459	**0.031**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.079	0.217	109.000	-0.366	0.715	
	SST	0.003	0.010	109.000	0.335	0.738	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	73.737	-141.470	0.000	1.000	1.000
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.001553199	0.001553199	1	109	0.1121109	0.7383971

Table S2 (cont): Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Symphodus ocellatus

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.509	0.164	177.550	-3.106	**0.002**	
	SST	0.022	0.007	190.197	3.078	**0.002**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	120.380	-234.760	18.553	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.142	0.142	1.000	190.197	9.474	**0.002**
PLD	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.629	0.144	170.319	4.354	**< 0.001**	
	SST	-0.028	0.006	183.660	-4.467	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	148.680	-291.360	12.412	1.000	**< 0.001**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.236	0.236	1.000	183.660	19.956	**< 0.001**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.120	0.155	183.978	0.772	0.441	
	SST	-0.006	0.007	197.200	-0.877	0.381	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	133.280	-260.570	19.290	1.000	**0.000**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.01010767	0.01010767	1	197.2005	0.7698787	0.3813223

Table S2 (cont): Detailed results of the species-specific linear mixed effects models of the influence of temperature on log-transformed Early life traits: Larval Growth Rate, Settlement size and pelagic larval duration (PLD). Bold p-values indicate significance at α=0.05.

Symphodus tinca

Larval Growth Rate	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	-0.847	0.125	5.455	-6.759	**0.001**	
	SST	0.043	0.006	5.734	6.859	**0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	90.498	-175.000	2.942	1.000	0.086
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.343	0.343	1.000	5.734	47.039	**0.001**
PLD	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.994	0.172	13.424	5.789	**< 0.001**	
	SST	-0.051	0.008	14.694	-6.082	**< 0.001**	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	75.194	-144.390	9.900	1.000	**0.002**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.339	0.339	1.000	14.694	36.993	**< 0.001**
Settlement size	**Fixed effects**	Estimate	Std. Error	df	t value	Pr(>|t|)	
	(Intercept)	0.026	0.203	34.169	0.126	0.900	
	SST	-0.002	0.010	44.286	-0.187	0.853	
	Random effects	npar	logLik	AIC	LRT	Df	Pr(>Chisq)
	(Locality)	3.000	63.625	-121.250	27.706	1.000	**0.000**
	ANOVA table	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
	SST	0.000328856	0.000328856	1	44.28604	0.03481302	0.8528401

Table S3 Present minimum, mean and maximum sea temperatures and estimated increases (in ºC) in the year 2100 in the West Mediterranean Sea (between 5.8ºW and 9ºE), under three different greenhouse gas concentration Representative Concentration Pathway (RCPs): RCP 2.6, RCP 4.5 and RCP 8.5. Estimations based on data from Bio-ORACLE database (Assis et al., 2018; Tyberghein et al., 2012).

	Present value	2100 estimated SST increase		
		RCP 2.6	RCP 4.5	RCP 8.5
Minimum temperature	10.43	1.01	1.41	2.80
Mean temperature	19.18	0.94	1.37	3.55
Maximum temperature	27.56	1.61	2.13	4.97

Capítulo 1.2

Impact of individual early life traits in larval dispersal: a multispecies approach using backtracking models.

Héctor Torrado[1, 2] *, Baptiste Mourre[3], Núria Raventós[1], Carlos Carreras[2], Joaquín Tintoré[3,4], Marta Pascual[2+], Enrique Macpherson[1+]

[1]Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Blanes, Girona, Spain.
[2]Department de Genètica, Microbiologia i Estadística and IRBio, Universitat de Barcelona, Barcelona, Spain.
[3]SOCIB, Balearic Islands Coastal Observing and Forecasting System. Palma, Illes Balears, Spain.
[4]Institut Mediterrani d'Estudis Avançats (IMEDEA-UIB-CSIC), Esporles, Illes Balears, Spain.
+Both authors contributed equally as senior researchers and should be considered to be at the same position.

Abstract

Dispersal is a key process shaping species population structure. In demersal marine fishes, which usually have sedentary adult phases, dispersion relies on drifting larval stages. However, the dynamics and seasonal variability of seawater masses can greatly determine the connectivity patterns of these species along the same geographic gradient. For this reason, detailed information on the release moment of larvae is needed to obtain accurate patterns of connectivity. In this study, we performed backtracking Lagrangian particle dispersion simulations, with individual-based early life traits data, obtained from otolith reading for 1,413 juveniles of nine fish species belonging to three families (Sparidae, Pomacentridae and Labridae). For each species, individuals had been sampled from four to seven localities in the western Mediterranean Sea between the Gulf of Lion to the Gibraltar Strait. These nine species reproduce in different seasons of the year and their pelagic larval duration (PLD) range from 7 to 43 days. We identified three hydrodynamic units separated by oceanographic discontinuities (Balearic Sea, West Algerian Basin and Alboran Sea) with low settler's exchange according to our

simulations, independently of the PLD and reproductive season of the species. Hatching date and PLD showed significant effects on larval dispersal distance and orientation, both at the intraspecific and interspecific levels, highlighting the importance of these variables in determining the geographic origin of individuals. Our multispecies modelling approach adds a step forward for an accurate description of larval dispersion and recruitment, key to understand population resilience and define management strategies.

Keywords: Fish larvae, Oceanographic models, Lagrangian particle dispersion, Individual-based model, Early life traits, Otolith reading.

Introduction

In benthic and pelagic marine habitats hydrodynamic processes, e.g. temperature, productivity gradients and turbulent oceanographic features, interact with biological processes affecting species distribution and communities (Cowen, 2002). Physical processes are usually highly variable and their role in generating and maintaining patterns in community structure are essential in marine ecology studies (Shanks & Brink, 2005; White et al., 2019). There is a close relationship between temporal and spatial scales of this physical variability and the apparent high levels of asymmetry and stochasticity in biological processes (Ayata, Lazure, & Thiébaut, 2010).

Most marine organisms, including most benthic fishes, have low-dispersive adult phases and high-dispersive pelagic larval stages. This dual life history makes early life processes especially pivotal to marine ecology, influencing not only dispersal but also settlement rates, with the resulting effects on community structure (Leis, 1991). In the plankton, the duration of the larval pelagic stage (PLD) determines the length of time that larvae are subject to movement by currents, winds or eddies, and other physical processes, influencing dispersal distances (Gaines, Gaylord, Gerber, Hastings, & Kinlan, 2007; Kinlan, Gaines, & Lester, 2005; Shanks, 2009). PLD is a key biological factor for larval dispersion since longer PLDs potentially allow larvae to

travel larger distances (K. A. Selkoe & Toonen, 2011; Shanks, 2009; Treml et al., 2012). PLD can also be influenced by the spawning periodicity or seasonality, potentially affecting dispersal and connectivity (Kough & Paris, 2015), suggesting that hatching time can also be important in dispersal patterns. The dispersive planktonic larval phase(s) is considered as a "black box" in part due to our limited understanding of the relationships between larvae and their environment (Cowen & Sponaugle, 2009; Paris & Cowen, 2004).

Numerous physical mechanisms, e.g. currents, eddies, waves, have been associated with the transport of larvae (Banks et al., 2007; Kimberly A. Selkoe et al., 2010), emphasising the importance of their temporal and spatial variability on the larval transport and settlement dynamics (S. Sponaugle & Cowen, 1996). On the other hand, the release of larvae in the plankton, e.g. hatching period, and their settlement time, show also a large temporal and spatial variability. The coupling of these physical and biological processes are difficult to identify, adding an element of stochasticity to these events (S. Sponaugle, Lee, Kourafalou, & Pinkard, 2005). Therefore, the difficulty to identify these fundamental physical-biological interactions, and how they change along space and time, remains scarcely studied, avoiding a more complete knowledge of dispersal processes.

The displacements produced in the course of this pelagic period have been frequently modelled to describe the potential dispersal capabilities of species using Lagrangian-based larval dispersal modelling (Andrello et al., 2017; Calò, Lett, Mourre, Pérez-Ruzafa, & García-Charton, 2018; Rossi, Ser-Giacomi, Lõpez, & Hernández-García, 2014; Schunter, Carreras-Carbonell, Macpherson, et al., 2011; Treml et al., 2012). Classic approaches use forward simulations, releasing particles from potential source areas and following the current fields to their settlement areas. This methodology has been used to study propagule dispersal and potential impact of climate change (Andrello, Mouillot, Somot, Thuiller, & Manel, 2015) or spatial and temporal variability of larval dispersion (Barbut et al., 2019; Di Franco et al., 2012; Ospina-Alvarez, Catalán, Bernal, Roos, & Palomera, 2015). Nevertheless, when the study focuses on particular sampling areas, an alternative approach consists in backtracking the particles from the

settlement areas by running the oceanographic model back in time in order to find their potential origin or source areas. Backtracking approximations have been used in larval dispersal modelling at population level to evaluate the effect of oceanographic structures in the area (Holliday et al., 2012), for potential hatch area identification (Calò et al., 2018; Christensen et al., 2007; Fraker et al., 2015), recruitment predictions for fisheries management (Allain, Petitgas, Lazure, & Grellier, 2007) and to reconstruct the environmental history associated to larval growth rate (Payne et al., 2013; Ross et al., 2012).

Recently, these models have included some aspects related to larval behaviour. The natatory capacities are reduced at the initial pelagic phases, and consequently their transport is mainly passive and driven by the ocean currents in the area (Leis, 2007). However, larvae experience an ontogenetic improvement of swimming capabilities and can modify their trajectories at later pelagic phases through vertical migrations (Leis, 2007; Paris & Cowen, 2004) and horizontal orientated swimming (Faillettaz, Durand, Paris, Koubbi, & Irisson, 2018; Staaterman, Paris, & Helgers, 2012). The mechanisms allowing larval orientation in open waters are badly known, although they could have a notable impact on their final destination, and some authors have suggested the use of odour, sun, or magnetic fields (Bottesch et al., 2016; Faillettaz, Blandin, Paris, Koubbi, & Irisson, 2015; Leis, Paris, Irisson, Yerman, & Siebeck, 2014; Mouritsen, Atema, Kingsford, & Gerlach, 2013; O'Connor & Muheim, 2017; Paris et al., 2013).

In fish larval dispersal studies, the high temporal and spatial variability of the oceanographic currents makes the accurate estimation of the dates and duration of the larval periods especially important. Otolith (ear bone) microstructure analysis provides a useful tool to obtain this information at individual level. Otoliths of most temperate fishes show daily growth rings and lay a clear settlement band when the individuals change from the pelagic to the benthic stage (Raventos & Macpherson, 2001, 2005; Wilson & McCormick, 1999). Using otolith reading, we can obtain the early life history of each settler, including the day of hatching and settlement, and the PLD. Therefore, the combination of otolith-inferred accurate early-life traits and high-

resolution models of oceanographic currents represents an interesting strategy to analyse the dispersal patterns of individual fish larvae.

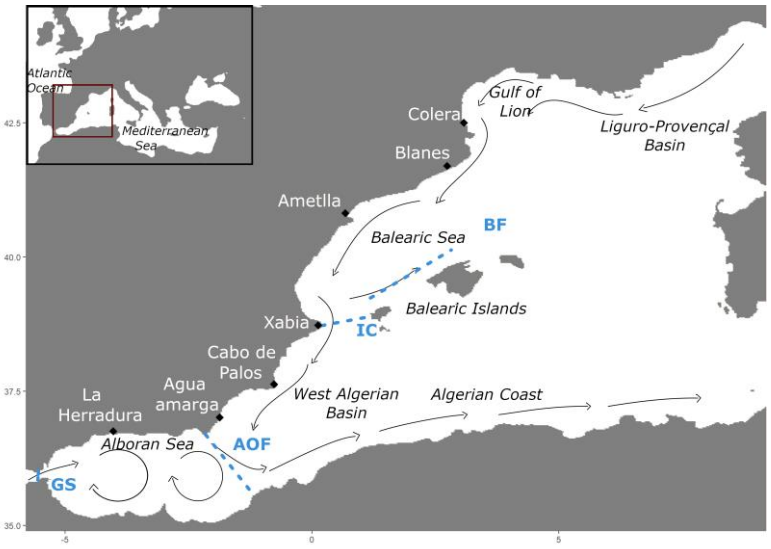


Figure 1: Map of the sampling locations and main schematic oceanographic currents in the Western Mediterranean Sea. The red square in the map of Europe correspond to the enlarged map of the analysed area. Diamonds: sampling localities; Dashed blue lines indicate oceanographic barriers; BF: Balearic front; IC: Ibiza channel; AOF: Almeria-Oran front; GS: Gibraltar Strait. Currents are represented as thin black lines with arrows identifying its main direction (Millot, 1999)

The Western Mediterranean Sea provides a good system to evaluate the effect of oceanographic processes on larval dispersal (Figure 1). It is connected with the Atlantic Ocean through the Strait of Gibraltar, where there is an inflow of surface Atlantic water and outflow of deeper Mediterranean water (Millot, 1999; Millot & Taupier-Letage, 2005). The circulation pattern and topography along the southern and eastern coasts of the Iberian Peninsula originate three main oceanographic discontinuities in the study area (Figure 1): the Almeria-Oran Front (AOF), the Ibiza Channel (IC) and the Balearic Front (BF). The AOF is a large scale density front located 400 km east of the Gibraltar Strait (GS) and formed by the convergence of the main jet of incoming Atlantic water and waters having recirculated in the Mediterranean Sea (Tintoré et al., 1988). The strength of the AOF varies with the changes affecting the Eastern Alboran Gyre (Renault, Oguz, Pascual, Vizoso, & Tintore, 2012). East of the AOF,

the main current carries Atlantic water eastward along the African coast. Part of this water continues towards the eastern basin through the Sicily Channel, while the other part flows through the Tyrrhenian Sea (Astraldi et al., 1999; Millot & Taupier-Letage, 2005). The circulation in the Western basin is mainly cyclonic, with the Northern Current flowing south-westward along the French and Spanish coasts and bifurcating when reaching the south of the Balearic Sea (Garcia Lafuente, Lopez Jurado, Cano Lucaya, Vargas Yanez, & Aguiar Garcia, 1995; Salat, 1996). The IC, 80 km wide and 800 m depth, corresponding to the passage intersecting the Balearic topographic ridge between Ibiza and the Iberian Peninsula at Cape La Nao, is a key choke point with important exchanges between waters of recent Atlantic origin in the south and saltier waters of Mediterranean characteristics in the northern side (Heslop et al., 2012; Pinot, López-Jurado, & Riera, 2002). Finally, the part of the flow directed north-eastward along the northern Balearic shelf edge forms a well-defined density front, the BF, present in the upper 200 m (Figure 1). The dynamic behaviour of these oceanographic discontinuities generate significant intra- and inter-annual variability (Tintoré et al., 2019) that can affect larval dispersal patterns (Su Sponaugle & Cowen, 1996).

In the present work, and using a multispecies approach, we evaluated how differing hatching dates and pelagic larval durations interact with variable oceanographic features to influence larval dispersal patterns throughout the Western Mediterranean. We modelled individual-based potential larval movement of 1,413 settlers of nine common coastal fish species along the western Mediterranean Sea, reproducing in different seasons and with different PLD. We used individual-based information on early life traits (day of hatching and settlement, and PLD) from otoliths readings (Raventos et al., Submitted). We backtracked each settler running Lagrangian trajectories back in time using currents from a high-resolution model of the Western Mediterranean Sea (Juza et al., 2016; Mourre et al., 2018), in order to find their potential origin and to evaluate the effect of the oceanographic variability. Our specific objectives were: a) to determine the effect of the currents and oceanographic structures and their seasonal variability on dispersal patterns and b) to assess the influence of interspecific and intraspecific variation of hatching date and PLD on

dispersal potential of the nine species. We hypothesized that the dispersal patterns of fishes would be influenced by variations in hatching date and PLD, between species as well as within species, coupled with changes in oceanographic conditions, and thus should be considered for accurately modelling dispersal.

Materials and Methods

Species data

In the present work we modelled the dispersal capacity of nine coastal co-occurring fish species from three different families along the western Mediterranean Sea: (1) Sparidae, the common two-banded seabream *Diplodus vulgaris*, the white seabream *D. sargus*, the sharpsnout seabream *D. puntazzo*, the salema *Sarpa salpa* and the saddled seabream *Oblada melanura*; (2) Pomacentridae, the damselfish *Chromis chromis*; (3) Labridae, the Mediterranean rainbow wrasse *Coris julis*, the ocellated wrasse *Symphodus ocellatus*, and the East Atlantic peacock wrasse *S. tinca*.

We based our study on the individualized early life traits information of 1,413 settlers from these nine species (Table S1) obtained in the study of (Raventos et al., Submitted): *Diplodus vulgaris* (n=174), *D. sargus* (n=175), *D. puntazzo* (n=206), *Sarpa salpa* (n=72), *Oblada melanura* (n=216), *Chromis chromis* (n=146), *Coris julis* (n=150), *Symphodus ocellatus* (n=177) and *S. tinca* (n=97). Settlers of all sizes were collected at the end of the settlement period of each species to ensure the representation of the entire hatching period. To avoid interannual variation, we restricted our analyses to individuals born during the same reproductive period across all sites for a given species. For each individual, the age at sampling was determined by counting the total number of bands in their otoliths, from core to margin. Their individual PLD and settlement day was determined by counting the number of daily rings visible between the core and the settlement mark, and ages were used to calculate their day of birth. The settlers of these nine species were sampled at seven locations of the Western Mediterranean Sea, from Colera in the Gulf of Lion (42°40'N, 3°11'E) to La Herradura (36°44'N, 3°44'W) in the Alboran Sea (Figure 1, Table S1). Some species were not

collected in all locations (e.g. *Sarpa salpa*, 4 locations) due to the lack of settlers' availability during the sampling periods in those localities. The nine selected species have different pelagic larval duration (PLD) and reproduce in different seasons of the year (Figure 2, Table S1).

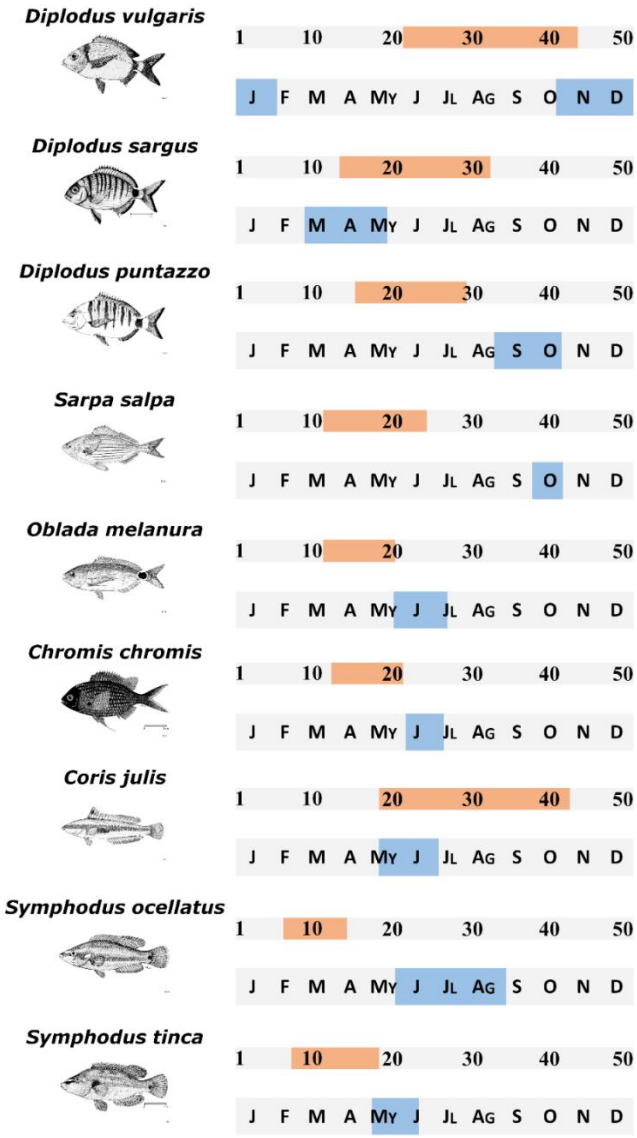


Figure 2: Ranges of Pelagic larval duration (PLD) in days (orange bar) and Hatching season in months (blue bar) for each species obtained from (Raventos et al., Submitted). Species' images obtained from (Bauchot, 1987; Schneider, 1990)

Modelling

For our modelling analyses we used the ocean current fields provided by the Western Mediterranean OPerational forecasting system (WMOP) (Juza et al., 2016; Mourre et al., 2018), developed at the Balearic Islands Coastal Observing and Forecasting System (www.socib.es, Tintoré et al., 2013). WMOP has a spatial coverage from Gibraltar strait to Sardinia Channel (35°N-44.5°N, 6°W-9°E) and a horizontal resolution of around 2 km and 32 sigma-levels in the vertical dimension. Detailed assessments of the WMOP model from the basin to the coastal scales have demonstrated that the model was able to represent the mean large-scale surface circulation and the associated modes of variability with a satisfactory level of realism (Aguiar et al., 2020; Juza et al., 2016; Mourre et al., 2018). Comparisons with satellite altimetry have shown that the model was generating realistic statistics of large eddies and accurate magnitude and spatial variability of eddy kinetic energy. At the coastal scales and based on surface drifter data, the model was found to represent small scale processes such as topographically driven coastal eddies or localized sea breeze effects close to Mallorca Island. High-Frequency (HF) radar data have demonstrated that the model was able to properly represent the position of the Northern Current and the coastal flow intensification in the coastal region off Ebro delta. In the Ibiza Channel, while some overestimation of the northward surface flow was highlighted based on HF radar measurements, the seasonal and interannual variability of the Western Intermediate Water was positively evaluated in the model using underwater glider time series (Juza et al., 2019). Daily updated comparisons with satellite SST, altimetry, ARGO profiles, mooring time series, glider sections and coastal currents from HF radars in 3 different locations (Gibraltar strait, Ibiza Channel and Ebro Delta) are available at https://socib.es/?seccion=modelling&facility=wmedvalidation.

Backtracking individual-based Lagrangian particle dispersion simulations were performed with the WMOP data using the TRACMASS software (Döös & Jönsson, 2013), implemented through the TracPy interface (Thyng & Hetland, 2014). Each individual simulation considered the following data: release point (sampling locality), date (settlement date) and simulation duration (pelagic larval duration) based on each

individual sample and otolith data. Simulations were then performed for different periods of the year in 2014, 2015 and 2017 according to individual data (Table S1, for full otolith data access, see Raventos et al., Submitted). For each individual we simulated the release and movement of 10^3 particles from their settlement site to their potential source origin during their settlement and hatch dates. We added random velocity fluctuations to simulations in order to account for model uncertainties and unresolved processes. The magnitude of the random turbulent velocities was the result of the specific calibration experiments performed in the framework of the development of the TRACMASS algorithm (Döös & Engqvist, 2007; Döös, Rupolo, & Brodeau, 2011), combined with our own experience with trajectory modelling using currents from the 2km resolution WMOP model when compared to real drifters (Mourre et al., 2018). The random coefficients were fixed so as to produce a standard deviation of random velocity fluctuations which reaches 0.8 times the standard deviation of the WMOP model velocities along the trajectories. During the last half of the larval period (first part of the backtracking simulation), we added a velocity vector towards the coastline, based on the larval Critical speed (Ucrit) of each species, representing larval swimming abilities. Ucrit values were obtained from laboratory studies, including the same or closely related species for each family, Sparidae and Pomacentridae (Faillettaz et al., 2018), and Labridae (Leis, Hay, & Gaither, 2011). We considered $\frac{1}{2}$ Ucrit as the mean velocity directed towards the coast when the particle was less than 5 km off the coastline. Therefore, we considered that within this distance larvae are capable of a good directional orientation due to the detection of more precise clues like odour or sound (Kingsford et al., 2002). For grid points more than 5km off the coastline, we assumed a velocity of $\frac{1}{4}$ Ucrit. Reducing velocities at further distances are the result of assuming a minor net displacement towards the coast due to reduced orientation capabilities and considering a bigger importance of individual stochasticity. Larvae of most of the studied species are usually concentrated in the upper 10 m of the water column, with very limited or negligible diel vertical migration (Sabatés & Olivar, 1996). Thus, particles representing fish larvae were simulated at

fixed depths for the full advection period (Table S1) following observed mean larval depth (Olivar et al., 2010; Sabatés & Olivar, 1996).

In order to get insights into the impact of the oceanographic variability and evaluate the extent to which the year of study represents a typical current field, we have analysed the interannual model transport time series across the two oceanographic discontinuities (IC and AOF) separating the three hydrodynamic units present in the area (see results) from 2014 to 2018. The transports were computed from coast to coast and from model surface to bottom, separating northward and southward flows and integrating the product of model cross-section velocities by the surface of 2km-long cells.

Data analyses

We generated a total of 1000 backtracking particle trajectories for each of the 1,413 sampled individuals of the nine species. Since the studied species do not live nor reproduce in open sea, we only kept for posterior analyses those particles originating in coastal waters (less than 2 km to the nearest land point) in the backtracking simulations, using nn2 function in the R package 'RANN' 2.6.1 (Arya, Mount, Kemp, & Jefferis, 2019). We plotted potential origin maps of settlers for each species in R 3.4.4 (R Core Team, 2018) using the following packages: 'ncdf4' 1.16 (Pierce, 2017) for netCDF files reading and processing, 'reshape' 0.8.7 (Hadley Wickham, 2007) and 'dplyr' 0.7.8 (H Wickham, François, Henry, & Müller, 2018) for data reorganization, 'magrittr' 1.5 (Bache & Wickham, 2014) for function construction, 'rgdal' 1.3-1 (Bivand, Keitt, & Rowlingson, 2018) for geoprocessing and 'ggplot2' 3.1.0 (Hadley Wickham, 2016) for plotting. Furthermore, we plotted the mean current velocity of the WMOP for each species across its planktonic period assessed from the otolith data with the same methodology, using a 'viridis' R package scale (Garnier, 2018) for colour scale construction.

For each individual, we calculated the mean dispersal distance and orientation considering the backtracking simulations originating in coastal waters. The angle of each particle was defined in relation to the Northwest in each sampling locality to avoid

that the range of potential orientations within any location included the zero value, and thus this parameter always increases as the direction changes clockwise. In order to evaluate the effect of PLD and hatching date on mean dispersal distances and orientations we performed generalized lineal mixed model (GLMM) tests with the R package 'lme4' 1.1-23 (Bates, Mächler, Bolker, & Walker, 2015). We first built an analysis with all data considering species and locality as random factors (Dispersal distance/Dispersal orientation ~ PLD + Hatching date + Species + Locality). Additionally, we carried out an analysis for each species separately to assess the mean dispersal distance and orientation as a function of PLD and hatching date while defining localities as a random factor (Dispersal distance/Dispersal orientation ~ PLD + Hatching date + Locality).

To assess the effect of oceanographic discontinuities on dispersal patterns we considered three sampled areas and seven source zones for simulated particles. We defined the three sampled areas as the three oceanographic regions separated by the oceanographic discontinuities along the study area (Figure 1): (1) Alboran Sea: from the Gibraltar Strait (GS) to the Almeria-Oran Front (AOF), (2) West Algerian Basin: Iberian Peninsula area from the AOF to the Ibiza Channel (IC), and (3) Balearic Sea: Iberian Peninsula area from IC to Gulf of Lion. On the other hand, we delimited the seven potential source coastal areas (Figure 1): (1) Atlantic Ocean (Atlantic particles arriving through GS), (2) Alboran Sea (from GS to AOF), (3) West Algerian Basin (Iberian peninsula from AOF to IC), (4) Algerian coast (northern Africa from AOF towards the east), (5) Balearic Sea (Iberian peninsula from IC to Gulf of Lion), (6) Balearic Islands and (7) Liguro-Provençal Basin (from the Gulf of Lion towards the northeast). For each species we tested for differences in individual mean dispersal distances between areas using a Kruskal-Wallis test and a Dunn post-hoc test in R. We evaluated with the same test the variability of oceanographic transport across the IC and AOF in different years (2014-2017) during the whole pelagic period for each species (considering Hatching date and PLD according to our data). We plotted in a chord diagram the links between the source and sample areas with the R package 'circlize' 0.4.9 (Gu, Gu, Eils, Schlesner, & Brors, 2014).

Results

Hydrographic setting and backtracking trajectories

We generated a total of 1,413,000 backtracking trajectories from which we kept a total of 410,627 trajectories originating in coastal waters (Figure 3), to be used in further analyses. Individual mean dispersal distance and orientation were calculated with these remaining trajectories, with standard deviations inside individuals around 48.3 km (min 0.3 km, max 370.6 km) for dispersal and 3.5 rad (min 0.4 rad, max 5.9 rad) for orientation. We observed that most settlers had their natal origin in the same oceanographic region where they were born (57.2%-99.9%), indicating a high level of self-recruitment at the regional level, with some differences across species and regions (Figure 4, Table S2). In the Balearic Sea, the simulations indicate that most of the settlers were from the same region (85.3% ± 15.3), but some could also have arrived from the Liguro-Provençal area (14.5% ± 15.0), and a small fraction from the Balearic Islands (0.3% ± 0.6), without any contribution from other source areas. The WMOP maps showed that the Northern Current (Figure 1) exhibited a constant pattern (Figures S1). This current followed the continental slope of the North-western Mediterranean Sea from the Ligurian Sea towards the Balearic Sea and recirculated when it reached the Ibiza Channel. As expected, the proportion of settlers from the Liguro-Provençal area varied across localities and decreased from the northernmost locality (e.g. Colera) to the southernmost one (e.g. Ametlla) (Figure 3). Differences among species were found from settlers originating in the Liguro-Provençal basin with a higher frequency in *D. vulgaris* and *C. chromis* (Figure 4, Table S2). Differences among species were also found for settlers originating in the Balearic Islands as for *D. puntazzo* and *D. vulgaris,* which were able to cross the Balearic Front from the Balearic Islands to the continental coast in higher frequency (Figure 3, Table S2).

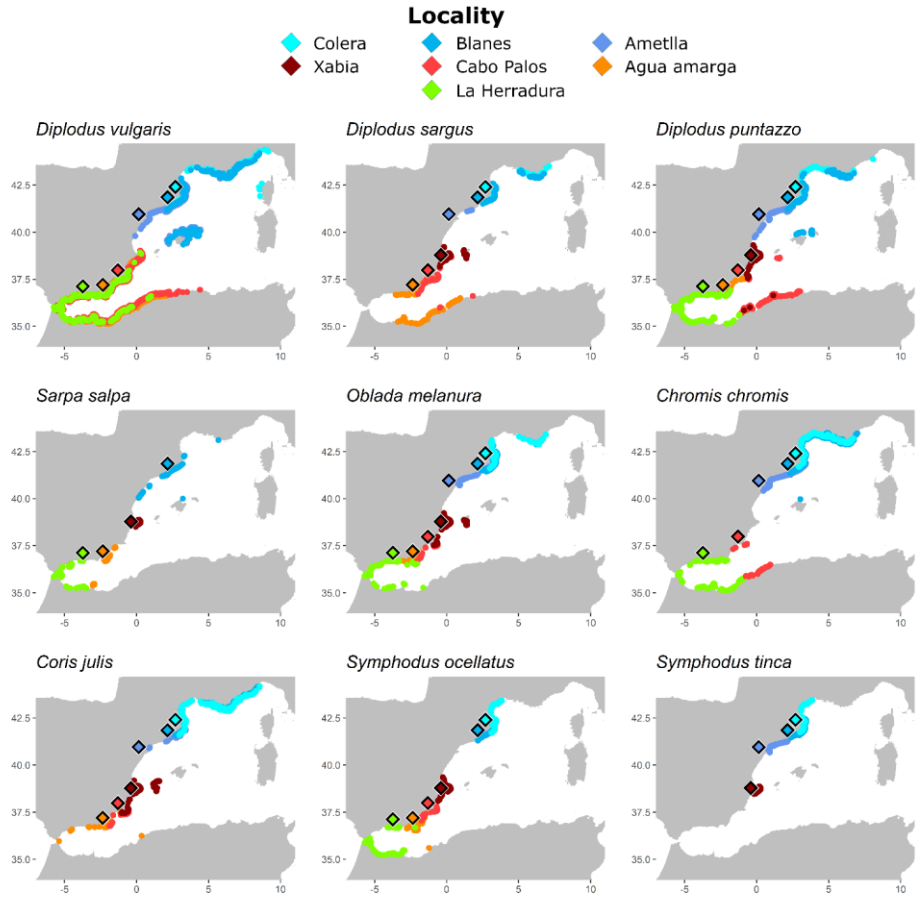


Figure 3: Potential coastal origin (dots) of the settlers captured in the different sampled localities (diamonds) as assessed by backtracking simulations for each species.

The Western Algerian Basin also had a high proportion of self-recruitment (80.1% ± 11.3) but coupled with some incomes from other source areas (Figures 3-4). The transport across the Ibiza channel connected the Balearic Sea and the Western Algerian basin. We found a considerable potential contribution of settlers from the Balearic Sea, crossing the Ibiza Channel (13.9%± 12.6). The higher pass of larvae from the northern to the southern side of the Ibiza channel was detected in *S. tinca* and *D. puntazzo* (Figure 4, Table S2). A small contribution from the Balearic Islands was observed (1.0% ± 1.8), being higher in *O. melanura* and *C. julis* (Figure 4). Moreover, for the three *Diplodus* species and *Chromis chromis* we detected a certain amount of settlers (3.4% ± 3) from the Algerian coast (Figure 4), specially arriving at Cabo de

Palos and Agua amarga sites (Figure 3). Finally, the potential proportion of settlers from the Alboran Sea through the Almeria-Oran Front was small (3.1% ± 5.3). However, Almeria-Oran Front decreased its strength when the eastern gyre faded allowing a transport of settlers in both directions, as it was observed during the larval period of *Diplodus vulgaris* where there is more communication between hydrodynamic units than, for example, in *D. puntazzo*, where the front is more apparent (Figures 4, S1).

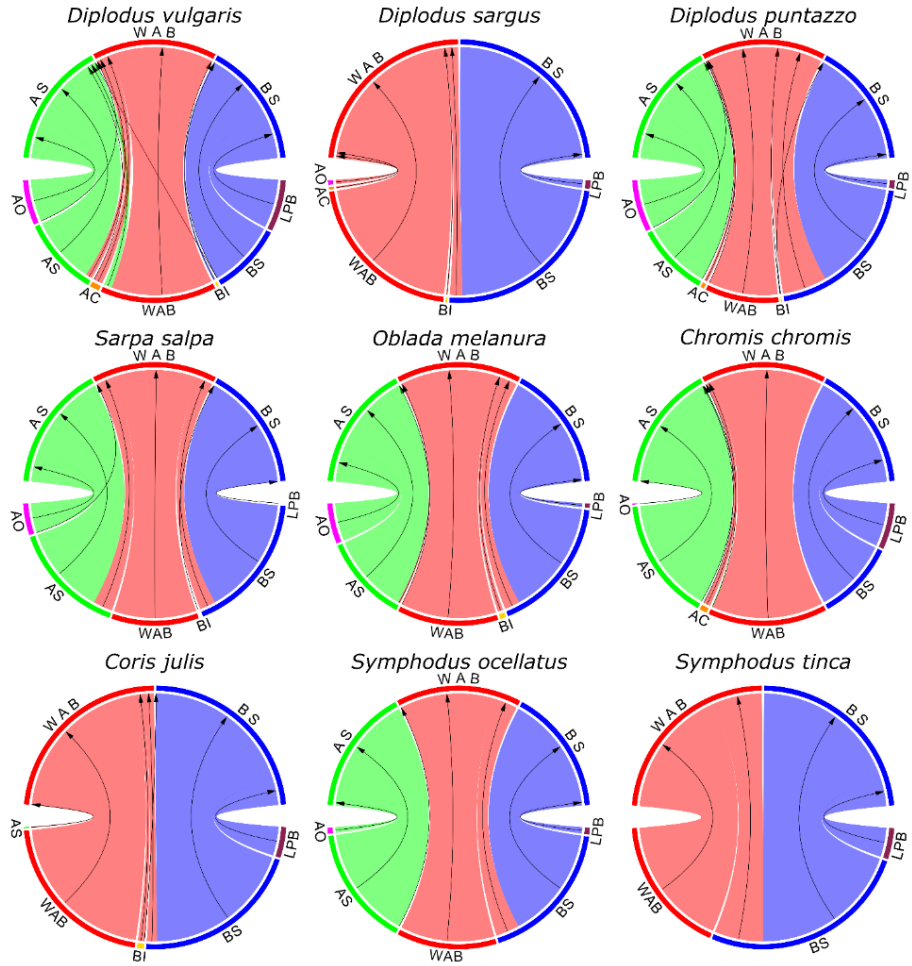


Figure 4 Dispersal plots for all nine species to the sampled areas of settlers (upper half of each circle) from their potential source coastal areas (lower half of each circle) according to the backtracking simulations. Arrows indicate larvae directionality and bottom bar colours delimit the source areas. AO: Atlantic Ocean (pink), AS: Alboran Sea (green), AC: Algerian Coast (orange), WAB: West Algerian Basin (red), BI: Balearic Islands (yellow), BS: Balearic Sea (blue), LPB: Liguro-Provençal Basin (purple). Circle and upper bar colours indicate sampled settlement areas.

The Alboran Sea was characterized by the presence of two anticyclonic gyres between Gibraltar Strait and Almeria-Oran Front. The western gyre was strong and present during the larval period of all species while the eastern gyre was milder (Figure S1). Therefore, the gyres and the AOF delimited a sub-basin that appeared to be badly connected with the rest of the Western Mediterranean Sea. A high number of settlers were self-recruits (75.3%± 17.8), nonetheless, we found a good number of potential incoming settlers from the Atlantic Ocean through the Strait of Gibraltar (23.3%± 16.9), common in most of the species (Figure 3). Thus, we can consider that the three sampling regions are three hydrodynamic units.

Influence of early life traits on dispersal distances and orientation

The global GLMM for dispersal distances explained a high proportion of the variance found in our data (R^2=0.814). In this model, all variables (PLD, hatching date, and the two random factors, locality and species) had a significant effect (Table 1). Most of the variability was

Table 1: Results of the generalized lineal mixed model (GLMM) for the dispersal distance and dispersal orientation We show the significance of the influence (p-value) of the fixed (PLD and Hatching date) and random (Species and Locality) factors on dispersal distance and orientation as well as the adjusted R^2 o the models including all factors (Full model) or only fixed factors.

Factors		Dispersal distance	Dispersal orientation
Fixed	PLD	**<0.001**	0.395
	Hatching date	**<0.001**	**<0.001**
Random	(Species)	**<0.001**	**<0.001**
	(Locality)	**<0.001**	**<0.001**
R^2	Fixed factors	0.045	0.007
	Full model	0.814	0.52

explained by the random factors, as the variance explained by the fixed factors was only R^2=0.045. The species-specific GLMM showed that PLD positively and significantly affected dispersal distance in all species (Table 2). The hatching date had a significant effect on dispersal distances in most species with a few exceptions as for *Diplodus vulgaris*, *Sarpa salpa* and *Coris julis* (Table 2). The sign of the correlation changed in some cases, for instance species such as *Diplodus sargus* and *Symphodus tinca*

presented significant negative correlations while the rest were positively correlated. Interestingly, these two species reproduce in spring suggesting that individuals hatching early in the season disperse over larger distances. The species with significant positive correlation (*Diplodus puntazzo*, *Oblada melanura*, *Chromis chromis* and *Symphodus ocellatus*) reproduce, at least partially, in summer, suggesting that individuals hatching later disperse larger distances. For all species differences between localities were significant and the full models presented high R^2 values. We observed different contribution of the random factors across species as compared with the variance explained by their fixed factors (Table 2). For instance, in *Symphodus tinca* the variance explained by PLD and hatching date was very high (R^2=0.72) while in *S. ocellatus* was very low (R^2=0.08).

Tabla 2: Results of the generalized lineal mixed model (GLMM) for the dispersal distance and dispersal orientation for each one of the studied species. We show the β coefficient of the influence of the fixed factors (PLD and Hatching date), in bold when there is a significant effect. We also indicate the influence (p-value) of Locality on dispersal distances and the adjusted R2 of the full model and only for fixed factors.

Species	Dispersal distance ~ PLD + Hatching date + (Locality)					Dispersal orientation ~ PLD + Hatching date + (Locality)				
	PLD	Hatching date	Locality	R^2 fixed	R^2 full	PLD	Hatching date	Locality	R^2 fixed	R^2 full
Diplodus vulgaris	**19.56**	0.31	**<0.001**	0.24	0.47	0.02	**0.55**	**<0.001**	0.03	0.58
Diplodus sargus	**20.85**	**-1.15**	**<0.001**	0.33	0.87	-0.02	0.44	**<0.001**	0.05	0.53
Diplodus puntazzo	**8.8**	**1.66**	**<0.001**	0.03	0.67	-0.09	0.5	**<0.001**	0.01	0.67
Sarpa salpa	**18.11**	2.21	**<0.001**	0.14	0.4	-0.14	-0.22	**<0.001**	0	0.54
Oblada melanura	**12.75**	**2.45**	**<0.001**	0.1	0.71	**-0.32**	**-1.64**	**<0.001**	0.14	0.69
Chromis chromis	**18.42**	**2.6**	**<0.001**	0.11	0.67	-0.2	-0.48	**<0.001**	0.01	0.57
Coris julis	**6.25**	-1.53	**<0.001**	0.13	0.62	0.11	-0.54	**<0.001**	0.03	0.53
Symphodus ocellatus	**18.47**	**0.54**	**<0.001**	0.08	0.85	0.01	0.43	**<0.001**	0.02	0.44
Symphodus tinca	**19.53**	**-1.88**	**0.04**	0.72	0.79	0.35	0.57	**<0.001**	0.01	0.52

Dispersal distances among hydrodynamic units were significantly different for most species, as assessed with the Kruskal-Wallis and post-hoc Dunn tests (Figure 5, Table S3). Individuals settling in the Alboran Sea showed the largest dispersal distances. The patterns in the other two basins varied across species. Settlers of the three *Diplodus* species had significantly longer dispersal distances in the West Algerian Basin than in the Balearic Sea, whereas *Oblada melanura, Chromis chromis*, *Coris julis* and *Symphodus ocellatus* had the reverse pattern (Figure 5). *Sarpa salpa* and

Symphodus tinca showed no differences in mean travelled distance by recruits from these two basins.

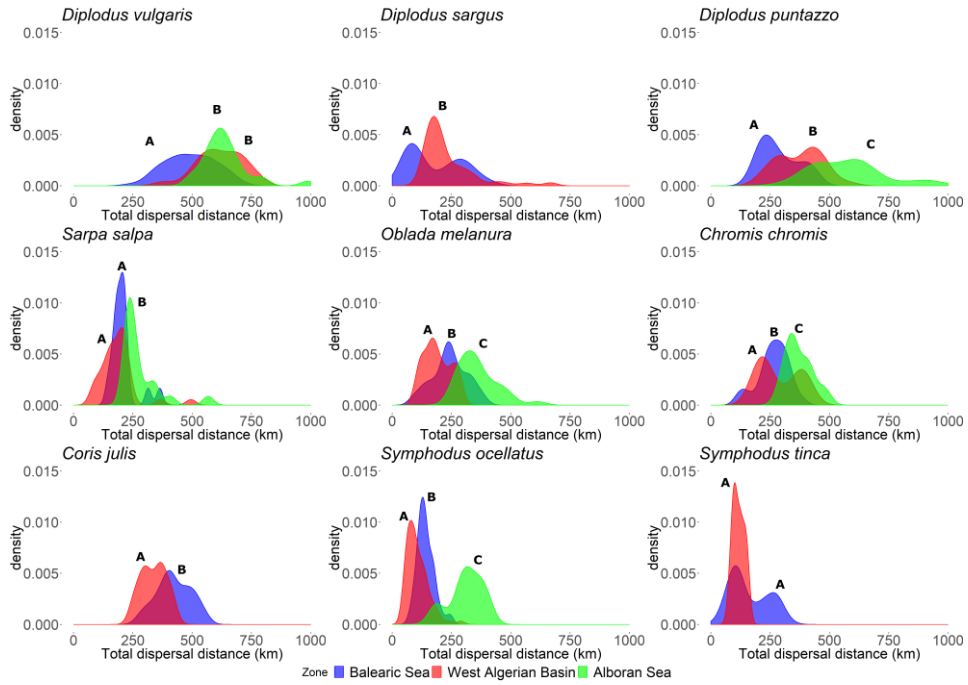


Figura 5: Dispersal distance distributions of the individuals sampled in the three hydrodynamic units based on the individual backtracking simulations for each species. Different letters within each species indicate statistically significant differences in larval dispersal distances between hydrodynamic units (post-hoc Dunn tests).

Dispersal orientation was significantly affected by hatching date and the two random factors (locality and species), but not by PLD (Table 1). The full model explained a good proportion of the variance (R^2=0.52) although mainly due to the random factors, since the effect of the fixed factors was small (R^2=0.007). For the species-specific GLMM significant differences in dispersal orientation were found among localities for each species (Table 2). However, significant effects for PLD were only found in *O. melanura,* the correlation being negative with dispersal orientation, suggesting that individuals with longer PLD disperse preferentially northwards. The hatching date in these analyses were only significant for *O. melanura* and *D. vulgaris,* with negative and positive correlations respectively (Table 2). Consequently, the variance explained by the fixed factors was small although the full model explained a

good proportion of the variance in all species with regressions ranging from R^2=0.44 in *S. ocellatus* to R^2=0.69 in *O. melanura*.

Overall, the variance explained by PLD and hatching date on dispersal distance and orientation was small within species. However, these two factors can have different effects in the final dispersal. For instance, individuals of *O. melanura* from Xabia with the same hatching date but different PLD had different displacement direction and longer dispersal distances with longer PLD as expected (Figure S2). Individuals from the same locality with the same PLD but different hatching date, changed also dispersal direction and distance (Figure S2). This pattern suggests that at a fine scale (*e.g.* locality scale) hatching date and PLD can have an important role in determining the origin of settlers mediated by changes in environmental conditions that at a larger scale remains undetectable.

The oceanographic transports during 2014-2017 showed monthly, seasonal and interannual variability across both the IC and AOF, with more stable net transports across the AOF and a higher variability across the IC (Figure S3). For each species, during its whole pelagic period, the K-W tests did not show significant differences across years in the AOF, while significant differences were observed in the IC (Table S4).

Discussion

Here, we measured the dispersal abilities (and thus, potential connectivity) in nine common fish species in the Mediterranean Sea by individual-based simulations. We show that the three sampled areas defined by the oceanographic discontinuities, Balearic Sea, West Algerian Basin and Alboran Sea, present low exchange of recruits in all species and should be considered three hydrodynamic units. Nonetheless, for some species we observed a higher directional exchange among them due to the temporal decrease in the front strength. We found significant influence of early life traits (PLD and hatching date) and sampling area in dispersal distance and orientation between and within species. Our results demonstrate that individual based information on life-

history traits is fundamental to model dispersal and evaluate connectivity at a fine and regional scale.

The low level of larval exchange among hydrodynamic units in all species suggests that self-recruitment is the common mechanism of larval replenishment in Mediterranean coastal fishes. Using molecular techniques, such as microsatellite loci, high levels of self-recruitment had been found in some fish species in the western Mediterranean (e.g. *Tripterygion delaisi*, *Serranus cabrilla*) suggesting that a high proportion of the larvae remained close to, or never left, their natal spawning area (Carreras-Carbonell, Macpherson, & Pascual, 2007; Schunter, Carreras-Carbonell, Macpherson, et al., 2011). Similar results were observed in coral-reef fishes (e.g. Buston and Elith, 2011; Jones et al., 2005, 1999; Swearer et al., 1999). For instance, Almany et al., 2007 found by stable isotope analyses that for species with both short (<2 weeks *Amphiprion percula*) and long (>1 month, *Chaetodon vagabundus*) PLD ca. 60% of settled juveniles were spawned at Kimbe island (Papua New Guinea), further confirmed by parentage analyses (Berumen et al., 2012; Planes, Jones, & Thorrold, 2009). Similar high level of self-recruitment was observed by parentage analyses within a network of marine reserves on the Great Barrier Reef in *Plectropomus maculatus* and *Lutjanus carponotatus* (Harrison et al., 2012). Nevertheless, with the same assignment technique, other species showed a smaller proportion of self-recruitment in the same area, e.g. *Amphiprion polymnus* (Saenz-Agudelo, Jones, Thorrold, & Planes, 2012) indicating a certain variability in these patterns probably species specific.

In our work, we found that the two oceanographic discontinuities present in the study area (Ibiza channel and Almeria-Oran front) determined the presence of three hydrodynamic units (Balearic Sea, West Algerian Basin and Alboran Sea) with low larval exchange in all species. Other studies have found different and larger hydrodynamic units in the same area, when simulating particle movement in different years and considering species-unspecific PLD models (Rossi et al., 2014). Temporal changes in currents and front strength in different species can account for changes in larval exchange. This seems the case of *D. vulgaris* reproducing in winter where the

AOF allowed more larval exchange in both directions according to our backtracking modelling. This front had been previously reported to decrease its strength or even disappear in winter (Fernández, Dietrich, Haney, & Tintoré, 2005; Renault et al., 2012; Tintore et al., 1988). In our study, we observed that the decrease in strength in winter was due to the fading of the eastern gyre in the Alboran Sea, allowing the transport of particles in both directions. Nonetheless, a previous genetic study with *D. vulgaris* found strong genetic differentiation across this front (Galarza et al., 2009) suggesting that interannual variation might be relevant. Oceanographic variability across years and seasons (Botsford et al., 2009) can modify the level of self-recruitment. Inter-annual changes in gene flow across the AOF have been reported for other species such as the crab *Liocarcinus depurator* (M. Pascual et al., 2016), or the sea urchins *Paracentrotus lividus* (Calderón, Pita, Brusciotti, Palacín, & Turon, 2012) and *Arbacia lixula* (Pérez-Portela et al., 2019), where temporal genetic population differentiation was associated to changes in water circulation patterns across years. Nonetheless, we did not observe significant changes in potential larval transportation across this oceanographic discontinuity during the analysed period (2014-2017) suggesting that this front is more permanent despite the reported interannual variation.

Oceanographic variability was also detected across the Ibiza Channel, where the meridional water transport was observed to change at weekly, seasonal and interannual scales (Balbín et al., 2014; Heslop et al., 2012; Pinot et al., 2002). We observed a high proportion of settlers in the West Algerian Basin coming from the Balearic Sea in *Symphodus tinca*, whereas in *S. ocellatus*, from the same genus and with similar PLD, the proportion was lower (Figure 4). Given that these two species reproduce at different seasons, temporal seasonality in front strength can account for the capacity to cross the Ibiza Channel. This higher capacity of *S. tinca* in crossing southwards the IC is in accordance with the lower genomic differentiation found in this species in comparison to that found in *S. ocellatus* (Torrado, Carreras, Raventos, Macpherson, & Pascual, 2020). Different patterns of genetic differentiation have been reported across IC for different fish species, for instance no genetic differentiation was found for *Epinephelus marginatus* (Schunter, Carreras-Carbonell, Planes, et al., 2011),

whereas strong genetic differentiation was detected for *Serranus cabrilla* (Schunter, Carreras-Carbonell, Macpherson, et al., 2011). Moreover, temporal variation has been detected in the northern site of the IC for *Liocarcinus depurator* further indicating that annual changes might promote differential exchange and connectivity (M. Pascual et al., 2016). This area is a highly dynamic transition area with variable transport in strength and direction as we have observed at different temporal scales (Figure S3) affecting dispersal patterns in all species. Therefore, the three hydrodynamic units defined in the western Mediterranean Sea by these two oceanographic discontinuities should not be considered as closed systems, but as a net with seasonally and yearly variability in connections.

Nevertheless, large-scale circulation patterns are not the unique factors influencing larval dispersal. In our study, we found a global relationship between early-life traits and dispersal distances and orientations. Our results demonstrate that species with longer PLD have greater dispersal distances, in accordance with previous studies (e.g. Pascual et al., 2017; Siegel et al., 2003). Moreover, hatching date also influenced dispersal distances as indicated in our global analysis, which could be related to temperature in modifying PLD, as reported in other species (Schunter et al., 2019). Nonetheless, different impact of hatching date was obtained at the intraspecific level, with no influence for the two species with larger PLD (*D. vulgaris* and *C. julis*) and for *S. salpa*. Moreover, the regression coefficient changed in sign across species with those reproducing in spring showing negative correlations with distance so that individuals hatching earlier experience colder temperatures, show longer PLD than individuals hatching later in the season and consequently disperse larger distances. On the other hand, those species reproducing in summer (*D. puntazzo*, *O. melanura*, *C. chromis* and *S. ocellatus*) have positive correlations, indicating that individuals hatching later have larger dispersal distances. Since temperature increases over time along this period with a corresponding decrease of PLD, the effect of hatching date on dispersal distance is probably due to interaction of other unknown factors.

The dispersal distances also varied at intraspecific level among our hydrodynamic units. These differences seemed to be related with two main factors: (1)

differences in circulation patterns and oceanographic currents strength to which the larvae were exposed in each area, and (2) the distance and direction of available source areas of larval hatching. For instance, the larvae in the Alboran Sea had for most species longer dispersal distances than larvae in the other two areas. These longer distances were mostly due to the presence of one permanent and one semipermanent eddy in this zone (Tintore et al., 1988; present study) making larvae to travel longer distances before reaching the coast. Thus, longer dispersal distances do not imply more connectivity across hydrodynamic units since eddies can strongly contribute to self-recruitment by promoting larval retention within these areas. Moreover, eddies can further impact on populations, since they can generate recruitment peaks in punctual moments (S. Sponaugle et al., 2005) and increase larval development rate and settlers' survivorship (Shulzitski et al., 2015; Shulzitski, Sponaugle, Hauff, Walter, & Cowen, 2016). Furthermore, the communication between Balearic Sea and Balearic Islands was probably mediated by temporal eddies occasionally formed in this area (e.g. Amores et al., 2013; Pascual et al., 2002) which can coincide with some species' larval periods as observed in our models.

We found that hatching date can also significantly affect dispersal orientation in the global model, and subsequently the geographic origin of settlers. However, at the intraspecific level this significance was only observed in two species, *D. vulgaris* and *O. melanura* with positive and negative correlations, respectively. Thus, individuals hatching early in the season show preferentially a southern origin in *O. melanura* while a northern origin in *D. vulgaris,* although this factor explained a small proportion of the model. For instance, we observed that individuals from the same locality (Xabia near the Ibiza channel) and the same PLD, but born at different dates, had a different hatching area, originating northwards or southwards. This variability could be related to daily current variations, common in coastal waters (Fernández et al., 2005), modifying dispersal orientation as observed in other species through direct dispersal assessment by parent-offspring analysis (Schunter, Pascual, Garza, Raventos, & Macpherson, 2014). Thus, the hatching date of the individual can be more important than previously thought at a fine scale.

Conclusions

The use of oceanographic models together with PLD general information have been used in previous studies at a regional (Andrello et al., 2013; Barbut et al., 2019; Rossi et al., 2014), and global scale (Andrello et al., 2017). These studies have generated conservational strategies recommending the use of these units or cells (Boero et al., 2016) and recommended their use to establish Marine Protected Areas (MPAs) network strategies and fisheries policies (Kerr, Cadrin, & Secor, 2010). In our study, both hatching date and PLD have been confirmed as factors to have in mind to design more precise larval dispersal models. We can conclude that having good individually otolith-inferred information about these parameters helps to a better definition of recruit's origin area and in defining hydrodynamic units. This more accurate information could be highly valuable for the identification of natural management units and can be useful when considering conservation strategies such as those establishing networks of Marine Protected Areas or in defining conservation measures at regional scales.

Acknowledgments

This work was supported by the Spanish Government project 'PopCOmics' (CTM2017-88080) (MCIU, AEI/FEDER, UE) and by the European FP7 CoCoNet project (Ocean 2011-4, grant agreement #287844). The authors CC and MP are members of the research group SGR2017-1120 and EM of SGR2017-378 (Catalan Government). HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities. The authors acknowledge the MEDCLIC project, funded by "La Caixa" Foundation, contributing to the development of the WMOP hydrodynamic model. We thank Maria Pilar Olivar and Ana Sabatés for their useful comments.

References

Aguiar, E., Mourre, B., Juza, M., Reyes, E., Hernández-Lasheras, J., Cutolo, E., … Tintoré, J. (2020). Multi-platform model assessment in the Western Mediterranean Sea: impact of downscaling on the surface circulation and mesoscale activity. *Ocean Dynamics*, *70*(2), 273–288. doi: 10.1007/s10236-019-01317-8

Allain, G., Petitgas, P., Lazure, P., & Grellier, P. (2007). Biophysical modelling of larval drift, growth and survival for the prediction of anchovy (*Engraulis encrasicolus*) recruitment in the Bay of Biscay (NE Atlantic). *Fisheries Oceanography*, *16*(6), 489–505. doi: 10.1111/j.1365-2419.2007.00443.x

Almany, G. R., Berumen, M. L., Thorrold, S. R., Planes, S., & Jones, G. P. (2007). Local replenishment of coral reef fish populations in a marine reserve. *Science*, *316*(5825), 742–744. doi: 10.1126/science.1140597

Amores, A., Monserrat, S., & Marcos, M. (2013). Vertical structure and temporal evolution of an anticyclonic eddy in the Balearic Sea (western Mediterranean). *Journal of Geophysical Research: Oceans*, *118*(4), 2097–2106. doi: 10.1002/jgrc.20150

Andrello, M., Guilhaumon, F., Albouy, C., Parravicini, V., Scholtens, J., Verley, P., … Mouillot, D. (2017). Global mismatch between fishing dependency and larval supply from marine reserves. *Nature Communications*, *8*(1), 1–9. doi: 10.1038/ncomms16039

Andrello, M., Mouillot, D., Beuvier, J., Albouy, C., Thuiller, W., & Manel, S. (2013). Low Connectivity between Mediterranean Marine Protected Areas: A Biophysical Modeling Approach for the Dusky Grouper Epinephelus marginatus. *PLoS ONE*, *8*(7). doi: 10.1371/journal.pone.0068564

Andrello, M., Mouillot, D., Somot, S., Thuiller, W., & Manel, S. (2015). Additive effects of climate change on connectivity between marine protected areas and larval supply to fished areas. *Diversity and Distributions*, *21*(2), 139–150. doi: 10.1111/ddi.12250

Arya, S., Mount, D., Kemp, S. E., & Jefferis, G. (2019). *RANN: Fast Nearest Neighbour Search (Wraps ANN Library) Using L2 Metric*. Retrieved from https://cran.r-project.org/package=RANN

Astraldi, M., Balopoulos, S., Candela, J., Font, J., Gacic, M., Gasparini, G. P., … Tintoré, J. (1999). The role of straits and channels in understanding the characteristics of Mediterranean circulation. *Progress in Oceanography*, *44*(1–3), 65–108. doi: 10.1016/S0079-6611(99)00021-X

Ayata, S. D., Lazure, P., & Thiébaut, É. (2010). How does the connectivity between populations mediate range limits of marine invertebrates? A case study of larval dispersal between the Bay of Biscay and the English Channel (North-East Atlantic). *Progress in Oceanography*, *87*(1–4), 18–36. doi: 10.1016/j.pocean.2010.09.022

Bache, S. M., & Wickham, H. (2014). magrittr: A Forward-Pipe Operator for R. R package version 1.5. *Vienna, Austria: The R Foundation. Retrieved from Https://CRAN. R-Project. Org/Package= Magrittr*.

Balbín, R., López-Jurado, J. L., Flexas, M. M., Reglero, P., Vélez-Velchí, P., González-Pola, C., … Alemany, F. (2014). Interannual variability of the early summer circulation

around the Balearic Islands: Driving factors and potential effects on the marine ecosystem. *Journal of Marine Systems*, *138*, 70–81. doi: 10.1016/j.jmarsys.2013.07.004

Banks, S. C., Piggott, M. P., Williamson, J. E., Bové, U., Holbrook, N. J., & Beheregaray, L. B. (2007). Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. *Ecology*, *88*(12), 3055–3064. doi: 10.1890/07-0091.1

Barbut, L., Groot Crego, C., Delerue-Ricard, S., Vandamme, S., Volckaert, F. A. M., & Lacroix, G. (2019). How larval traits of six flatfish species impact connectivity. *Limnology and Oceanography*, *64*(3), 1150–1171. doi: 10.1002/lno.11104

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Berumen, M. L., Almany, G. R., Planes, S., Jones, G. P., Saenz-Agudelo, P., & Thorrold, S. R. (2012). Persistence of self-recruitment and patterns of larval connectivity in a marine protected area network. *Ecology and Evolution*, *2*(2), 444–452. doi: 10.1002/ece3.208

Bivand, R., Keitt, T., & Rowlingson, B. (2018). *rgdal: Bindings for the "geospatial" data abstraction library (version 1.3-1) https://CRAN.R-project.org/package=rgdal*.

Boero, F., Foglini, F., Fraschetti, S., Goriup, P., Macpherson, E., Planes, S., … Consortium, C. (2016). CoCoNet: towards coast to coast networks of marine protected areas (from the shore to the high and deep sea), coupled with sea-based wind energy potential. *SCIRES-IT-SCIentific RESearch and Information Technology*, *6*, 1–95. doi: 10.2423/i22394303v6SpI

Botsford, L. W., White, J. W., Coffroth, M. A., Paris, C. B., Planes, S., Shearer, T. L., … Jones, G. P. (2009, June 11). Connectivity and resilience of coral reef metapopulations in marine protected areas: Matching empirical efforts to predictive needs. *Coral Reefs*, Vol. 28, pp. 327–337. doi: 10.1007/s00338-009-0466-z

Bottesch, M., Gerlach, G., Halbach, M., Bally, A., Kingsford, M. J., & Mouritsen, H. (2016). A magnetic compass that might help coral reef fish larvae return to their natal reef. *Current Biology*, Vol. 26, pp. R1266–R1267. doi: 10.1016/j.cub.2016.10.051

Buston, P. M., & Elith, J. (2011). Determinants of reproductive success in dominant pairs of clownfish: A boosted regression tree analysis. *Journal of Animal Ecology*, *80*(3), 528–538. doi: 10.1111/j.1365-2656.2011.01803.x

Calderón, I., Pita, L., Brusciotti, S., Palacín, C., & Turon, X. (2012). Time and space: Genetic structure of the cohorts of the common sea urchin *Paracentrotus lividus* in Western Mediterranean. *Marine Biology*, *159*(1), 187–197. doi: 10.1007/s00227-011-1799-z

Calò, A., Lett, C., Mourre, B., Pérez-Ruzafa, Á., & García-Charton, J. A. (2018). Use of Lagrangian simulations to hindcast the geographical position of propagule release zones in a Mediterranean coastal fish. *Marine Environmental Research*, *134*, 16–27. doi: 10.1016/j.marenvres.2017.12.011

Carreras-Carbonell, J., Macpherson, E., & Pascual, M. (2007). High self-recruitment levels in a Mediterranean littoral fish population revealed by microsatellite markers. *Marine Biology*, *151*(2), 719–727. doi: 10.1007/s00227-006-0513-z

Christensen, A., Daewel, U., Jensen, H., Mosegaard, H., St John, M., & Schrum, C. (2007). Hydrodynamic backtracking of fish larvae by individual-based modelling. *Marine*

Ecology Progress Series, *347*, 221–232. doi: 10.3354/meps06980

Cowen, R. K. (2002). *Oceanographic Influences on Larval Dispersal and Retention and Their Consequences for Population Connectivity*.

Cowen, R. K., & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, *1*, 443–466. doi: 10.1146/annurev.marine.010908.163757

Di Franco, A., Gillanders, B. M., de Benedetto, G., Pennetta, A., de Leo, G. A., & Guidetti, P. (2012). Dispersal patterns of coastal fish: Implications for designing networks of marine protected areas. *PLoS ONE*, *7*(2). doi: 10.1371/journal.pone.0031681

Döös, K., & Engqvist, A. (2007). Assessment of water exchange between a discharge region and the open sea – A comparison of different methodological concepts. *Estuarine, Coastal and Shelf Science*, *74*(4), 709–721. doi: 10.1016/j.ecss.2007.05.022

Döös, K., & Jönsson, B. (2013). TRACMASS—A lagrangian trajectory model. In *Preventive Methods for Coastal Protection: Towards the Use of Ocean Dynamics for Pollution Control* (pp. 225–249). doi: 10.1007/978-3-319-00440-2_7

Döös, K., Rupolo, V., & Brodeau, L. (2011). Dispersion of surface drifters and model-simulated trajectories. *Ocean Modelling*, *39*(3–4), 301–310. doi: 10.1016/j.ocemod.2011.05.005

Faillettaz, R., Blandin, A., Paris, C. B., Koubbi, P., & Irisson, J. O. (2015). Sun-compass orientation in mediterranean fish larvae. *PLoS ONE*, *10*(8), e0135213. doi: 10.1371/journal.pone.0135213

Faillettaz, R., Durand, E., Paris, C. B., Koubbi, P., & Irisson, J. O. (2018). Swimming speeds of Mediterranean settlement-stage fish larvae nuance Hjort's aberrant drift hypothesis. *Limnology and Oceanography*, *63*(2), 509–523. doi: 10.1002/lno.10643

Fernández, V., Dietrich, D. E., Haney, R. L., & Tintoré, J. (2005). Mesoscale, seasonal and interannual variability in the Mediterranean Sea using a numerical ocean model. *Progress in Oceanography*, *66*(2–4), 321–340. doi: 10.1016/j.pocean.2004.07.010

Fraker, M. E., Anderson, E. J., Brodnik, R. M., Carreon-Martinez, L., DeVanna, K. M., Fryer, B. J., … Ludsin, S. A. (2015). Particle backtracking improves breeding subpopulation discrimination and natal-source identification in mixed populations. *PLoS ONE*, *10*(3). doi: 10.1371/journal.pone.0120752

Gaines, S. D., Gaylord, B., Gerber, L. R., Hastings, A., & Kinlan, B. P. (2007). Connecting places: The ecological consequences of dispersal in the sea. *Oceanography*, *20*(SPL.ISS. 3), 90–99. doi: 10.5670/oceanog.2007.32

Galarza, J. A., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G. F., & Rico, C. (2009). The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences*, *106*(5), 1473–1478. doi: 10.1073/pnas.0806804106

Garcia Lafuente, J., Lopez Jurado, J. L., Cano Lucaya, N., Vargas Yanez, M., & Aguiar Garcia, J. (1995). Circulation of water masses through the Ibiza Channel. *Oceanologica Acta*, *18*(2), 245–254. Retrieved from https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=2916044

Garnier, S. (2018). *viridis: Default Color Maps from "matplotlib."* Retrieved from https://cran.r-project.org/package=viridis

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*, *30*(19), 2811–2812. doi: 10.1093/bioinformatics/btu393

Harrison, H. B., Williamson, D. H., Evans, R. D., Almany, G. R., Thorrold, S. R., Russ, G. R., … Jones, G. P. (2012). Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Current Biology*, *22*(11), 1023–1028. doi: 10.1016/j.cub.2012.04.008

Heslop, E. E., Ruiz, S., Allen, J., López-Jurado, J. L., Renault, L., & Tintoré, J. (2012). Autonomous underwater gliders monitoring variability at choke points in our ocean system: A case study in the Western Mediterranean Sea. *Geophysical Research Letters*, *39*(20), 1–6. doi: 10.1029/2012GL053717

Holliday, D., Beckley, L., Millar, N., Olivar, M., Slawinski, D., Feng, M., & Thompson, P. (2012). Larval fish assemblages and particle back-tracking define latitudinal and cross-shelf variability in an eastern Indian Ocean boundary current. *Marine Ecology Progress Series*, *460*, 127–144. doi: 10.3354/meps09730

Jones, G. P., Millcich, M. J., Emsile, M. J., & Lunow, C. (1999). Self-recruitment in a coral fish population. *Nature*, *402*(6763), 802–804. doi: 10.1038/45538

Jones, Geoffrey P., Planes, S., & Thorrold, S. R. (2005). Coral reef fish larvae settle close to home. *Current Biology*, *15*(14), 1314–1318. doi: 10.1016/j.cub.2005.06.061

Juza, M., Escudier, R., Vargas-Yáñez, M., Mourre, B., Heslop, E., Allen, J., & Tintoré, J. (2019). Characterization of changes in Western Intermediate Water properties enabled by an innovative geometry-based detection approach. *Journal of Marine Systems*, *191*(July 2018), 1–12. doi: 10.1016/j.jmarsys.2018.11.003

Juza, M., Mourre, B., Renault, L., Gómara, S., Sebastian, K., López, S. L., … others. (2016). Operational SOCIB forecasting system and multi-platform validation in the Western Mediterranean. *J. Oper. Oceanogr*, *9*, 9231.

Kerr, L. A., Cadrin, S. X., & Secor, D. H. (2010). Simulation modelling as a tool for examining the consequences of spatial structure and connectivity on local and regional population dynamics. *ICES Journal of Marine Science*, *67*(8), 1631–1639. doi: 10.1093/icesjms/fsq053

Kingsford, M. J., Leis, J. M., Shanks, A., Lindeman, K. C., Morgan, S. G., & Pineda, J. (2002). Sensory environments, larval abilities and local self-recruitment. *Bulletin of Marine Science*, Vol. 70, pp. 309–340. Retrieved from https://www.ingentaconnect.com/content/umrsmas/bullmar/2002/00000070/A00101s1/art00005

Kinlan, B. P., Gaines, S. D., & Lester, S. E. (2005). Propagule dispersal and the scales of marine community process. *Diversity and Distributions*, *11*(2), 139–148. doi: 10.1111/j.1366-9516.2005.00158.x

Kough, A. S., & Paris, C. B. (2015). The influence of spawning periodicity on population connectivity. *Coral Reefs*, *34*(3), 753–757. doi: 10.1007/s00338-015-1311-1

Leis, J. M. (1991). The pelagic stage of reef fishes: the larval biology of coral fishes. *The Ecology of Fishes on Coral Reefs*.

Leis, J. M. (2007). Behaviour as input for modelling dispersal of fish larvae: Behaviour, biogeography, hydrodynamics, ontogeny, physiology and phylogeny meet hydrography. *Marine Ecology Progress Series*, Vol. 347, pp. 185–193. doi: 10.3354/meps06977

Leis, J. M., Hay, A. C., & Gaither, M. R. (2011). Swimming ability and its rapid decrease at settlement in wrasse larvae (Teleostei: Labridae). *Marine Biology*, *158*(6), 1239–1246. doi: 10.1007/s00227-011-1644-4

Leis, J. M., Paris, C. B., Irisson, J. O., Yerman, M. N., & Siebeck, U. E. (2014). Orientation of fish larvae in situ is consistent among locations, years and methods, but varies with time of day. *Marine Ecology Progress Series*, *505*, 193–208. doi: 10.3354/meps10792

Millot, C. (1999). Circulation in the Western Mediterranean Sea. *Journal of Marine Systems*, *20*(1–4), 423–442. doi: 10.1016/S0924-7963(98)00078-5

Millot, C., & Taupier-Letage, I. (2005). Circulation in the Mediterranean Sea. In A. Saliot (Ed.), *The Mediterranean Sea* (pp. 29–66). doi: 10.1007/b107143

Mouritsen, H., Atema, J., Kingsford, M. J., & Gerlach, G. (2013). Sun Compass Orientation Helps Coral Reef Fish Larvae Return to Their Natal Reef. *PLoS ONE*, *8*(6). doi: 10.1371/journal.pone.0066039

Mourre, B., Aguiar, E., Juza, M., Hernandez-Lasheras, J., Reyes, E., Heslop, E., … Tintoré, J. (2018). Assessment of High-Resolution Regional Ocean Prediction Systems Using Multi-Platform Observations: Illustrations in the Western Mediterranean Sea. In *New Frontiers in Operational Oceanography* (pp. 663–694). doi: 10.17125/gov2018.ch24

O'Connor, J., & Muheim, R. (2017). Pre-settlement coral-reef fish larvae respond to magnetic field changes during the day. *Journal of Experimental Biology*, *220*(16), 2874–2877. doi: 10.1242/jeb.159491

Olivar, M. P., Emelianov, M., Villate, F., Uriarte, I., Maynou, F., Álvarez, I., & Morote, E. (2010). The role of oceanographic conditions and plankton availability in larval fish assemblages off the Catalan coast (NW Mediterranean). *Fisheries Oceanography*, *19*(3), 209–229. doi: 10.1111/j.1365-2419.2010.00538.x

Ospina-Alvarez, A., Catalán, I. A., Bernal, M., Roos, D., & Palomera, I. (2015). From egg production to recruits: Connectivity and inter-annual variability in the recruitment patterns of European anchovy in the northwestern Mediterranean. *Progress in Oceanography*, *138*, 431–447. doi: 10.1016/j.pocean.2015.01.011

Paris, C. B., Atema, J., Irisson, J. O., Kingsford, M., Gerlach, G., & Guigand, C. M. (2013). Reef Odor: A Wake Up Call for Navigation in Reef Fish Larvae. *PLoS ONE*, *8*(8), e72808. doi: 10.1371/journal.pone.0072808

Paris, C. B., & Cowen, R. K. (2004). Direct evidence of a biophysical retention mechanism for coral reef fish larvae. *Limnology and Oceanography*, *49*(6), 1964–1979. doi: 10.4319/lo.2004.49.6.1964

Pascual, A., Buongiorno Nardelli, B., Larnicol, G., Emelianov, M., & Gomis, D. (2002). A case of an intense anticyclonic eddy in the Balearic Sea (western Mediterranean). *Journal of Geophysical Research C: Oceans*, *107*(11), 4–1. doi: 10.1029/2001jc000913

Pascual, M., Palero, F., García-Merchán, V. H., Macpherson, E., Robainas-Barcia, A., Mestres, F., … Abelló, P. (2016). Temporal and spatial genetic differentiation in the crab *Liocarcinus depurator* across the Atlantic-Mediterranean transition. *Scientific Reports*, *6*(1), 1–10. doi: 10.1038/srep29892

Pascual, M., Rives, B., Schunter, C., & Macpherson, E. (2017). Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in

the Mediterranean Sea. *PLoS ONE*, *12*(5), 1–20. doi: 10.1371/journal.pone.0176419

Payne, M. R., Ross, S. D., Clausen, L. W., Munk, P., Mosegaard, H., & Nash, R. D. M. (2013). Recruitment decline in North Sea herring is accompanied by reduced larval growth rates. *Marine Ecology Progress Series*, *489*, 197–211. doi: 10.3354/meps10392

Pérez-Portela, R., Wangensteen, O. S., Garcia-Cisneros, A., Valero-Jiménez, C., Palacín, C., & Turon, X. (2019). Spatio-temporal patterns of genetic variation in *Arbacia lixula*, a thermophilous sea urchin in expansion in the Mediterranean. *Heredity*, *122*(2), 244–259. doi: 10.1038/s41437-018-0098-6

Pierce, D. (2017). *ncdf4: interface to Unidata netCDF (version 4 or earlier) format data files.--R package ver. 1.16*.

Pinot, J. M., López-Jurado, J. L., & Riera, M. (2002, November 1). The CANALES experiment (1996-1998). Interannual, seasonal, and mesoscale variability of the circulation in the Balearic Channels. *Progress in Oceanography*, Vol. 55, pp. 335–370. doi: 10.1016/S0079-6611(02)00139-8

Planes, S., Jones, G. P., & Thorrold, S. R. (2009). Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5693–5697. doi: 10.1073/pnas.0808007106

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/

Raventos, N., & Macpherson, E. (2001). Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Marine Biology*, *138*(6), 1115–1120. doi: 10.1007/s002270000535

Raventos, N., & Macpherson, E. (2005). Effect of pelagic larval growth and size-at-hatching on post-settlement survivorship in two temperate labrid fish of the genus *Symphodus*. *Marine Ecology Progress Series*, *285*, 205–211. doi: 10.3354/meps285205

Raventos, N., Torrado, H., Arthur, R., Alcoverro, T., & Macpherson, E. (n.d.). *Temperature reduces fish dispersal as larvae grow faster to their settlement size*.

Renault, L., Oguz, T., Pascual, A., Vizoso, G., & Tintore, J. (2012). Surface circulation in the Alboran Sea (western Mediterranean) inferred from remotely sensed data. *Journal of Geophysical Research: Oceans*, *117*(8). doi: 10.1029/2011JC007659

Ross, S. D., Payne, M. R., Worsøe Clausen, L., Munk, P., Mosegaard, H., & Nash, R. D. M. (2012). Coupling otolith microstructure analysis and hydrographic backtracking suggests a mechanism for the 2000s North Sea herring recruitment failure. *ICES CM 2012/J*, *14*, 33. Retrieved from https://orbit.dtu.dk/files/123044807/J1412.pdf

Rossi, V., Ser-Giacomi, E., Lõpez, C., & Hernández-García, E. (2014). Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves. *Geophysical Research Letters*, *41*(8), 2883–2891. doi: 10.1002/2014GL059540

Sabatés, A., & Olivar, M. P. (1996). Variation of larval fish distributions associated with variability in the location of a shelf-slope front. *Marine Ecology Progress Series*, *135*(1–3), 11–20. doi: 10.3354/meps135011

Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R., & Planes, S. (2012). Patterns and persistence of larval retention and connectivity in a marine fish metapopulation. *Molecular Ecology*, *21*(19), 4695–4705. doi: 10.1111/j.1365-294X.2012.05726.x

Salat, J. (1996). Review of hydrographic environmental factors that may influence anchovy habitats in northwestern Mediterranean. *Scientia Marina*, *60*(SUPPL. 2), 21–32.

Schunter, C., Carreras-Carbonell, J., Macpherson, E., Tintoré, J., Vidal-Vijande, E., Pascual, A., … Pascual, M. (2011). Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, *20*(24), 5167–5181. doi: 10.1111/j.1365-294X.2011.05355.x

Schunter, C., Carreras-Carbonell, J., Planes, S., Sala, E., Ballesteros, E., Zabala, M., … Pascual, M. (2011). Genetic connectivity patterns in an endangered species: The dusky grouper (Epinephelus marginatus). *Journal of Experimental Marine Biology and Ecology*, *401*(1–2), 126–133. doi: 10.1016/j.jembe.2011.01.021

Schunter, C., Pascual, M., Garza, J. C., Raventos, N., & Macpherson, E. (2014). Kinship analyses identify fish dispersal events on a temperate coastline. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1785). doi: 10.1098/rspb.2014.0556

Schunter, C., Pascual, M., Raventos, N., Garriga, J., Garza, J. C., Bartumeus, F., & Macpherson, E. (2019). A novel integrative approach elucidates fine-scale dispersal patchiness in marine populations. *Scientific Reports*, *9*(1), 1–10. doi: 10.1038/s41598-019-47200-w

Selkoe, K. A., & Toonen, R. J. (2011). Marine connectivity: A new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series*, *436*, 291–305. doi: 10.3354/meps09238

Selkoe, Kimberly A., Watson, J. R., White, C., Horin, T. Ben, Iacchei, M., Mitarai, S., … Toonen, R. J. (2010). Taking the chaos out of genetic patchiness: Seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. *Molecular Ecology*, *19*(17), 3708–3726. doi: 10.1111/j.1365-294X.2010.04658.x

Shanks, A. L. (2009). Pelagic Larval Duration and Dispersal Distance Revisited. *The Biological Bulletin*, *216*(3), 373–385. doi: 10.1086/bblv216n3p373

Shanks, A. L., & Brink, L. (2005). Upwelling, downwelling, and cross-shelf transport of bivalve larvae: Test of a hypothesis. *Marine Ecology Progress Series*, *302*(Rumrill 1990), 1–12. doi: 10.3354/meps302001

Shulzitski, K., Sponaugle, S., Hauff, M., Walter, K., D'Alessandro, E. K., & Cowen, R. K. (2015). Close encounters with eddies: Oceanographic features increase growth of larval reef fishes during their journey to the reef. *Biology Letters*, *11*(1). doi: 10.1098/rsbl.2014.0746

Shulzitski, K., Sponaugle, S. S., Hauff, M., Walter, K. D., & Cowen, R. K. (2016). Encounter with mesoscale eddies enhances survival to settlement in larval coral reef fishes. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(25), 6928–6933. doi: 10.1073/pnas.1601606113

Siegel, D. A., Kinlan, B. P., Gaylord, B., & Gaines, S. D. (2003). Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series*, *260*, 83–96. doi: 10.3354/meps260083

Sponaugle, S., & Cowen, R. K. (1996). Larval supply and patterns of recruitment for two

caribbean reef fishes, Stegastes partitus and Acanthurus bahianus. *Marine and Freshwater Research*, *47*(2), 433–447. doi: 10.1071/MF9960433

Sponaugle, S., Lee, T., Kourafalou, V., & Pinkard, D. (2005). Florida Current frontal eddies and the settlement of coral reef fishes. *Limnology and Oceanography*, *50*(4), 1033–1048. doi: 10.4319/lo.2005.50.4.1033

Sponaugle, Su, & Cowen, R. K. (1996). Larval supply and patterns of recruitment for two caribbean reef fishes, *Stegastes partitus* and *Acanthurus bahianus*. *Marine and Freshwater Research*, *47*(2), 433–447. doi: 10.1071/MF9960433

Staaterman, E., Paris, C. B., & Helgers, J. (2012). Orientation behavior in fish larvae: A missing piece to Hjort's critical period hypothesis. *Journal of Theoretical Biology*, *304*, 188–196. doi: 10.1016/j.jtbi.2012.03.016

Swearer, S. E., Caselle, J. E., Lea, D. W., & Warner, R. R. (1999). Larval retention and recruitment in an island population of a coral-reef fish. *Nature*, *402*(6763), 799–802. doi: 10.1038/45533

Thyng, K., & Hetland, R. (2014). TracPy: Wrapping the Fortran Lagrangian trajectory model TRACMASS. *Proceedings of the 13th Python in Science Conference*, 79–84. doi: 10.25080/majora-14bd3278-00d

Tintore, J., La Violette, P. E., Blade, I., & Cruzado, A. (1988). A Study of an Intense Density Front in the Eastern Alboran Sea: The Almeria–Oran Front. *Journal of Physical Oceanography*, *18*(10), 1384–1397. doi: 10.1175/1520-0485(1988)018<1384:asoaid>2.0.co;2

Tintoré, J., Pinardi, N., Álvarez-Fanjul, E., Aguiar, E., Álvarez-Berastegui, D., Bajo, M., … Zodiatis, G. (2019). Challenges for Sustained Observing and Forecasting Systems in the Mediterranean Sea. *Frontiers in Marine Science*, *6*(September). doi: 10.3389/fmars.2019.00568

Tintoré, J., Vizoso, G., Casas, B., Heslop, E., Pascual, A., Orfila, A., … Manriquez, M. (2013). SOCIB: The Balearic Islands Coastal Ocean Observing and Forecasting System Responding to Science, Technology and Society Needs. *Marine Technology Society Journal*, *47*(1), 101–117. doi: 10.4031/MTSJ.47.1.10

Torrado, H., Carreras, C., Raventos, N., Macpherson, E., & Pascual, M. (2020). Individual-based population genomics reveal different drivers of adaptation in sympatric fish. *Scientific Reports*, Vol. 10. doi: 10.1038/s41598-020-69160-2

Treml, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., & Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integrative and Comparative Biology*, *52*(4), 525–537. doi: 10.1093/icb/ics101

White, J. W., Carr, M. H., Caselle, J. E., Washburn, L., Woodson, C. B., Palumbi, S. R., … Milligan, K. (2019). Connectivity, dispersal, and recruitment: Connecting benthic communities and the coastal ocean. *Oceanography*, *32*(3), 50–59. doi: 10.5670/oceanog.2019.310

Wickham, H, François, R., Henry, L., & Müller, K. (2018). *Dplyr: A Grammar of Data Manipulation; R Package Version 0.7. 8.*

Wickham, Hadley. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12), 1–20.

Wickham, Hadley. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Retrieved from http://ggplot2.org

Wilson, & McCormick, M. I. (1999). Microstructure of settlement-marks in the otoliths of tropical reef fishes. *Marine Biology*, *134*(1), 29–41. doi: 10.1007/s002270050522

Supplementary material

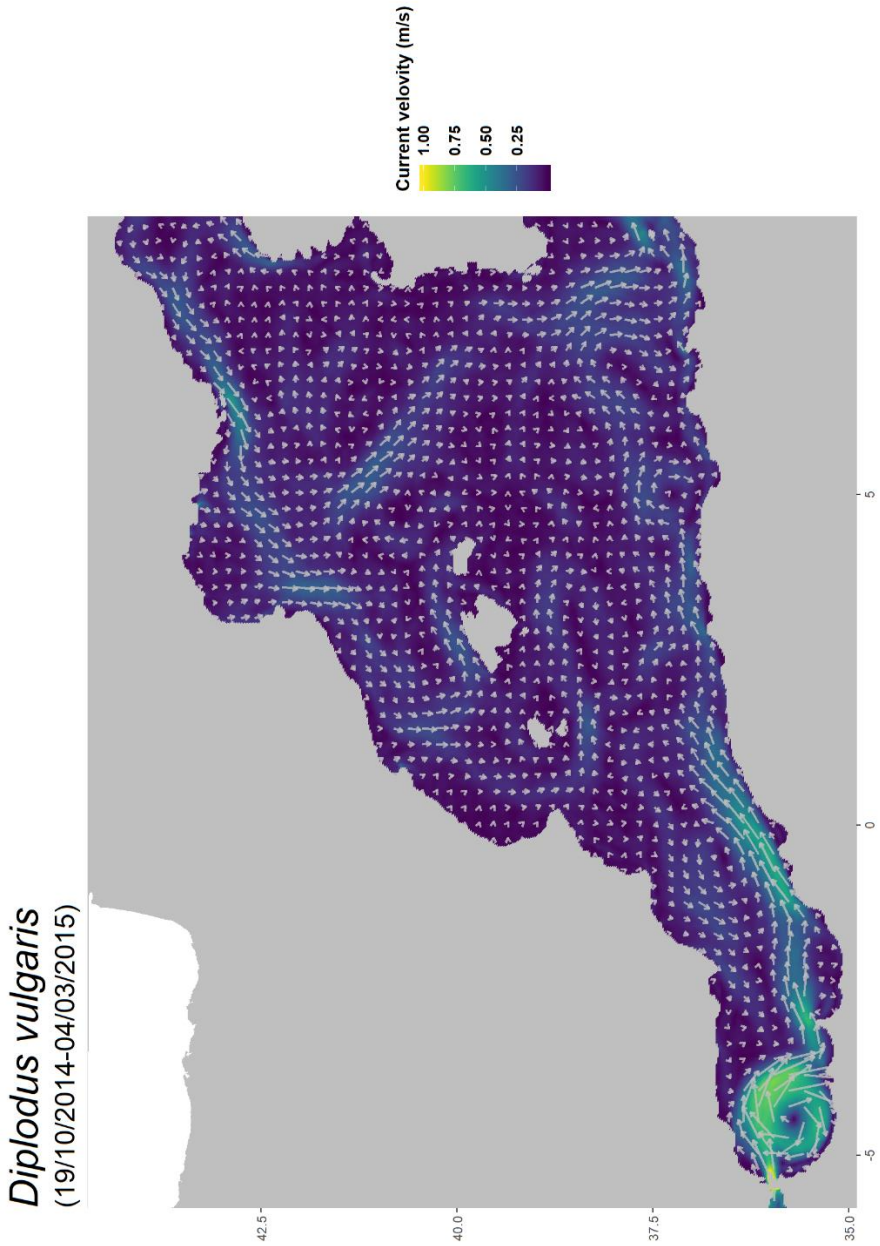


Figure S1 Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

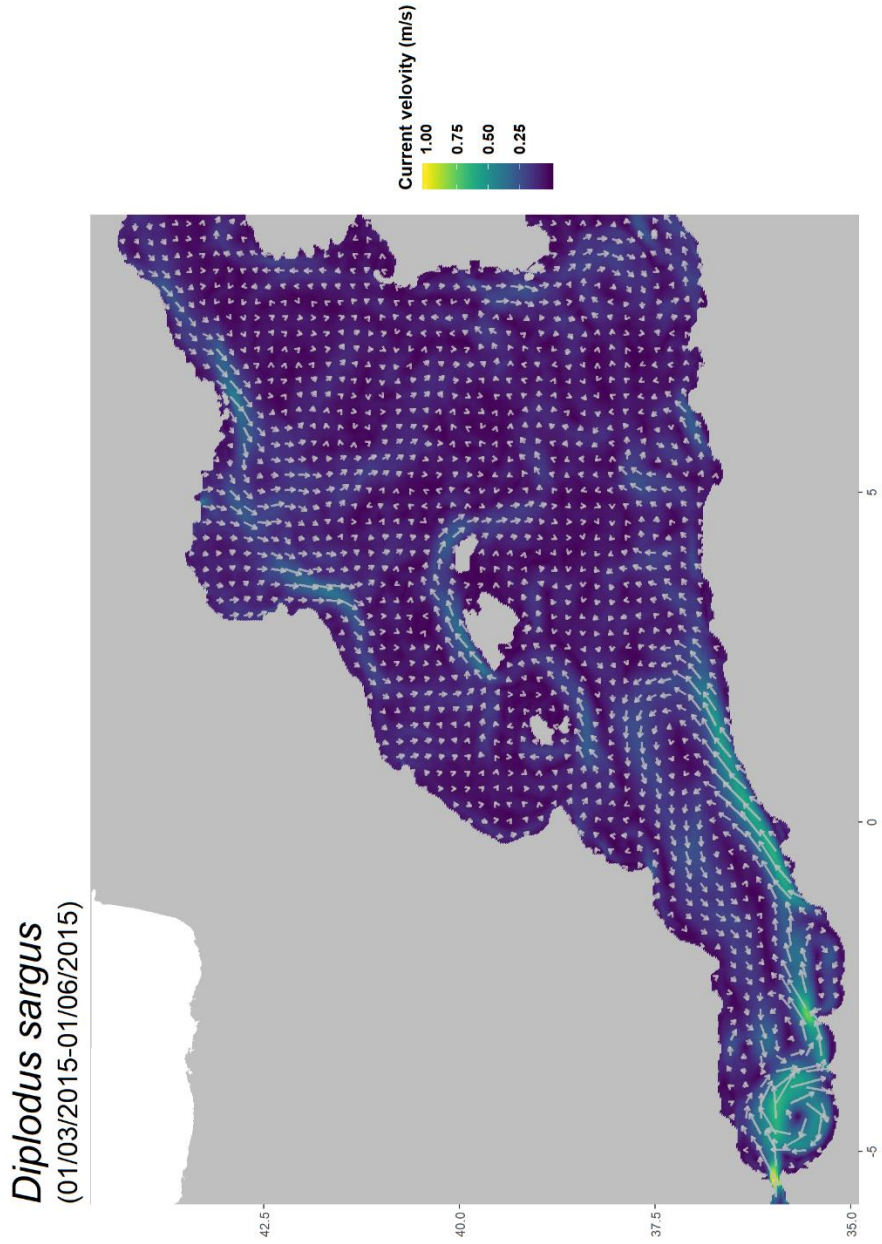


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

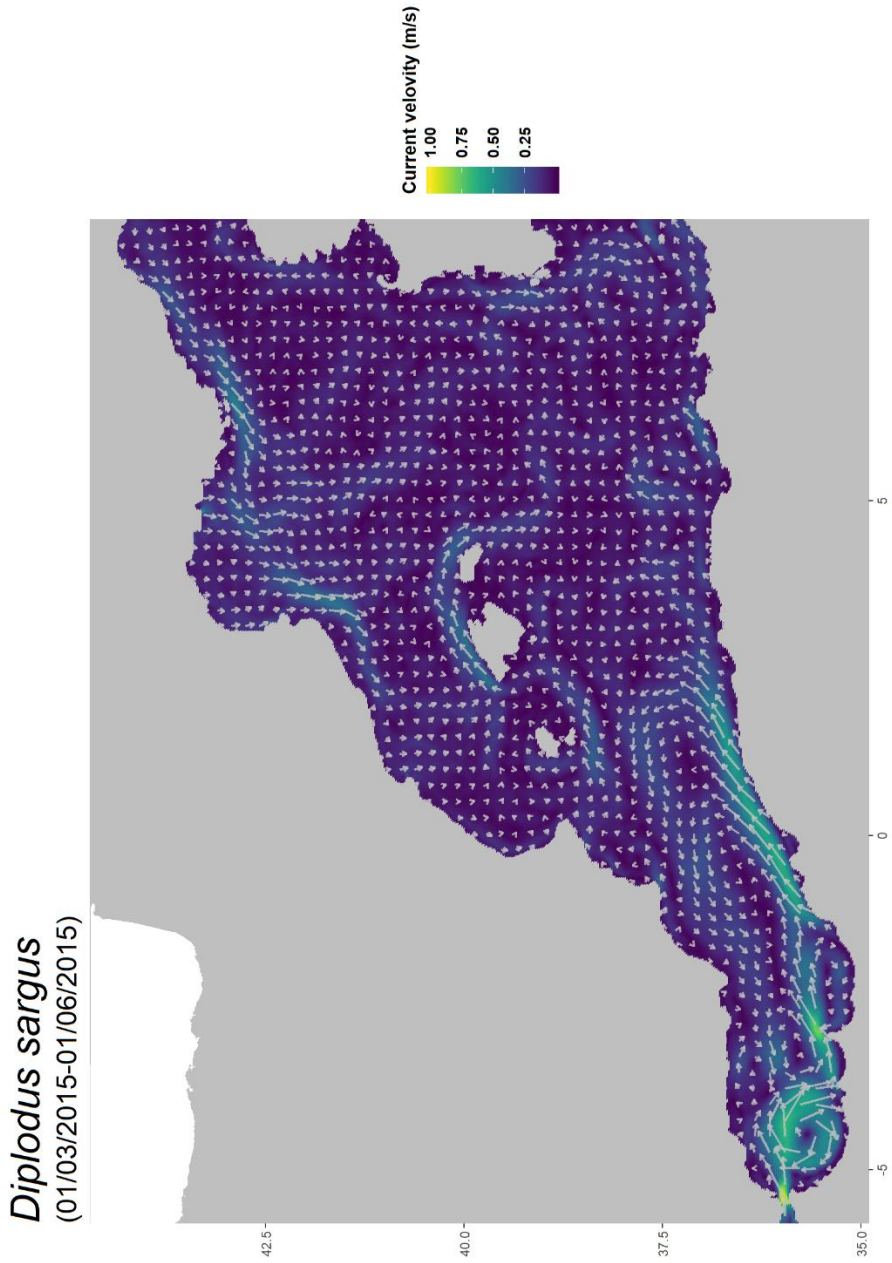


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

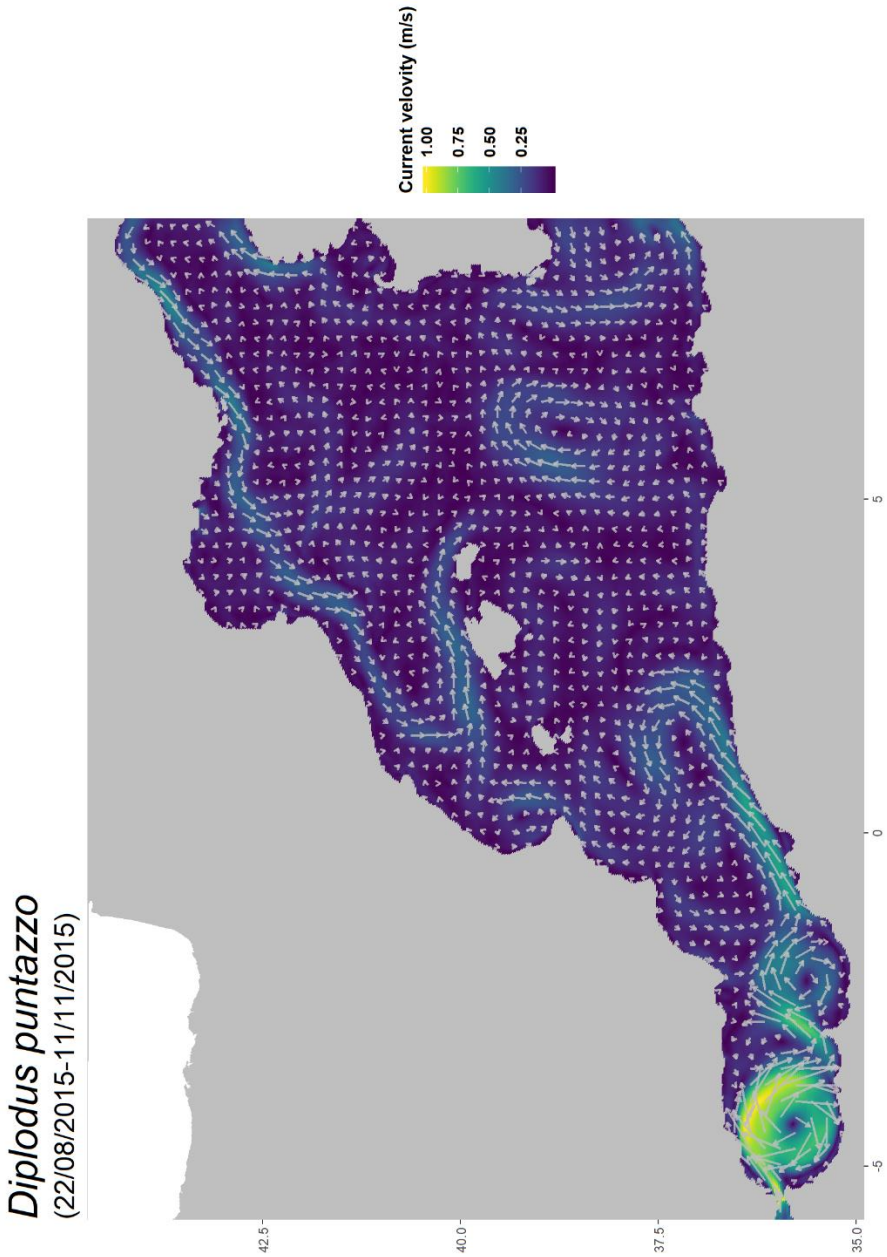


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

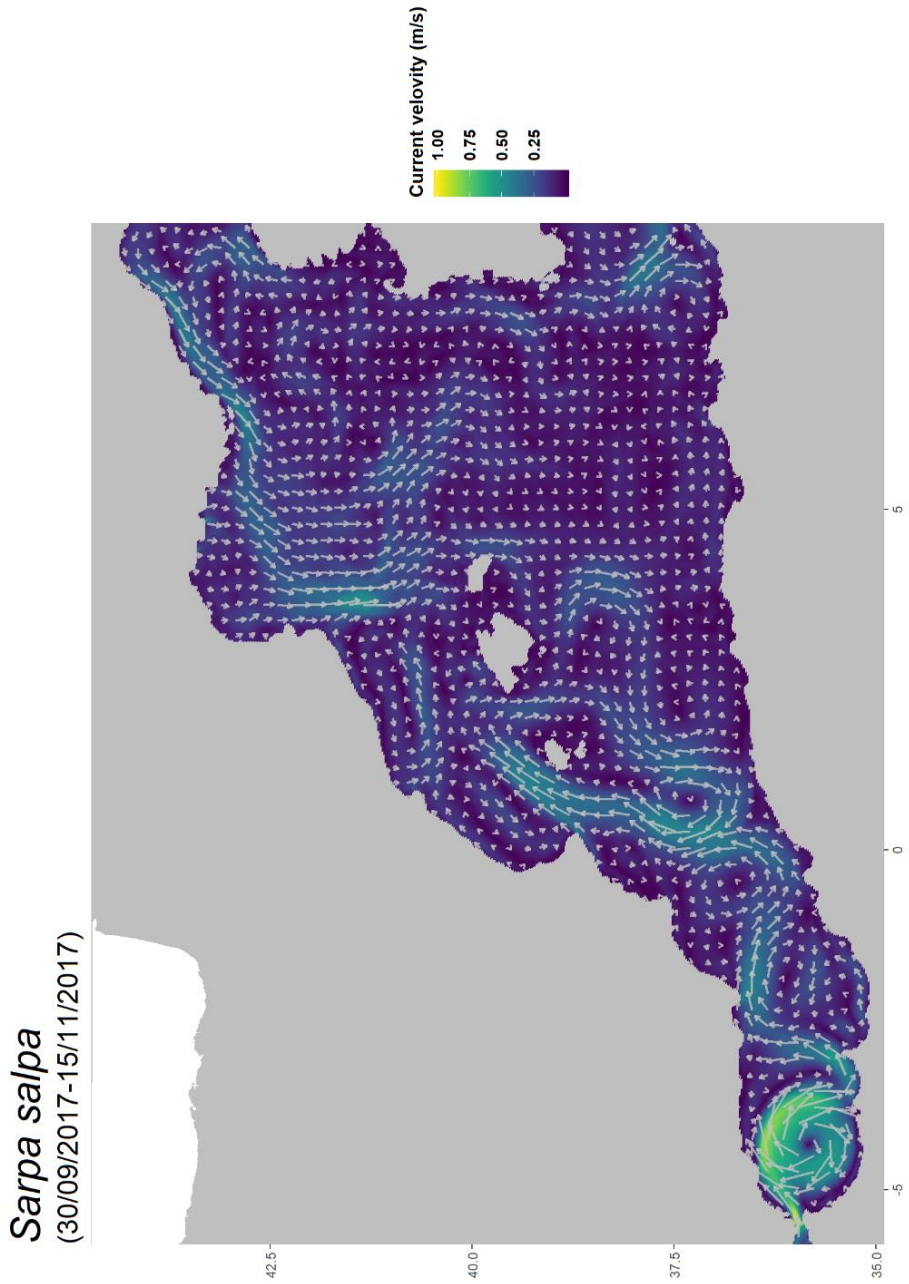


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

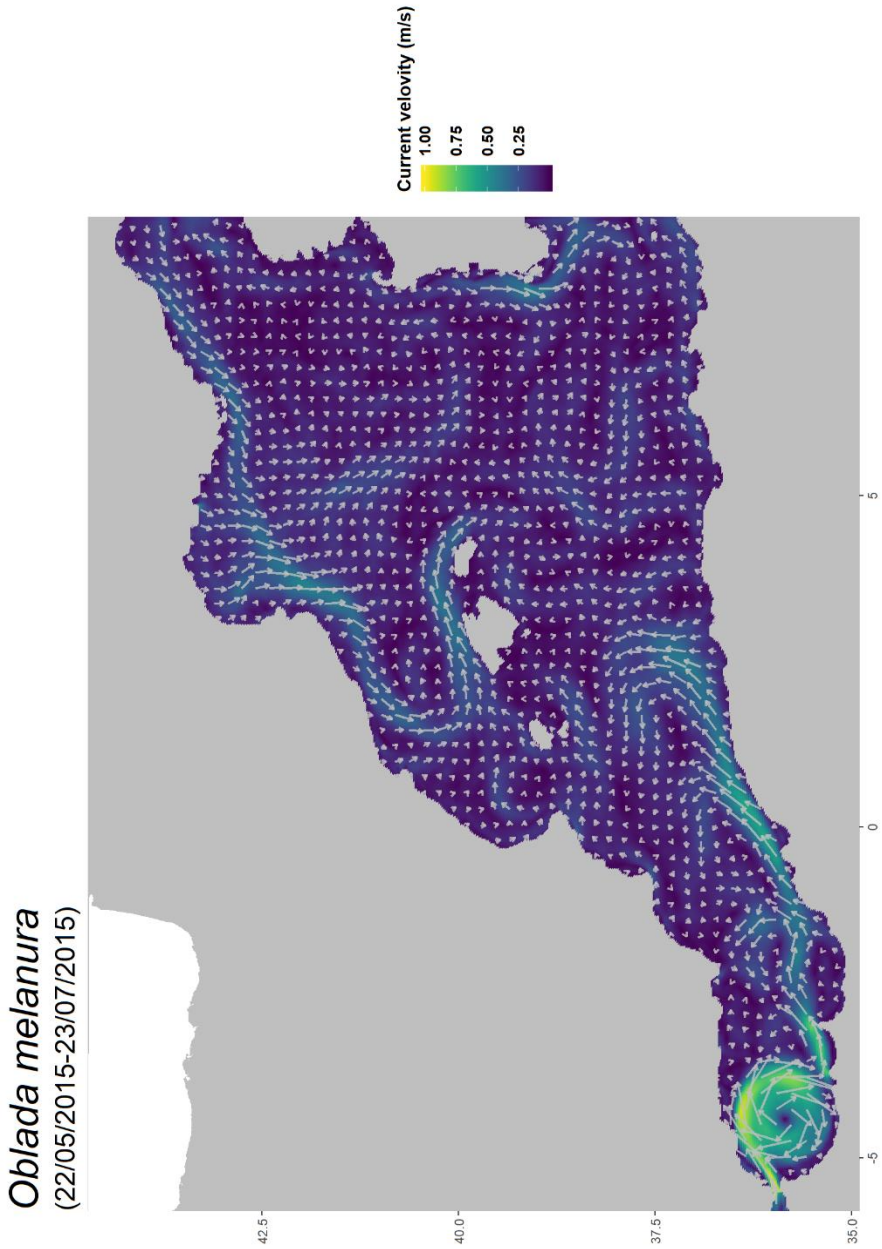


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

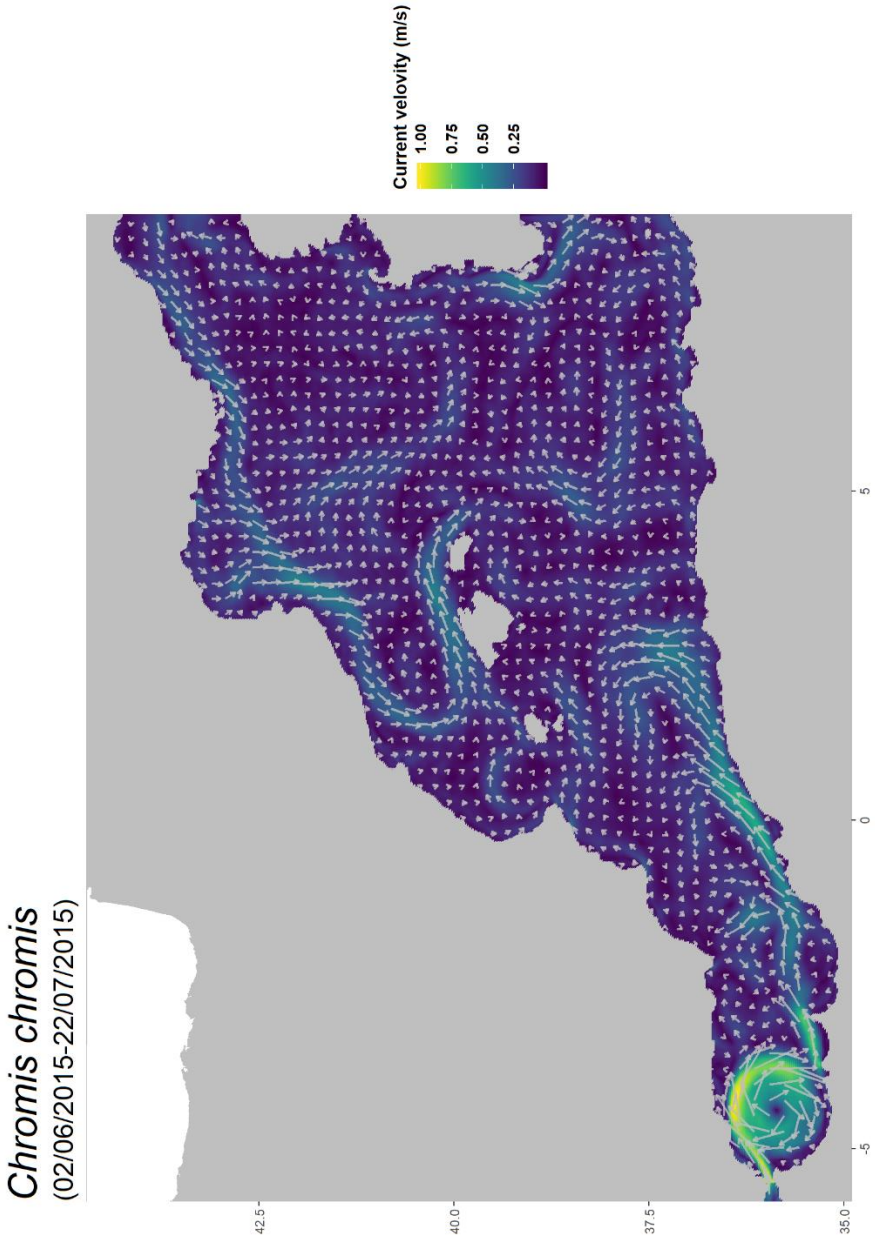


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

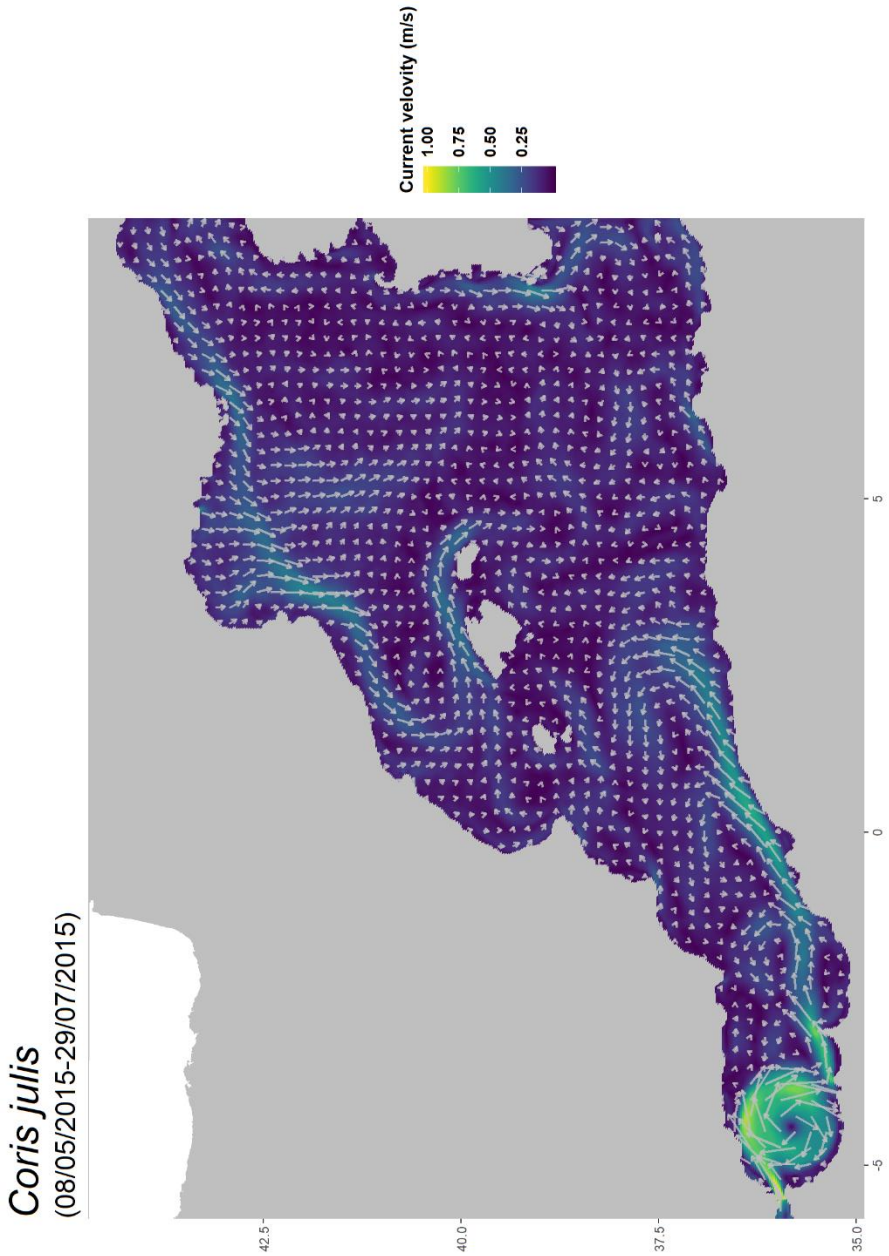


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

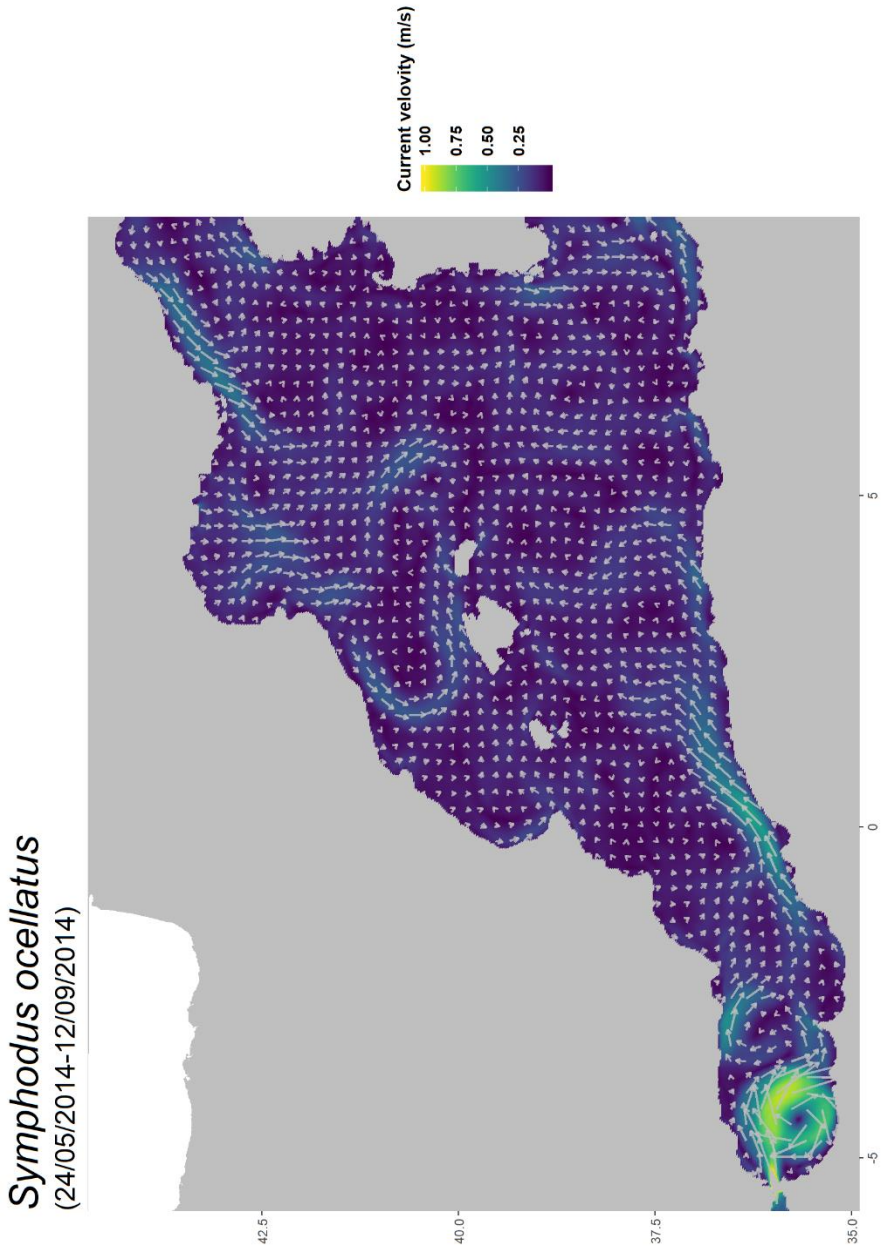


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

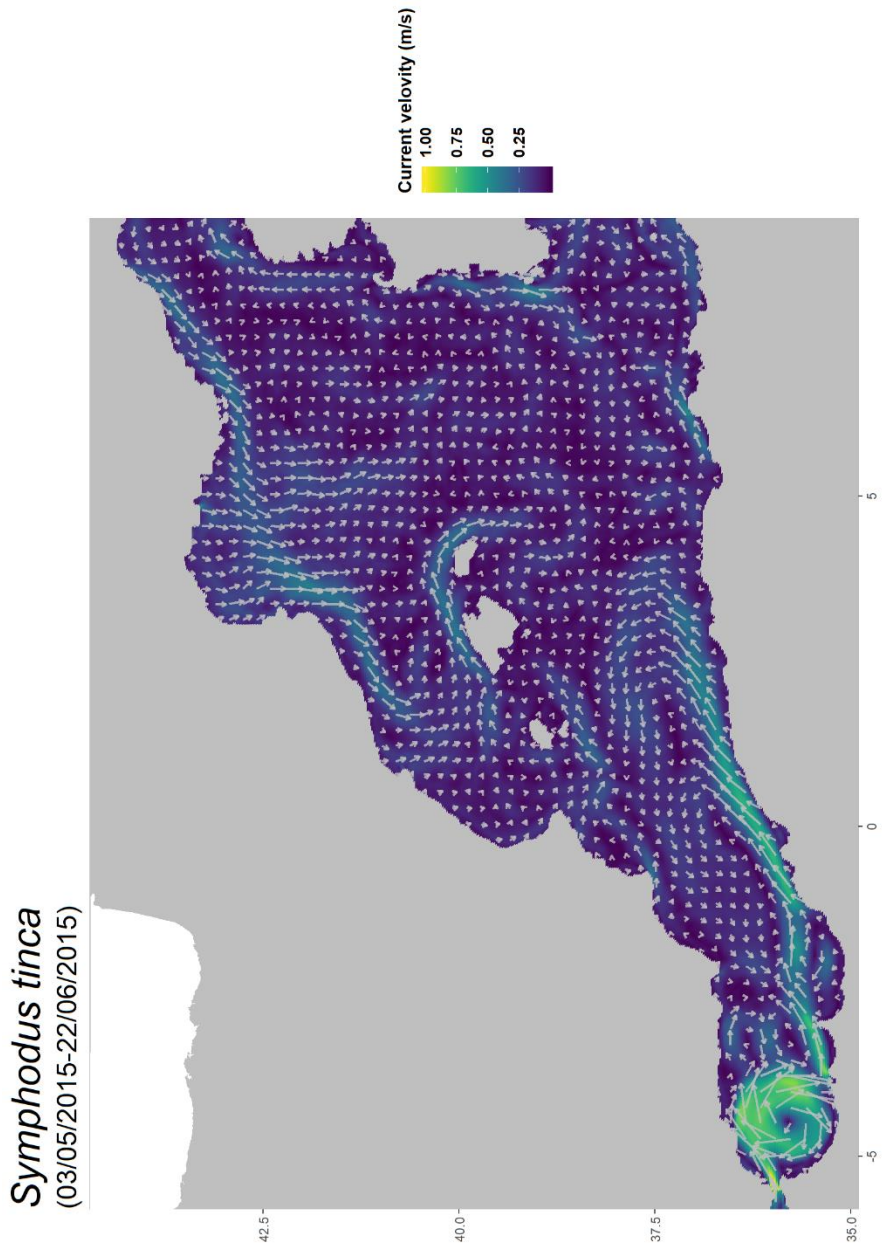


Figure S1(cont.) Average current velocity (colour) and current fields (arrows) modelled during the larval periods (in parentheses) of *Diplodus vulgaris*, *D. sargus*, *D. puntazzo*, *Sarpa salpa*, *Oblada melanura*, *Chromis chromis*, *Coris julis*, *Symphodus ocellatus* and *S. tinca*. Maximum velocity has been limited to 1m/s to improve visualization of slower currents.

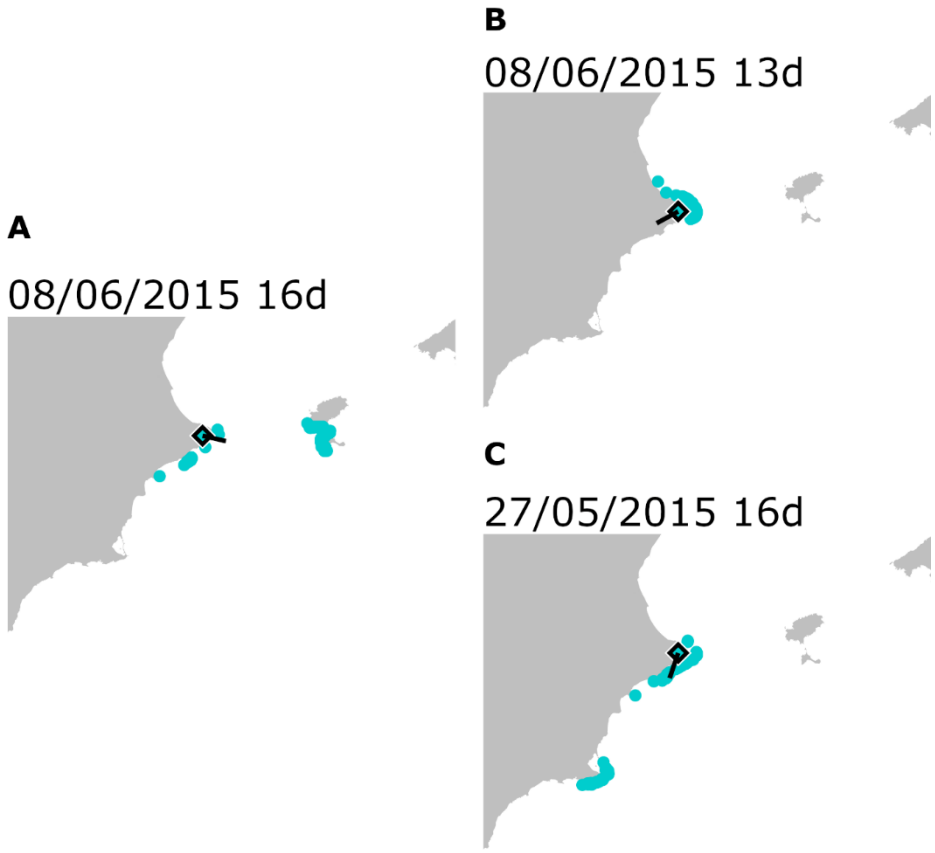


Figure S2. Detail of backtracking simulations for three individuals of *Oblada melanura* collected in Xabia. Dots represent simulated particles, diamonds represent the sampling locality used as starting point of backtracking simulations, and the black line indicates the average dispersal orientation of the simulated particles (*i.e.*, the individual's mean orientation). Particles are released in the nearest ocean point in the model to our sampling locality, so the land-directed orientation in B means high proportion of potential origins near this releasing point. (A) Individual hatching on 08/06/2015 with PLD 16 days (B) Individual hatching on 08/06/2015 with 13 days (C) hatching on 27/05/2015 with PLD 16 days.

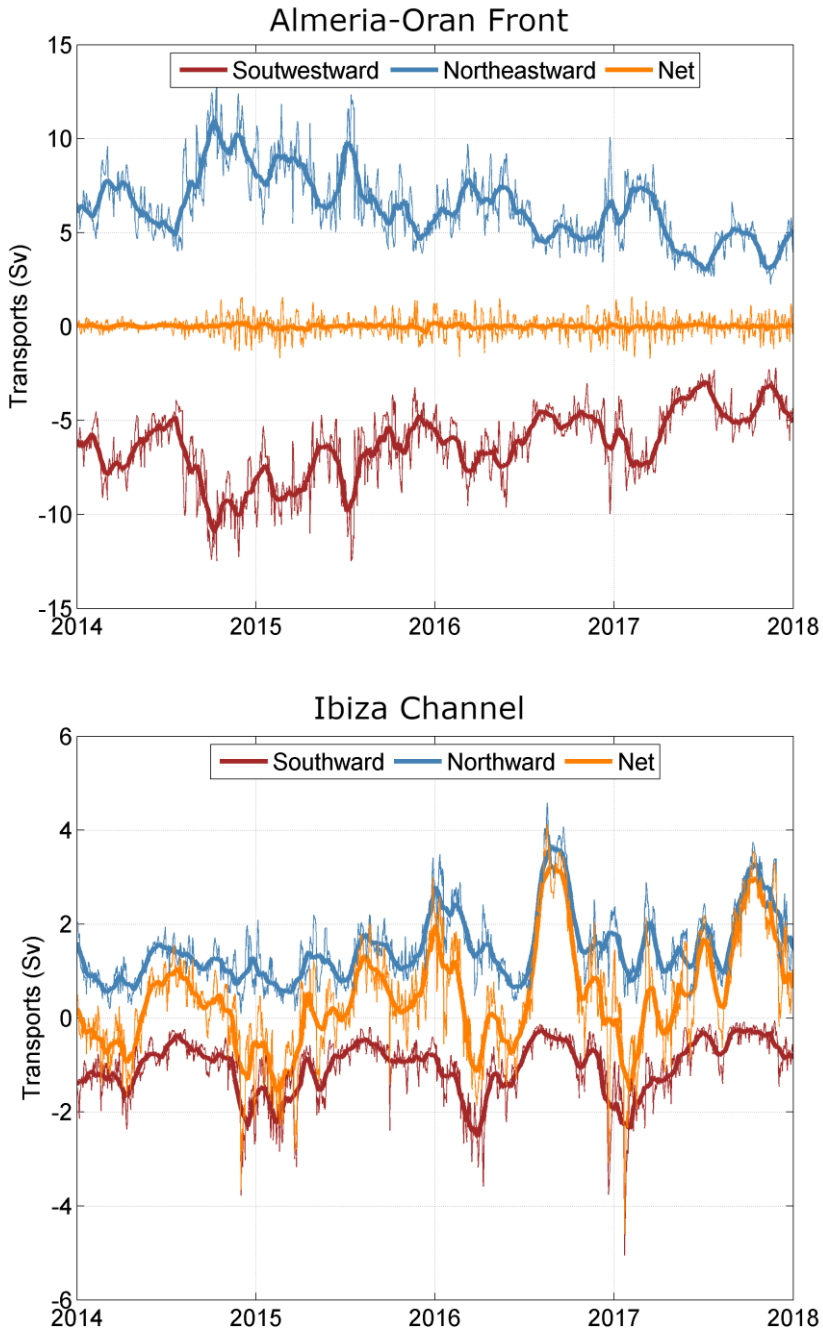


Figure S3. Model transports across Almeria-Oran front (a) and Ibiza Channel (b) between 01/01/2014 and 31/12/2017. While the thin line represents the daily variability, the thick line represents the time series after applying a 30-day moving average.

Table S1. Summary information of the nine modelled species. Number of total individuals used per species (N) in the backtracking simulations. Early life traits of the used individuals from otolith readings: Estimated minimum and maximum hatching date and PLD range (in days) per species. Mean larval depth in the water column (Depth). Larval critical swimming speed (Ucrit). Sampled localities and number of settlers used in each location. NA = samples not available.

Family	Species	N	Hatch dates (min-max) (DD/MM/YYYY)	PLD range	Depth (m)	Ucrit (cm s^{-1})	Colera	Blanes	Ametlla	Xabia	Cabo de Palos	Agua amarga	La Herradura
Sparidae	*Diplodus vulgaris*	174	19/10/2014-31/01/2015	22-43	13	6	29	31	29	NA	31	24	30
	Diplodus sargus	175	01/03/2015-15/05/2015	14-32	13	6	28	26	30	31	32	28	NA
	Diplodus puntazzo	206	22/08/2015-22/10/2015	16-29	13	6	29	30	29	30	30	30	28
	Sarpa salpa	72	30/09/2017-27/10/2017	12-24	13	6	NA	18	NA	18	NA	16	20
	Oblada melanura	216	22/05/2015-09/07/2015	12-20	13	6	32	32	30	31	29	30	32
Pomacentridae	*Chromis chromis*	146	02/06/2015-06/07/2015	13-21	0	12	29	31	29	NA	28	NA	29
Labridae	*Coris julis*	150	08/05/2015-01/07/2015	19-42	11.4	7.4	32	29	9	21	29	30	NA
	Symphodus ocellatus	177	02/06/2014-12/09/2014	7-14	0	7.4	36	42	NA	30	30	33	6
	Symphodus tinca	97	18/05/2015-22/06/2015	8-18	0	7.4	26	29	27	15	NA	NA	NA

Table S2. Frequency of backtracking simulations from each settlement area to each source coastal area for each species. The location of each area is represented in Figure 1. Acronyms of source areas are given in parentheses.

Species	Settlement areas	Source areas of the backtracking simulations						
		Liguro-Provençal Basin (LPB)	Balearic Sea (BS)	Balearic Islands (BI)	West Algerian Basin (WAB)	Algerian Coast (AC)	Alboran Sea (AS)	Atlantic Ocean (AO)
Diplodus vulgaris	Balearic Sea	0.4126	0.5703	0.0171	0	0	0	0
	West Algerian Basin	0	0.0034	0	0.8613	0.0739	0.0568	0.0047
	Alboran Sea	0	0.0001	0	0.0694	0.0033	0.5725	0.3546
Diplodus sargus	Balearic Sea	0.0453	0.9547	0	0	0	0	0
	West Algerian Basin	0	0.0708	0.0079	0.8845	0.0127	0.0242	0
Diplodus puntazzo	Balearic Sea	0.0664	0.9277	0.0059	0	0	0	0
	West Algerian Basin	0	0.3788	0.0016	0.5886	0.0290	0.0020	0.0001
	Alboran Sea	0	0	0	0.0009	0	0.5828	0.4163
Sarpa salpa	Balearic Sea	0.0002	0.9997	0.0002	0	0	0	0
	West Algerian Basin	0	0.1362	0	0.6986	0	0.1645	0.0008
	Alboran Sea	0	0	0	0	0	0.7501	0.2499
Oblada melanura	Balearic Sea	0.0338	0.9662	0	0	0	0	0
	West Algerian Basin	0	0.1299	0.0466	0.8064	0	0.0171	0
	Alboran Sea	0	0	0	0.0004	0	0.6787	0.3209
Chromis chromis	Balearic Sea	0.3669	0.6331	0	0	0	0	0
	West Algerian Basin	0	0	0	0.9440	0.0559	0	0
	Alboran Sea	0	0	0	0.0093	0.0052	0.9787	0.0067
Coris julis	Balearic Sea	0.1626	0.8373	0.0001	0	0	0	0
	West Algerian Basin	0	0.0583	0.0369	0.8985	0	0.0063	0
Symphodus ocellatus	Balearic Sea	0.0532	0.9468	0	0	0	0	0
	West Algerian Basin	0	0.1991	0	0.7966	0	0.0043	0
	Alboran Sea	0	0	0	0	0	0.9529	0.0471
Symphodus tinca	Balearic Sea	0.1632	0.8368	0	0	0	0	0
	West Algerian Basin	0	0.2716	0	0.7284	0	0	0

Table S3. Dispersal distances in Km for each species and hydrodynamic unit. In bold P-values correspond to significant differences in dispersal distances among hydrodynamic units assessed with Kruskal-Wallis tests.

Species	Hydrodynamic unit	min	max	mean	sd	Kruskal-Wallis test
Diplodus vulgaris	Balearic sea	242.3	709.6	492.7	105.6	**<0.001**
	Western Algerian basin	363.8	792.7	617.5	96.6	
	Alboran sea	494.4	986.3	638.8	99.1	
Diplodus sargus	Balearic sea	57.5	407.5	181.2	111.9	**0.007**
	Western Algerian basin	120.1	668.4	234.3	113.4	
Diplodus puntazzo	Balearic sea	158.4	463.1	288.3	83.3	**<0.001**
	Western Algerian basin	172.1	612.7	375.6	97.3	
	Alboran sea	326.5	1023.6	610.1	186.8	
Sarpa salpa	Balearic sea	144.2	363.7	209.2	52.7	**<0.001**
	Western Algerian basin	82.6	495.2	187.5	77.3	
	Alboran sea	222.6	568.1	279.9	84.5	
Oblada melanura	Balearic sea	82.8	395.5	242.3	73.2	**<0.001**
	Western Algerian basin	99.9	285	189.6	55.9	
	Alboran sea	251.7	613.4	361.6	81.4	
Chromis chromis	Balearic sea	106.4	391.6	263.5	63.9	**<0.001**
	Western Algerian basin	174.8	430.5	287	85.8	
	Alboran sea	281.4	493.5	371.4	56.0	
Coris julis	Balearic sea	274	558.2	422.8	71.7	**<0.001**
	Western Algerian basin	232.1	437.4	336.7	52.9	
Symphodus ocellatus	Balearic sea	71.3	248.4	141.1	35.4	**<0.001**
	Western Algerian basin	44.2	289.4	106	45.5	
	Alboran sea	187.7	396.7	311.5	73.6	
Symphodus tinca	Balearic sea	40.1	332.1	162.5	81.0	0.177
	Western Algerian basin	78.4	153.9	117	24.1	

Table S4. Results for interannual variability in oceanographic transport during 2014-2017. Bold P-values correspond to significant diferences among years, assessed with Kruskal-Wallis tests.

		AOF		IC	
	df	K-W chi-squared	p-value	K-W chi-squared	p-value
Diplodus vulgaris	3	2.642	0.45	104.129	**<0.001**
Diplodus sargus	3	3.675	0.299	53.9	**<0.001**
Diplodus puntazzo	3	1.825	0.61	161.811	**<0.001**
Sarpa salpa	3	1.231	0.746	116.035	**<0.001**
Oblada melanura	3	3.916	0.271	107.626	**<0.001**
Chromis chromis	3	2.229	0.526	91.722	**<0.001**
Coris julis	3	1.572	0.666	94.385	**<0.001**
Symphodus ocellatus	3	1.569	0.666	10.058	**0.018**
Symphodus tinca	3	3.088	0.378	69.631	**<0.001**

Capítulo 2:

Estudio genómico de las poblaciones de peces en el Mediterráneo occidental: estructuración poblacional, adaptación y metodologías

Capítulo 2.1

Individual-based population genomics reveal different drivers of adaptation in sympatric fish

Héctor Torrado[1, 2] *, Carlos Carreras [2], Núria Raventós [1], Enrique Macpherson [1+], Marta Pascual [2+]

[1]Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Car. Acc. Cala St. Francesc 14, 17300 Blanes, Girona, Spain.
[2]Department de Genètica, Microbiologia i Estadística and IRBio, Universitat de Barcelona, Av.Diagonal 643, 08028 Barcelona, Spain.
+Both authors contributed equally as senior researchers and should be considered to be at the same position

Published as:
Héctor Torrado, Carlos Carreras, Núria Raventós, Enrique Macpherson, Marta Pascual (2020) Individual-based population genomics reveal different drivers of adaptation in sympatric fish. Scientific Reports 10:12683. https://doi.org/10.1038/s41598-020-69160-2

Abstract

Connectivity and local adaptation are two contrasting evolutionary forces highly influencing population structure. To evaluate the impact of early-life traits and environmental conditions on genetic structuring and adaptation, we studied two sympatric fish species in the Western Mediterranean Sea: *Symphodus tinca* and *S. ocellatus*. We followed an individual-based approach and measured early-life history traits from otolith readings, gathered information on environmental variables and obtained genome-wide markers from genotyping-by-sequencing (GBS). The two species presented contrasting population structure across the same geographic gradient, with high and significant population differentiation in *S. ocellatus*, mostly determined by oceanographic fronts, and low differentiation and no front effect in *S. tinca*. Despite their different levels of genetic differentiation, we identified in both species candidate regions for local adaptation by combining outlier analysis with

environmental and phenotypic association analyses. Most candidate loci were associated to temperature and productivity in *S. ocellatus* and to temperature and turbulence in *S. tinca* suggesting that different drivers may determine genomic diversity and differentiation in each species. Globally, our study highlights that individual-based approach combining genomic, environmental and phenotypic information is key to identify signals of selection and the processes mediating them.

Keywords: adaptation, genotyping-by-sequencing, early-life traits, population genomics, association analyses, individual-based analyses

Introduction

Population structure is highly influenced by the level of connectivity among localities, driven by dispersal potential and barriers to gene flow, but also shaped by local adaptation. In this context, environmental conditions can impose spatially differential genomic selective pressures on phenotypic traits and their combined study can allow to forecast the species adaptive capacity [1]. The use of environmental information coupled with the genetic structure of species can be essential for understanding how they are adapted to their habitats [2]. Some studies have investigated the relationship between local adaptation and environmental conditions in several marine species with temperature, salinity and productivity values being important factors affecting different biological processes and genetic differentiation at the population level [3–6]. For instance, thermal gradients observed at sea can explain the distributional patterns of numerous species, their diversification processes and the potential influence of climate changes [7].

Population genomic studies are fundamental to identify patterns of genetic structure and the processes determining them. Genotyping-by-sequencing (GBS) [8], provides a cost-effective approach for population genomic studies on non-model species. This methodology provides high density of markers that may allow identifying non-neutral genomic signatures related to adaptation processes, with the potential for

annotation [6,9]. The assessment of genomic signals of adaptation on non-model organisms is usually performed in two different ways, outlier analyses (OAs) and genotype-environment association analyses (EAAs) [2]. On one hand, OAs are based on the search of genome-wide markers showing outlier values of genetic differentiation (F_{ST}) among the studied populations [10]. On the other hand, EAAs using population genomic data combined with environmental data have been successfully used to discover polymorphisms involved in adaptation to environmental conditions [3,6]. However, individual-based information could be used to further refine those associations. Individual-based information is used in genome-wide association studies (GWAS), which can identify signals of adaptation to both environmental conditions (EAAs) and phenotypic traits through phenotype association analysis (PAAs). This methodology has proven to be a powerful tool to study interesting phenotypic traits and adaptation in model species such as humans, plants and captive fish [11–13]. While the use of individual-based information is common in model species, the application in wild populations is rare and has been only used in a few studies combined with environmental data information [14]. Up to now, there is a lack of individual-based studies in natural marine populations, preventing an accurate knowledge of the relationships between genomic data, environment and early life phenotypes in determining population structure in environmental gradients. This approach can be undertaken in species were individual based information during dispersal phases can be obtained, such as benthic fishes.

In numerous Mediterranean benthic fishes, the pelagic larval stage allows dispersal and maintains connectivity between populations, whereas adults are territorial and present a high site fidelity [15–18]. The length of the pelagic larval duration (PLD) can not only affect dispersal but also other fitness characteristics such as larval growth, size at settlement and mortality rate, modulating the year-class strength in numerous species [19,20]. Moreover, environmental conditions can have major effects during the larval and settlement periods, where genetic and phenotypic polymorphism can be affected by selection and/or stochastic factors [15]. Fortunately, in most fishes, individual-based information on both phenotypic and environmental early-life

variables can be obtained from otoliths (bones in the inner part of the ear) and has been successfully combined with genomic data to infer fine-scale dispersal [21]. Therefore, the study of individually based genomic variation associated to early-life history traits across geographic gradients in sympatric fish can provide insights into the adaptive potential of species while considering both connectivity and selection.

Here we use an individual-based approach to evaluate the phenotypic and environmental drivers shaping population genomic structure in two congeneric fish, *Symphodus tinca* and *S. ocellatus,* that are sympatric across a geographic/environmental gradient in the Western Mediterranean Sea. We selected two closely related species to ensure that any difference found between the two species on the drivers shaping population structure is not the result of a phylogenetic signal of divergent lineages, but a species-specific differential adaptation signal. On the other hand, adaptation signals on marine organisms may change depending on the environmental gradient present on the sampling sites and the scale of analysis[6]. For this reason, a sympatric distribution is key to check for a differential adaptation strategy, as both species are going to be under the same environmental gradient and face the same oceanographic fronts. We analysed 303 individuals from these two species with genotyping by sequencing, measured from otolith readings some individualized key early-life history traits (e.g. date of hatching, pelagic larval duration, size at settlement, growth rate) and gathered information of environmental variables during their pelagic larval duration (e.g. surface temperature, turbulence, and productivity). In order to test if phenotypic and environmental variables differentially influence the population structure in these two species, we i) gathered individual-based information on early-life phenotypic and environmental variables, ii) assessed population differentiation, iii) identified loci putatively under selection using different methodological approaches and iv) evaluated the effect of phenotypic and environmental factors in structuring populations across geographic gradients. We expected to find similar associations and genomic structuring for both species due to their similar distribution and phylogenetic closeness. Nevertheless, considering their

differences in reproduction times, temporal environmental variation and behaviour could change population connectivity and the drivers of population structure.

Results

Phenotypic and environmental analyses

We obtained reliable information on early-life history traits from otolith reading (Figure S1) for 93% and 89% of the juvenile individuals from most of the localities in *S. ocellatus* and *S. tinca*, respectively (Figure 1, Table 1). With these data we acquired environmental and phenotypic individually based values (Table 2). Significant differences among sampling locations were detected with PERMANOVA for all phenotypic and environmental variables with three exceptions: PLD growth rate and Hatching date in *Symphodus ocellatus* and PLD in *S. tinca* (Table S1). The temperature during the larval stage was lower in northern localities for both species, even considering a delay on the hatching date in some of these localities to encounter warmer dates (Table 2).

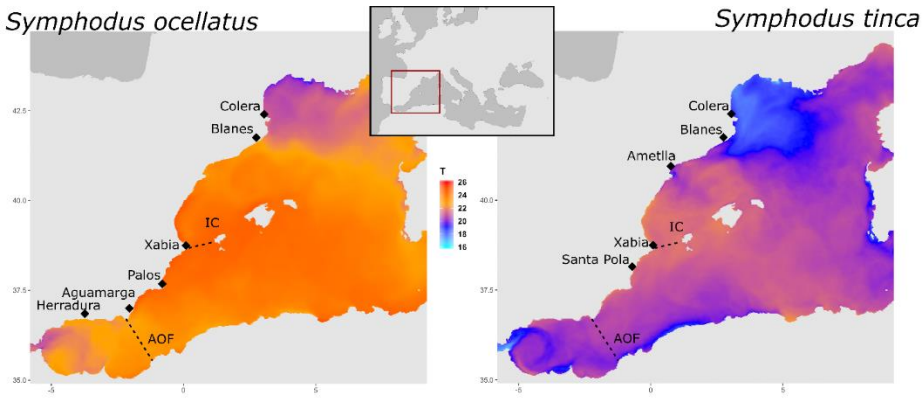


Figura 1: Map of the sampled locations in the Western Mediterranean for *Symphodus ocellatus* and *S. tinca*. Dashed lines indicate oceanographic discontinuities: Ibiza channel (IC) and Almeria-Oran front (AOF). The colour gradient of the sea identifies the mean sea surface temperature (ºC) during the larval period of each species. The reduced grey map shows the location of our study area within the Mediterranean Sea. Data obtained from Western Mediterranean OPerational forecasting system (WMOP) [63].Maps were plotted using the R[60] packages 'ncdf4'[61] and 'ggplot2' [62].

For *S. ocellatus,* individuals from the colder areas (Figure 1), the Alboran Sea and the two northernmost localities, delayed hatching date and thus experienced milder

temperatures during PLD than otherwise expected (Table 2). In the case of *S. tinca*, the individuals of colder localities (Figure 1) did not always delay their hatching date. For instance, in Blanes, individuals hatched early in the season and thus experienced low temperatures and had longer PLD (Table 2). The information of multiple variables was synthesized with principal component analysis (PCA). The PCA yielded several components for the phenotypic (PC) and environmental (EC) variables, with different contribution from each variable, and overall explaining a large proportion of the variance (Table S2).

Population genomics

We analysed 303 individuals from the two sympatric *Symphodus* species with GBS (Table 1, Figure 1). We obtained a mean (±SD) of 2.6±0.8 and 2.6±0.9 million reads per individual for *S. ocellatus* and *S. tinca,* respectively. After filtering, we found 3978 polymorphic haplotype loci and 5123 SNP loci for *S. ocellatus* and 5276 polymorphic haplotype loci and 6833 SNP loci for *S. tinca*. The mean depth per locus was 19.5±11.4 for *S. ocellatus* and 19.5±11.3 for *S. tinca*. The maximum number of alleles per haplotype locus was 17 for *S. ocellatus* and 15 for *S. tinca*. Observed and expected heterozygosities per locality were similar in the two species (Table 1).

For *S. ocellatus*, pairwise F_{ST} values (Table S3) as well as the DAPC plot (Figure 2) and k-means clustering suggest the same structure identifying three groups. One group was formed by the northern localities (Colera – Blanes), another included the localities between the IC and AOF oceanographic discontinuities (Xabia, Palos and Aguamarga) as a central group, and the last one was constituted by the southernmost locality (Herradura) sampled in the Alboran Sea (Figure 1). It is worthwhile noticing that Herradura clusters with the localities of the central group in the first axis. However, we did not find isolation by distance with a Mantel test (r=0.45, p-value=0.08). For *S. tinca* no clear structuring was detected (Figure 2, Table S4), and only two significant F_{ST} values were found, when Santa Pola was compared to Blanes and to Ametlla. When comparing only the three localities sampled in both species, we observed similar groupings in the DAPC analysis as when considering all sampled localities, with a clear

effect of the IC discontinuity genetically differentiating Xabia from the northern localities in *S. ocellatus* but not in *S. tinca* (Figure 2). As for *S. ocellatus* we did not find isolation by distance with a Mantel test (r= -0.38, p-value=1).

Table 1: Sampling information and population diversity values for the locations of *Symphodus ocellatus* and *S. tinca*. Sampling date, Number of genotyped individuals (N) and in parentheses the number of juvenile individuals with otoliths information, Allelic richness (Ar), Mean observed (H_o) and expected (H_e) heterozygosities, Inbreeding coefficient (F_{IS}). All F_{IS} values are significant. The names of localities in parentheses indicate the toponymal names of the corresponding localities. Diversity values were obtained from the haplotype loci datafiles.

Species	Location	Sampling date	N	Ar	Ho	He	F_{IS}
Symphodus ocellatus	Colera 42°24'26.9"N 3°09'23.8"E	June-August 2014	30 (30)	2.283	0.302	0.362	0.164
	Blanes 41°40'44.5"N 2°48'28.8"E	June-August 2014	37 (37)	2.284	0.296	0.360	0.177
	Xabia 38°47'46.1"N 0°11'01.4"E	May-July 2014	28 (26)	2.255	0.281	0.354	0.207
	Palos (Cabo de Palos) 37°38'05.4"N 0°41'31.6"W	May-August 2014	26 (24)	2.229	0.285	0.350	0.185
	Aguamarga 36°56'20.8"N 1°55'53.1"W	May-August 2014	28 (27)	2.242	0.296	0.352	0.158
	Herradura (La Herradura) 36°43'40.0"N 3°44'07.9"W	July-August 2016	13 (6)	2.224	0.284	0.352	0.193
Symphodus tinca	Colera 42°24'26.9"N 3°09'23.8"E	May-June 2015	30 (24)	2.600	0.298	0.356	0.164
	Blanes 41°40'44.5"N 2°48'28.8"E	May 2014	31 (27)	2.588	0.293	0.355	0.175
	Ametlla (L'Ametlla de Mar) 40°52'27.1"N 0°47'42.3"E	May-June 2015	26 (23)	2.597	0.301	0.356	0.155
	Xabia 38°47'46.1"N 0°11'01.4"E	May-June 2015	28 (28)	2.601	0.293	0.356	0.178
	Santa Pola 38°11'27.9"N 0°33'57.2"W	2015	26 (0)	2.592	0.293	0.355	0.173

Tabla 2: Mean values and standard deviation of individually based environmental and phenotypic variable values by location. No environmental and phenotypic variable is available for Santa Pola in *Symphodus tinca*, since juveniles with otolith information were not sampled in this locality.

Species	Location	Planktonic Larval Duration (PLD) (days)	Settlement size (µm)	PLD Growth rate (µm/day)	Hatching date (day of the year)	Surface temperature (ºC)	Productivity (ChlA mg/m³)	Turbulence (m)
Symphodus ocellatus	Colera	10.1 ± 1.0	45.4 ± 4.6	4.5 ± 0.6	192.8 ± 10.8	21.8 ± 0.4	0.211 ± 0.005	1.40 ± 0.24
	Blanes	9.3 ± 1.0	48.7 ± 6.2	5.3 ± 0.6	193.4 ± 13.5	22.5 ± 0.9	0.164 ± 0.009	0.67 ± 0.18
	Xabia	9.3 ± 1.2	43.3 ± 7.1	4.7 ± 0.6	186.0 ± 15.8	24.1 ± 1.2	0.171 ± 0.004	0.87 ± 0.08
	Palos	8.6 ± 0.8	42.0 ± 3.8	4.9 ± 0.6	184.2 ± 10.2	23.8 ± 0.9	0.163 ± 0.016	0.97 ± 0.08
	Aguamarga	9.3 ± 1.3	43.8 ± 4.8	4.8 ± 0.6	188.6 ± 19.0	23.1 ± 1.9	0.158 ± 0.003	0.83 ± 0.11
	Herradura	9.3 ± 0.8	46.0 ± 5.6	5.0 ± 0.6	229.7 ± 11.7	23.9 ± 0.3	0.261 ± 0.030	0.90 ± 0.22
Symphodus tinca	Colera	9.3 ± 0.9	33.0 ± 2.4	3.6 ± 0.4	159.9 ± 4.9	20.1 ± 0.9	0.306 ± 0.039	1.02 ± 0.23
	Blanes	12.4 ± 1.4	40.9 ± 4.0	3.3 ± 0.2	135.7 ± 5.7	16.5 ± 0.1	0.335 ± 0.011	0.72 ± 0.11
	Ametlla	9.0 ± 1.0	35.8 ± 4.0	4.0 ± 0.4	157.0 ± 5.2	22.0 ± 1.3	0.333 ± 0.066	0.52 ± 0.14
	Xabia	9.9 ± 1.1	42.6 ± 4.6	4.3 ± 0.4	149.0 ± 3.6	21.0 ± 0.6	0.201 ± 0.019	0.59 ± 0.10

In the outlier analysis with Bayescan after correcting for multiple comparisons, we identified 95 outlier haplotype loci potentially under positive selection for *S. ocellatus* and 5 for *S. tinca* (Tables S5 and S6).

Genomic associations to phenotypic and environmental variables

The variance partition analysis using the SNP loci datafile showed a small contribution of environmental and phenotypic variables in explaining the whole genomic variance in both species, although higher for environmental variables (Table

S7). The redundancy analysis (RDA) for *S. ocellatus*, using the individual environmental values and the genotypes from the SNP loci datafile, identified two clusters of individuals along the first axis, explaining 58.0% of the variation (Figure 3). All individuals of the central group (Xabia, Palos and Aguamarga) clustered together and were positively associated with temperature and negatively associated with productivity while those from the other three localities (Colera, Blanes and Herradura) clustered also together but presented the opposite trend. Overall, 155 haplotype loci had SNPs significantly correlated with environmental predictors (Table S8) with only one locus (20878) being associated to more than one predictor, although with different SNPs (Table S5).

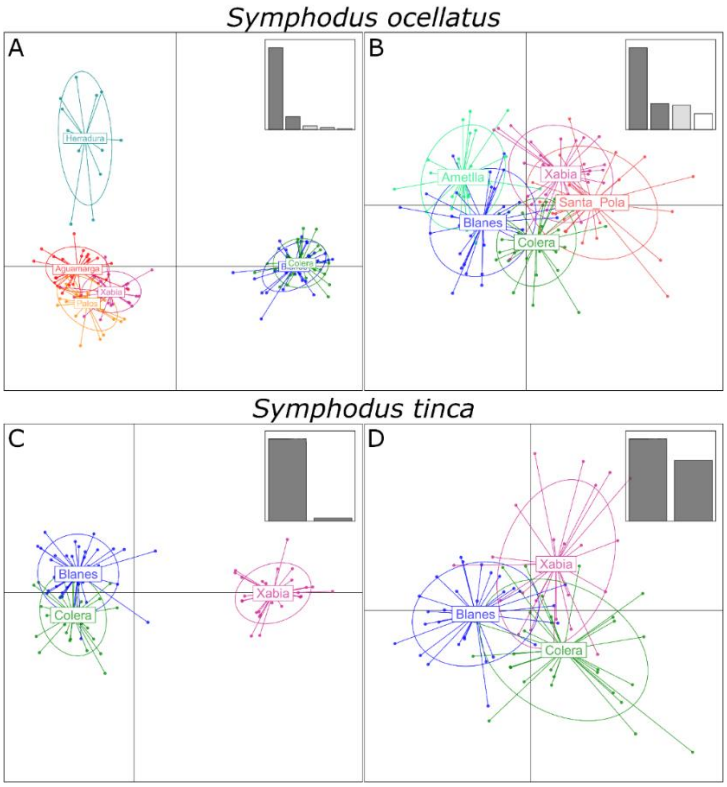


Figura 2: Discriminant Analysis of Principal Components (DAPC). DAPC plots with DA eigenvalues (top right corner) using all analysed localities for each species (A and B) or only the localities sampled in both species (C and D).

For *S. tinca*, the two axes of the RDA explained 67.4% of the variation and separated the individuals by locality (Figure 3). The first axis was associated mainly to

productivity while the second axis was associated to temperature and turbulence with a different sign. Consequently, Ametlla and Xabia were associated to higher temperature and low turbulence, while Xabia and Blanes seemed to have less productivity. Overall, 166 haplotype loci had SNPs significantly correlated with environmental predictors (Table S8) with only three haplotype loci (12921, 24436, 9609) associated to more than one predictor, although in all cases with different SNPs (Table S6) as in *S. ocellatus*.

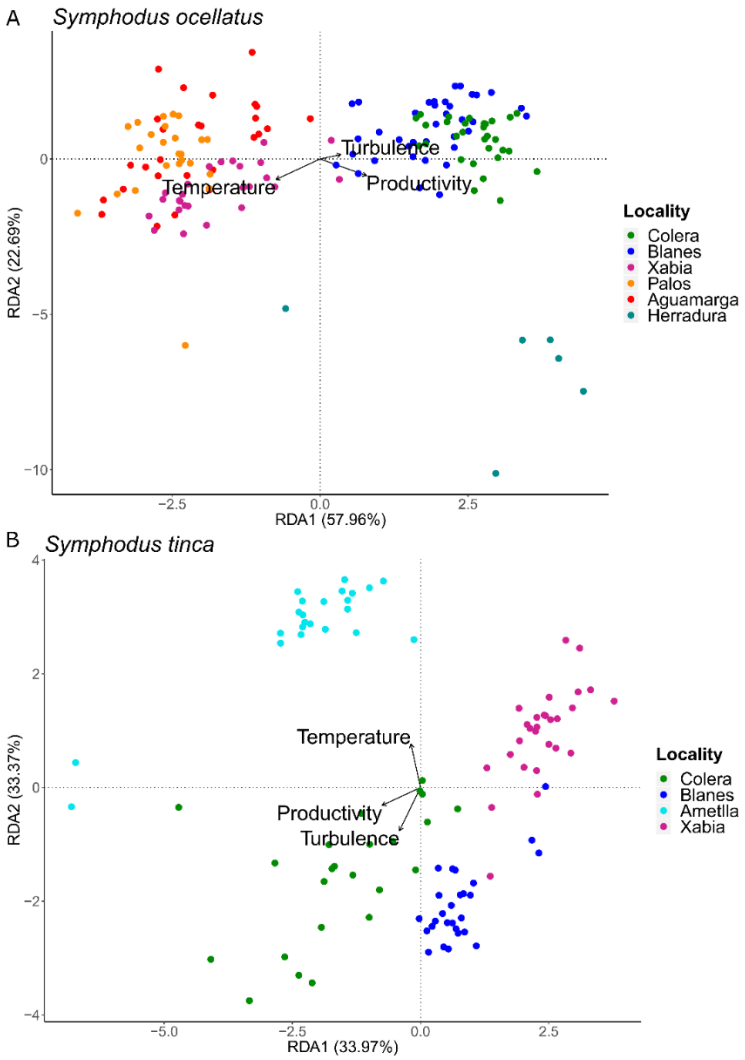


Figura 3: Redundancy analysis (RDA). RDA performed with individual environmental data and SNPs at all loci. Coloured points represent individual fish coded by locality. Vectors indicate environmental predictors of the two first RDA components. A) *Symphodus ocellatus* and B) *S. tinca*.

The GWAS using as covariates the two first components of the MDS analyses (GWAS-Cov) identified for *S. ocellatus* only one SNP associated to the first environmental component (Table S8). However, when performing the analysis without covariates 207 haplotype loci presented SNPs significantly associated to environmental and phenotypic variables. Most of the SNPs were associated to environmental variables, particularly productivity and temperature and 63 of them were related to more than one variable (Table S5). For phenotype-genotype associations, 21 SNPs were associated to the composite phenotypic variable PC2, positively related to PLD and settlement size (Table S2). In the case of *S. tinca*, GWAS-Cov found six SNPs with associations to phenotypic and environmental variables (Table S8). The same SNPs were associated to the same variables when covariates were not used in the GWAS, with the exception of SNP 12307_17 where the association to EC3 was not detected when using covariates (Table S6).

Combining all methodologies (RDA, GWAS-Cov, GWAS and Outliers), a total of 292 and 168 haplotype loci were identified as potentially candidates to selection in *S. ocellatus* and *S. tinca*, respectively (Figure 4). A low number of loci were shared by all methodologies being larger in *S. ocellatus* which presented higher population differentiation. The percentage of loci with SNPs associated to environmental predictors differed between the two species. Similar numbers of loci were associated to each environmental variable for *S. tinca*, although slightly larger to turbulence (Figure S2). However, for *S. ocellatus* most loci were associated to temperature and productivity. We used all candidate loci to cluster localities based on their observed heterozygosities and the frequency of their major allele from the haplotype loci dataset. In *S. ocellatus* both heatmaps clustered Colera and Blanes, the two northernmost localities, with Herradura, the southernmost locality (Figure 5). Similar groupings were also observed when only loci related to temperature and productivity were used but differed for loci associated to turbulence (Figure S3). In *S. tinca* different population clustering was observed depending on the variable used and the SNPs associated to different environmental predictors (Figure 5, Figure S4). Nonetheless, in

both species, contrasting frequencies of heterozygotes and major alleles were found among localities (Figure 5).

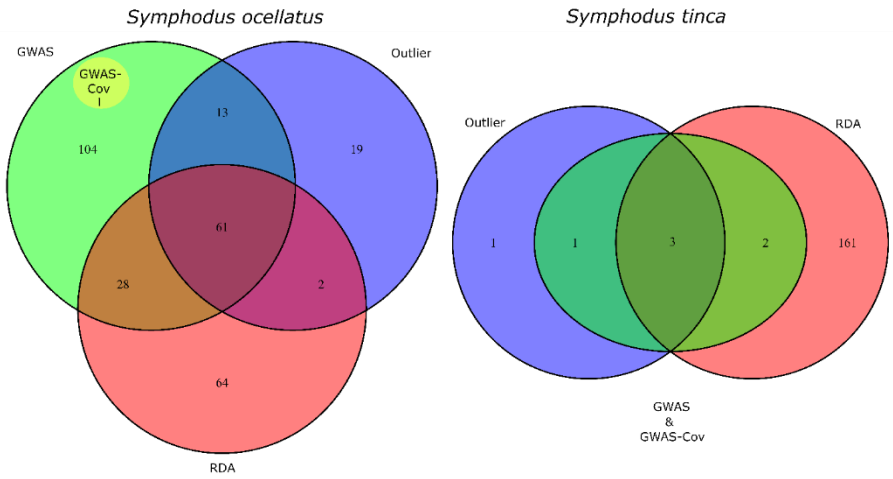


Figura 4: Venn diagram of the candidate haplotype loci showing significant values with different methodologies. Haplotype loci were considered to have a significant value for RDA and GWAS if containing SNPs with significant values for these analyses. Blue: outlier loci; Green: loci associated to environmental and/or phenotypic variables according to GWAS; Yellow: GWAS loci with the two first MDS components as covariates (GWAS-Cov); Pink: loci associated to environmental variables by RDA.

The Blast of all loci potentially under selection against the Ensemble database provided 172 hits for *S. ocellatus*, 93 of them inside genes (Table S5) and for *S. tinca* 103 loci had a blast hit, 43 of them inside genes (Table S6). The compilation of the biological functions using the GO terms in REVIGO showed that the genes identified in *S. ocellatus* and *S. tinca* had different functions (Figure 6). On one hand, *S. ocellatus* genes were associated to embryonic caudal fin morphogenesis, receptor localisation to synapse, response to stimuli, protein stability, autophagosome assembly and cell adhesion (Figure 6). On the other hand, *S. tinca* genes were associated to intracellular protein transport, heart contraction, thigmotaxis and swimming behaviour, regulation of GTPase activity, dephosphorylation, nucleobase catabolism and cytoskeleton organisation (Figure 6). Furthermore, within each species the genes associated to each of the environmental predictors showed different biological functions (Figure S5 and Figure S6).

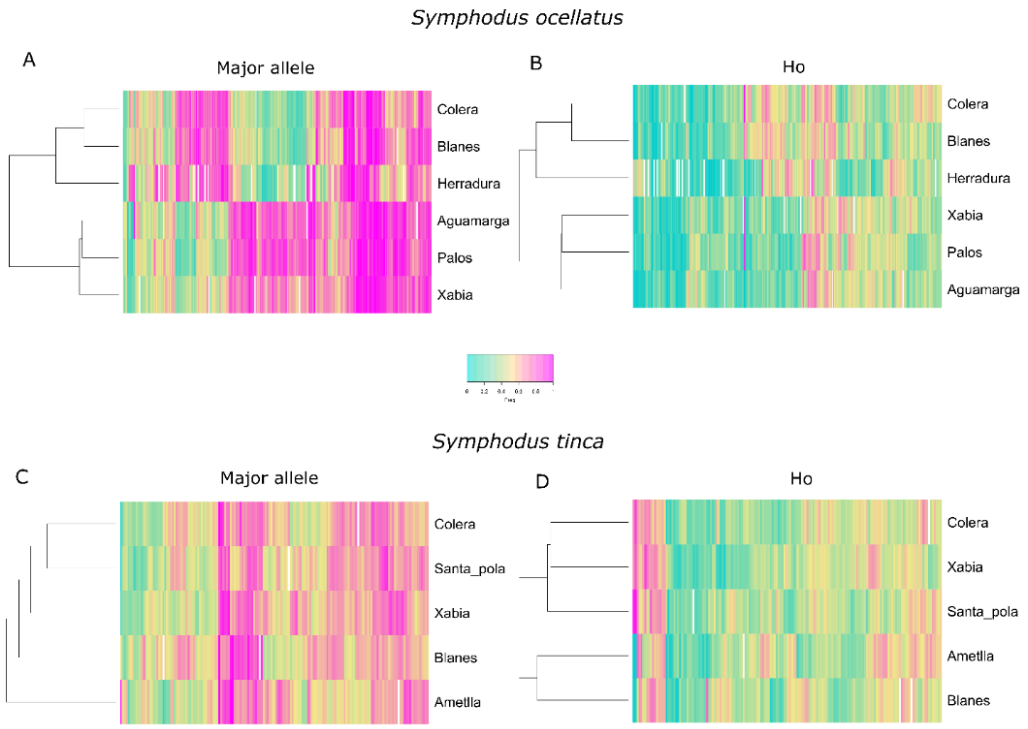


Figura 5: Haplotype loci values showing significant associations. Dendrogram and heatmap grouping the localities based on all candidate loci showing significant values with the different methodologies and using the values from the haplotype loci dataset: *Symphodus ocellatus* (A) frequency of the major allele and (B) observed heterozygosity. *Symphodus tinca* (C) frequency of the major allele and (D) observed heterozygosity

Discussion

The interaction of different evolutionary forces, such as dispersal and selection, defines the actual distribution of the genomic diversity within species [4,22]. We suggest that differences in environmental and phenotypic variables at early-life stages can produce contrasting genomic response that can be revealed by comparative studies on closely related species. Our two studied species presented very contrasting patterns of genetic structure, nonetheless the use of several methodologies for assessing signals of selection revealed multiple loci associated to environmental and phenotypic variables in both of them. Our individual level approach integrating genomic data with

environmental and phenotypic information was fundamental for identifying the factors and processes affecting population structuring.

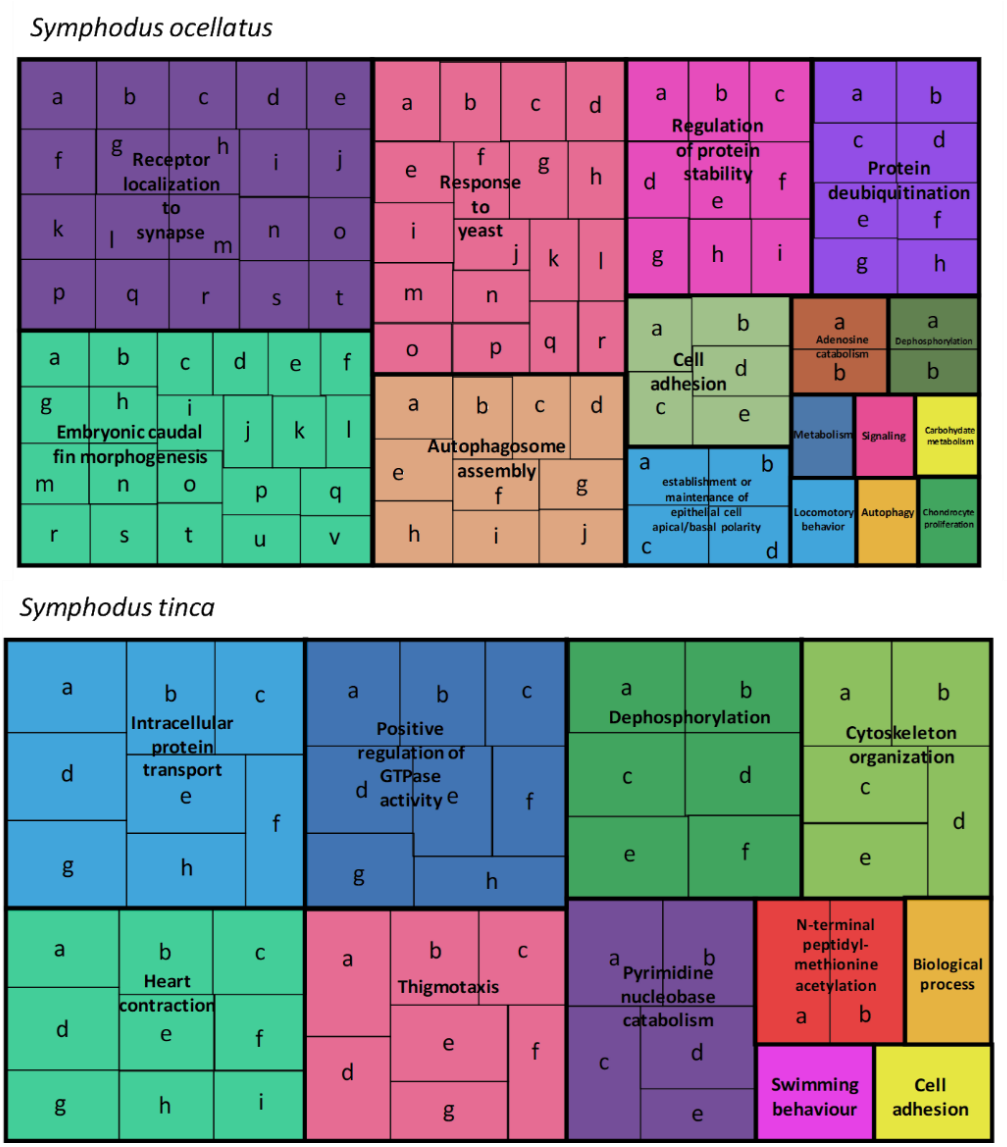


Figura 6: Gene ontology treemap of the candidate genes to selection combining all methodologies in *S. ocellatus* and *S. tinca*. The box size correlates to the –log10 p-value of the GO-term. Boxes with the same colour can be grouped together as correspond to the same upper-hierarchy GO-term which is found in the middle of each box. Description of letters and detailed information of the treemap can be found in supplementary tables S9 and S10.

Population structure and timing of reproduction

Our genome-wide analysis of the two *Symphodus* species, *S. tinca* and *S. ocellatus*, showed different levels of connectivity across the same environmental gradient. Two contrasting patterns are found despite their potentially similar dispersal capabilities, as both species build nests with algae [23,24] and present similar early-life traits such as planktonic larval duration (PLD), size at settlement and growth rate [25]. However, they have different life span and reproductive season [26,27]. If some adults move to neighbouring areas, species with longer life span such as *S. tinca* could present more connected populations than its congener *S. ocellatus* that has a shorter life span [28]. Adults are considered to be sedentary and remain close to their settling area [26] and thus other variables such as the time of reproduction could influence population structure.

In *Symphodus ocellatus*, we found a strong population structure and a clear effect of the two main oceanographic barriers in the area, the Ibiza channel and Almeria-Oran front, genetically clustering our sampling locations in three groups. These two oceanographic barriers seem to reduce gene flow across a great variety of taxa with different dispersal capabilities [29]. On the contrary, we observed very low genetic differentiation for *S. tinca* and a null effect of the Ibiza channel, in agreement with a previous study using microsatellite loci [30]. Seasonal variation in the reproductive and larval periods could probably influence the effect of oceanographic fronts on connectivity. For instance, the Ibiza channel although presenting inter-annual variations is known to break more often in spring than in summer [31,32] thus promoting connectivity for species reproducing in this season [16,33]. In the western Mediterranean, in spring there are more extreme turbulence values and higher frequency of unpredictable energy fluxes and episodic events than in summer [34–36] which could enhance connectivity. Contrarily to the results we found in the western Mediterranean, a significant population genomic structure was unveiled for *S. tinca* in the Adriatic Sea [10]. Interestingly, during the reproductive season of *S. tinca*, lower turbulence is observed in the Adriatic Sea in comparison to the western Mediterranean [34].

Some fronts, such as the AOF, seem to reduce connectivity for many marine species across all taxa, although many other species seem not to be influenced [29,37]. This could be attributed to temporal differences. Annual variation in the current pattern across the oceanographic fronts seems to account for genetic differentiation in neighbouring locations in crabs and sea urchins [33,38,39]. Seasonal variation in the hydrodynamic conditions across the fronts [31,40] could also differentially affect population structure in different species according to their reproductive season, as found in the present study. Therefore, we hypothesize that more stable oceanographic conditions could reduce genetic connectivity, facilitate long term local adaptation and leave more detectable genomic footprints of selection. Further population genomic studies across the same barriers in a wider number of species considering the reproductive season could allow testing this hypothesis.

Genomic associations to phenotypic and environmental variables

Local adaptation might be shaping population structure along with connectivity reduction mediated by oceanographic barriers [41]. Different genotypes can be under diverse selective pressures among localities, rendering individual-based information highly valuable. The use of several methodologies was fundamental to capture the signatures of adaptation in species with contrasting degree of population structuring. We found, in both species, small contribution of our studied variables to genomic structuring with the variation partitioning of the SNPs datafile at the individual level. These results confirm that the effect of external factors in genomic structuring is small across the whole genome, but does not preclude to be important in some specific loci. For instance, in *S. ocellatus* the higher population differentiation resulted in a higher number of outlier loci and potential for detecting local adaptation in comparison to *S. tinca*. Nonetheless, for both species of *Symphodus,* the combination of the results of GWAS and RDA revealed a larger number of candidate loci, putatively under selection, as well as the selection drivers of most of the outlier loci.

Redundancy analyses have been successfully used to detect signals of selection by comparing population allele frequencies to their environmental variables [4,6]. As we

had individually based environmental data, we assessed how each single individual was distributed along the RDA axes based on its genotype, and according to its environmental predictors [42]. In *S. ocellatus* the individuals were separated in two groups along the first RDA axis, which was explained by surface temperature and productivity. This analysis grouped geographically separated localities, that present low temperature and high productivity, thus demonstrating that population differentiation is not only influenced by oceanographic barriers but also respond to environmental selection pressures. For, *S. tinca* individuals were grouped by locality in the RDA analysis, but their clustering was less clear, with temperature, productivity and turbulence having similar impact. Thus, despite the low number of outlier loci and population structure in that species, environmental variables can nonetheless contribute to adaptation. These apparently contradictory results rely on the power of different methodologies to detect signals of selection. The analysis of outlier loci is based on the detection of loci that show divergent patterns of genetic differentiation among localities. Thus, this analysis is highly dependent on the basal levels of population structuring as indicated in previous studies [4,42]. For instance, the abundance of outlier loci in *S. tinca* found in our study area (0.1% of all loci) was lower than that obtained for the same species in the Adriatic Sea (0.5%) [10], suggesting that a more structured region can favour the power to detect local adaptation by outlier analysis. For this reason, RDA should be also used, as it can detect signals of selection even on conditions of low structuring [42]. Moreover, this approximation can be done both at the individual and population level and can thus be implemented in species where individual based information is difficult to obtain.

The analyses of phenotypic and environmental variables associated to individual genotypes are necessary to unveil candidate regions of the genome determining these traits [3,6]. However, when there is population structure, signals of selection within them can be confounded by either demographic or selective processes among populations [43]. In both *Symphodus* species we identified only a few candidate loci associated to environmental variables by using GWAS after removing the population effect using the two first MDS axis as covariates. This methodology has been

suggested to control for hierarchical structure to avoid identifying spurious associations due to population differentiation. However, this approach is very conservative and can underestimate phenotype-genotype associations in loci under local adaptation [4,42]. For instance, in *S. ocellatus*, the high level of population structuring undermines the power to detect selection using GWAS with covariates, while GWAS without covariates and RDA revealed a high number of candidate loci to selection. In the case of *S. tinca*, the low level of population structuring limited the resolution of the two types of GWAS while many loci under selection were found with RDA. This highlights the need to assess selection using different methodologies as the degree of population structuring may underestimate the role of selection depending on the methodology used.

Environmental factors, such as temperature or salinity, are known to impose differential spatial signatures of selection on marine organisms [6,44,45]. The biological functions of the genes associated to the candidate loci of selection in our two *Symphodus* species suggest different mechanisms of adaptation in response to the environmental factors although both of them showed strong influence of temperature and productivity. The importance of these two variables in marine organisms is not unprecedented. Temperature is a crucial factor in many marine organisms as it may impact metabolism, larval survival and development among others [46]. Interestingly, turbulence was found to be more important in *S. tinca* than in *S. ocellatus*. Although the turbulence values can be similar during the reproduction period of both species, the frequency of storms in our study area is much higher in spring than in summer [34,47]. This temporal instability may explain the importance of this variable on the spring breeder *S. tinca* with 40% of candidate loci associated to this factor while only 11% where found to be associated in *S. ocellatus.*

In both species, we found several polymorphisms to be significantly associated with phenotypic and environmental variables, indicating a possible genotype-phenotype-environment interaction. In *Symphodus tinca*, we identified two loci associated to hatching date and also to environmental variables, being one of them related to turbulence and the other to temperature. Hatching date seems to be under

selections in several marine species. For instance, in *S. tinca* reproducing females appear to prefer males with nests early in the reproductive season, and hatching success is consistently higher for eggs placed in early nests [48]. In the sunfish, *Lepomis macrochirus*, individuals born early in the season have higher survivorship than those born late [49]. In the peacock blenny, *Salaria pavo*, males select a different reproductive tactic depending of the birth date [50]. For the common triplefin, *Forsterygion lapillum*, more successful males hatch early in the hatching period [51]. Overall, these studies suggest that individuals born early could attain larger sizes at the reproductive season, rank more highly in social hierarchy and have better chances to reproduce. In *S. ocellatus,* 22 loci were related to settlement size or to PC2, highly influenced by PLD and settlement size, and 70% of them were also associated to temperature. A high proportion (50%) of those loci were also associated to productivity and EC1. Size at settlement, and other early-life traits, clearly increase post-settlement survival in some species, e.g. *Pomatomus saltatrix* [52], *Thalassoma bifasciatum* [53], *Symphodus* spp. [54] demonstrating the existence of a size-selective mortality. However, the absence of a consistent trend for selective mortality [53] and annual variation among settlement intensity, recruitment level and year class strength [55] suggest an important role of random events on survival. Therefore, natural selection could increase the frequency of alleles favouring reproduction early in the season as well as size at settlement and high stochasticity in survival maintain genetic polymorphism.

In conclusion, the results of this study show that individual-based information combining genomics, environmental and phenotypic values is key to unveil detailed patterns of evolutionary genomic signatures in living organisms and thus to glimpse the potential response to future changes. Despite the study of early life traits through otolith readings is only applicable to bone fishes, the combination of the different methodologies conducted in this paper to assess population structuring and local adaptation is useful for all species with individualized genotypic, environmental and phenotypic data. Furthermore, this approach has demonstrated to be applicable at the population level and in the absence of individual based information [4,6]. Our findings emphasize that candidate regions to local adaptation, mediated by productivity and

temperature, can even be detected in species presenting low population structuring. This holistic approach should be considered in scientifically informed management actions since, in spatially and temporally variable environments, as found along the distribution ranges of most species, environmental and phenotypic factors could be imposing previously concealed selective pressures.

Materials and methods

Study species

Symphodus ocellatus (Linnaeus, 1758) and *Symphodus tinca* (Linnaeus, 1758) live in shallow waters (0-30 m), have similar larval duration (7-13 days) and are considered to disperse only during the larval stages, since their adults exhibit territorial behaviour [26]. These two species are representative components of the rocky, sea-grass beds and soft bottom communities of the Western Mediterranean [56,57]. They have a narrow bathymetric range (<100 m), are found ubiquitously across the studied area and produce a clear settlement band in their otoliths [25,58]. The reproductive season of *S. tinca* extends during spring, whereas *S. ocellatus* reproduces in summer (Lejeune, 1985; Raventos & Macpherson, 2005b). *S. ocellatus* individuals live near 5 years, reach a maximum length of 12 cm and attain sexual maturity between 1 and 2 years [27]. *S. tinca* individuals live up to 15 years [28], can reach 44 cm, and attain sexual maturity in 2 or 3 years [27]. Due to the phylogenetic relatedness and shared biological traits and distribution, these species have a high value for a comparative population genomic study.

Sampling

A total of 162 individuals of *Symphodus ocellatus* (Linnaeus, 1758) and 141 of *Symphodus tinca* (Linnaeus, 1758) were sampled and analysed from six and five different locations respectively along the Spanish Mediterranean coast between 2014 and 2016 (Table 1, Figure 1). This area contains two oceanographic discontinuities, the Almeria-Oran Front (AOF) and the Ibiza Channel (IC), that can affect connectivity

among populations [29,30]. From Santa Pola, adult fin clips were provided by professional fishermen, while for the remaining localities juveniles were captured during recruitment using hand nets. The whole juvenile individuals and adult fin clips were taken and stored in 96% Ethanol. The collection of fish samples met the Spanish and European regulations. According to article 3.1 of the European Union directive (2010/63/UE) from the 22/9/2010, no approval is needed for fish sacrification with the purpose of tissue or organ analyses. Furthermore, the study species *Symphodus ocellatus* and *S. tinca* are not listed in CITES. A map of the sampling localities (Figure 1) was plotted using the R[60] packages 'ncdf4'[61] and 'ggplot2' [62]. The colour gradient in the map represents the mean temperature across the larval period of each species in a 2X2 Km grid obtained from Western Mediterranean OPerational forecasting system (WMOP) [63].

Phenotypic and environmental variables

At the laboratory, the otoliths (lapilli) of all juveniles were extracted and mounted in an oil droplet on a microscope slide. Age readings and the presence of settlement marks were determined using a light microscope following the standard methodology described in the literature [25,54]. In brief, we counted the age of the individual with the number of otolith marks along the longest radius, and with this information, we inferred the hatch date of each individual (Figure S1). PLD was calculated with the number of otolith marks before the settlement mark. From otolith readings, four early-life traits were considered for each individual: hatching date (expressed as the number of day of the year), pelagic larval duration (PLD, in days), growth rate during the pelagic larval duration (increment in otolith size in µm divided by PLD) and otolith settlement size (in µm). We measured the otolith width six times and used their mean to minimize measurements errors.

For each individual, the hatching date and its PLD value were used to determine the exact time of its pelagic period, over which the individual-based environmental variables were averaged. Three environmental variables were considered: two from remote sensing, surface temperature (in ºC) and productivity (measured as

Chlorophyll a concentration (ChlA) in mg·m^{-3}), and one from modelling, significant wave height (in meters) as a proxy for turbulence. The water surface temperature was obtained from SOCIB from a daily gridded dataset with a resolution of 0.02 degrees (Balearic Islands coastal ocean observatory and forecasting system, http://www.socib.eu/). Productivity data used in this paper were produced with the Giovanni online data system, developed and maintained by the NASA GES DISC [64], obtaining monthly mean values of ChlA with a spatial resolution of 4km. Significant wave height was obtained from SIMAR model from the Spanish state ports agency, with hourly values and a spatial resolution of 0,25 degrees. Individual-based environmental and phenotypic variables were only available for individuals with otolith information (Table 1).

In order to evaluate the possible differences of early-life traits between locations, Permutational Multivariate Analyses of Variance (PERMANOVA) were performed for each phenotypic and environmental variable, as implemented in the R function 'adonis', from the 'vegan' package v 2.5-2 [65]. Two principal components analysis (PCA) using 'dudi.pca' function from 'ade4' package v.1.7-8 [66] were performed for each species after standardizing each variable, one with all phenotypic and the other with all environmental variables. We kept the minimum number of principal components necessary to explain at least 99% of the variance.

DNA extraction and genotyping

DNA was individually extracted using the QIAamp DNA Mini Kit (QIAGEN) extraction kit, following manufacturer's instructions. DNA integrity was checked by gel electrophoresis and quantified by NanoDrop. To build individual genomic libraries by GBS, 1.5-3 µg of DNA for each sample were sent to the Cornell University Biotechnology Resource Centre (BRC). DNA was digested with EcoT22I restriction enzyme and ligated to barcode and common adaptors with appropriated sticky ends as in a previous study [10]. For each plate, one blank sample and 95 samples were pooled and cleaned using QIAGEN PCR cleanup kit (following manufacturer's instructions). An amplification by PCR was applied to the resulting 96plex libraries, using generic primers matching the

adaptors. The PCR was made under the following conditions: 5 minutes at 72ºC, 30 seconds at 98ºC, 18 cycles of 10 seconds at 98ºC, 30 seconds at 65ºC and 30 seconds at 72ºC and a final extension of 5 minutes at 72ºC. The PCR product was cleaned with the QIAquick PCR Purification Kit, diluted and single-end sequenced in an Illumina HiSeq 2500 platform at BRC, by using one lane per plate and the HiSeq v4 reagents kit. Individuals from the same locality were distributed in different lanes.

Raw Illumina sequences from all samples were used for genotyping with STACKS vs 1.47 software [67]. Good reads with a barcode were trimmed to 59bp, excluding barcodes and primers. To construct a catalogue loci two mismatches were allowed between stacks within and between individuals, and a minimal stack depth of three was required. Individual genotypes were outputted both as haplotype loci and SNP loci VCF files. Additional filters were applied using VCFtools vs 1.12 [68]. Individual genotypes with a depth below 5X were not considered. Loci with a missingness value higher than 30% were removed from the loci datafiles. Additionally, haplotype loci with the major allele frequency equally or higher than 0.95 (*i.e.* monomorphic at that level) were identified by the R function 'isPoly' from the package 'adegenet' [69] and removed from the vcf file by VCFtools. For SNP loci, those with a minimum allele frequency (MAF) of 0.05 were also removed from the SNP dataset by VCFtools. Finally, the loci in HW disequilibrium at more than 60% of the sampling sites [10] were eliminated of the analysis from both the haplotype and SNP loci datafiles. This filter was used step to eliminate possible paralogous loci.

Population genomic analyses

We used the haplotype loci datafiles for the population genomic analyses. For each locality, observed (Ho) and expected (He) heterozygosities, allelic richness (Ar), F_{ST} pairwise values and their significance (computed by 999 permutations) were obtained by R software [60] 'hierfstat' package vs 0.04-2 [70]. A FDR correction for multiple comparisons was applied in all required analyses to estimate the threshold of differentiation [71]. Isolation by distance was evaluated between Euclidean geographic distances and F_{ST} distances with Mantel tests, as implemented by the 'mantel' function

in the 'vegan' R package. Discriminant Analysis of Principal Components (DAPC) were performed retaining a number of PCAs equal to one third of the number of individuals. To detect the optimal number of genetic clusters, we performed K-means clustering from k=1 to k=10, using 10^5 iterations, retaining a number of PCs equal to sample size and compared the results using the Akaike Information Criterion (AIC). The analyses were performed with the R software package 'adegenet' and represented with the 'ggplot2' package [62]. To identify outlier loci (OAs), we used BayeScan vs 2.1 [72], with individuals grouped by sampling locality. To minimize false positives, 100,000 simulations were run and a prior odd of 10,000 was specified [73]. Those haplotype loci with a q-value (FDR analogue of the p-value) below 0.05 were considered as statistically significant outliers.

We used the SNP datafile for association analysis with environmental (EAAs) and phenotypic (PAAs) variables. We kept all polymorphic SNPs in each locus after filtering. Missing genotypes were imputed by replacing them with the most common genotype across all individuals. We performed a variance partition analysis for the complete matrix of SNPs, the matrix of the three environmental variables and the matrix of the four phenotypic variables using the 'varpart' function in the 'vegan' R package. We looked for SNPs strongly correlated with environmental variables by using a Redundancy Analysis (RDA) with the SNP loci datafiles following the method described by Forester and co-authors [42]. We used as response variables a genotypic matrix with the SNPs of all loci and the three environmental variables as predictors. All environmental variables were retained because they were not highly correlated ($r < 0.75$).

A genome-wide association analysis (GWAS) was performed with the linear model implemented in the software PLINK vs 1.9 [74]. For each individual, we used its genotype from the SNP loci datafile and its environmental and phenotypic variables as well as the composite PCA variables previously detailed. Moreover, we performed a Multidimensional Scaling analysis (MDS) with the SNP datafile per species with the same software. The effect of genetic differentiation among localities is often removed in associated studies with the use of covariates. Thus, we carried out two association

studies with PLINK: one using as covariates the two first components of the MDS (GWAS-Cov), and the other without covariates (GWAS).

For all candidate loci potentially under selection, identified as population outlier loci and/or including SNPs associated to environmental or phenotypic variables, we computed the frequency of the major allele and their observed heterozygosity for the haplotype dataset in each sampling location. Both values were represented in heatmaps and hierarchical dendrograms, using the heatmap.2 function of the package 'gplots' vs 3.0.1 [75].

BLAST searches were conducted with the 59 bp sequences of all candidate loci against the *Labrus bergylta* Ascanius, 1767 genome [76], available in BLASTn search tool of Ensembl website (www.ensembl.org). Sensitivity was set to "Distant homologies" to maximise matches' length considering the phylogenetic distance between species. A maximum E-value of 10^{-3} was allowed and only matches of at least half of the locus sequence were considered. When sequences matched within a gene, the function of that gene was searched at the UniProt database (www.uniprot.org). For the list of annotated genes their GO-terms were introduced in REVIGO to produce treemaps of the biological functions [77].

Acknowledgments

This work was supported by the Spanish Government projects 'ChallenGen' (CTM2013-48163) and 'PopCOmics' (CTM2017-88080) (MCIU, AEI/FEDER, UE) and by the European FP7 CoCoNet project (Ocean 2011-4, grant agreement #287844). The authors CC and MP are members of the research group SGR2017-1120 and EM of SGR2017-378 (Catalan Government). HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities (FPU15/02390).

Data Availability Statement

Raw read data from all individuals will be available from an NCBI SRA Bioproject upon acceptance. A list of genotypes of all individuals, including sampling location, will be available as Supplementary Dataset upon acceptance.

Author Contributions

HT, CC, EM and MP conceived and designed the study. NR and EM sampled all juveniles. NR did the otolith readings. HT conducted the laboratory and data analysis with inputs from CC, EM and MP. HT, CC, EM and MP wrote the manuscript.

References

1. Grummer, J. A. *et al.* Aquatic Landscape Genomics and Environmental Effects on Genetic Variation. *Trends Ecol. Evol.* 34, 641–654 (2019).
2. Ahrens, C. W. *et al.* The search for loci under selection: trends, biases and progress. *Mol. Ecol.* 27, 1342–1356 (2018).
3. Benestan, L. *et al.* Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Mol. Ecol.* 25, 5073–5092 (2016).
4. Xuereb, A., Kimber, C. M., Curtis, J. M. R., Bernatchez, L. & Fortin, M. J. Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. *Mol. Ecol.* 27, 5035–5048 (2018).
5. Hasan, M. M. *et al.* Sticklebacks adapted to divergent osmotic environments show differences in plasticity for kidney morphology and candidate gene expression. *J. Exp. Biol.* 220, 2175–2186 (2017).
6. Carreras, C. *et al.* East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). *Divers. Distrib.* 26, 382–398 (2020).
7. Hansen, M. M., Olivieri, I., Waller, D. M. & Nielsen, E. E. Monitoring adaptive genetic responses to environmental change. *Mol. Ecol.* 21, 1311–1329 (2012).
8. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6, e19379 (2011).
9. Stapley, J. *et al.* Adaptation genomics: The next generation. *Trends in Ecology and Evolution* 25, 705–712 (2010).
10. Carreras, C. *et al.* Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Sci. Rep.* 7, 43417 (2017).
11. Babbucci, M. *et al.* An integrated genomic approach for the study of mandibular

prognathism in the European seabass (*Dicentrarchus labrax*). *Sci. Rep.* 6, 1–11 (2016).
12. Gonzalez-Pena, D. *et al.* Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Front. Genet.* 7, (2016).
13. Zhong, X. *et al.* Genome-Wide Association Study Reveals Multiple Novel QTL Associated with Low Oxygen Tolerance in Hybrid Catfish. *Mar. Biotechnol.* 19, 379–390 (2017).
14. Schweizer, R. M. *et al.* Genetic subdivision and candidate genes under selection in North American grey wolves. *Mol. Ecol.* 25, 380–402 (2016).
15. Planes, S. & Lenfant, P. Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. *Mol. Ecol.* 11, 1515–1524 (2002).
16. Schunter, C. *et al.* Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Mol. Ecol.* 20, 5167–5181 (2011).
17. Treml, E. A. *et al.* Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. in *Integrative and Comparative Biology* 52, 525–537 (Narnia, 2012).
18. Aspillaga, E. *et al.* Ordinary and extraordinary movement behaviour of small resident fish within a mediterranean marine protected area. *PLoS One* 11, 1–19 (2016).
19. Garrido, S. *et al.* Born small, die young: Intrinsic, size-selective mortality in marine larval fish. *Sci. Rep.* 5, 17065 (2015).
20. Raventos, N., Macpherson, E. & García-Rubiés, A. Effect of brine discharge from a desalination plant on macrobenthic communities in the NW Mediterranean. *Mar. Environ. Res.* 62, 1–14 (2006).
21. Schunter, C. *et al.* A novel integrative approach elucidates fine-scale dispersal patchiness in marine populations. *Sci. Rep.* 9, 10796 (2019).
22. Warner, R. R. Evolutionary ecology: how to reconcile pelagic dispersal with local adaptation. *Coral Reefs* 16, S115–S120 (1997).
23. Warner, R. R. & Lejeune, P. Sex change limited by paternal care: a test using four Mediterranean labrid fishes, genus *Symphodus*. *Mar. Biol.* 87, 89–99 (1985).
24. Raventos, N. Nest site characteristics and nesting success of the five- spotted wrasse *Symphodus roissali* in the north-western Mediterranean Sea. *J. Fish Biol.* 68, 305–309 (2006).
25. Raventos, N. & Macpherson, E. Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Mar. Biol.* 138, 1115–1120 (2001).
26. Lejeune, P. Le comportement social des Labridés méditerranéens. *Cah. d'Ethologie Appliqué* 5, 1–208 (1985).
27. Froese, R. & Pauly, D. FishBase. *World Wide Web electronic publication.* www.fishbase.org, version (10/2016) (2016).
28. Gordoa, A., Molí, B. & Raventos, N. Growth performance of four wrasse species on the north-western Mediterranean coast. *Fish. Res.* 45, 43–50 (2000).
29. Pascual, M., Rives, B., Schunter, C. & Macpherson, E. Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PLoS One* 12, 1–20 (2017).

30. Galarza, J. A. *et al.* The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proc. Natl. Acad. Sci.* 106, 1473–1478 (2009).
31. Balbín, R. *et al.* Interannual variability of the early summer circulation around the Balearic Islands: Driving factors and potential effects on the marine ecosystem. *J. Mar. Syst.* 138, 70–81 (2014).
32. Fernández, V., Dietrich, D. E., Haney, R. L. & Tintoré, J. Mesoscale, seasonal and interannual variability in the Mediterranean Sea using a numerical ocean model. *Prog. Oceanogr.* 66, 321–340 (2005).
33. Pascual, M. *et al.* Temporal and spatial genetic differentiation in the crab *Liocarcinus depurator* across the Atlantic-Mediterranean transition. *Sci. Rep.* 6, 1–10 (2016).
34. Izaguirre, C., Méndez, F. J., Menéndez, M. & Losada, I. J. Global extreme wave height variability based on satellite data. *Geophys. Res. Lett.* 38, 1–6 (2011).
35. Álvarez, I. *et al.* Drivers of larval fish assemblage shift during the spring-summer transition in the coastal Mediterranean. *Estuar. Coast. Shelf Sci.* 97, 127–135 (2012).
36. Guillén, J. *et al.* Coastal oceanographic signatures of heat waves and extreme events of dense water formation during the period 2002-2012 (Barcelona, NW Mediterranean). *Sci. Mar.* 82, 189–206 (2019).
37. Patarnello, T., Volckaert, F. & Castilho, R. Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? *Mol. Ecol.* 16, 4426–4444 (2007).
38. Pérez-Portela, R. *et al.* Spatio-temporal patterns of genetic variation in *Arbacia lixula*, a thermophilous sea urchin in expansion in the Mediterranean. *Heredity (Edinb).* 122, 244–259 (2019).
39. Calderón, I., Pita, L., Brusciotti, S., Palacín, C. & Turon, X. Time and space: genetic structure of the cohorts of the common sea urchin *Paracentrotus lividus* in Western Mediterranean. *Mar. Biol.* 159, 187–197 (2012).
40. Renault, L., Oguz, T., Pascual, A., Vizoso, G. & Tintore, J. Surface circulation in the Alborn Sea (western Mediterranean) inferred from remotely sensed data. *J. Geophys. Res. Ocean.* 117, 1–12 (2012).
41. Bradbury, I. R. *et al.* Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proc. R. Soc. London B Biol. Sci.* 277, 3725–34 (2010).
42. Forester, B. R., Lasky, J. R., Wagner, H. H. & Urban, D. L. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.* 27, 2215–2233 (2018).
43. Excoffier, L., Hofer, T. & Foll, M. Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb).* 103, 285–298 (2009).
44. Franch-Gras, L. *et al.* Genomic signatures of local adaptation to the degree of environmental predictability in rotifers. *Sci. Rep.* 8, 16051 (2018).
45. Lamichhaney, S. *et al.* Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19345–50 (2012).
46. Sanford, E. & Kelly, M. W. Local Adaptation in Marine Invertebrates. *Ann. Rev. Mar.*

Sci. 3, 509–535 (2010).
47. Font, J., Garcialadona, E. & Gorriz, E. The seasonality of mesoscale motion in the Northern Current of the western Mediterranean: several years of evidence. *Oceanol. Acta* 18, 207–219 (1995).
48. Warner, R. R., Wernerus, F., Lejeune, P. & Van Den Berghe, E. Dynamics of female choice for parental care in a fish species where care is facultative. *Behav. Ecol.* 6, 73–81 (1995).
49. Cargnelli, L. M. & Gross, M. R. The temporal dimension in fish recruitment: birth date, body size, and size-dependent survival in a sunfish (bluegill: *Lepomis macrochirus*). *Can. J. Fish. Aquat. Sci.* 53, 360–367 (1996).
50. Fagundes, T. *et al.* Birth date predicts alternative life-history pathways in a fish with sequential reproductive tactics. *Funct. Ecol.* 29, 1533–1542 (2015).
51. Moginie, B. F. & Shima, J. S. Hatch date and growth rate drives reproductive success in nest-guarding males of a temperate reef fish. *Mar. Ecol. Prog. Ser.* 592, 197–206 (2018).
52. Hare, J. A. & Cowen, R. K. Size, growth, development, and survival of the planktonic larvae of *Pomatomus saltatrix* (Pisces: Pomatomidae). *Ecology* 78, 2415–2431 (1997).
53. Searcy, S. & Sponaugle, S. Selective Mortality during the Larval-Juvenile Transition in Two Coral Reef Fishes. *Ecology* 82, 2452–2470 (2001).
54. Raventos, N. & Macpherson, E. Effect of pelagic larval growth and size-at-hatching on post-settlement survivorship in two temperate labrid fish of the genus *Symphodus*. *Mar. Ecol. Prog. Ser.* 285, 205–211 (2005).
55. Raventos, N. Relationships between adult population size, recruitment, and year-class strength in a labrid fish in the Mediterranean Sea. *Estuar. Coast. Shelf Sci.* 85, 167–172 (2009).
56. García Rubies, A. & Zabala, M. Effects of total fishing prohibition on the rocky fish assemblages of Medes Islands marine reserve (NW Mediterranean). *Scientia Marina* 54, 317–328 (1990).
57. Macpherson, E., Gordoa, A. & García-Rubies, A. Biomass Size Spectra in Littoral Fishes in Protected and Unprotected Areas in the NW Mediterranean. *Estuar. Coast. Shelf Sci.* 55, 777–788 (2002).
58. Wilson & McCormick, M. I. Microstructure of settlement-marks in the otoliths of tropical reef fishes. *Mar. Biol.* 134, 29–41 (1999).
59. Raventos, N. & Macpherson, E. Environmental influences on temporal patterns of settlement in two littoral labrid fishes in the Mediterranean Sea. *Estuar. Coast. Shelf Sci.* 63, 479–487 (2005).
60. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
61. Pierce, D. ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. R package version 1.16.1. (2019).
62. Wickham, H. Elegant graphics for data analysis (ggplot2). (2009).
63. Juza, M. *et al.* Operational SOCIB forecasting system and multi-platform validation in the Western Mediterranean. *J. Oper. Ocean.* 9, 9231 (2016).
64. Acker, J. G. & Leptoukh, G. Online analysis enhances use of NASA Earth Science Data. *Eos (Washington. DC).* 88, (2007).
65. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.5-2.

(2018).
66. Dray, S. & Dufour, A.-B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* 22, (2015).
67. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140 (2013).
68. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158 (2011).
69. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405 (2008).
70. Goudet, J. & Jombart, T. hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04-22. (2015).
71. Narum, S. R. Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conserv. Genet.* 7, 783–787 (2006).
72. Foll, M. & Gaggiotti, O. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 180, 977 LP – 993 (2008).
73. Lotterhos, K. E. & Whitlock, M. C. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol. Ecol.* 23, 2178–2192 (2014).
74. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575 (2007).
75. Warnes, M. G. R., Bolker, B., Bonebakker, L. & Gentleman, R. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. (2016).
76. Lie, K. K. *et al.* Loss of stomach, loss of appetite? Sequencing of the ballan wrasse (*Labrus bergylta*) genome and intestinal transcriptomic profiling illuminate the evolution of loss of stomach function in fish. *BMC Genomics* 19, 186 (2018).
77. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS One* 6, e21800 (2011).

Supplementary material

Figure S1: *Symphodus tinca* otolith with dairy increment marks. The central arrow indicates the settlement ring. White dots and arrows are used to indicate the age of the individual (in the rectangle). The solid line is used to measure the otolith size in µm.

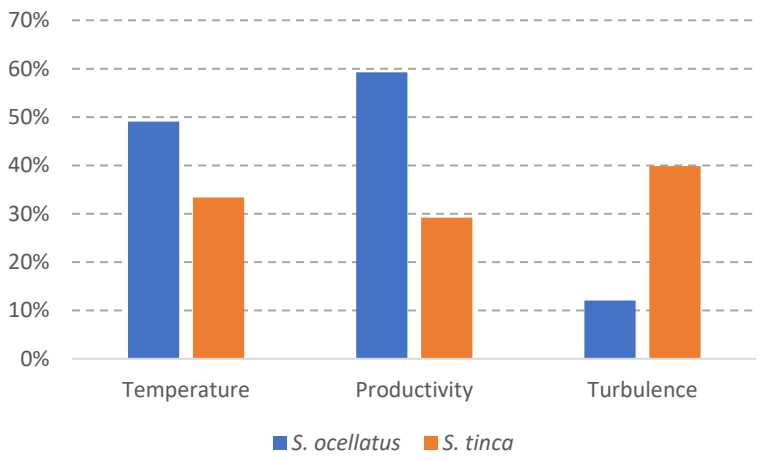


Figure S2: Percentage of haplotype loci with SNPs associated to environmental variables combining the results of different methodologies (RDA and GWAS). In *S. ocellatus* 54 loci were associated to both temperature and productivity, while only four were associated to more than one of these variables in *S. tinca*.

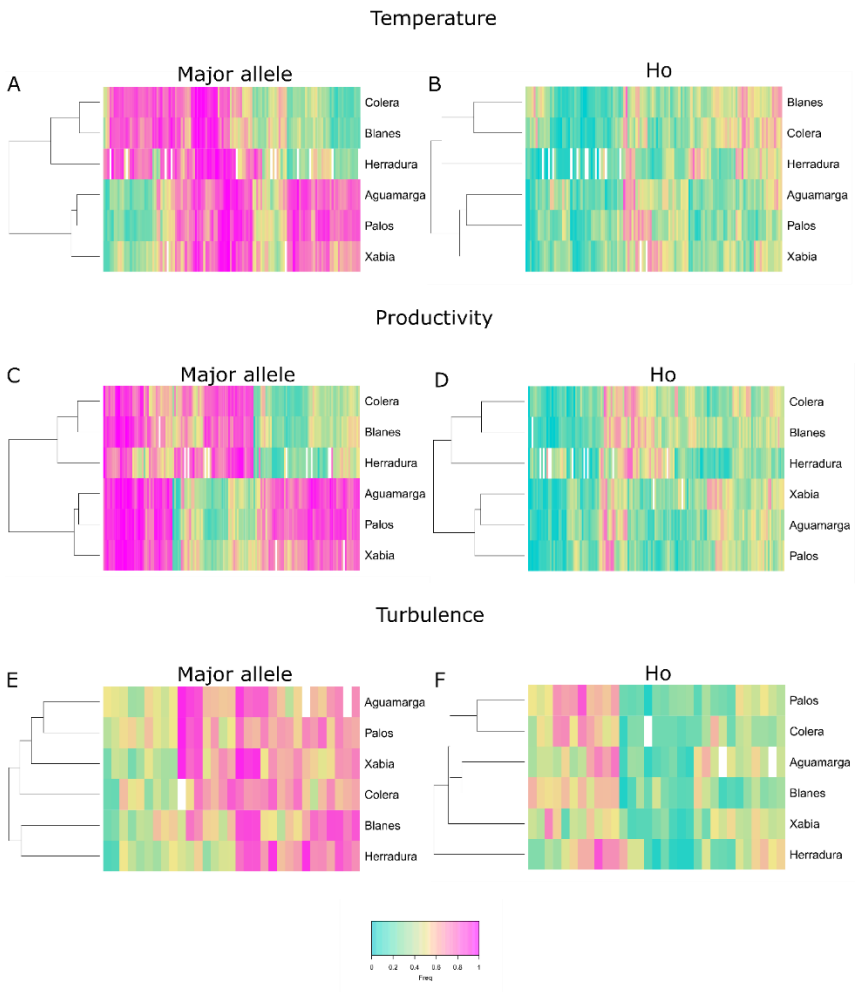


Figure S3: Dendrogram and heatmap grouping the localities based on the frequency of the major allele and observed heterozygosities of loci showing significant values in *Symphodus ocellatus* associated to (A and B) Temperature, (C and D) Productivity and (E and F) Turbulence.

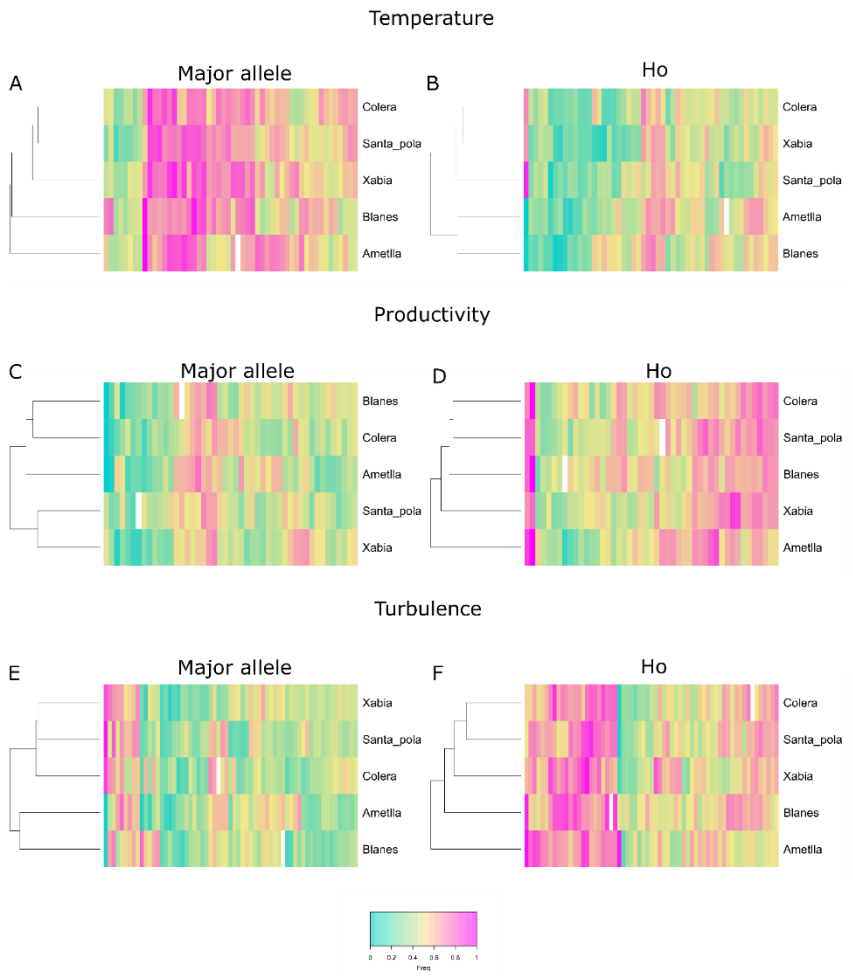


Figure S4: Dendrogram and heatmap grouping the localities based on the frequency of the major allele and observed heterozygosities of loci showing significant values in *Symphodus tinca* associated to (A and B) Temperature, (C and D) Productivity and (E and F) Turbulence.

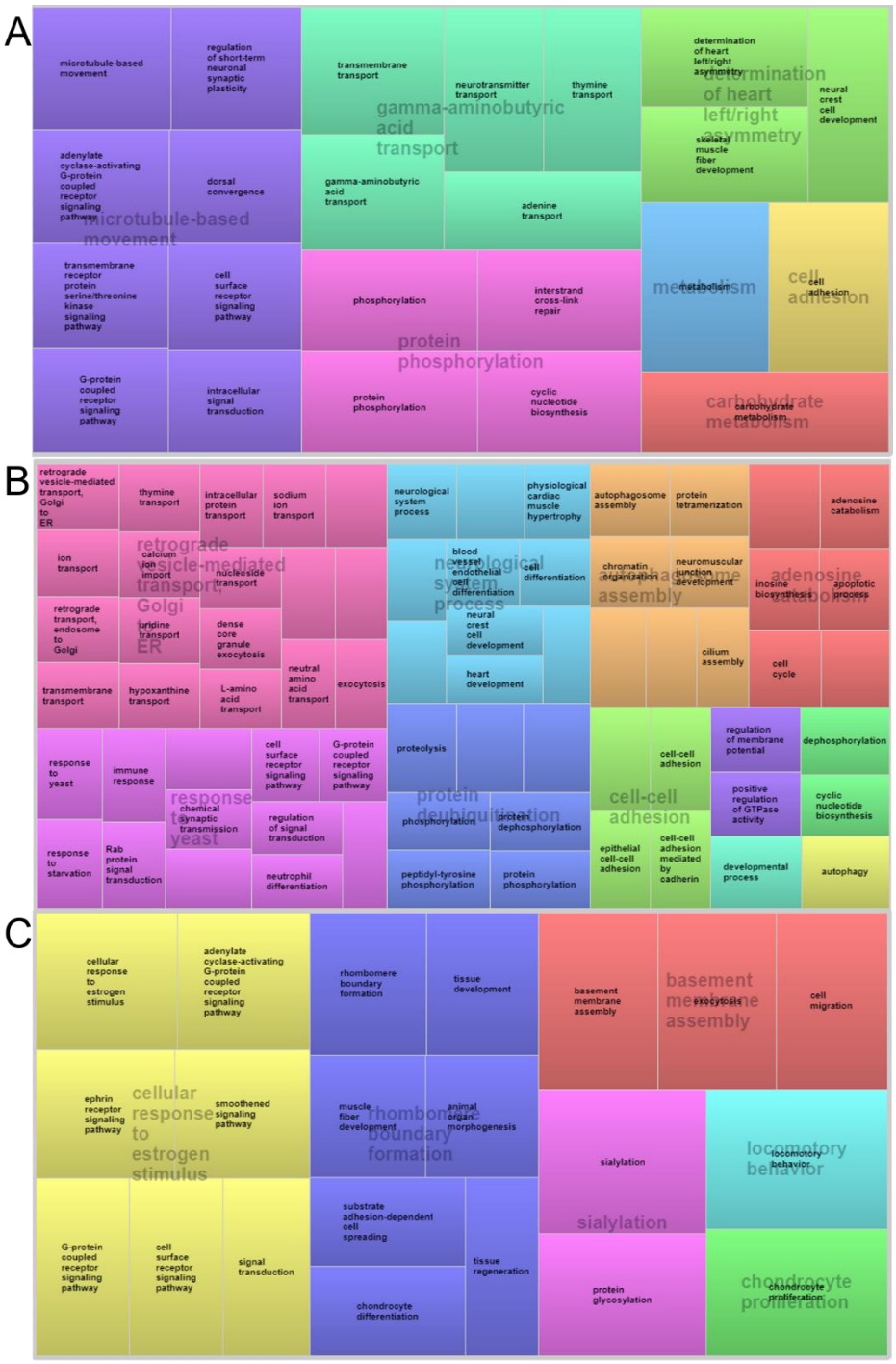


Figure S5: Gene ontology treemap of the candidate genes to selection combining all methodologies in *Symphodus ocellatus* associated to (A) Temperature, (B) Productivity and (C) Turbulence.

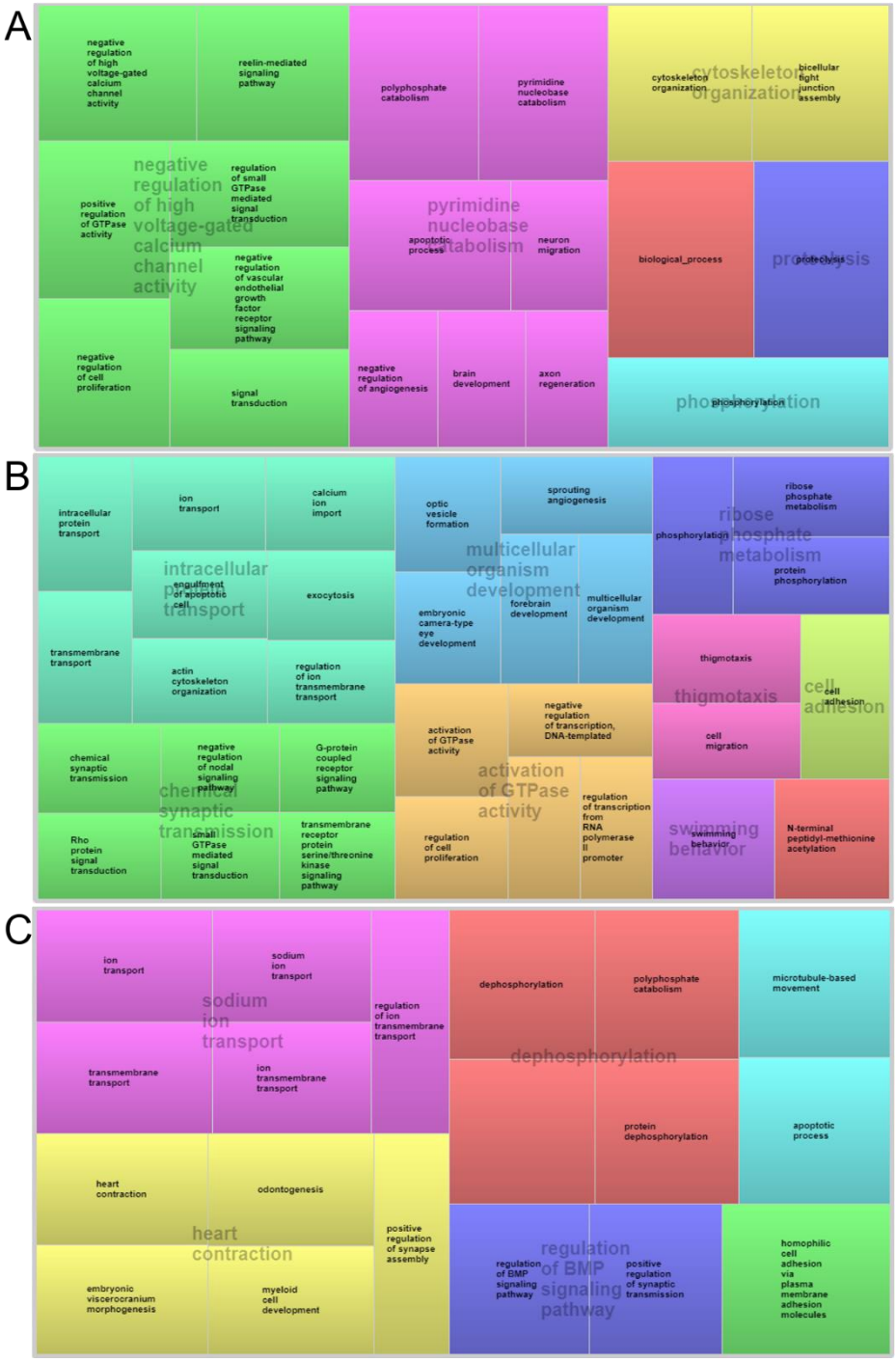


Figure S6: Gene ontology treemap of the candidate genes to selection combining all methodologies in *Symphodus tinca* associated to (A) Temperature, (B) Productivity and (C) Turbulence.

Table S1. Differences in phenotypic and environmental values between localities of *Symphodus ocellatus* and *S. tinca* assessed with a PERMANOVA test. Bold values indicate significant differences among sampling locations after a FDR correction.

		S. ocellatus		*S. tinca*	
		F	p	F	p
Phenotypic	PLD	15.464	**0.001**	2.463	0.098
	PLD Growth rate	1.764	0.163	47.686	**0.001**
	Settlement size	6.864	**0.008**	31.236	**0.001**
	Hatching date	0.702	0.432	6.609	**0.004**
Environmental	Surface Temperature	39.971	**0.001**	14.454	**0.003**
	Productivity	32.474	**0.001**	25.164	**0.001**
	Turbulence	46.058	**0.001**	105.570	**0.001**

Table S2. Weighting values of the phenotypic and environmental variables contributing to the phenotypic (PC) and environmental (EC) PCA components.

		Symphodus ocellatus			***Symphodus tinca***		
		PC1	**PC2**	**PC3**	**PC1**	**PC2**	**PC3**
Phenotypic	**PLD**	0.256	0.770	-0.219	-0.624	-0.070	0.440
	Settlement size	-0.550	0.596	0.006	-0.397	0.664	0.331
	PLD Growth rate	-0.760	-0.111	0.215	0.320	0.744	-0.222
	Hatching date	-0.234	-0.199	-0.952	0.592	-0.030	0.805
	Proportion of variance (%)	40.53	35.68	23.60	60.88	33.38	5.56
		EC1	**EC2**	**EC3**	**EC1**	**EC2**	**EC3**
Environmental	**Productivity**	0.610	0.441	0.659	0.640	-0.240	-0.730
	Surface temperature	-0.464	0.873	-0.154	-0.616	0.407	-0.674
	Turbulence	0.643	0.211	-0.736	0.459	0.881	0.113
	Proportion of variance (%)	55.83	27.50	16.67	54.06	27.96	17.98

Table S3. Pairwise genetic distances (F_{ST}) among populations of *Symphodus ocellatus* for all 3978 haplotype loci (below diagonal) and their p-values (above diagonal). Bold: Significant p-values after FDR correction (p <0.0204).

	Colera	Blanes	Xabia	Palos	Aguamarga	Herradura
Colera		0.798	**0.001**	**0.001**	**0.001**	**0.001**
Blanes	-0.001		**0.001**	**0.001**	**0.001**	**0.001**
Xabia	**0.029**	**0.028**		0.039	0.022	**0.001**
Palos	**0.041**	**0.041**	0.003		0.114	**0.001**
Aguamarga	**0.038**	**0.038**	0.004	0.001		**0.001**
Herradura	**0.022**	**0.023**	**0.036**	**0.049**	**0.042**	

Table S4. Pairwise genetic distances (F_{ST}) among populations of *Symphodus tinca* for all 5276 haplotype loci (below diagonal) and their p-values (above diagonal). Bold: Significant p-values after FDR correction (p< 0.0219).

	Colera	Blanes	Ametlla	Xabia	Santa Pola
Colera		0.027	0.125	0.297	0.235
Blanes	0.001		0.457	0.030	**0.002**
Ametlla	0.001	0.000		0.067	**0.003**
Xabia	0.000	0.001	0.001		0.773
Santa Pola	0.000	**0.002**	**0.002**	0.000	

Table S5 (*Symphodus ocellatus* loci, and their sequences, showing significant values with different methodologies) and Table S6 (*Symphodus tinca* loci, and their sequences, showing significant values with different methodologies) are available at:

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-020-69160-2/MediaObjects/41598202069160*MOESM2ESM.xlsx*

Table S7. Adjusted R^2 and degrees of freedom (Df) for environmental and phenotypic variables as a result of the variance partition analysis.

	Df	*Symphodus ocellatus*	*Symphodus tinca*
Environmental	3	0.0135	0.0016
Phenotypic	4	0.0073	-0.0003
Environmental+Phenotypic	7	0.0204	0.0006

Table S8. Number of haplotype loci containing SNPs detected as significantly associated to the different environmental and phenotypical variables considering the three used methodologies: Genome-wide associated studies without (GWAS) and with covariates (GWAS-cov) and redundancy analysis (RDA). NA indicate that this variable was not included in the analysis.

		S. ocellatus			*S. tinca*		
		RDA	GWAS-Cov	GWAS	RDA	GWAS-Cov	GWAS
Environmental	Temperature	54	0	110	55	0	0
	Productivity	70	0	140	48	1	1
	Turbulence	32	0	0	66	3	3
	EC1	NA	1	35	NA	0	0
	EC2	NA	0	1	NA	3	3
	EC3	NA	0	0	NA	0	1
Phenotypic	Hatching date	NA	0	0	NA	2	2
	PLD	NA	0	1	NA	0	0
	Settlement size	NA	0	5	NA	0	0
	PC2	NA	0	21	NA	0	0

Table S9 (Functions of the GO terms of the genes associated to the loci potentially under selection for *Symphodus ocellatus*) and Table S10 (Functions of the GO terms of the genes associated to the loci potentially under selection for *Symphodus tinca*) are available at:

https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-020-69160-2/MediaObjects/41598202069160*MOESM2ESM.xlsx*

Capítulo 2.2

Helping decision making for a reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species

Anna Barbanti[1]*, Hector Torrado[1,2]*, Enrique Macpherson[2], Luca Bargelloni[3], Rafaella Franch[3], Carlos Carreras[1§], Marta Pascual[1§].

[1]Department of Genetics, Microbiology and Statistics and IRBio, University of Barcelona, Diagonal 643, 08028 Barcelona, Spain.
[2]Center for Advanced Studies of Blanes (CEAB-CSIC). C/ d'accés a la Cala St. Francesc, 14, 17300 Blanes, Girona, Spain.
[3]Department of Comparative Biomedicine and Food Science. University of Padova, Viale dell'Universita' 16, I-35020 Legnaro, Italy.
*These authors contributed equally to the manuscript
§ These author are considered senior authors

Published as:

Anna Barbanti, Héctor Torrado, Enrique Macpherson, Luca Bargelloni, Rafaella Franch, Carlos Carreras, Marta Pascual (2020) Helping decision making for a reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. Molecular Ecology Resources (2020) 20:795– 806. https://doi.org/10.1111/1755-0998.13144

Abstract

High-throughput sequencing has revolutionized population and conservation genetics. RAD sequencing methods, such as 2b-RAD, can be used on species lacking a reference genome. However, transferring protocols across taxa can potentially lead to poor results. We tested two different IIB enzymes (AlfI and CspCI) on two species with different genome sizes (the loggerhead turtle Caretta caretta and the sharpsnout seabream Diplodus puntazzo) to build a set of guidelines to improve 2b-RAD protocols on non-model organisms while optimising costs. Good

results were obtained even with degraded samples, showing the value of 2b-RAD in studies with poor DNA quality. However, library quality was found to be a critical parameter on the number of reads and loci obtained for genotyping. Resampling analyses with different number of reads per individual showed a trade-off between number of loci and number of reads per sample. The resulting accumulation curves can be used as a tool to calculate the number of sequences per individual needed to reach a mean depth ≥ 20 reads to acquire good genotyping results. Finally, we demonstrated that selective-base ligation does not affect genomic differentiation between individuals, indicating that this technique can be used in species with large genome sizes to adjust the number of loci to the study scope, to reduce sequencing costs and to maintain suitable sequencing depth for a reliable genotyping without compromising the results. Finally, we provide a set of guidelines to improve 2b-RAD protocols on non-model organisms with different genome sizes, helping decision-making for a reliable and cost-effective genotyping.

Keywords: Conservation genomics, high-throughput sequencing, *Caretta caretta*, *Diplodus puntazzo*, genotyping-by-sequencing, sequencing simulations

Introduction

High-throughput sequencing technologies have revolutionized the fields of population and conservation genetics during the last ten years by providing easy access to genomic data from virtually any taxonomic group (Andrews & Luikart, 2014; Bellin et al., 2009; Davey & Blaxter, 2011; Hudson, 2008). Many studies have explored the potential of genomic analysis to address a variety of questions, such as population structuring (Girault, Blouin, Vergnaud, & Derzelle, 2014), inbreeding depression (Hoffman et al., 2014), local adaptation (Savolainen, Lascoux, & Merilä, 2013) or hybridization (Hohenlohe, Amish, Catchen, Allendorf, & Luikart, 2011). Restriction site associated techniques (RAD) are based on massive sequencing after enzymatically

reducing the fraction of the genome being analysed and can identify and score thousands of genetic markers, randomly distributed across the genome in many individuals simultaneously (Davey & Blaxter, 2011; Pecoraro et al., 2016). The advantage of these methodologies is that they can be carried out with no or limited previous sequence knowledge, since RAD tags can be analysed using pipelines for *de novo* loci identification if a reference genome is not available (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Davey & Blaxter, 2011; Hapke & Thiele, 2016; Lu, Glaubitz, Harriman, Casstevens, & Elshire, 2012). These methods allow parallel and multiplexed sample sequencing of tag libraries, with a rapid and very cost-effective procedures resulting in high genome coverage (Baird et al., 2008; Pecoraro et al., 2016). The RAD marker approach has the flexibility to assay different number of markers depending on the restriction enzyme of choice (Baird et al., 2008).

Many studies focusing on population structure in non-model organisms have implemented different RAD technologies, such as RADseq (e.g. Lim et al., 2017; Xu et al., 2014), ddRAD (e.g. Lavretsky, DaCosta, Sorenson, McCracken & Peters, 2019; Portnoy et al., 2015), GBS (e.g. Carreras et al., 2017; Hess et al., 2015), and 2b-RAD (e.g. Boscari et al., 2019; Galaska, Sands, Santos, Mahon, & Halanych, 2017). By shifting the realms of genomics from laboratory-based studies of model species towards studies of natural populations of ecologically well-characterized organisms, researchers can now start to address important ecological and evolutionary questions on a scale and precision that, only a few years ago, was unrealistic (Ekblom & Galindo, 2011). As for all genotyping-by-sequencing methodologies, the mean number of reads per locus (mean depth of coverage) is crucial to consider reliable the quality of markers and their genotypes (Sims, Sudbery, Ilott, Heger, & Ponting, 2014). Some recent population studies prioritised the number of sequenced individuals over depth of coverage or used improved bioinformatics pipelines to extract information from low coverage data (Buerkle & Gompert, 2013; Maruki & Lynch, 2014). However, when depth is generally low, statistical uncertainty of individual sequence data is high and calling of genotypes is difficult (Maruki & Lynch, 2017). Although probabilistic genotyping methods are thought to overcome shortcomings of low-depth sequencing data, they may behave

unpredictably when compared to high-depth data (Hendricks et al. 2018). Thus, any analysis involving individual genotypes is going to be limited by coverage (Chow, Anderson & Shedlock, 2019). For this reason, RAD sequencing techniques and laboratory protocols should be adjusted to target enough sequencing depth to obtain reliable genotypes while optimising sequencing costs.

2b-RAD is a RAD methodology that uses IIB restriction endonucleases, which cleave genomic DNA upstream and downstream of the target sites producing 32-34 bp fragments (Wang, Meyer, McKay, & Matz, 2012). This method is simple and provides a cost-effective alternative to existing reduced representation genotyping methods, allowing its use in routine experimental laboratory (Baird et al., 2008; Luo et al., 2017; Wang et al., 2012). One of the most interesting features of 2b-RAD is that the number of loci/marker density can be adjusted by using selective adaptors (Wang et al., 2012) to reduce the number of expected markers and increase the coverage per locus for a given sequencing effort. This RAD sequencing technique has been used to identify candidate genes associated with specific traits (Luo et al., 2017), to construct ultra-high density genetic maps (Fu, Liu, Yu, & Tong, 2016), to identify genomic regions under selection in population genetic studies (Pecoraro et al., 2016), and to perform genomic prediction for relevant traits in agricultural species (Palaiokostas, Ferraresso, Franch, Houston & Bargelloni, 2016). It has also been extended to microbial ecology (Pauletto et al., 2016).

In this paper, we provide a protocol for laboratory and bioinformatic analyses to optimise studies using 2b-RAD sequencing on different non-model organisms. We focused our study on the sharpsnout seabream *Diplodus puntazzo* Walbaum, 1792 and the loggerhead turtle *Caretta caretta* Linnaeus, 1758 characterized by very different genome sizes. This study aims to unveil key elements to adapt library building of non-model organisms to best profit from this genomic method. Specifically, we focused our analyses on five main objectives. 1) Assess the effect of initial DNA quality and concentration on sequencing results. 2) Evaluate the performance of different IIB enzymes (i.e. AlfI and CspCI) on genomic library construction in the two species. 3) Calculate the optimum number of raw sequences needed per each combination of

species and enzyme in order to achieve the maximum number of loci with an optimum depth per locus for a correct genotyping. 4) Assess if selective base ligation protocols have an impact on genetic differentiation among individuals. 5) Set guidelines for new population genomic studies on non-model organisms. Our study provides useful information for future studies on non-model species with different genome sizes, helping decision-making to obtain a reliable and cost-effective genotyping.

Methods

Samples

We analysed two species with approximately three-fold different genome sizes. We consider the sharpsnout seabream (*Diplodus puntazzo*) genome size to be similar to that of *Diplodus anularis* (0.9Gb), its closest relative's sequenced genome (www.genomesize.com). The loggerhead turtle (*Caretta caretta*) genome size was considered to be similar to the genome of *Chelonia mydas* (Wang et al., 2013), which measures 2,24Gb.

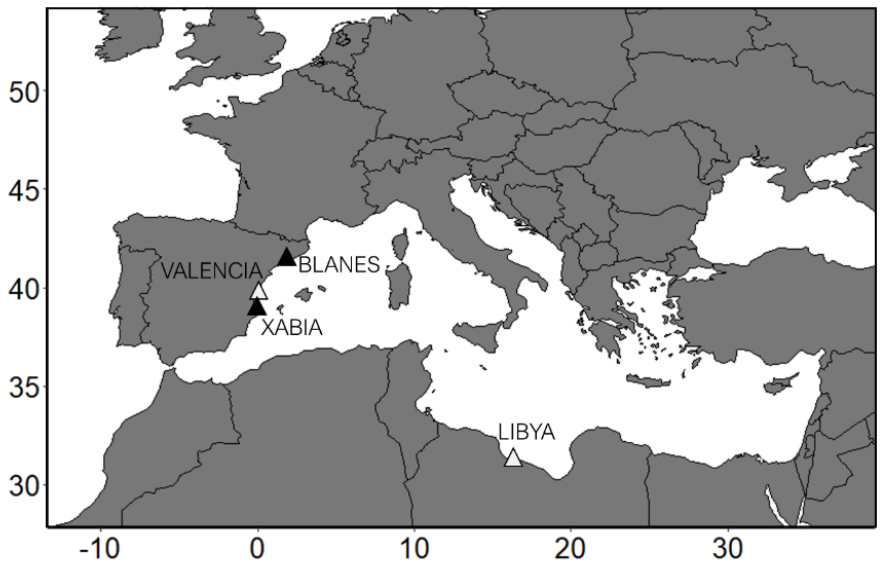


Figura 1: Sampling sites. White triangles show sampling sites for *C. caretta*, Libya is a nesting ground while Valencia is a foraging ground. Black triangles show sampling sites for *D. puntazzo*.

Juveniles of *D. puntazzo* were collected in Blanes (N=12) and Xabia (N=12) (Spain) during recruitment using hand nets (Figure 1). Samples of *C. caretta* were taken from bycaught juveniles at the foraging ground off Valencia (Spain) (N=9) (Figure 1) and from dead hatchlings at the nesting ground west of Sirte (Lybia) (N=14) (Clusa et al., 2018). We also added a sample collected from a live female turtle nesting in Pulpí (Spain) as positive control (Carreras et al., 2018). Consequently, our study included 24 samples per species. All samples were stored in 96% ethanol.

DNA extraction and library construction

Genomic DNA was extracted using Qiagen® Qiamp blood and tissue kit following the manufacturer's protocol. DNA concentration was measured with Nanodrop® or Qubit®, and DNA degradation assessed in 1% agarose gels. This information was recorded to be used in further statistical analysis. We coded the level of degradation as 'high' if the DNA was mostly located at the bottom of the run in the agarose gel or the smear intensity increased in direction top-to-bottom, and as 'low' if the DNA was mostly located at the top of the gel or the smear intensity faded in direction top-to-bottom. When possible we included samples that presented degraded DNA or low DNA concentration to test 2b-RAD efficiency for population genomics, as DNA degradation is a common issue when sampling non-model organisms (*e.g.* marine turtle studies targeting stranded individuals or dead embryos found after excavation of nests). A total of 24 individual libraries were constructed with each enzyme per species. Individual libraries were prepared adjusting the protocol from Wang et al. (2012). In brief, the construction of 2b-RAD libraries consisted of four main steps (for detailed protocol, see Annex I). i) Genomic DNA was digested by a IIB restriction enzyme providing short (32-34 bp) sequences. Each individual sample was digested with *AlfI* and *CspcI* enzymes separately. ii) During ligation, adaptors were attached to the sticky ends of the digested sequences. This step is crucial in the library preparation process because at this point, adaptors can be customised to attach to any sticky end or to attach only to sticky ends with specific sequences, based on the last two bases of the adaptor. For this study we used degenerated bases (5'-NN-3') for our adaptors

(Figure 2). iii) In the amplification step, barcodes and Illumina primers were attached to the adaptors and sequences were amplified by PCR. At the end of this step the resulting fragment is expected to measure ~165 bp. Library products were run through a 1.8% agarose gel to check amplification success. The library DNA quality of each sample was coded as 'good' when the band of the agarose gel was bright or 'bad' when it was faint (Figure S1). iv) Purification was performed using magnetic beads to remove primers and sequences longer and shorter than 165 bp. At the end of this step, 2b-RAD libraries were ready to be sequenced. The DNA concentration of purified libraries was quantified using a Real Time PCR. The 48 libraries of each species were pooled for SR50 single read sequencing (one species per lane) with a HiSeq 4000 Illumina at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center.

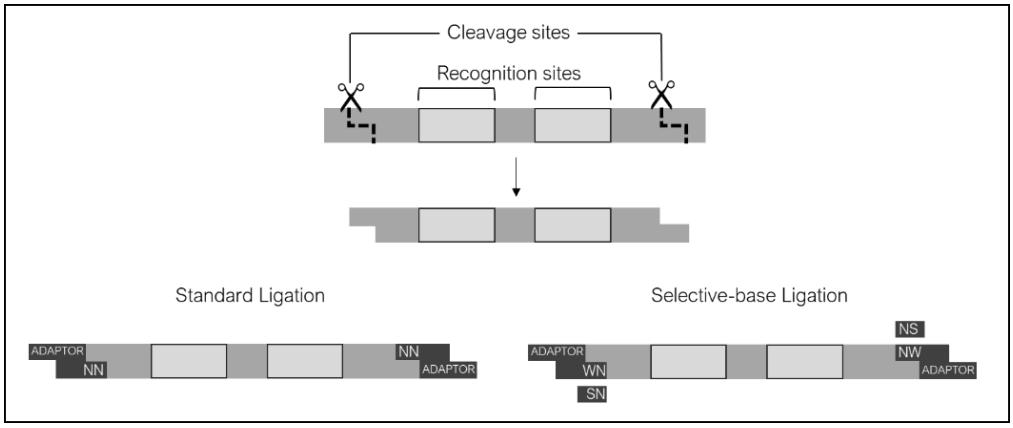


Figura 2: Selective-base ligation. In 2b-RAD protocol, after IIB enzyme digestion, specific fragments can be selected to reduce the density of markers to be amplified by designing customised adaptors with one fully degenerated base (N) and one partially degenerated base (S = G and C bases, W = A and T bases).

Genotyping

Sequences were processed using customized scripts (Annex II). First, raw sequences were trimmed to eliminate ligation adaptors and then cut down to the same length (*i.e.* 32bp for CspCI and 34bp for AlfI). Processed sequences were used for genotyping using the STACKS vs 1.47 pipeline (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen et al., 2013). To construct a loci catalogue we used Stacks

function *denovo_map.pl* setting the following parameters: a minimum depth of three reads to consider a stack within an individual (m = 3), up to three mismatches allowed between stacks (putative alleles) to merge them into a putative locus within an individual (M = 3), and two mismatches allowed between stacks (putative loci) during construction of the catalogue (n = 2). Individual genotypes were outputted as haplotype loci VCF files. We used 5 main filters to process loci found in our samples. We removed individual genotypes based on less than 5 reads, loci present in less than 70% of individuals, loci with outlier values of mean depth across all individuals (those above the upper whisker of the R 'boxplot' , corresponding to 1.5 times the interquartile range from the data), loci with a major allele frequency higher than 99% and loci out of Hardy-Weinberg equilibrium (HWE) in at least one of the populations. In the case of *C. caretta* HWE was considered only for Libya, since Valencia is a feeding aggregation of individuals from different populations, and thus deviations from HWE are expected (Clusa et al, 2014). Filtering was performed with VCFtools vs 1.12 (Danecek et al., 2011), with the exception of loci with a major allele frequency higher than 99%, which were identified by the function isPoly from the package 'adegenet' (Jombart, 2008) and the assessment of HWE, computed with the function hw.test from the package 'pegas' (Paradis, 2010) in R (R Core Team, 2018). We performed linear regression and Wilcoxon-Mann-Whitney test in R to assess whether initial and library DNA concentrations, initial DNA degradation and library quality influenced the number of total sequences and the final number of loci of each sample.

Resampling analysis

We used bioinformatic simulations for each species and enzyme to obtain several sample sets, each one presenting a different number of reads per individual. We used a customised script to create new sample sets with different number of reads per sample by performing a random selection with replacement of the real data up to different target numbers of raw reads per sample (Annex II). We performed 10 iterations for each target number. Target numbers varied for each species to accommodate the data points to the expected accumulation curve results for the

different genome sizes. For *D. puntazzo* we simulated 0.5, 1, 2, 4, 8 and 10 million raw reads per sample for CspCI and AlfI enzymes. For *C. caretta* we simulated 4, 8, 12, 16 and 20 million raw reads per sample for each enzyme. Each resampled set underwent the same process of loci identification and filtering as explained above with the exception of the filter removing loci out of Hardy-Weinberg equilibrium. This filter was not applied because loci genotyping could be biased in the low depth datasets, artificially creating loci out of H-W equilibrium, since resampling was done with replacement. For this reason, this technique should not be used to artificially increase locus depth for a proper genotyping, as these genotyping errors are going to persist in the extended datasets. We calculated the formula that best fitted the accumulation curve for each species and enzyme and plotted the curve with R package 'ggplot2' (Wickham, 2016). We calculated the number of reads per individual needed to obtain a mean depth of coverage of 20x, since this threshold of quality is used in 2b.RAD studies (Resh, Galaska & Mahon, 2018; Whelan et al., 2019). We also estimated values for a coverage of 25x (Warmuth & Ellegren, 2019) to evaluate if with higher coverage we can detect an improvement in the number of total loci.

Selective-base ligation simulation

We assessed the potential impact of reducing the number of loci by selective-base-ligation in population genomic analyses. We bioinformatically selected trimmed reads of the corresponding combination of nucleotides to simulate the use of customised adaptors for selective-base ligation on each combination of species and enzyme (Annex II). This type of ligation is usually performed in the laboratory by designing adaptors that will attach only to reads having the target base at both sticky ends (Figure 2). The simulation of a selective-base-ligation aims to test whether the processing of a proportionally lower number of loci per individual results in the same genetic differentiation as for the whole sample set. We removed from this analysis all samples that had a final mean depth per locus < 10 to eliminate errors given by low depth of coverage. For *D. puntazzo* no samples were removed, while for *C. caretta* 5 samples were removed from the AlfI sample set and 7 form CspCI sample set. We used

a customized script simulating the effects of building libraries with adaptors ending in 5'-WN-3' (W = A and T) or 5'-SN-3' (S = G and C) instead of 5'-NN-3'. These simulations aimed to select trimmed sequences by their first and last base and allocate them in separate folders. These selected sequences were then analysed with Stacks and loci were filtered with the same process as explained above for the whole dataset. We calculated the genetic differentiation between pairs of individuals using Prevosti distance with the R function prevosti.dist from the package 'poppr' 2.8.0 (Kamvar, Tabima, & Grünwald, 2014; Kamvar, Brooks, & Grünwald, 2015) for the dataset containing all combinations (NN) and for the two simulated selective-base-ligation datasets. The pairwise genetic distance matrixes among individuals for each selective-base-ligation subset were compared to the original NN matrix with a Mantel test using Genalex v6.503 (Peakall & Smouse, 2012), then for each matrix we ran a Principal Coordinate Analysis (PCoA) to evaluate whether individuals maintained the same clustering pattern among subsets, using the same program. To detect the eventual decrease of heterozygosity in the subsets compared to their original set of loci we calculated individual observed heterozygosities for the three datasets with VCFtools and used R to perform a Kruskal-Wallis test for each species and enzyme.

Results

Library construction and loci identification in *C. caretta*

In *C. caretta* extracted DNA ranged from 17.3 to 133.5 ng/µl, and showed high level of degradation in 38% of the samples probably due to the bad condition of the tissue used (Table S1). After adaptor ligation and amplification by PCR we observed generally good results with AlfI but much lower amplification success with CspCI with 46% of faint bands, as assessed with gel electrophoresis (Tables S1). After purification, library DNA concentration was similar for the two enzymes ranging between 6.7 and 52.3 ng/µl. The mean number of reads per sample was higher for AlfI digested samples, $7.6x10^6$ reads per sample (max $10.1x10^6$, min $4.0x10^6$), than for CspCI digested samples, $6.6x10^6$ reads per sample (max $10.7x10^6$, min $2.6x10^6$) mostly because some

samples had low number of reads (Table S1). The trimming process discarded all the sequences that were shorter than 34bp for AlfI and 32bp for CspCI or missed the chosen restriction site, with an average (±SE) lower loss per sample in AlfI (19.2±2.1%) than in CspCI (41.9±4.7%) (Table 1). After the loci calling, *C.caretta* showed higher total number of loci with AlfI (66907 loci) than CspCI (25416 loci). The mean number of loci retained after all filtering steps were slightly higher for AlfI (72.9±0.4%) than for CspCl (69.4±0.9%), although their final mean depth was smaller (Table 1).

Tabla 1: Summary of sequencing outcome. Mean (±SE) values per individual are given for TR: total number of reads, TMR: number of trimmed reads, IL: initial number of loci, FL: final number of loci after filtering, RL: percentage of loci retained after filtering, and FMD: final mean depth of coverage per locus.

Species	*C. caretta*		*D. puntazzo*	
Enzyme	AlfI	CspcI	AlfI	CspcI
TR	7.6±0.3x10^6	6.6±0.4x10^6	7.1±0.3x10^6	6.5±0.3x10^6
TMR	6.2±0.4x10^6	4.2±0.5x10^6	5.3±0.2x10^6	4.3±0.2x10^6
IL	48740±1489	17811±1010	75971±130	27989±40
FL	35576±1124	12455±732	68978±115	25421±27
RL	72.9±0.4%	69.4±0.9%	90.6±0.1%	90.8±0.1%
FMD	11.5±0.7	19.3±2.4	29.2±1.4	52.2±2.3

Library construction and loci identification in *D. puntazzo*

In *D. puntazzo* starting concentrations ranged from 22.3 to 43.1 ng/µl and none of the samples was degraded. Adaptor ligation and amplification yielded successful amplifications with both enzymes although 17% of the samples digested with CspCl had faint bands (Table S2). After purification, library DNA concentration was slightly higher for Alfl ranging between 13.6 and 109.63 ng/µl. As for *C. caretta* the sequencing of AlfI in *D. puntazzo* resulted in slightly higher mean number of reads per sample than for CspCI (Table 1). After the loci calling and filtering higher number of loci were also found for *D.puntazzo* for AlfI (84382 loci) than for CspCI (31111 loci). The mean number of loci retained after all filtering steps was similar for AlfI (90.6±0.1%) than for CspCl (90.8±0.1%), although their final mean depth was almost double in the latter (Table 1).

Quality predictors of sequencing success

In the two species analysed and for both restriction enzymes the number of raw reads was significantly correlated to the final number of loci (Table 2). For *D. puntazzo*, initial DNA concentration, DNA degradation and library DNA quality had no significant effect in the number of raw reads or number of loci. However, for CspCI in *C. caretta,* the initial DNA concentration showed a significant impact on number of reads and loci, and on library concentration (Table 2). The library DNA concentration explained sequencing success in both species since the regression between library DNA concentration and the number of reads and loci was significant in most cases, with the exception of AlfI in *C. caretta* and the number of loci with CspCl in *D. puntazzo* (Table 2). The impact of DNA degradation on sequencing success was only assessed in *C. caretta* since in *D. puntazzo* DNA had initial good quality (Tables S1 and S2). Interestingly, initial DNA degradation was not a good predictor of neither the number of reads nor loci (Table 2). However, library DNA quality and thus amplification success assessed in an agarose gel significantly increased the number of raw reads and final number of loci (Table 2).

Resampling analysis

We simulated the sequencing of different target number of reads per sample set and we obtained the total number of loci and mean depth for each simulation (Figure 3, Table S3). In all simulations, the mean depth of coverage was highly correlated to the number of reads per individual with an R^2 >0.99. Based on the accumulation curve (Figure 3) we estimated the mean number of reads per individual and the corresponding number of loci for two mean depth of coverage, 20x and 25x (Table 3). For both species, AlfI needed a much higher number of reads per individual than CspCI to reach the desired coverage of 20x, due to the higher number of loci obtained with this enzyme. We found that, using a coverage of 25x, the total number of final loci improved in AlfI by 4% and by 7% for *D. puntazzo* and *C. caretta* respectively, and by 9% in CspCI for both species.

Table 2: Statistical analyses of potential quality predictors. In bold are shown significant p-values after FDR correction. na: tests not available due to insufficient samples with bad initial DNA quality or low library DNA quality.

			Caretta caretta				***Diplodus puntazzo***			
			CspCI		*AlfI*		*CspCI*		*AlfI*	
Explanatory variable	**Response Variable**	**Test**	**F or W**	**p value**	**F or W**	**p value**	**F or W**	**p value**	**F or W**	**p value**
Raw reads	Final loci	Linear Regression	17.7	**0.000**	30.4	**0.000**	4.7	**0.041**	34.4	**0.000**
Initial DNA concentration	Raw reads	Linear Regression	15.4	**0.001**	2.7	0.115	0.5	0.469	0.4	0.522
	Final loci	Linear Regression	5.2	**0.032**	2.1	0.159	0.0	0.959	1.5	0.236
	Library DNA concentration	Linear Regression	15.8	**0.001**	0.0	0.986	3.2	0.087	0.3	0.611
Initial DNA degradation	Raw reads	Wilcoxon-Mann-Whitney	60.0	0.682	61.0	0.726	na	na	na	na
	Final loci	Wilcoxon-Mann-Whitney	44.0	0.170	45.0	0.194	na	na	na	na
	Library DNA concentration	Wilcoxon-Mann-Whitney	34.0	**0.048**	44.0	0.174	na	na	na	na
Library DNA concentration	Raw reads	Linear Regression	14.2	**0.001**	2.0	0.174	22.6	**0.000**	6.3	**0.020**
	Final loci	Linear Regression	20.3	**0.000**	3.2	0.086	0.4	0.559	6.2	**0.021**
Library DNA quality	Raw reads	Wilcoxon-Mann-Whitney	19.0	**0.002**	13.0	**0.037**	na	na	25.0	0.261
	Final loci	Wilcoxon-Mann-Whitney	12.5	**0.001**	6.0	**0.005**	na	na	9.0	**0.018**

Selective-base ligation simulation

The selective-base ligation subsets obtained from *C. caretta* retained between 22.2% and 31.5% of the total loci from their original sample sets (Table S4). In *D.puntazzo* the amount of loci retained was more variable between the two tested subsets (Table S4), ranging from 19.8% to 43.4%. In this species we also found that for CspCI enzyme the subsets presented lower coverage than the original set, which could be a consequence of the base composition of the regions where this enzyme is cutting and related with the characteristics of the genomes that make the results species specific (Seetharam & Stuart, 2013). Mantel tests in both species showed high correlation between the pairwise genetic distances among individuals assessed with all loci and assessed with a selective base ligation, for both CspCI and AlfI enzymes (Figure 4). This was also reflected in the PCoA, as *C. caretta* samples do not have the exact same pattern among subsets whereas *D. puntazzo* patterns match perfectly despite the lower number of loci retained in the different datasets (Figure S2). The Kruskal-Wallis test showed no significant differences in observed heterozygosity among any of the subsets and the original set of loci for both species and enzymes (Table S5).

Table 2: Estimated number of loci and reads needed to obtain different mean depth per locus as derived from the accumulation curve. The table shows the number of reads per individual and the total number of loci per set corresponding to a mean depth of coverage of 20x and 25x for each species and enzyme.

		Caretta caretta		*Diplodus puntazzo*	
		AlfI	**CspCI**	**AlfI**	**CspCI**
20x	Reads (10^6)	13.5	6.1	3.5	1.7
	Loci	142910	49588	68079	22225
25x	Reads (10^6)	17.4	7.9	4.6	2.2
	Loci	152998	53842	70571	24173

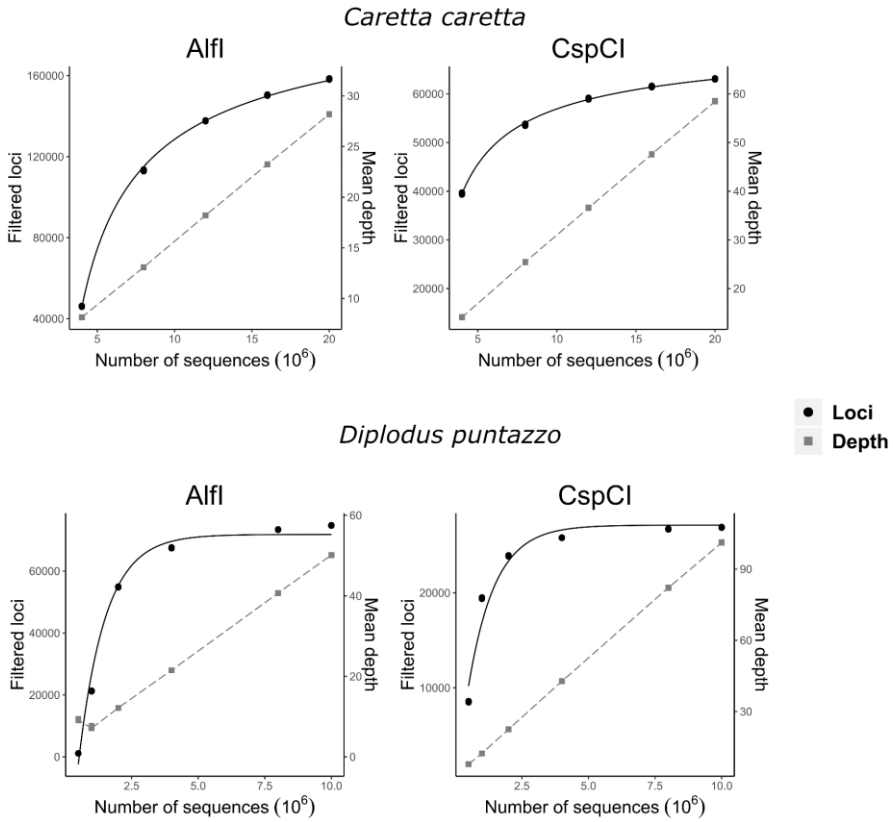


Figura 3 Accumulation curves resulting from the resampling analysis. The graphs show the number of final loci (circles) and the mean depth per locus (squares) obtained after filtering, for *C. caretta* (top) and *D. puntazzo* (bottom).

Protocol optimization

We used the results obtained from these simulations to refine the laboratory protocol for 2b-RAD libraries preparation and sequencing. In fact, given the mean value of depth of coverage, the optimum number of loci and the size of the studied species genome, we can calculate the number of samples to be sequenced in one lane to optimize costs without compromising the results. To facilitate the decision-making process, based on our results, we constructed a flowchart (Figure 5) and a set of guidelines (Box 1) to help future studies design the most efficient and cost effective protocol to reach their goals.

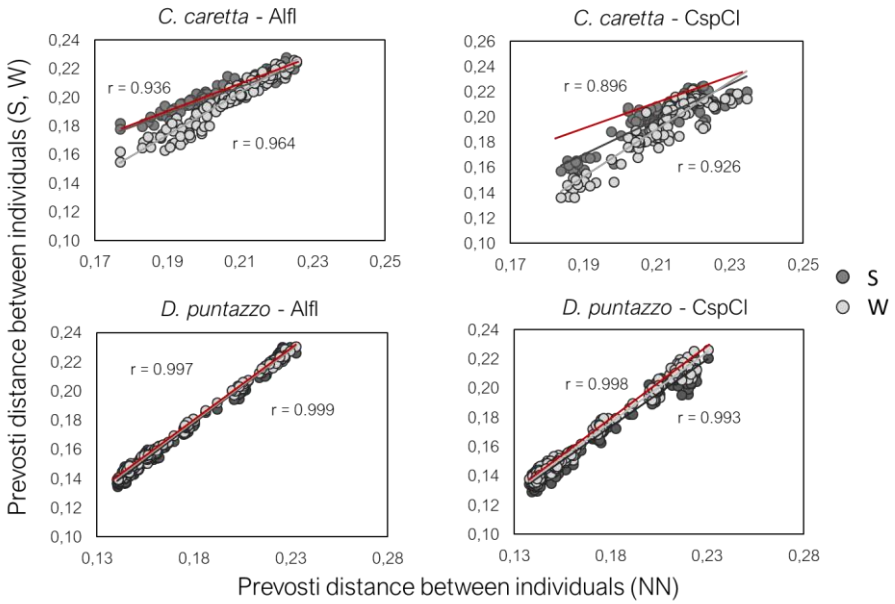


Figura 4 Mantel test of genetic differentiation between selective-base subsets and original sets. X-axes show Prevosti distance between pairs of individuals for each one of the four original sample sets (with fully degenerated bases –NN-). Y-axis show Prevosti distance between the same pairs of individuals for subsets obtained from bioinformatic simulations of selective base ligation (either –SN- or –WN-) for each species and enzyme. Dark grey shows genetic differentiation for S (G and C bases) subsets and light grey for W (A and T) subsets. Correlation coefficient (r) is given for each test above the lines for S and below for W. The red line represents the expected correlation function when no deviation in genetic distances is found in the selective-base subsets compared to NN.

Discussion

In this study, we have shown that 2b-RAD protocol provides efficient results even with degraded samples and we demonstrated how this protocol can be optimised for population genomics of non-model species with different genome sizes. To prove this point, we analysed the sharpsnout seabream *D. puntazzo* and the loggerhead turtle *C. caretta* with two different enzymes, AlfI and CspCI, and performed bioinformatic simulations. Our simulations allow estimating the mean number of reads needed per individual to obtain a reliable genotyping and the corresponding expected number of loci. Moreover, our results indicate that selective-base ligation can be used without compromising pairwise genetic distances among individuals.

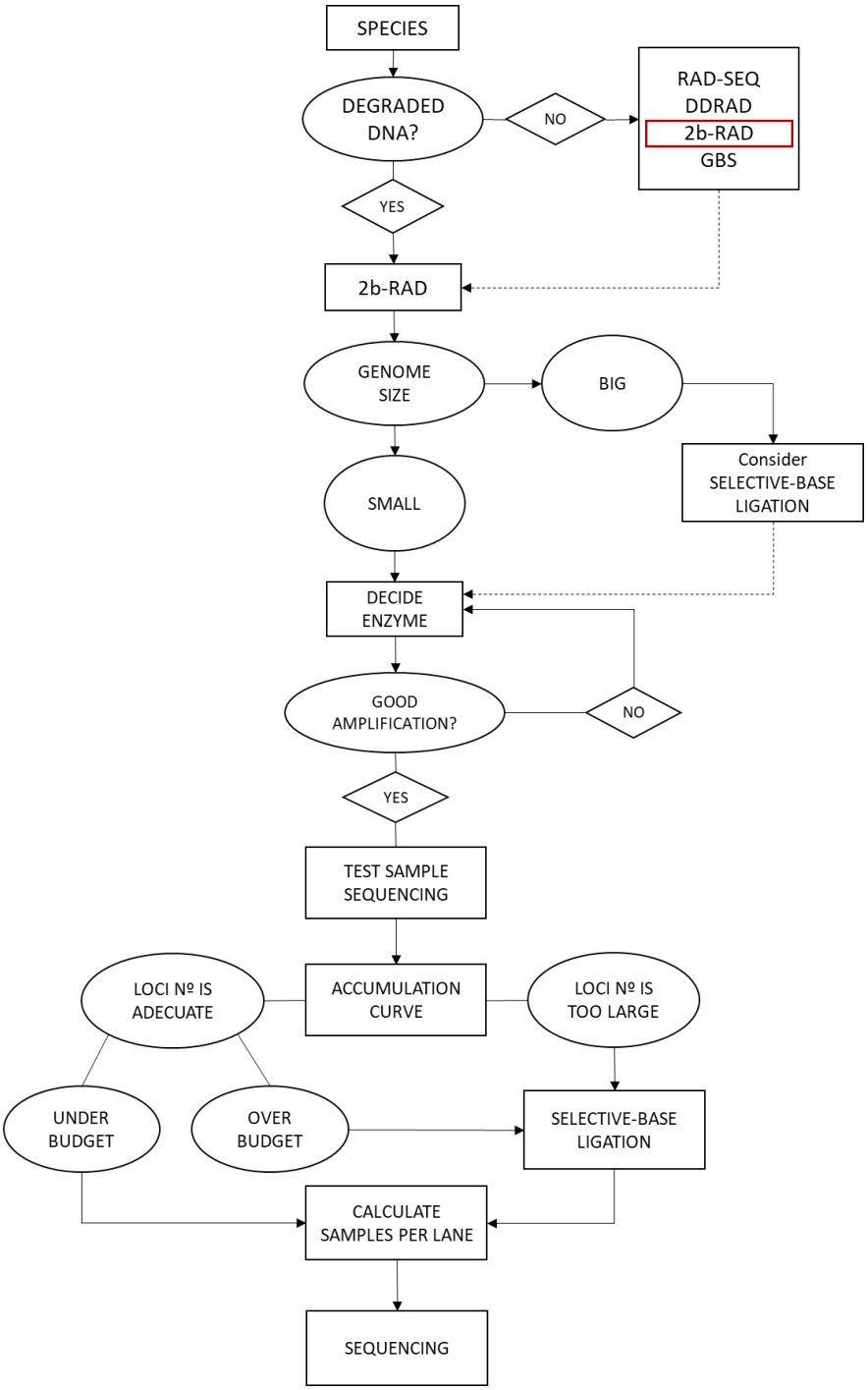


Figura 5 Flowchart for 2b-RAD laboratory protocol. This flowchart is meant to aid decision making for 2b-RAD laboratory protocols when studying non-model species. Together with the guidelines listed above this chart aims to make 2b-RAD studies not only easier but also more cost-effective.

In the case of the loggerhead turtle, where several samples had highly degraded DNA, we found that the quality of the initial DNA did not affect the number of raw reads nor the final number of loci, for both enzymes. In fact, the DNA short length for proper IIB enzyme functioning (i.e. 32-34bp digested fragment) reduces the probability of missing loci even in highly degraded samples. This is a highly valuable characteristic of 2b-RAD methodology, since not all studies can easily access high quality samples. For instance, marine turtle genetic studies usually rely on sampling of stranded individuals (Clusa et al., 2016) or dead embryos found at nests after excavation (Clusa et al., 2018), due to the complexity of their behaviours and the paucity of individuals. In such cases, a genomic protocol capable of providing optimal results with degraded samples is invaluable.

The library quality after adaptor ligation and amplification was a good predictor of sequencing success. The electrophoresis gel after the library amplification of the loggerhead turtle clearly showed that AlfI resulted in a better amplification than CspCI, which failed to yield a clear band in 46% of individuals. Moreover, the sequencing success was poor for samples with faint amplification bands, which resulted in lower number of reads per individual and thus lower number of loci. We thus suggest discarding samples with poor library DNA quality to help optimising sequencing costs. In the case of the sharpsnout seabream, both enzymes showed good results after the amplification, although a few individuals yielded poorer amplification that resulted in significantly lower number of loci, as observed also in the loggerhead turtle were the difference in library quality with the two enzymes was even greater. Moreover, AlfI provided higher number of loci than CspCI in both species as expected, since AlfI recognition sequence has six fixed nucleotides, while CspCI has seven fixed nucleotides. Therefore, AlfI is expected to have a greater density of restriction sites across any genome than CspCI, and potentially yield more loci as observed in the kissing bug *Rhodnius ecuadoriensis* (Hernandez-Castro et al. 2017).

Obtaining more loci, though, reduces depth of coverage per locus for the same mean number of reads per individual. As expected, when using CspCI enzyme our sample sets showed higher values of mean depth then when using AlfI in both species,

despite poorer amplification success for CspCl in the loggerhead turtle. A low mean depth per locus leads to less accurate genotype calling and thus higher percentage of missing data across loci (Casso, Turon & Pascual, 2019; Maruki & Lynch, 2017; Chow et al., 2019), and for this reason a good depth coverage is important to consider data reliable. Since library construction and sequencing produces a variable number of reads per locus, a mean depth of 20x would guarantee that the minimum of five reads per genotype is consistently achieved across most loci for each sample. This would result in fewer genotypes lost and thus more loci retained over all samples. Our simulations on resampling analyses, allowed the construction of the accumulation curve relating the number of reads per sample and the resulting number of loci as well as the linear correlation between the mean depth per locus and the number of reads per individual. Based on the combination of these two functions the number of individuals to be sequenced in one lane can be calculated easily, simplifying decision-making and analysis design for optimizing population genomic studies at the lowest cost. The amount of reads per individual required by the sharpsnout seabream would allow including a fair number of individuals per lane for each enzymes, since both yielded good library DNA quality across samples. However, in the case of the loggerhead turtle, only AlfI enzyme should be used according to library DNA quality. In this case, the amount of reads needed to achieve an adequate coverage would be very large and the number of loci obtained very high, due to the size of the genome. Under these circumstances, the number of individuals of loggerhead turtle to be included in one sequencing lane would be too small and not affordable by most research groups.

The difference between the two species is mostly related to the crucial role played by the genome size. Species with large genomes will likely produce more loci (due to a greater number of regions yielding the enzyme recognition site) and would need a greater sequencing effort to reach the suitable number of reads per sample for an adequate genotyping. Using a selective-base ligation the number of individuals can be adjusted to the needs of the study considering the number of loci projected by the accumulation curve. Our simulations of customized adaptors with selective base ligation, which extremities would end in –WN or –SN, proved that this type of reduction

in the number of loci does not affect genetic differentiation between pairs of individuals. Therefore, the use of a selection of sequences for each sample instead of the whole set, would allow reducing costs by fitting more samples in one lane without compromising overall genetic differentiation. In both species we found that the subsets from the simulated selective-base ligation had a proportionally similar lower number of raw sequences and final loci than the original sets (~25%). However, some differences were observed according to the base and enzyme used in each species suggesting that the species' genome base composition may affect the outcome. Nonetheless, the high levels of correlation that we found between the subsets and the original sets, regardless of the number of loci retained, indicate that they are reliable sources of information. In fact, the slightly lower correlation in genetic distances of *C. caretta* and its differences in PCoAs patterns among subsets were probably a consequence of the bigger genome size of the species, resulting in a lower coverage. This type of selective ligation would be particularly interesting in the case of species with large genomes such as *C. caretta*. Considering the size of this species genome (2.24Gb) and referring to our resampling simulation, we would need 13.5-17.4 million reads per sample to achieve 20x-25x of coverage, therefore only 20-25 samples could be sequenced in the same lane of a platform providing 340 million reads per run as in the present study. A selective-base ligation would allow reducing the costs of sequencing while ensuring good loci coverage, without influencing the outcome. In fact, since the selective-base ligated set would need only ~25% of the original set, between 3.4 and 4.4 million reads per sample are expected to reach the adequate coverage (Warmuth & Ellegren, 2019). Therefore, as much as 78-100 samples could fit in the same Illumina lane, greatly reducing costs without compromising genetic differentiation between individuals. Nevertheless, the number of loci required for a study depends on the scope, the type of analysis performed, and the target species. For instance, selective-base ligation would be less powerful for studies aiming to identify adaptation, since the probability of finding candidate genes can decrease when analysing only a small fraction of the genome (Ahrens et al., 2018).

Finally, we show that 2b-RAD methodologies can be reliable even for degraded DNA samples. Following our set of guidelines, researchers can optimize effort, time, and sequencing cost of 2b-RAD library building for non-model species while maintaining good sequencing depth for a proper genotyping (Box 1, Figure 5).

Conclusions

Genomic population studies are increasing in species without reference genomes that rely on restriction-site associated DNA sequencing techniques, although some protocols require good quality DNA. Moreover, transferring protocols across taxa can potentially lead to poor results, such as low number of recovered markers or inadequate genotyping due to differential genomic features. Researchers working with species with large genome sizes or needing lower number of markers can adjust the number of loci by performing selective-base ligation, allowing the sequencing of a larger number of samples, without altering genomic differentiation between individuals as observed by our simulations. The optimal number of samples per lane can, therefore, be adjusted as a trade off with the desired target number of loci and the species genome size for an adequate mean depth of coverage for a correct genotyping. Our results and guidelines aim to improve 2b-RAD protocols on non-model organisms with different genome sizes, helping initial decision-making for a reliable, faster and cost-effective genotyping for population genomic studies.

Acknowledgements

M. Babucci (University of Padova) for helping with the initial trimming and preliminary analysis of sequenced data and J. Abril (University of Barcelona) for helping with script building for the simulation analysis with replacement. L. Cardona (University of Barcelona), J.Tomás (University of Valencia) and A.A. Hamza (Alfateh University and University Malaysia Terengganu) for providing tissue samples of *Caretta caretta* from the Valencian feeding ground and the Libyan rookery.

This work was supported by the project CTM2017-88080 from the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER). CC and MP are part of the research group SGR2017-1120, and EM of SGR2017-378 (Catalan Government). HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities (FPU15/02390). AB was supported by grant 2017 FI_B 00997 of the Catalan Government-AGAUR.

BOX 1

Guidelines for the optimisation of a 2b-RAD protocol with non-model species.

- Use 2b-RAD instead of other RAD sequencing techniques if you have degraded samples.
- If the target species has a big genome size, consider performing a selective-base ligation to retain 20-40% of total loci.
- If the species genome is small, proceed without selective base-ligation.
- Test different IIB enzymes with the target species.
- Use library quality and concentration as predictors of sequencing success.
- Sequence the test samples with conservative conditions to obtain good coverage.
- Calculate an accumulation curve in a preliminary analysis with the test samples to identify the number of reads needed per individual and the total number of loci corresponding to a coverage ≥20x.
- If the total number of loci is adequate for the selected type of study, proceed to sequence the rest of your samples to obtain the mean number of reads needed according to the curve.
- If the total number of loci is too high for the selected study, use a selective base ligation for library building to reduce the amount of loci.
- The number of samples to be sequenced in the same lane is a trade-off between the number of reads per individual, the number of reads provided per lane and available budget.
- If the total number of loci is adequate but the cost of sequencing is over budget, use a selective base ligation for further 2b-RAD library building to reduce the amount of reads needed per sample and therefore fit more samples in one lane.

References

Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D., & Dudaniec, R. Y. (2018). The search for loci under selection: trends, biases and progress. *Molecular Ecology*, 27(6), 1342-1356.

Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. *Molecular ecology*, 23(7), 1661-1667.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one*, *3*(10), e3376.

Bellin, D., Ferrarini, A., Chimento, A., Kaiser, O., Levenkova, N., Bouffard, P., & Delledonne, M. (2009). Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species. *BMC genomics*, *10*(1), 555.

Boscari, E., Abbiati, M., Badalamenti, F., Bavestrello, G., Benedetti-Cecchi, L., Cannas, R., ... & Fraschetti, S. (2019). A population genomics insight by 2b-RAD reveals populations' uniqueness along the Italian coastline in Leptopsammia pruvoti (Scleractinia, Dendrophylliidae). *Diversity and Distributions*, 25, 1101-1117.

Buerkle, C. A., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: how low should we go? *Molecular ecology*, 22(11), 3028-3035.

Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E., & Pascual, M. (2017). Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Scientific reports*, 7, 43417.

Carreras, C., Pascual, M., Tomás, J., Marco, A., Hochscheid, S., Castillo, J. J., ... & Cardona, L. (2018). Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (*Caretta caretta*). *Scientific reports*, 8(1), 1435.

Casso M., Turon X., Pascual M. 2019. Single zooids, multiple loci: independent colonisations revealed by population genomics of a global invader. *Biological Invasions*. https://doi.org/10.1007/s10530-019-02069-8

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics*, 1(3), 171-182.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology*, *22*(11), 3124-3140.

Chow, J. C., Anderson, P. E., & Shedlock, A. M. (2019). Sea turtle population genomic discovery: Global and locus-specific signatures of polymorphism, selection, and adaptive potential. *Genome biology and evolution*, 11(10), 2797-2806.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Giacoma, C., ... & Maffucci, F. (2014). Fine-scale distribution of juvenile Atlantic and Mediterranean loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. *Marine biology*, 161(3), 509-519.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Avolio, D., ... & Aguilar, A. (2016). Potential bycatch impact on distinct sea turtle populations is dependent on fishing ground rather than gear type in the Mediterranean Sea. *Marine biology*, 163(5), 122.

Clusa, M., Carreras, C., Cardona, L., Demetropoulos, A., Margaritoulis, D., Rees, A. F., ... &

Aguilar, A. (2018). Philopatry in loggerhead turtles *Caretta caretta*: beyond the gender paradigm. *Marine Ecology Progress Series*, 588, 201-213.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

Davey, J., & Blaxter, M. L. (2011). RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, 9, 108.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1.

Fu, B., Liu, H., Yu, X., & Tong, J. (2016). A high-density genetic map and growth related QTL mapping in bighead carp (*Hypophthalmichthys nobilis*). *Scientific reports*, *6*, 28679.

Galaska, M. P., Sands, C. J., Santos, S. R., Mahon, A. R., & Halanych, K. M. (2017). Geographic structure in the Southern Ocean circumpolar brittle star *Ophionotus victoriae* (Ophiuridae) revealed from mt DNA and single-nucleotide polymorphism data. *Ecology and evolution*, *7*(2), 475-485.

Girault, G., Blouin, Y., Vergnaud, G., & Derzelle, S. (2014). High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC genomics*, *15*(1), 288.

Hapke, A., & Thiele, D. (2016). GI b PS s: a toolkit for fast and accurate analyses of genotyping-by-sequencing data without a reference genome. *Molecular ecology resources*, 16(4), 979-990.

Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., ... & Sethuraman, A. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, 11(8), 1197-1211.

Hernandez-Castro, L. E., Paterno, M., Villacís, A. G., Andersson, B., Costales, J. A., De Noia, M., ... & Llewellyn, M. S. (2017). 2b-RAD genotyping for population genomic studies of Chagas disease vectors: Rhodnius ecuadoriensis in Ecuador. *PLoS neglected tropical diseases*, 11(7), e0005710.

Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., ... & Wildbill, A. J. (2015). Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, 15(1), 187-202.

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. ... & Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, 201318945.

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular ecology resources*, *11*, 117-122.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403-1405.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281.

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in genetics*, 6, 208.

Lavretsky, P., DaCosta, J. M., Sorenson, M. D., McCracken, K. G., & Peters, J. L. (2019). ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Molecular ecology*, 28, 2594-2609.

Lim, H. C., Gawin, D. F., Shakya, S. B., Harvey, M. G., Rahman, M. A., & Sheldon, F. H. (2017). Sundaland's east–west rain forest population structure: variable manifestations in four polytypic bird species examined using RAD-Seq and plumage analyses. *Journal of biogeography*, 44(10), 2259-2271.

Lu, F., Glaubitz, J., Harriman, J., Casstevens, T., & Elshire, R. (2012). TASSEL 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline documentation. *White Paper*, 2012, 1-12.

Luo, X., Shi, X., Yuan, C., Ai, M., Ge, C., Hu, M., ... & Yang, X. (2017). Genome-wide SNP analysis using 2b-RAD sequencing identifies the candidate genes putatively associated with resistance to ivermectin in *Haemonchus contortus*. *Parasites & vectors*, 10(1), 31.

Maruki, T., & Lynch, M. (2014). Genome-wide estimation of linkage disequilibrium from population-level high-throughput sequencing data. *Genetics*, 197(4), 1303-1313.

Maruki, T., & Lynch, M. (2017). Genotype calling from population-genomic sequencing data. *G3: Genes, Genomes, Genetics*, 7(5), 1393-1404.

Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R.D., & Bargelloni, L. (2016). Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3 (Bethesda)*, 8;6(11), 3693-3700. doi:10.1534/g3.116.035220

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics*, 26(3), 419-420.

Pauletto, M., Carraro, L., Babbucci, M., Lucchini, R., Bargelloni, L., & Cardazzo B. (2016). Extending RAD tag analysis to microbial ecology: a comparison between MultiLocus Sequence Typing and 2b-RAD to investigate Listeria monocytogenes genetic structure. *Molecular Ecology Resources,* 16(3):823-35. doi: 10.1111/1755-0998.12495.

Peakall, R. & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, 28, 2537-2539.

Pecoraro, C., Babbucci, M., Villamor, A., Franch, R., Papetti, C., Leroy, B., ... & Murua, H. (2016). Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*). *Marine genomics*, *25*, 43-48.

Portnoy, D. S., Puritz, J. B., Hollenbeck, C. M., Gelsleichter, J., Chapman, D., & Gold, J. R. (2015). Selection and sex-biased dispersal in a coastal shark: the influence of philopatry on adaptive variation. *Molecular ecology*, *24*(23), 5877-5885.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-

project.org/.
Resh, C. A., Galaska, M. P., & Mahon, A. R. (2018). Genomic analyses of Northern snakehead (*Channa argus*) populations in North America. *PeerJ*, 6, e4581.
Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807.
Seetharam, A.S. and Stuart, G.W. (2013). Whole genome phylogeny for 21 Drosophila species using predicted 2b-RAD fragments. *PeerJ* *1*:e226 https://doi.org/10.7717/peerj.226
Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), 121.
Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature methods*, *9*(8), 808.
Warmuth, V. M., & Ellegren, H. (2019). Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADSeq data. *Molecular Ecology Resources*, 19, 586–596.
Whelan, N. V., Galaska, M. P., Sipley, B. N., Weber, J. M., Johnson, P. D., Halanych, K. M., & Helms, B. S. (2019). Riverscape genetic variation, migration patterns, and morphological variation of the threatened Round Rocksnail, *Leptoxis ampla*. *Molecular Ecology*, 28(7), 1593-1610.
Wickham, H. (2016). ggplot2: elegant graphics for data analysis. *Springer*.
Xu, P., Xu, S., Wu, X., Tao, Y., Wang, B., Wang, S., ... & Li, G. (2014). Population genomic analyses from low-coverage RAD-Seq data: a case study on the non-model cucurbit bottle gourd. *The Plant Journal*, 77(3), 430-442.

Data accessibility

Raw reads from all individuals, including information of location of all samples, will be stored in a SRA Bioproject upon acceptance. All customised scripts (.sh files) can be found in the Supplementary File customised_scripts.zip.

Author contributions

AB, HT, EM, CC and MP conceived and designed the study. AB and HT did the laboratory analysis with inputs from LB and RF. AB and HT conducted the data analysis. AB and HT wrote the manuscript with input from all authors.

Supplementary information

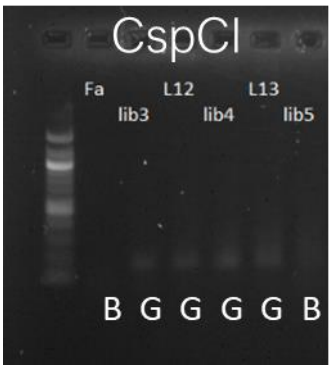


Figure S1. Example of library DNA quality identification. This agarose gel of *C. caretta* shows how library DNA was labelled as 'Good' (G) if the band was bright or 'Bad' (B) if the band was faint.

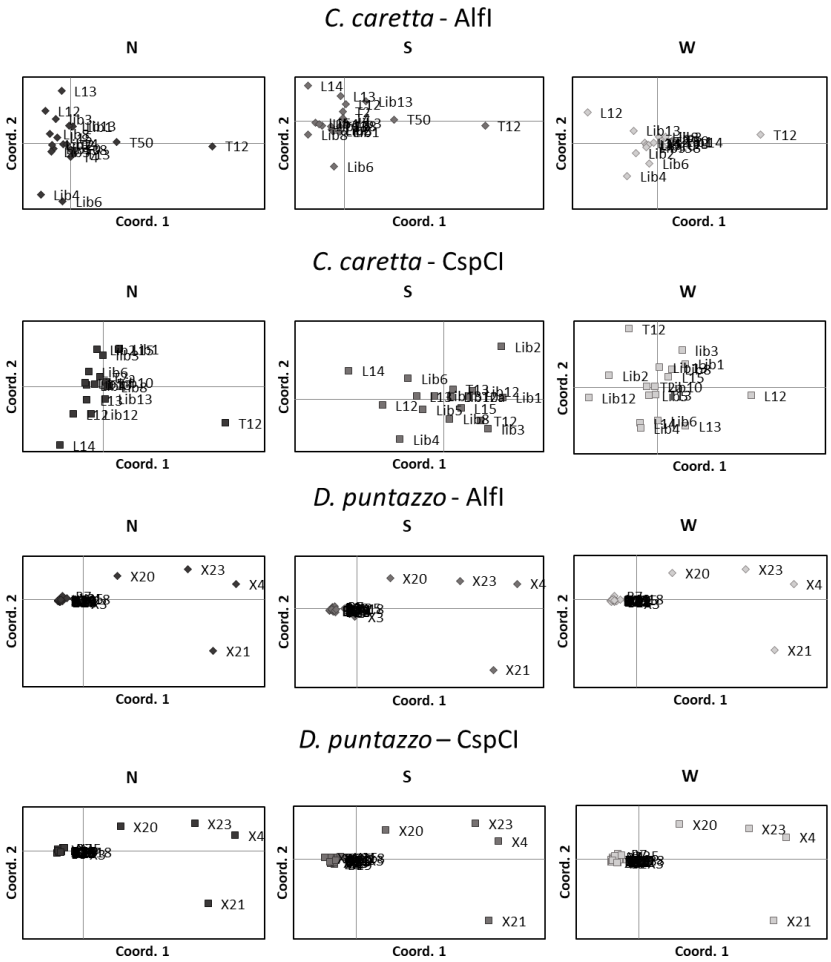


Figure S2. Principal Coordinates Analysis of different sets of loci. The PCoAs show the distribution of genetic distances of subsets and original sets of loci. In the case of *D. puntazzo* the distribution follows the same pattern among all enzymes and data sets. In *C. caretta* patterns do not match probably due to the low depth of coverage of the original sequences. Diamonds show data for AlfI and squares for CspCI enzyme. Individuals' points for the original data set (N) are in black, For CG selection (S) in dark Grey and for AT selection (W) in light grey.

Tables S1-S3 are available at :

https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13144

Table S4. Summary of selective-base ligation simulation for *Caretta caretta* and *Diplodus puntazzo* for the 2b-RAD enzymes (AlfI and CspcI). The symbols of the selective bases simulated for ligation of adapters correspond to the IUPAC notation: S=C+G and W=A+T, N=all bases. TR: total number of reads matching the base criteria. TMR: number of trimmed reads. IL: initial number of loci. FL: final number of loci. FMD: mean depth of coverage per locus after all the filtering process.

Species	Enzyme	Base	TR	TMR	IL	FL	FMD
Caretta caretta	Alf	N	182215898	149242071	66907	40319	11.5
		S	60699692	32301347	17065	9903	11.0
		W	69557740	40145869	15756	8943	10.4
	CspcI	N	158900658	100831134	25416	16332	19.3
		S	28296438	23381824	10442	4965	15.4
		W	29059846	24998483	10910	5149	15.3
Diplodus puntazzo	Alf	N	170651088	129508843	84382	66786	29.5
		S	26519654	25397694	17340	13209	27.2
		W	37242552	36219043	25703	19761	27.7
	CspcI	N	157140351	103657668	31111	24401	52.9
		S	22561494	19068008	11772	7372	29.2
		W	31887386	28060229	16177	10600	32.6

Table S5. Kruskal-Wallis test of heterozygosity values. We found no significant difference between heterozygosity values of selective-base ligation subsets and their original set of loci.

	Caretta caretta		*Diplodus puntazzo*	
	AlfI	CspCI	AlfI	CspCI
Kruskal-Wallis test	4.512	3.232	1.647	3.229
p-value	0.105	0.198	0.438	0.198

Annex I

Library building protocol

Genomic DNA of each sample is diluted with DNA-free water to obtain 7µl containing 180ng of DNA.

i)Digestion

Digestion of each sample is prepared using one of the following protocols depending on the enzyme used. a) 1µl Enzyme Buffer R (10x), 1µl SAM (S-adenosyl-methionine, 100µM), 1µl AlfI enzyme (2u/µl), 3µl of sterile water and 4µl of DNA, or b) 1µl Enzyme CutSmart (10x), 1µl SAM (100µM), 1µl CspCI enzyme (5u/µl), 3µl of sterile water and 4µl of DNA. All steps are performed with a GenAmp PCR System 2700 (Applied Biosystems®). Digesting conditions are 37.0 °C for 60 min, 65.0 °C for 20 min.

Enzyme	**Fragment length and recognition sequence (5'-3')**
AlfI	(N_{10-12}) **GCA** (N_6) **TGC** (N_{12-10})
CspCI	(N_{10-11}) **CAA** (N_5) **GTGG** (N_{12-13})

ii)Ligation

The two adaptors are prepared by hybridization of different oligos: adaptor 2 is obtained by hybridising the oligos 1 and 2 and adaptor 3 is obtained by hybridizing the oligos 1 and 3. Mix 22µl of oligo 1 (100µM) with 22µl of oligo 2 (or 3) (100µM) and 206µl sterile water. Annealing conditions are: 65.0 °C for 30 min, 83 cycles of 65.0 °C for 20 sec and 64.7 °C for 20 sec, with temperature decreasing by 0.3°C for each cycle, and a final hold of 15.0 °C for 10 min.

Oligo	**Sequence (5'-3')**
1	AGA TCG GAA GAGC
2	CTA CAC GAC GCT CTT CCG ATC TNN
3	CAG ACG TGT GCT CTT CCG ATC TNN

Ligation is prepared using the following protocol: 3µl sterile water, 1.5µl T4 Ligase Buffer (10x), 2.5µl adaptor 2 (4µM), 2.5µl adaptor 3 (4µM), 0.5µl ATP (10mM), 5µl T4 Ligase (200u/µl) and 10µl of digested DNA. Ligation conditions are: 16.0 °C for 180 min and 65.0 °C for 10 min.

iii)Amplification

Amplification is prepared using the following protocol: 25.15µl sterile water, 12µl Taq HF Buffer (5x), 0.75µl dNTPs (25mM), 1.2µl Amplification primer F (10µM), 1.2µl Amplification primer R (10µM), 3µl primer FOR (10µM), 1.2µl Taq phusion, 3µl Barcode primer (10µM) and 12.5µl of ligated DNA. Each 60 µl sample amplification is split in 3 microplate wells containing 20µl of amplification mix to optimise amplification outcome. Amplification PCR conditions are: 98.0 °C for 5 min, 14 cycles of 98.0 °C for 5 sec, 60.0 °C for 20 sec and 72.0 °C for 5 sec, and a final extension step of 72.0 °C for 5 min.

Primer	**Sequence (5'-3')**
Amplification primer F	AAT GAT ACG GCG ACC ACC GA
Amplification primer R	CAA GCA GAA GAC GGC ATA CGA
Primer FOR	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATCT
Barcode primer (N_7)	CAA GCA GAA GAC GGC ATA CGA GAT**(N_7)**GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC

iv)Purification

Purification is carried out mixing 60µl of amplified product with 90µl of magnetic beads (Beckman Coulter®) and placing the plate on a magnetic plate. After removing the supernatant, the product was washed with 180µl of 85% ethanol. The ethanol was removed and the beads with the attached DNA were left to dry. Finally the DNA was resuspended in 25µl of sterile water and 20µl of supernatant without beads were stored for sequencing.

v) Pool

Prepare the pool by pipetting 180ng of DNA for each purified library, based on each sample's library concentration.

Annex II

All customised scripts (.sh files) can be found in the Supplementary file customised_scripts.zip

AlfI trimming

Cut the adapter form the raw data using **cut_adapter_AlfI.sh.**

The input files are raw reads in fastq format. This script cuts the adapter from the sequences using the program cutadapt-1.5 (Martin, 2011) and finally converts the sequences in fasta format. The output files must be used as input files for the next step.

Trim raw sequences using **bRad_AlfI.sh**.

The input files are raw reads in fasta format. This script identifies the enzyme recognition site, cut all sequences down to a same length, flip sequences to have them all orientated in the same direction and check for duplicates. This script uses the following programs: TruncateFastq.pl, 2b_Extract.pl (http://people.oregonstate.edu/~meyere/tools.html), fnafile (NEWBLERtools), SHRiMP v2.2.3 (David, Dzamba, Lister, Ilie, & Brudno, 2011), revcompl.pl (https://github.com/KorfLab/Perl_utils) and CD-HIT (Fu, Niu, Zhu, Wu, & Li, 2012). The output file includes the sequences ready to be used in Stacks.

CspCI trimming

Cut the adapter form the raw data using **cut_adapter_CspCI.sh.**

The input files are raw reads in fastq format. This script cuts the adapter from the sequences and finally converts the sequences in fasta format. The output files must be used as input files for the next step.

Trim raw sequences using **recsite_CspCI_subset.sh**.

The input files are raw reads in fasta format. This script identifies the enzyme recognition site, cut all sequences down to a same length, flip sequences to have them all orientated in the same direction and check for duplicates. The output file includes the sequences ready to be used in Stacks.

Resampling simulation for accumulation curve

Perform a resampling simulation with replacement using **resampling_fasta.sh**.

The input files are raw reads in fasta format. This script resamples a given number of raw sequences per individual (N_{SEQ}), with replacement.

Selective-base Ligation

Remove the adaptor sequence from each read using either **cut_adaptor_AlfI.sh** or **cut_adaptor_CspCI.sh.** The input files are raw reads in fasta format. The output file is used as input for the simulation of selective-base ligation.

Perform a simulation of selective-base ligation using **select_bases_fasta_2.0.sh.**

This script extracts sequences selected by first and last base to simulate a selective-base ligation. The output has to be trimmed with one of the scripts mentioned above depending on the enzyme.

Each script is annotated with information on how to set the script parameters.

References

David, M., Dzamba, M., Lister, D., Ilie, L., & Brudno, M. (2011). SHRiMP2: sensitive yet practical short read mapping. *Bioinformatics*, 27(7), 1011-1012.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150-3152.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1), 10-12.

Capítulo 2.3

Born late, die late: Genomic basis for early-life selective mortality in *Diplodus puntazzo*.

Héctor Torrado[1, 2] *, Núria Raventós [1], Carlos Carreras [2], Enrique Macpherson [1+], Marta Pascual [2+]

[1]Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Car. Acc. Cala St. Francesc 14, 17300 Blanes, Girona, Spain.
[2]Department de Genètica, Microbiologia i Estadística and IRBio, Universitat de Barcelona, Av.Diagonal 643, 08028 Barcelona, Spain.
+Both authors contributed equally as senior researchers and should be considered to be at the same position

Abstract

In this study, we perform a study of mortality based on the larval early-life traits of Diplodus puntazzo in three localities, between settlers and six-months survivors. We found clear signals of selection for hatch date and lower for growth rate and pelagic larval duration (PLD). We confirmed these results with a phenotype-genotype association study with 2b-RAD SNP loci, finding a large number of loci related with the traits studied. Furthermore, we performed a blast with Sparus aurata in which some loci matched inside coding regions, some of them related with development. All this associations between loci and these early-life traits, suggest a genetic basis of the differences between settlers and survivors.

Keywords: 2b-RAD, populations genomics, selective mortality, early-life traits, hatch date, *Diplodus puntazzo*, phenotype-genotype association study.

Introduction

Numerous Mediterranean benthic fishes have a complex life cycle with a pelagic larval stage, which allows dispersal, whereas adults present high site fidelity (Aspillaga et al., 2016; Planes & Lenfant, 2002; Schunter, Carreras-Carbonell, MacPherson, et al.,

2011; Treml et al., 2012). Some larval characteristics, as the length of the pelagic larval duration (PLD), the larval growth, the size at hatch or the size at settlement can affect the mortality rate, modulating the year-class strength in numerous species (Garrido et al., 2015; Hamilton, Regetz, & Warner, 2008; Johnson, Christie, Stallings, Pusack, & Hixon, 2015). The effect of larval traits in survival does not stop with recruitment, but can also have remnant effects in the performance and survival of subsequent stages(Crean, Monro, & Marshall, 2011). For example, traits like larval growth rate, PLD or hatching size have been demonstrated to enhance post-metamorphic survival of marine species (Hamilton et al., 2008; Maldonado & Young, 1999; Raventos & Macpherson, 2005; Searcy & Sponaugle, 2001). Moreover, some studies have found changes in allele frequency related to growth-related selective mortality (Lenfant & Planes, 2002; Planes & Romans, 2004; Vigliola et al., 2007).

High-throughput sequencing technologies and the new genomic tools based on the massive sequencing of restriction enzyme-directed genomic subsets (RAD-seq) have revolutionized the field of population genetics. This is because the first facilitates the acquisition of genome-wide data at low cost and the second one has opened the possibility of working with non-model but ecologically relevant species, since it allows genotyping without a reference genome. Thus, this techniques provide a cost-effective approach to genomic studies of populations in non-model species, with a high density of markers that can identify non-neutral genomic signatures related to adaptation processes, with annotation potential (Carreras et al., 2020; Stapley et al., 2010). The implementation of these techniques has been broadly used in studies focus in a wide taxonomic array of non-model organisms (Babbucci et al., 2016; Barbanti et al., 2020; Benestan et al., 2016; Carreras et al., 2020; Lu et al., 2013; Torrado, Carreras, Raventos, Macpherson, & Pascual, 2020; Xuereb et al., 2018). Among the different RAD-seq techniques, 2b-RAD has been tested successfully in *D. puntazzo* (Barbanti et al., 2020). This technique produce tags of uniform length, using type IIB restriction enzymes, which cleave genomic DNA upstream and downstream of the target site (Wang, Meyer, Mckay, & Matz, 2012), and produce good results even for degraded DNA samples (Barbanti et al., 2020).

Furthermore, in most fishes individual-based information on early life history traits can be obtained from otoliths (bones in the inner part of the ear) (Raventos & Macpherson, 2001a). This data has been successfully combined with genomic data to infer fine-scale dispersal (Schunter et al., 2019) and for genome-wide phenotype and environmental associations (Torrado et al., 2020). Thus, the study of individually otolith-inferred larval traits, combined with genomic data can allow investigating selective mortality and identify the genetic basis for traits showing trends in associated mortality across multiple locations.

In this work, we evaluate the effect of selective mortality at early-life phases and identify associated candidate genes. A total of 61 settlers (individuals with few days after settling from pelagic phase) and 44 survivors (individuals with 6 months after settlement) were sampled in three different localities in the Western Mediterranean. Specifically, we 1.- Study the potential selective mortality occurring in early settlement stage of *D. puntazzo*. 2.-Infer the potential genetic bases of this mortality applying 2b-RAD. Our study provides the first clue on genomic basis for selective mortality in our studied candidate species.

Methods

Study species

Diplodus puntazzo (Walbaum, 1792) is a shallow-water (0-30 m) coastal species. It has a larval duration of 16-29 days and is a representative components of the rocky and sea-grass beds communities of the Western Mediterranean (García Rubies & Zabala, 1990; Macpherson, Gordoa, & García-Rubies, 2002). They have a narrow bathymetric range (<100 m), are found ubiquitously across the studied area and produce a clear settlement band in their otoliths (Raventos & Macpherson, 2001a). Their reproductive season takes place from August to October (Raventós, Torrado, Arthur, Alcoverro, & Macpherson, n.d.).

Sampling

A total of 61settlers (individuals with few days after settling from pelagic phase) and 44 survivors (individuals with 6 months after settlement) of *D. puntazzo* were sampled from Blanes, Xabia and Aguamarga in 2016 using hand nets and stored in 96% Ethanol (Figure 1). The collection of fish samples met the Spanish and European regulations. According to article 3.1 of the European Union directive (2010/63/UE) from the 22/9/2010, no approval is needed for fish sacrification with the purpose of tissue or organ analyses. Furthermore, the study species is not listed in CITES.

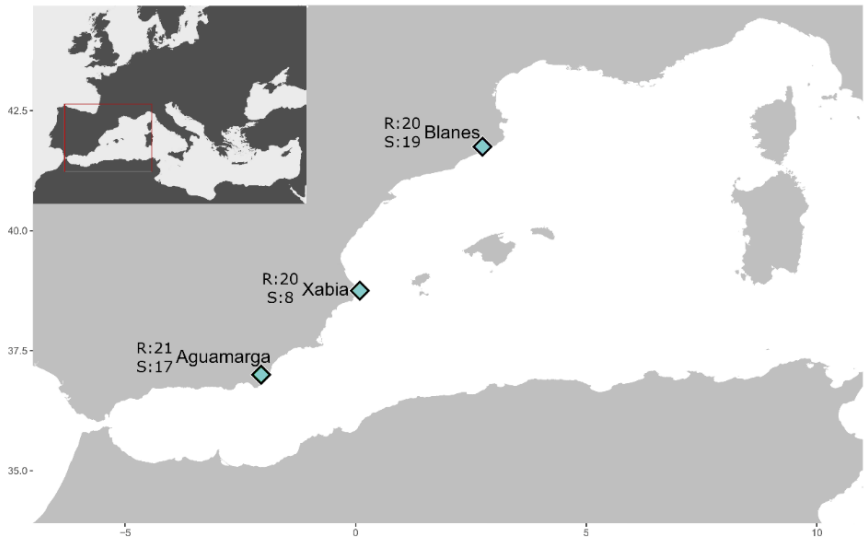


Figure 1 Map of the sampling area and their localization in Europe. Diamonds indicate sampling localities. R = number of settlers and S = number of survivors.

At the laboratory, the otoliths (lapilli) of all individuals were extracted and mounted in an oil droplet on a microscope slide. Five early-life traits were considered for each individual: hatching date (expressed as the number of day of the year), pelagic larval duration (PLD, in days), growth rate during the pelagic larval duration (increment in otolith size in µm divided by PLD), otolith hatch size and otolith settlement size (in µm). These traits were determined using a light microscope following the standard methodology described in the literature (Raventos &

Macpherson, 2001b, 2005). The age of the individual was estimate counting the number of otolith marks along the longest radius, and this information was used to calculate the hatch date. PLD was calculated with the number of otolith marks before the settlement mark. We measured the otolith sizes six times and used their mean to minimize measurements errors.

In order to evaluate the possible differences of early-life traits between settlers and survivors, five Permutational Multivariate Analyses of Variance (PERMANOVA) were performed, one for each phenotypic variable, as implemented in the R function 'adonis', from the 'vegan' package v 2.5-2 (Oksanen et al., 2018). Those analyses were carried out for all samples combined and for each locality independently.

DNA extraction and genotyping

DNA was individually extracted using the QIAamp DNA Mini Kit (QIAGEN) extraction kit, following manufacturer's instructions. DNA integrity was checked by gel electrophoresis and quantified by NanoDrop or Qubit.

To build individual genomic libraries with 2b-RAD we followed the protocol described in Barbanti et al., 2020. Some of the settlers of Blanes (N=12) and Xabia (N=12) were sequenced previously (Barbanti et al., 2020) and data used in the present study. Additionally, libraries of 37 settlers and 44 survivors were constructed. In brief, the Genomic DNA of each sample was individually digested by AlfI restriction enzyme. During ligation, adaptors with degenerated bases (5′-NN-3′) were attached to the sticky ends of the digested sequences. During amplification, barcodes and Illumina primers were attached to the adaptors and sequences were amplified by a PCR. Library products were run through a 1.8% agarose gel to check amplification success. Purification was performed using magnetic beads. DNA concentration of purified libraries was quantified using PicoGreen® and samples pooled (one age group per lane) and single end sequenced (50bp) with a HiSeq 2500 Illumina at the Center for Genomic Regulations (CRG) of Barcelona.

Raw sequences of all individuals used in the analysis were processed simultaneously previous to genotyping using the scripts in Barbanti et al., 2020. Those files were used for genotyping with STACKS vs 2.53 software (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). To construct a catalogue loci two mismatches were allowed between stacks within (M=2) and between (n=2) individuals, and a minimal stack depth of three was required (m=3). Individual genotypes were outputted both as haplotype loci and SNP loci VCF files. Additional filters were applied using VCFtools vs 1.12 (Danecek et al., 2011). Individual genotypes with a depth below 5X were not considered. Loci with a missingness value higher than 30% were removed from the loci datafiles. Additionally, loci with the major allele frequency equally or higher than 0.95 (*i.e.* monomorphic at that level) were identified by the R function 'isPoly' from the package 'adegenet' (Jombart, 2008) and removed from the vcf file by VCFtools. For SNP loci, those with a minimum allele frequency (MAF) of 0.05 were also removed from the SNP dataset by VCFtools. Finally, the loci in HW disequilibrium at two of the three localities were eliminated of the analysis from both the haplotype and SNP loci datafiles to remove possible paralogous loci. Genotyping and filtering processes were performed twice, once only with settlers' data for the population genomic analysis and a second one with both settlers and survivors' data to identify genetic changes associated with selective mortality.

Population genomic analyses

We used the settlers' haplotype loci datafiles for the population genomic analyses, as they are the age group commonly use in this kind of studies (e.g. Torrado et al., 2020). F_{ST} pairwise values and their significance (computed by 999 permutations) were obtained by R software (R Core Team, 2020) 'hierfstat' package vs 0.04-2 (Goudet & Jombart, 2015). Also, a Discriminant Analysis of Principal Components (DAPC) was performed retaining a number of PCAs equal to one third of the number of individuals. The analyses were performed with the R software package 'adegenet' (Jombart, 2008) and represented with the 'ggplot2' package (Wickham, 2009).

Selective mortality

In order to look for genomic basis of this selective mortality, we performed for each locality a Redundancy Analysis (RDA) with the SNP loci datafiles following the method described by Forester and co-authors (Forester, Lasky, Wagner, & Urban, 2018), identifying as outliers those loci with more than 3 sd from mean loading (equivalent to a two-tailed p-value = 0.0027). We used as response variables a genotypic matrix with the SNPs of all loci and the five phenotypic variables as predictors. In order to visualize the number of SNPs associated to each predictor in each locality, we plotted a set of Venn diagrams using R package "VennDiagram" (Chen, 2018).

BLAST searches were conducted with the 34 bp sequences of all outlier loci against the genome of *Sparus aurata* Linnaeus, 1758, available in Ensembl website (www.ensembl.org). Sensitivity was set to "Distant homologies" to maximise matches' length. A maximum E-value of 10-3 was allowed and only matches of at least half of the locus sequence were considered. When sequences matched within a gene, the function of that gene was searched at the UniProt database (www.uniprot.org). The GO-terms list of annotated genes were introduced in REVIGO to produce treemaps of the biological functions (Supek, Bošnjak, Škunca, & Šmuc, 2011).

Results

Population genomics

Mean number of raw reads per individual was 7.7E6 ± 1.5E6, after processing we kept 6.2E6 ± 1.5E6 reads per individual (Table 1). After filtering, we obtained 58906 haplotype loci and 85031 SNP loci, with good mean depth per locus (29.67±13.54) (Table 1).

Table 1: Mean number of raw reads per individual, Number of haplotype loci, mean depth per SNP, observed (Ho) and expected (Hs) heterozigosities for haplotype loci.

Mean number of raw reads	7.7E6 ± 1.5E6
Mean number of usable reads	6.2E6 ± 1.5E6
Polymorphic Haplotype loci	58906
Polymorphic SNPs	85031
Mean depth locus	29.67±13.54

In the DAPC analysis (Figure 2), we observed three groups but with a few individuals genetically distributed between the individuals of the other localities. In the F_{ST} analysis we can see a significant differentiation between Blanes (northernmost) and Xabia and Aguamarga localities with and no difference between Xabia (central) and Aguamarga (southernmost) (Table 2).

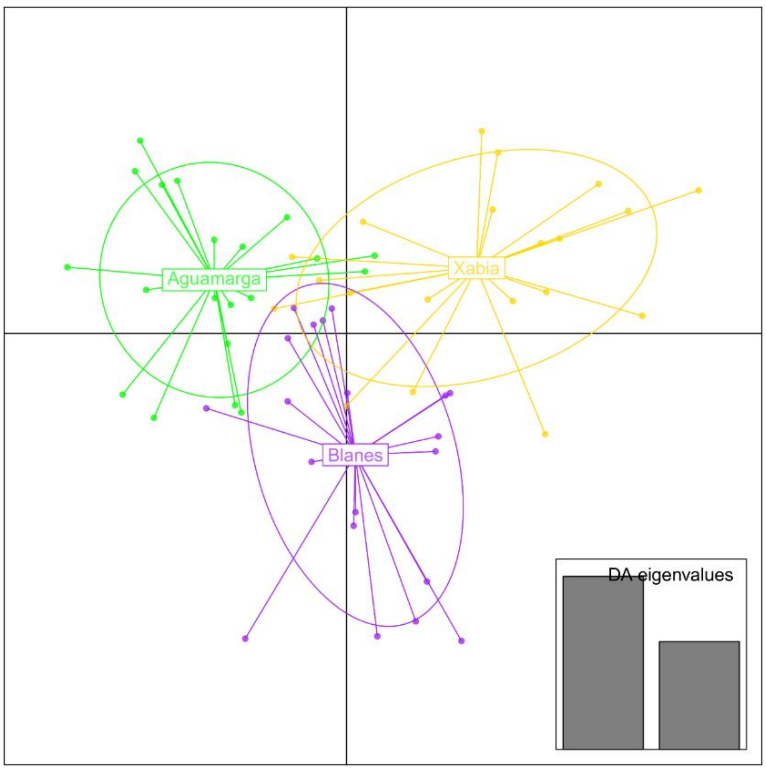


Figure 2: Discriminant Analysis of Principal Components (DAPC) plots for haplotype loci with DA eigenvalues (top right corner).

Table 2: Pairwise F_{ST} analysis using haplotype loci. Bold p-values mean significance at α=0.05.

	Blanes	Xabia	Aguamarga
Blanes			
Xabia	**0.008**		
Aguamarga	**0.012**	-0.002	

Selective mortality

PERMANOVA analysis revealed significant differences between both age groups for PLD, growth rate and hatch date (Table 3). PLD was higher in survivors than settlers, as well as the hatching date, but larval growth rate was lower (i.e. survivors born later, grow slower and live longer time as larvae) (Figure 3). R^2 values were higher in GRPLD and hatch date than in PLD (Table 3).

Furthermore, we found significant differences among localities for all traits but settlement size, and a significant effect of the interaction of both factors in hatch date and growth rate (Table 3). Due to the significant effect of Locality, we analysed differences among settlers and survivors in each locality separately (Table 4). We found significant differences among age groups in all localities for hatch date, with survivors born later in all cases (Figure 4). For growth rate, survivors significantly grow slower in all but one locality (Xabia) (Figure 4). For PLD, significantly higher values in survivors were found only in Aguamarga (Table 4).

Table 3: PERMANOVA pseudo-F and R^2 for the effect of Development stage and Locality in the studied early life traits. Bold pseudo-F values indicate significant effects at α=0.05. Full model details are provided in Table S1.

		Development	Locality	Development: Locality
PLD	F.Model	**8.21**	**18.81**	2.69
	R2	0.06	0.26	0.04
GRPLD	F.Model	**59.69**	**26.93**	**5.56**
	R2	0.27	0.25	0.05
Settlement size	F.Model	0.50	0.52	2.51
	R2	0.00	0.01	0.05
hatch size	F.Model	2.11	**3.05**	3.07
	R2	0.02	0.06	0.06
hatch date	F.Model	**72.78**	**34.83**	**7.23**
	R2	0.29	0.28	0.06

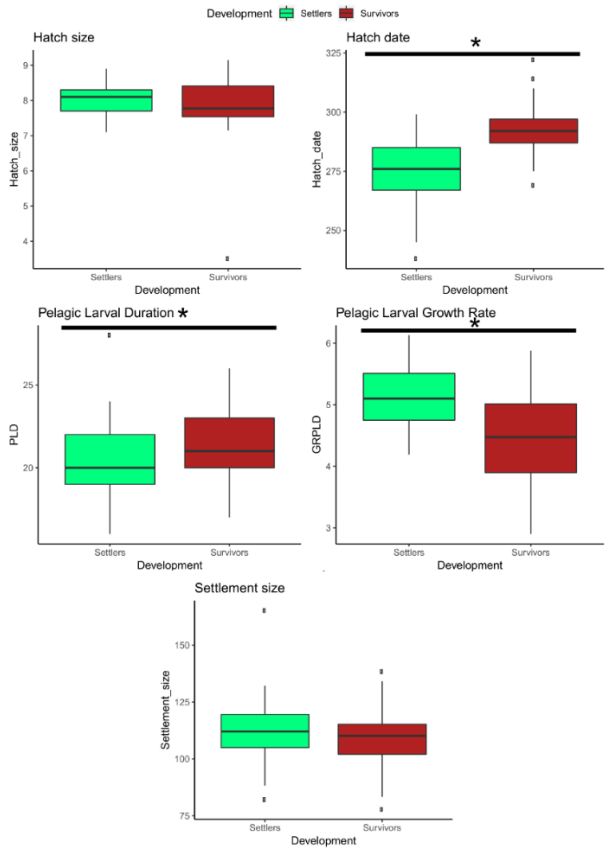


Figure 3: Boxplot of the early-life traits studied grouped by developmental degree. Figures marked with a bar and asterisk have significative differentiation with PERMANOVA.

In the Redundancy Analysis (RDA) the two axes of the RDA clearly separated the individuals by age group in all the localities. Nevertheless, the percentage of explanation varied, with a 49.5% in Blanes, 61.49% in Xabia and a 52.91% in Aguamarga (Figure 5). Overall, 9134 SNPs corresponding to 7270 haplotypes were significantly correlated with phenotypic predictors, some of them with more than one predictor or in more than one locality (Figure 6). Four SNPs were associated to the same predictor in all three localities: 100380:24_C with GRPLD, 102878:11_A and 102878:31_C with hatch date and 8948:2_A with hatch size. Nonetheless, a fair number of loci was associated to the same predictor in two localities (Figure 6).

Table 4: PERMANOVA pseudo-F and R^2 for the influence of Development stage in PLD, GRPLD and hatch date for the locality-specific analysis. Bold pseudo-F values indicate significant effects at α=0.05. Full model details are provided in Table S2.

		Blanes	Xabia	Aguamarga
PLD	F.Model	3.06	1.01	**9.74**
	R2	0.08	0.04	0.21
GRPLD	F.Model	**42.17**	0.10	**12.67**
	R2	0.53	0.00	0.26
Settlement Size	F.Model	3.56	0.19	0.79
	R2	0.09	0.01	0.02
hatch Size	F.Model	0.07	3.58	1.52
	R2	0.00	0.13	0.04
hatch Date	F.Model	**74.34**	**14.16**	**14.58**
	R2	0.67	0.37	0.30

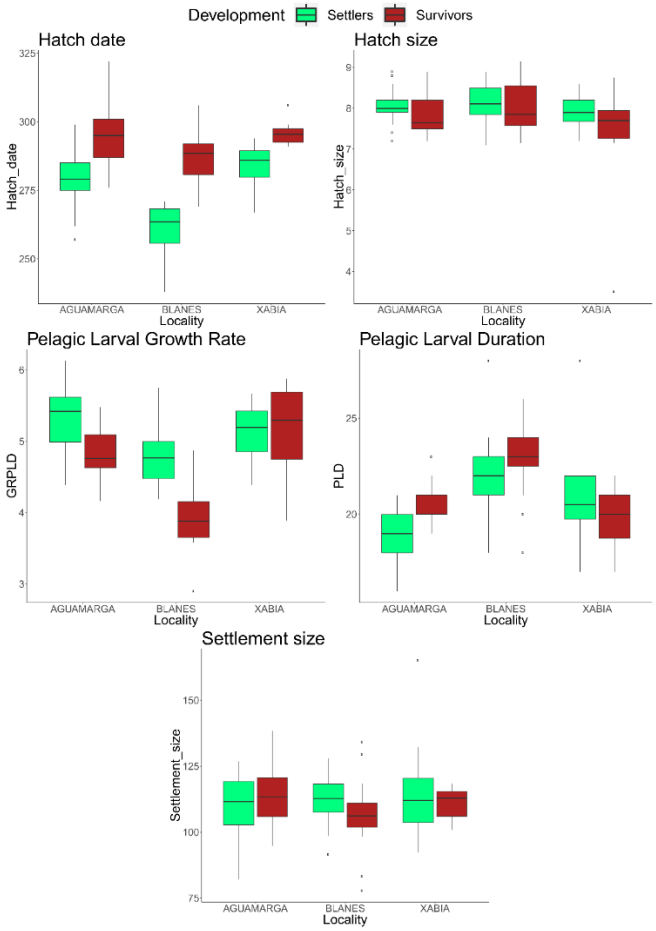


Figure 4: Boxplot of the early-life traits studied grouped by developmental degree separated by sampling locality. Figures marked with a bar and asterisk have significative differentiation with PERMANOVA.

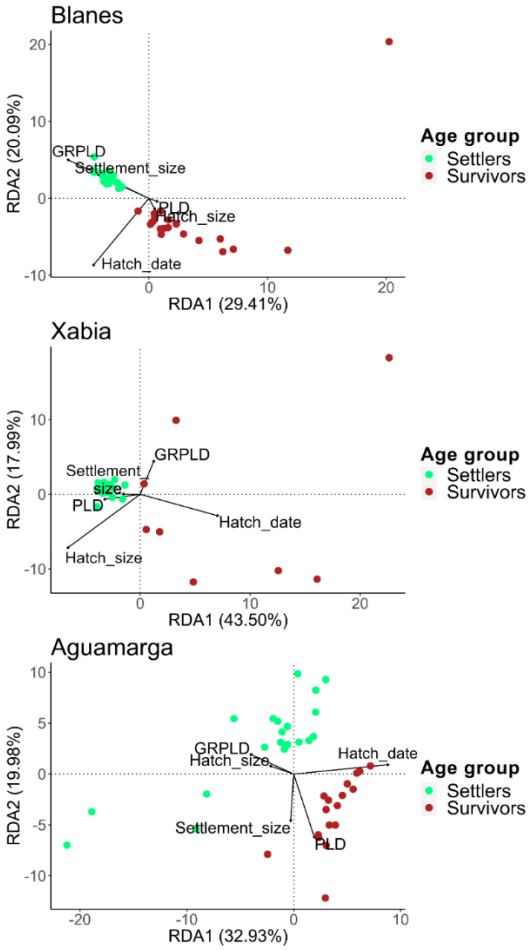


Figure 5: RDA performed with individual phenotypic data and SNPs. Coloured points represent individual fish coded by age group. Vectors indicate phenotypic predictors of the two first RDA components.

For all the loci associated to the same predictor in at least two localities (N=373), a BLAST search with the genome of *Sparus aurata* was carried out. We found 260 matches, 153 of them inside a coding gene (Table S3). 81.3% of the functions were associated to only one predictor (Table S4). Hatching date had 87 functions associated and growth rate had 50, while PLD had 46, hatch size 33 and settlement size had 9 (Table S4). Those associated with hatch date had functions mainly related to functions as the transition of mitotic cell cycle, olfactory nerve formation or cardiac muscle action potential (Figure 7A, Table S4). GRPLD-related had some loci related directly with development as otic placode formation and modification of dendritic spine (Figure 7B, Table S4). Hatch size had functions related to wound healing, hemidesmosome

assembly and left/right axis specification (Figure S1A, Table S4). PLD-related loci had functions mainly related with endoplasmatic reticulum organization, endocytosis or chemotaxis (Figure S1B, Table S4). Settlement size-related loci had functions associated to peptide transport original transduction (Figure S1C, Table S4).

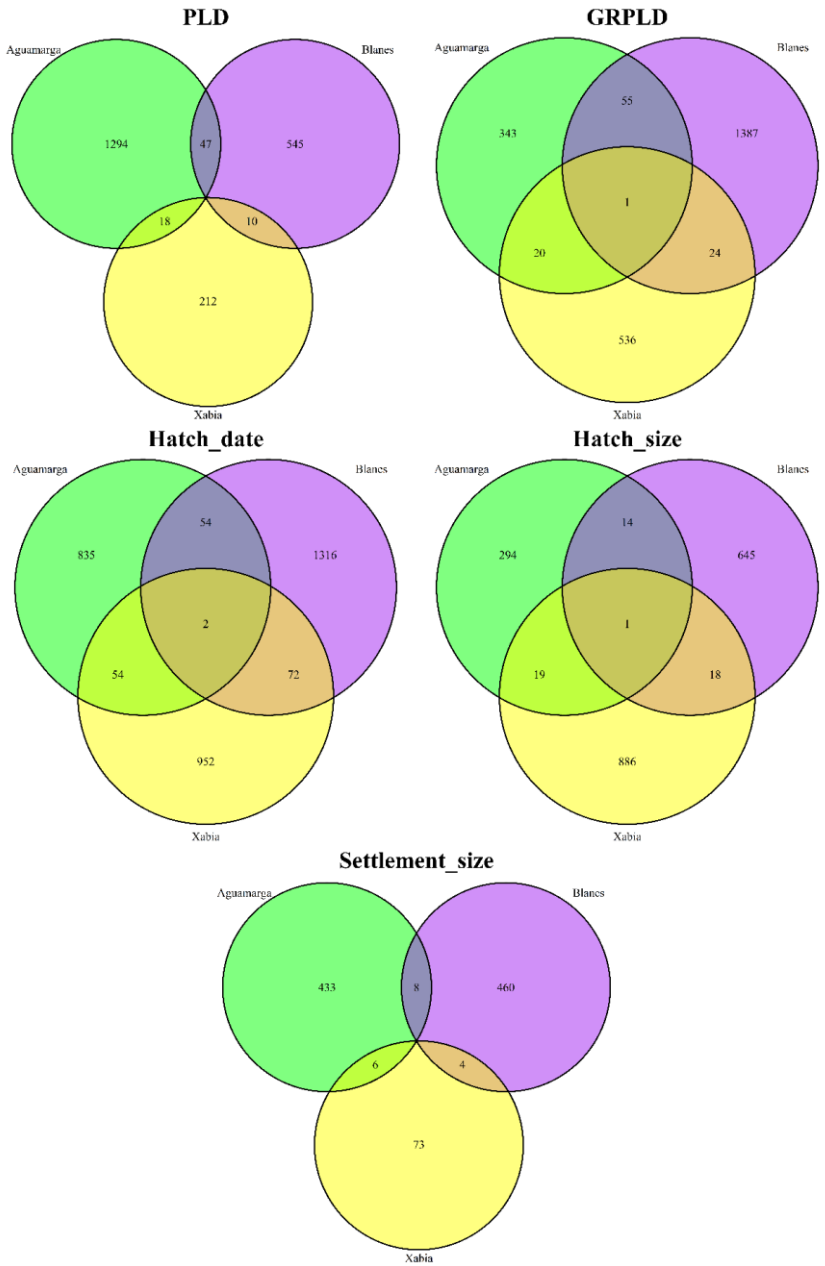


Figure 6: Venn diagrams for loci associated by the RDA with the early-life traits. Figures marked with an asterisk have significative differentiation with PERMANOVA.

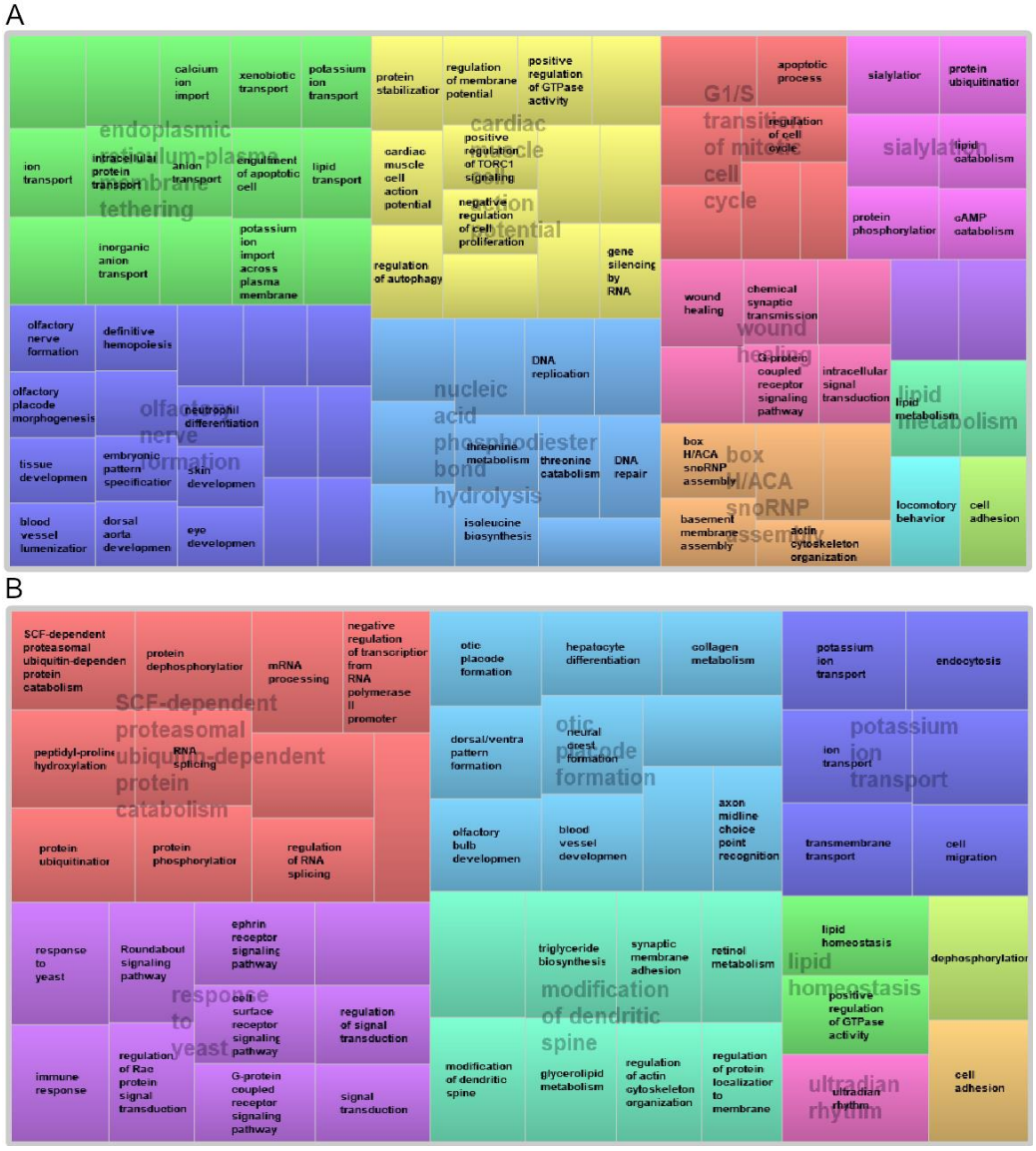


Figure 7. Gene ontology treemap of the genes associated with hatch date (A) and GRPLD (B). The box size correlates to the uniqueness of the GO-term. Boxes with the same colour can be grouped together as correspond to the same upper-hierarchy GO-term which is found in the middle of each box.

Discussion

In this study we compare settlers and survivors of *Diplodus puntazzo* at three different localities and we observe differences in mortality mediated by early life characteristics of the individuals. In *D. puntazzo* survivors hatched significantly later than settlers in all localities and grew more slowly in the two northern localities. We detected several loci associated to this selective mortality process in a replicated manner in multiple localities.

The genomic structure marks a separation between Blanes and Xabia and Aguamarga. This structuring seems logic as it divides the localities by their position respect to the Ibiza Channel, an oceanographic front located in the area. Blanes is situated northern to it and the other two localities southern. The effect of this front in genomic structuring have been seen in some species as *Serranus cabrilla* (Schunter, Carreras-Carbonell, Macpherson, et al., 2011) or *Symphodus ocellatus* (Torrado et al., 2020). Furthermore, oceanographic larval dispersion models for *D. puntazzo* have shown, despite some pass of larvae through this front, a little communication between these localities and have been clustered it separately (Torrado, Mourre, et al., (Submitted); Torrado, Treml, (Submitted)).

We found clear signals of selective mortality in early post-settlement phases of *D. puntazzo* mostly affecting two larval characteristics; larval growth rate was lower in survivors, whereas hatch date was higher. These changes through development are supported by the phenotype-genotype associations found by the RDA associated to early life traits and especially remarkable in the hatching date and partially to larval growth rate. Thus, our results suggest that the differences between settlers and survivors have a genetic basis and, therefore, the current selection in phenotypes can potentially affect the genotypes and phenotypes of next generations. Furthermore, genetic differences among age groups related with growth-related pressures had been observed using different markers. For example, allelic selection in allozymes related to growth was found in *Diplodus sargus* after 4-months of settlers' monitoring (Planes & Romans, 2004). Significant genetic structure was found between settlers and three

months-old survivors as a result of changes in frequency of mtDNA haplotypes in *Neopomacentrus filamentosu*s, consistent in four cohorts (Vigliola et al., 2007). Experiments in *Dascyllus aruanus* indicated an influence of size-selective predatory exposure in the genetic composition with 10 microsatellite loci (Ciotti & Planes, 2019).

The unique trait maintaining this selective effect in all localities is the hatching date. Hatching date can be a source of selection for some reasons, both in larval and in post-settlement phases. This mortality could cause a shift in the spawning dates along the years, as seen in *Theragra chalcogramma* (Yoklavich & Bailey, 1990). Nevertheless, this pattern is not constant among species, and some studies have shown an opposite pattern in *Lepomis macrochirus* and *Forsterygion lapillum* (Cargnelli & Gross, 1996; Moginie & Shima, 2018). This suggest a non-directional time-depending pressure, that could be due to different factors among species or populations of the same species. Also, several loci are associated to hatching date in other fishes, as *Symphodus tinca*. (Torrado et al., 2020). In *D. puntazzo*, the genes matching with the hatching date loci have not a general function clearly related with that trait, and additional data would be necessary, including parental information, to explain why the survivors hatch and settle later than others.

Growth rate is a trait highly study in selective mortality research, and has well known effects in pre- and post-settlement phases. In our case, survival's growth rate was lower, which coincides with other studies that have found evidence of selection acting against faster growing and/or larger individuals (Biro, Abrahams, Post, & Parkinson, 2004; Litvak & Leggett, 1992; Sponaugle, Boulay, & Rankin, 2011). Nevertheless, there are other studies showing a positive selection of fast-growing individuals (e.g. D'Alessandro et al., 2013; Meekan et al., 2006; Takasuka et al., 2003). Moreover, in *Diplodus sargus* growth rate-related loci have some matches with loci involved in development (Planes & Romans, 2004). In our study, we have also found a set of genes related with development, some involved in nervous system development, which could give an idea of what kind of selective pressures could act in this species.

The relationship between PLD and selective mortality is not clear. In our study, only the southernmost locality (Aguamarga) showed significant differences between

settlers and survivors. This variability could be related with differences in the adaptive pressures suffered by larvae in each locality. The mortality could be related to a less selectively environment, as it was observed in nearshore versus offshore larvae (Hamilton et al., 2008). Nonetheless, other studies defend the hypothesis that expending less time as larvae is advantageous ('stage duration' hypothesis; Houde, 1987), as this phase is the most vulnerable of the whole cycle. Numerous studies on selective mortality indicate that this process can be affected by the environmental variability and it is not constant along time (Grorud-Colvert & Sponaugle, 2011; Rankin & Sponaugle, 2011; Searcy & Sponaugle, 2001).

In conclusion, we find clear clues of a selective mortality in settlers of *Diplodus puntazzo* associated with the hatch date of the individuals in all localities. There are a large number of loci associated to these early-life traits, suggesting a genetic basis of the differences between settlers and survivors. In any case, our results show a large number of loci associated to these early-life traits, suggesting a genetic basis of the differences between settlers and survivors.

Acknowledgments

This work was supported by the Spanish Government project 'PopCOmics' (CTM2017-88080) (MCIU, AEI/FEDER, UE). The authors CC and MP are members of the research group SGR2017-1120 and EM of SGR2017-378 (Catalan Government). HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities (FPU15/02390).

References

Aspillaga, E., Bartumeus, F., Linares, C., Starr, R. M., López-Sanz, À., Díaz, D., … Hereu, B. (2016). Ordinary and extraordinary movement behaviour of small resident fish within a mediterranean marine protected area. *PLoS ONE*, *11*(7), 1–19. doi: 10.1371/journal.pone.0159813

Babbucci, M., Ferraresso, S., Pauletto, M., Franch, R., Papetti, C., Patarnello, T., … Bargelloni, L. (2016). An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*). *Scientific Reports*, *6*(August), 1–11. doi: 10.1038/srep38673

Barbanti, A., Torrado, H., Macpherson, E., Bargelloni, L., Franch, R., Carreras, C., &

Pascual, M. (2020). Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. *Molecular Ecology Resources*, (October 2019), 1–12. doi: 10.1111/1755-0998.13144

Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., … Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, *25*(20), 5073–5092. doi: 10.1111/mec.13811

Biro, P. A., Abrahams, M. V., Post, J. R., & Parkinson, E. A. (2004). Predators select against high growth rates and risk-taking behaviour in domestic trout populations. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1554), 2233–2237. doi: 10.1098/rspb.2004.2861

Cargnelli, L. M., & Gross, M. R. (1996). The temporal dimension in fish recruitment: birth date, body size, and size-dependent survival in a sunfish (bluegill: *Lepomis macrochirus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *53*(2), 360–367.

Carreras, C., García-Cisneros, A., Wangensteen, O. S., Ordóñez, V., Palacín, C., Pascual, M., & Turon, X. (2020). East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). *Diversity and Distributions*, *26*(3), 382–398. doi: 10.1111/ddi.13016

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. doi: 10.1111/mec.12354

Chen, H. (2018). *VennDiagram: Generate High-Resolution Venn and Euler Plots*. Retrieved from https://cran.r-project.org/package=VennDiagram

Ciotti, B. J., & Planes, S. (2019). Within-generation consequences of postsettlement mortality for trait composition in wild populations: An experimental test. *Ecology and Evolution*, *9*(5), 2550–2561. doi: 10.1002/ece3.4911

Crean, A. J., Monro, K., & Marshall, D. J. (2011). Fitness consequences of larval traits persist across the metamorphic boundary. *Evolution*, *65*(11), 3079–3089. doi: 10.1111/j.1558-5646.2011.01372.x

D'Alessandro, E. K., Sponaugle, S., & Cowen, R. K. (2013). Selective mortality during the larval and juvenile stages of snappers (Lutjanidae) and great barracuda Sphyraena barracuda. *Marine Ecology Progress Series*, *474*(McCormick 1998), 227–242. doi: 10.3354/meps10114

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, *27*(9), 2215–2233. doi: 10.1111/mec.14584

García Rubies, A., & Zabala, M. (1990). Effects of total fishing prohibition on the rocky fish assemblages of Medes Islands marine reserve (NW Mediterranean). *Scientia Marina*, Vol. 54, pp. 317–328.

Garrido, S., Ben-Hamadou, R., Santos, A. M. P., Ferreira, S., Teodósio, M. A., Cotano, U., … Ré, P. (2015). Born small, die young: Intrinsic, size-selective mortality in

marine larval fish. *Scientific Reports*, *5*, 17065. Retrieved from https://doi.org/10.1038/srep17065

Goudet, J., & Jombart, T. (2015). *hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04-22*. Retrieved from https://cran.r-project.org/package=hierfstat

Grorud-Colvert, K., & Sponaugle, S. (2011). Variability in water temperature affects trait-mediated survival of a newly settled coral reef fish. *Oecologia*, *165*(3), 675–686. doi: 10.1007/s00442-010-1748-4

Hamilton, S. L., Regetz, J., & Warner, R. R. (2008). Postsettlement survival linked to larval life in a marine fish. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(5), 1561–1566. doi: 10.1073/pnas.0707676105

Houde, E. D. (1987). Fish Early Life Dynamics and Recruitment Variability. *American Fisheries Society Symposium*, 2: 17-29.

Johnson, D. W., Christie, M. R., Stallings, C. D., Pusack, T. J., & Hixon, M. A. (2015). Using post-settlement demography to estimate larval survivorship: a coral reef fish example. *Oecologia*, *179*(3), 729–739. doi: 10.1007/s00442-015-3368-5

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. doi: 10.1093/bioinformatics/btn129

Lenfant, P., & Planes, S. (2002). Temporal genetic changes between cohorts in a natural population of a marine fish, *Diplodus sargus*. *Biological Journal of the Linnean Society*, *76*(1), 9–20. doi: 10.1046/j.1095-8312.2002.00041.x

Litvak, M. K., & Leggett, W. C. (1992). Age and size-selective predation on larval fishes: the Bigger-Is-Better hypothesis revisited. *Marine Ecology Progress Series*, *81*, 13–24.

Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., … Costich, D. E. (2013). Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genetics*, *9*(1), e1003215. doi: 10.1371/journal.pgen.1003215

Macpherson, E., Gordoa, A., & García-Rubies, A. (2002). Biomass Size Spectra in Littoral Fishes in Protected and Unprotected Areas in the NW Mediterranean. *Estuarine, Coastal and Shelf Science*, *55*(5), 777–788. doi: 10.1006

Maldonado, M., & Young, C. M. (1999). Effects of the duration of larval life on postlarval stages of the demosponge *Sigmadocia caerulea*. *Journal of Experimental Marine Biology and Ecology*, *232*(1), 9–21. doi: 10.1016/S0022-0981(98)00076-8

Meekan, M. G., Vigliola, L., Hansen, A., Doherty, P. J., Halford, A., & Carleton, J. H. (2006). Bigger is better: Size-selective mortality throughout the life history of a fast-growing clupeid, Spratelloides gracilis. *Marine Ecology Progress Series*, *317*(Anderson 1988), 237–244. doi: 10.3354/meps317237

Moginie, B. F., & Shima, J. S. (2018). Hatch date and growth rate drives reproductive success in nest-guarding males of a temperate reef fish. *Marine Ecology Progress Series*, *592*, 197–206. doi: 10.3354/meps12506

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … others. (2018). *vegan: Community Ecology Package. R package version 2.5-2*. Retrieved from https://cran.r-project.org/package=vegan

Planes, S., & Lenfant, P. (2002). Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. *Molecular Ecology*, *11*(8), 1515–1524.

Planes, S., & Romans, P. (2004). Evidence of genetic selection for growth in new recruits of a marine fish. *Molecular Ecology*, *13*(7), 2049–2060. doi: 10.1111/j.1365-294X.2004.02202.x

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/

Rankin, T. L., & Sponaugle, S. (2011). Temperature influences selective mortality during the early life stages of a coral reef fish. *PLoS ONE*, *6*(5). doi: 10.1371/journal.pone.0016814

Raventos, N., & Macpherson, E. (2001a). Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Marine Biology*, *138*(6), 1115–1120. doi: 10.1007/s002270000535

Raventos, N., & Macpherson, E. (2001b). Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Marine Biology*, *138*(6), 1115–1120. doi: 10.1007/s002270000535

Raventos, N., & Macpherson, E. (2005). Effect of pelagic larval growth and size-at-hatching on post-settlement survivorship in two temperate labrid fish of the genus *Symphodus*. *Marine Ecology Progress Series*, *285*, 205–211. doi: 10.3354/meps285205

Raventós, N., Torrado, H., Arthur, R., Alcoverro, T., & Macpherson, E. (n.d.). *Temperature reduces fish dispersal as larvae grow faster to their settlement size*.

Schunter, C., Carreras-Carbonell, J., Macpherson, E., Tintoré, J., Vidal-Vijande, E., Pascual, A., … Pascual, M. (2011). Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, *20*(24), 5167–5181. doi: 10.1111/j.1365-294X.2011.05355.x

Schunter, C., Carreras-Carbonell, J., MacPherson, E., TintorÉ, J., Vidal-Vijande, E., Pascual, A., … Pascual, M. (2011). Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, *20*(24), 5167–5181. doi: 10.1111/j.1365-294X.2011.05355.x

Schunter, C., Pascual, M., Raventos, N., Garriga, J., Garza, J. C., Bartumeus, F., & Macpherson, E. (2019). A novel integrative approach elucidates fine-scale dispersal patchiness in marine populations. *Scientific Reports*, *9*(1), 10796. doi: 10.1038/s41598-019-47200-w

Searcy, S. P., & Sponaugle, S. (2001). Selective mortality during the larval-juvenile transition in two coral reef fishes. *Ecology*, *82*(9), 2452–2470. doi: 10.1890/0012-9658(2001)082[2452:SMDTLJ]2.0.CO;2

Sponaugle, S., Boulay, J., & Rankin, T. (2011). Growth- and size-selective mortality in pelagic -larvae of a common reef fish. *Aquatic Biology*, *13*(3), 263–273. doi: 10.3354/ab00370

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., … Slate, J. (2010, December). Adaptation genomics: The next generation. *Trends in Ecology and Evolution*, Vol. 25, pp. 705–712. doi: 10.1016/j.tree.2010.09.002

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, *6*(7), e21800. doi:

10.1371/journal.pone.0021800

Takasuka, A., Aoki, I., & Mitani, I. (2003). Evidence of growth-selective predation on larval Japanese anchovy Engraulis japonicus in Sagami Bay. *Marine Ecology Progress Series*, *252*, 223–238. doi: 10.3354/meps252223

Torrado, H., Carreras, C., Raventos, N., Macpherson, E., & Pascual, M. (2020). Individual-based population genomics reveal different drivers of adaptation in sympatric fish. *Scientific Reports*, Vol. 10. doi: 10.1038/s41598-020-69160-2

Torrado, H., Mourre, B., Raventós, N., Carreras, C., Tintoré, J., Pascual, M., & Macpherson, E. (n.d.). *Impact of individual early life traits in larval dispersal: a multispecies approach using backtracking models*.

Torrado, H., Treml, E., Carreras, C., Macpherson, E., & Pascual, M. (n.d.). *Matching genomics with biophysical-based networks to evaluate connectivity*.

Treml, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., & Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integrative and Comparative Biology*, *52*(4), 525–537. doi: 10.1093/icb/ics101

Vigliola, L., Doherty, P. J., Meekan, M. G., Drown, D. M., Jones, M. E., & Barber, P. H. (2007). Genetic identity determines risk of post-settlement mortality of a marine fish. *Ecology*, *88*(5), 1263–1277. doi: 10.1890/06-0066

Wang, S., Meyer, E., Mckay, J. K., & Matz, M. V. (2012). 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, *9*(8), 808–810. doi: 10.1038/nmeth.2023

Wickham, H. (2009). *Elegant graphics for data analysis (ggplot2)*. New York, NY: Springer-Verlag.

Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M. R., Bernatchez, L., & Fortin, M. J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology*, *27*(10), 2347–2364. doi: 10.1111/mec.14589

Yoklavich, M., & Bailey, K. (1990). Hatching period, growth and survival of young walleye pollock Theragra chalcogramma as determined from otolith analysis. *Marine Ecology Progress Series*, *64*(Fao 1988), 13–23. doi: 10.3354/meps064013

Supplementary material

Table S1: Full model details on PERMANOVA for the effect of Development stage and Locality in the studied early life traits. Bold values indicate significant effects at α=0.05.

		Development	Locality	Development: Locality	Residuals	Total
PLD	Df	1	2	2	94	99
	SumsOfSqs	0.016	0.074	0.011	0.184	0.284
	MeanSqs	0.016	0.037	0.005	0.002	
	F.Model	8.213	18.805	2.687		
	R2	0.057	0.259	0.037	0.647	1
	Pr(>F)	**0.006**	**0.001**	0.08		
GRPLD	Df	1	2	2	94	99
	SumsOfSqs	0.129	0.117	0.024	0.203	0.473
	MeanSqs	0.129	0.058	0.012	0.002	
	F.Model	59.692	26.929	5.559		
	R2	0.273	0.246	0.051	0.430	1
	Pr(>F)	**0.001**	**0.001**	**0.007**		
Settlement size	Df	1	2	2	94	99
	SumsOfSqs	0.002	0.003	0.015	0.284	0.304
	MeanSqs	0.002	0.002	0.008	0.003	
	F.Model	0.501	0.523	2.513		
	R2	0.005	0.010	0.050	0.935	1
	Pr(>F)	0.479	0.59	0.095		
Hatch size	Df	1	2	2	94	99
	SumsOfSqs	0.005	0.014	0.014	0.215	0.248
	MeanSqs	0.005	0.007	0.007	0.002	
	F.Model	2.111	3.046	3.066		
	R2	0.019	0.056	0.057	0.868	1
	Pr(>F)	0.133	**0.032**	0.062		
Hatch date	Df	1	2	2	94	99
	SumsOfSqs	0.019	0.019	0.004	0.025	0.067
	MeanSqs	0.019	0.009	0.002	0.000	
	F.Model	72.780	34.827	7.231		
	R2	0.290	0.278	0.058	0.375	1
	Pr(>F)	**0.001**	**0.001**	**0.004**		

Table S2: Full model details on PERMANOVA of Development stage in PLD, GRPLD and Hatch date for the locality-specific analysis.

		Blanes			Xabia			Aguamarga		
		Development	Residuals	Total	Development	Residuals	Total	Development	Residuals	Total
PLD	Df	1	37	38	1	26	27	1	36	37
	SumsOfSqs	0.007	0.079	0.085	0.002	0.064	0.067	0.012	0.046	0.058
	MeanSqs	0.007	0.002		0.002	0.002		0.012	0.001	
	F.Model	3.064			1.009			9.738		
	R2	0.076	0.924	1	0.037	0.963	1	0.213	0.787	1
	Pr(>F)	0.099			0.352			**0.004**		
GRPLD	Df	1	37	38	1	26	27	1	36	37
	SumsOfSqs	0.103	0.091	0.194	0.000	0.068	0.068	0.020	0.058	0.079
	MeanSqs	0.103	0.002		0.000	0.003		0.020	0.002	
	F.Model	42.172			0.097			12.668		
	R2	0.533	0.467	1	0.004	0.996	1	0.260	0.740	1
	Pr(>F)	**0.001**			0.781			**0.002**		
Settlement Size	Df	1	37	38	1	26	27	1	36	37
	SumsOfSqs	0.010	0.102	0.112	0.001	0.088	0.089	0.002	0.106	0.109
	MeanSqs	0.010	0.003		0.001	0.003		0.002	0.003	
	F.Model	3.559			0.187			0.789		
	R2	0.088	0.912	1	0.007	0.993	1	0.021	0.979	1
	Pr(>F)	0.064			0.716			0.399		
Hatch Size	Df	1	36	37	1	24	25	1	34	35
	SumsOfSqs	0.000	0.044	0.044	0.021	0.141	0.161	0.001	0.031	0.032
	MeanSqs	0.000	0.001		0.021	0.006		0.001	0.001	
	F.Model	0.069			3.576			1.519		
	R2	0.002	0.998	1	0.130	0.870	1	0.043	0.957	1
	Pr(>F)	0.784			0.082			0.219		
Hatch Date	Df	1	36	37	1	24	25	1	34	35
	SumsOfSqs	0.023	0.011	0.033	0.002	0.004	0.006	0.004	0.010	0.015
	MeanSqs	0.023	0.000		0.002	0.000		0.004	0.000	
	F.Model	74.341			14.161			14.579		
	R2	0.674	0.326	1	0.371	0.629	1	0.300	0.700	1
	Pr(>F)	**0.001**			**0.002**			**0.004**		

Table S3. Genomic information of Blastn searches matching a gene with the genome of *Sparus aurata* at the Ensembl website.

haplo	Trait	Genomic Location	Overlapping Gene(s)
2990	GRPLD	15:7528867-7528903	ENSSAUG00010026015
5610	Hatch_date	19:26927648-26927681	DIP2C
8337	Hatch_date	8:5769302-5769330	LDLRAD3
11103	Hatch_date	4:37046601-37046634	HOMER2
12238	GRPLD	2:32151881-32151914	ulk2
13965	Settlement_size	20:21851048-21851081	tap2t
16942	Hatch_size	12:1101598-1101631	dcc
21274	Hatch_date	15:25800128-25800161	meis1b
23237	Hatch_date	11:9125068-9125101	pigc
33734	GRPLD	21:24574342-24574375	gpc1b
35638	PLD	11:1686750-1686783	mast2
44217	Hatch_date	12:25000422-25000455	hapln1b
44246	GRPLD	5:33027511-33027544	cux2b
46136	Hatch_date	22:25421016-25421047	PLCB1
49955	PLD	20:6654846-6654879	fdxr
55716	GRPLD	6:32068007-32068040	ENSSAUG00010027048
56437	Hatch_date	12:21906144-21906177	nedd4l
57512	PLD	18:28478793-28478826	efemp2a // atl3
63398	GRPLD	20:21959938-21959971	arhgap44
63931	GRPLD	3:1508222-1508255	ENSSAUG00010001116
75964	GRPLD	12:6056094-6056127	efna5a
78525	GRPLD	2:8802984-8803017	robo1
82964	Hatch_date	12:19872175-19872208	st6galnac6
84396	GRPLD	19:24409501-24409534	ENSSAUG00010021920
86666	Hatch_date	4:11031791-11031824	itfg1
87820	Hatch_date	2:3025744-3025776	fam222bb
94259	Hatch_date	21:34024583-34024616	ENSSAUG00010018443
100903	Hatch_date	16:19854318-19854348	sdsl
101331	Hatch_date	8:32231475-32231505	sv2bb
102878	Hatch_date	1:10226531-10226564	OLFM2
110098	GRPLD	8:8313086-8313119	mafb
116363	Hatch_date	6:31462416-31462449	shq1
119319	GRPLD	:27249114-27249147	kcnk9
119469	GRPLD	2:6620034-6620067	lhfpl6
125611	Hatch_date	19:6760018-6760051	ENSSAUG00010025264
127447	Settlement_size	22:19663520-19663553	si:dkey-119f1.1
131082	Hatch_date	8:7773492-7773523	gnaia
152023	GRPLD	11:25551861-25551894	lrp8

154666	Hatch_date	21:29557780-29557813	esyt2a
155564	Hatch_date	3:25528995-25529024	fndc5b
164417	GRPLD	3:1131049-1131082	ENSSAUG00010001053
205053	GRPLD	20:351785-351818	rbfox3a
270778	Hatch_date	12:6655091-6655124	cacna1bb
993	GRPLD	5:20967108-20967143	znrf3
3357	Hatch_date	3:4064295-4064328	fam131ba
3664	Hatch_date	20:24559317-24559350	sdk2b
4184	Settlement_size	6:6683332-6683365	ptprt
6326	Hatch_date	9:23587254-23587287	SH3RF3
6331	PLD	18:28760295-28760325	kdrl
6560	PLD	17:20880876-20880909	ENSSAUG00010002022
6999	Hatch_date	12:16896824-16896857	MED13L
7560	GRPLD	2:1583625-1583658	ENSSAUG00010002434
7764	Hatch_date	8:7598388-7598421	polr3b
8195	PLD	10:6676817-6676850	ENSSAUG00010001465
8508	PLD	9:6817065-6817098	ENSSAUG00010027244
11078	PLD	15:5318920-5318953	pcdh15a
11079	Hatch_date	7:35427238-35427271	tbc1d25
12231	GRPLD	18:32052159-32052190	enox2
12456	PLD	13:7331368-7331397	ENSSAUG00010011563
12811	Hatch_date	2:6946392-6946425	cadm2b
16568	GRPLD	11:20588687-20588720	ptprsa
16656	PLD	3:31150049-31150080	ENSSAUG00010005147
21643	Hatch_date	2:14541799-14541830	sipa1l3
22044	Hatch_date	12:1529658-1529691	dcc
24193	Hatch_date	6:6212332-6212365	grip2b
24480	Hatch_date	12:25449447-25449480	sez6l
24795	GRPLD	11:16637133-16637166	p3h2
26515	Hatch_date	20:3951661-3951694	ENSSAUG00010012102
27686	PLD	15:28618261-28618294	atrnl1a
29625	Hatch_date	1:16367322-16367354	ENSSAUG00010018719
31754	PLD	4:29699473-29699506	rufy2
33920	Hatch_date	11:14843873-14843906	elavl4
34535	Hatch_date	23:4840534-4840567	slc18a2
34829	GRPLD	19:19030241-19030274	si:ch73-173p19.1
37945	Hatch_date	7:3866771-3866802	xkr7
38687	PLD	20:4435518-4435549	RAB37
38849	PLD	13:3250011-3250044	ENSSAUG00010011937

40229	Hatch_date	4:36407219-36407252	CHST8
43866	Hatch_size	4:37081598-37081631	atp6v0d1, FSD2
45802	Hatch_date	5:12339621-12339654	ENSSAUG00010023542
46698	GRPLD	7:7423434-7423467	hnf4a
47299	Hatch_date	4:36694616-36694649	LRRC49
48100	GRPLD	6:37874697-37874730	ENSSAUG00010005424
48188	GRPLD	11:3417352-3417381	NAALADL2
50881	Hatch_size	9:12812666-12812699	klf12b
51576	Hatch_date	20:25245947-25245980	ENSSAUG00010017660
52605	GRPLD	14:3152659-3152693	large1
52668	Hatch_date	5:28414660-28414693	pde4d
54457	GRPLD	22:25604068-25604101	trmt6
57699	Hatch_date	12:4815371-4815404	fech
59910	PLD	15:5055660-5055693	prkg1b
60272	Hatch_date	7:29586738-29586771	klhdc8b
62922	GRPLD	0:32621510-32621543	KCNIP4
67288	Hatch_size	3:2186831-2186864	col14a1b
67863	Hatch_date	12:24258292-24258325	dab2ipb
70501	GRPLD	17:21067716-21067746	dgat1b
71610	PLD	4:37588073-37588106	ENSSAUG00010014880
72250	GRPLD	18:33977760-33977789	cxcl14
73099	Hatch_size	22:27433662-27433695	ches1
74086	PLD	2:1578727-1578759	ENSSAUG00010002434
74557	Hatch_date	7:24370564-24370597	soga1
77344	Hatch_size		
79400	PLD	4:34517053-34517086	slc6a2
85704	Hatch_date	3:7117165-7117198	pleca
85704	Hatch_size	3:7117165-7117198	pleca
87707	PLD	22:25599715-25599747	fermt1
88056	Hatch_size	9:26895980-26896012	wnt6b
89392	Hatch_date	14:15894949-15894982	iqsec3a
91096	Hatch_date	9:6177929-6177962	wu:fj59g11
92198	Hatch_date	22:16266789-16266822	lpgat1
93664	Hatch_date	2:30130369-30130402	LRRC75A
98224	Hatch_date	1:33851494-33851523	ENDOV
103503	Hatch_date	1:30006718-30006750	xylt1
104110	Hatch_size	20:4092648-4092681	ENSSAUG00010012584
105266	PLD	9:5181303-5181336	kcnh3
107511	Hatch_date	11:33198252-33198285	ENSSAUG00010019424
109563	Hatch_date	16:14052453-14052486	UNC79
109762	Hatch_size	15:28718175-28718208	ENSSAUG00010019103

110035	Hatch_date	9:29637589-29637621	tmeff2a
116506	Hatch_date	9:5180797-5180830	kcnh3
120573	PLD	7:5411480-5411513	plxna2
122796	Hatch_date	3:7311210-7311240	pkhd1l1
123621	Hatch_date	24:9533734-9533767	gpr137c
124528	Hatch_date	14:7287341-7287374	ano2b
126364	PLD	11:4422317-4422350	znf648, dipk1ab
128042	Hatch_date	6:34214806-34214838	iqsec1b
130270	PLD	17:34101922-34101953	sema6e
136608	Hatch_size	5:36297909-36297940	ENSSAUG00010005760
137139	Hatch_date	21:21998579-21998612	dspa
137260	Hatch_date	21:7598125-7598158	ctnnd2b
137854	Hatch_date	13:35588626-35588658	ENSSAUG00010011967
137984	Hatch_date	22:20040424-20040457	efr3bb
139690	Hatch_date	11:3280908-3280941	NAALADL2
141459	Hatch_date	20:5942164-5942191	RYR2
143141	PLD	13:299487-299520	gabra1
146966	GRPLD	10:19414059-19414092	tbc1d19
154351	PLD	19:6464710-6464743	GFOD1
156998	Hatch_size	19:22699405-22699438	CDH7
157066	Hatch_date	1:4598828-4598861	ENSSAUG00010002524
157114	GRPLD	17:35492079-35492112	bmper
158449	GRPLD	15:4492731-4492764	fbxw4
158474	Hatch_date	14:16757429-16757462	snd1
159543	Hatch_date	9:30534581-30534614	slc4a3
160291	Settlement_size	22:25493766-25493799	PLCB1
161920	Hatch_date	6:29330256-29330288	taf4a
162395	PLD	8:1283667-1283698	si:ch211-284k5.2
164583	Settlement_size	13:32160161-32160194	dbn1
166198	Settlement_size	8:15306669-15306700	PDE3A
166382	PLD	12:10777062-10777095	ENSSAUG00010013618
166731	Hatch_size	CAAHFC010000086:10079-10109	EFR3A
167499	Hatch_size	22:6780507-6780538	snx14
169381	Hatch_size	1:7897014-7897041	cacng2a
169496	Hatch_date	12:4353400-4353433	cdk9
171427	GRPLD	4:3505413-3505446	CDH8
172618	Hatch_date	19:1973999-1974032	ENSSAUG00010016423

Capítulo 3:

Conectividad de los peces litorales del Mediterráneo Occidental

Capítulo 3.1

Matching genomics with biophysical-based networks to evaluate connectivity

Héctor Torrado1, 2 *, Eric Treml3, 4, Carlos Carreras 2, Enrique Macpherson 1+, Marta Pascual 2+

[1]Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Car. Acc. Cala St. Francesc 14, 17300 Blanes, Girona, Spain.
[2]Department de Genètica, Microbiologia i Estadística and IRBio, Universitat de Barcelona, Av.Diagonal 643, 08028 Barcelona, Spain.
[3]School of BioSciences, The University of Melbourne, Melbourne, Victoria, Australia.
[4]School of Life and Environmental Sciences, Centre for Integrative Ecology, Deakin University, Geelong, Victoria, Australia
+Both authors contributed equally as senior researchers and should be considered to be at the same position

Abstract

Many benthic fishes have a brief but critical period dispersing through the plankton and a sedentary adult phase, and thus the arrival of larvae is the main mechanism to maintain connectivity across populations. Therefore, dispersal modelling and network reconstruction may help detecting the most important areas to maintain the connectivity network, crucial for a good spatial management of the species. In the present work, we model larval dispersal of three common coastal fish species along the western Mediterranean Sea. For each species, we develop connectivity estimates and construct networks representing dispersal pathways among all possibly suitable habitat patches. Furthermore, we use genomic data to evaluate the connectivity results of the dispersal model and detect high concordant results in all species between the connecting clusters found by modelling and genomic data. We locate the main nodes of network connectivity (sources, sinks and stepping-stones) and evaluate their current protection status. Thus, our study helps identifying relevant areas to ensure the resilience of species by maintaining sustainable networks of protected areas.

Keywords: Connectivity, West Mediterranean, network analysis, MPAs, *Diplodus puntazzo*, *Symphodus ocellatus*, *Symphodus tinca*.

Introduction

Connectivity is the base of long-term population's survival in all ecosystems. Marine populations are generally considered more connected due to the scarce physical barriers. Nevertheless, lots of studies have found barriers to connectivity in different species (e.g. Carreras et al., 2020; García-Merchán et al., 2012; Torrado et al., 2020). Benthic fishes usually show low mobility in adult phases (e.g. Aspillaga et al., 2016; Lejeune, 1985), therefore, the arrival of larvae from other locations is the main mechanism that maintains connectivity. The population structure is highly conditioned by the larval dispersal and barriers to gene flow, both strongly influenced by the oceanographic conditions to which they are subjected in the pelagic habitat (Kinlan et al., 2005; Scheltema, 1986). Furthermore, larval settlement is limited by the availability of suitable habitat (García-Rubies & Macpherson, 1995), which further limits the potential connectivity. As a consequence, marine fish species tend to be structured in metapopulations, with subpopulations of adults distributed in patches of benthic habitats, connected almost exclusively by movements that occur during the larval phase (Cowen & Sponaugle, 2009). Thus, conservation strategies for species with this metapopulation structure should strongly consider connectivity networks integrating habitat suitability in order to design a proper system of protected areas.

Population connectivity for a given species is difficult to measure empirically, due to the influence of spatial and temporal scales and the impact of life history traits (Kool & Nichol, 2015). In order to quantify connectivity, an assortment of techniques have been used, such as otolith chemistry analysis (Almany et al., 2007; Di Franco et al., 2012), direct genetic movement through parentage analysis (Berumen et al., 2012; Planes et al., 2009; Schunter et al., 2014), measures of genetic population structuring (Carreras-Carbonell et al., 2007; Pascual et al., 2016), biophysical models of dispersal (Andrello et al., 2013; Kough & Paris, 2015; Treml et al., 2012), or a combination of several of them (Crandall et al., 2014; Schunter et al., 2011). Among the different

technologies applied, biophysical models have demonstrated to be very useful, as they provide a mechanistic understanding of larval dispersal and connectivity (Roberts et al., 2020). Furthermore, this method allows to produce connectivity matrices which, analysing them as ecological networks, provide an intuitive way to visualize the connectivity structure and their emergent properties, such key nodes, network resilience or community clustering (Cumming et al., 2010; Treml & Halpin, 2012; Urban et al., 2009). Furthermore, contrasting this information with genetics studies to evaluate the results and find common patterns is very useful, as it allows to assess in the same study short-term (modelled) and long-term (genomic) connectivity.

Understanding the patterns and detecting the important connectivity hubs is crucial for a good spatial management of the species. Characterizing population connectivity for a given species and region is key for understanding species responses to progressive habitat fragmentation and climate change (Heller & Zavaleta, 2009). This is particularly important in areas such the West Mediterranean, where the presence of oceanographic fronts limit the connectivity among hydrodynamic units (Boero et al., 2016; Torrado, Mourre, et al., n.d.). Among the spatial management figures, Marine protected areas (MPAs) are indisputably the flagship tools to protect marine biodiversity. Nevertheless, some studies have suggested connectivity problems for some species in current marine protected area (MPA) networks (Boero et al., 2016), for example in Australia (Coleman et al., 2011) or the Mediterranean (Andrello et al., 2013). Thus, study connectivity in different species is crucial in MPAs design, to ensuring their utility and species long-term persistence, maintaining not only self-persistent isolated populations, but a well-connected network which make species more resilient to local stochastic perturbations.

In the present work we model larval dispersal of three common coastal fish species along the western Mediterranean Sea. This biophysical model is based on three components: the physical oceanography, the biological attributes of each species and the availability of their preferred habitats in the studied zone. For each species, we develop connectivity estimates and construct networks representing dispersal pathways among all possible habitat patches. Furthermore, we use data from

population genomics for these three species previously gathered (Torrado et al., 2020; Torrado, Raventós, et al., n.d.) to evaluate the connectivity results from different methodologies. Our specific objectives are: a) to model the main connectivity structure in the area; b) compare the information from network clustering with genomic patterns of population structure; and c) use network information to improve the conservation strategies for those species.

Material and methods

Species

In the present work, we study the connectivity patterns in three representative coastal fishes of the Western Mediterranean Sea: the sharpsnout seabream *Diplodus puntazzo*, the ocellated wrasse *Symphodus ocellatus*, and the East Atlantic peacock wrasse *S. tinca* (Table 1). These three species are representative components of the rocky, sea-grass beds and soft bottom communities of the Western Mediterranean (García-Rubies & Zabala 1991, Macpherson et al. 2002). They share the same kinds of habitats, have a narrow bathymetric range (<100 m) and their variability on early-life traits and abundances in the study area are detailed in (Raventós et al., n.d.). The selected species expend different times in the plankton as larvae (PLD) and reproduce in different times (Table 1) and for these reasons, they constitute a representative set of the benthic fish communities of the western Mediterranean.

Table 1. Studied species' attributes: Hatch date and mean PLD. Data from (Raventós et al., n.d.)

Species	Hatch dates (min-max) (DD/MM)	PLD mean
Diplodus puntazzo (Walbaum, 1792)	15/08-30/10	21
Symphodus ocellatus (Linnaeus, 1758)	24/05-15/09	10
Symphodus tinca (Linnaeus, 1758)	12/05-22/06	10

Study area

The Mediterranean Sea is a hotspot of biodiversity (Myers et al., 2000). It is a semi-enclosed marine basin surrounded by large continental masses and connected with the Atlantic Ocean through the Strait of Gibraltar. The patterns of water circulation in the Western Mediterranean is characterized by the inflow of surface Atlantic water and outflow of deeper Mediterranean water (Millot & Taupier-Letage, 2005). The circulation pattern and topography along the southern and eastern coasts of the Iberian Peninsula originate three main oceanographic discontinuities (Figure 1) which can act as barriers to dispersion (Pascual et al., 2016; Torrado, Mourre, et al., n.d.) (1) the Almeria-Oran Front, (2) the Ibiza Channel and (3) the Balearic Front.

The three species used in the present study live in shallow infralittoral rocky habitats and occasionally in seaweeds extensions (Sala et al., 2012), and thus habitats with Rocky and *Posidonia* beds infralittoral bottoms (Figure 1) have been selected from databases. Habitat data has been derived from EMODnet broad-scale seabed habitat map for Europe (2016), licensed under CC-BY 4.0 from the European Marine Observation and Data Network (EMODnet) Seabed Habitats initiative (www.emodnet-seabedhabitats.eu), funded by the European Commission. That is made available under the European Marine Observation Data Network (EMODnet) Seabed Habitats project (http://www.emodnet-seabedhabitats.eu/), funded by the European Commission's Directorate-General for Maritime Affairs and Fisheries (DG MARE). The habitats with Rocky and Posidonia beds infralittoral bottoms where our selected species are known to live are A3, A3.2, A3.3, A5.535 and A5.5353 in EUNIS habitat classification.

Modelling

Larval dispersal was simulated following a biophysical modelling approach composed by three components: oceanic currents' velocities data, a gridded map of the seascape (coastlines and potential habitat patches, see above), and biological parameters for each species.

The ocean current fields were provided by the Western Mediterranean OPerational forecasting system (WMOP) (Juza et al., 2016), developed at the Balearic

Islands Coastal Observing and Forecasting System (www.socib.es). WOMP is a regional configuration of the ROMS model implemented over the Western Mediterranean Sea, with a spatial coverage from Gibraltar strait to Sardinia Channel (6°W, 9°E, 35°N, 44.5°N) and a horizontal resolution of around 2 km and 32 sigma-levels in the vertical dimension.

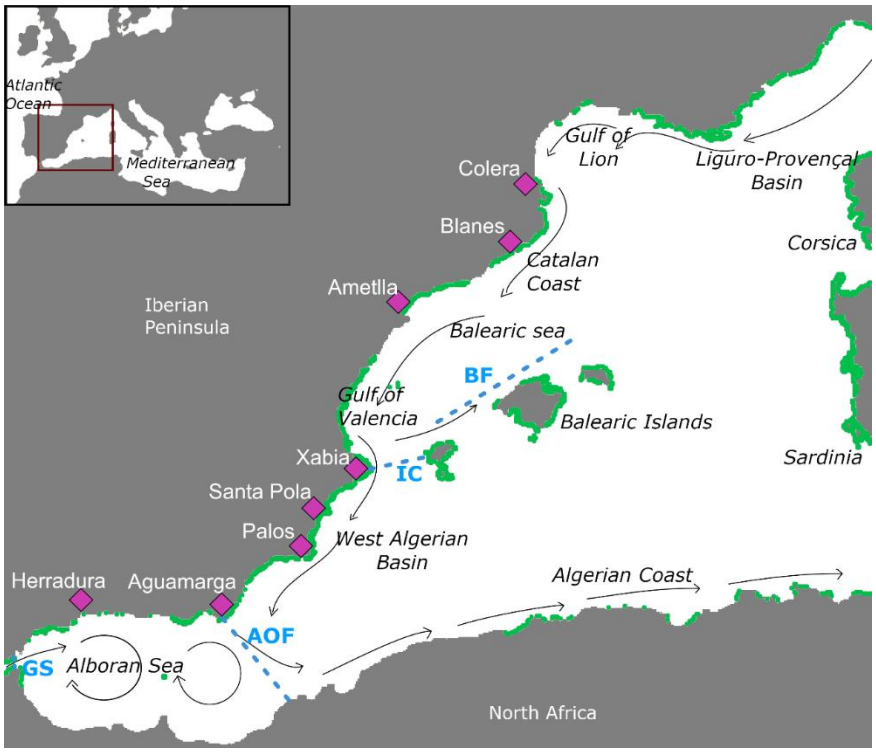


Figure 1 Location of potential habitats (green) and sampling points for genomic analyses (purple diamonds). The main geographic areas (black text), main oceanographic circulation (arrows) and oceanographic barriers (blue) are provided. GS: Gibraltar strait, AOF: Almeria-Oran front, IC: Ibiza channel, BF: Balearic front. Circulation patterns adapted from (Millot, 1999).

Both ocean current fields and habitat layers were reprojected to a Lambert Azimuthal Equidistant projection using ArcGIS. Biological parameters (Table 1) were estimated based on otolith data from (Raventós et al., n.d.). We modelled larval dispersal along the potential suitable habitat for each species during their reproductive time considering five years (2014-2018). The simulation consisted in following the

dispersal of a cloud of virtual larvae (equivalent to more than 10^7 larvae) released from each habitat patch (green dots in Figure 1), during the whole reproductive period (Table 1), recording the quantity of larvae settled in each suitable habitat after the mean PLD as in (Treml & Halpin, 2012). Larval settlement information was used to build connectivity networks showing the geographic structure of the populations' connectivity following (Cowen et al., 2006; Treml et al., 2008), using Igraph R package (Csardi et al., 2006). We analysed these networks to find densely connected subgraphs (*i.e.,* clusters of connectivity) and to calculate for all the nodes (*i.e.* each habitat patch) three main parameters for each species, using Igraph functions. First, the total number of incoming paths (indegree), *i.e.* the number of arriving connections from other nodes to the node of interest. Secondly, the number of outcoming paths (outdegree), *i.e.* the number of leaving connections from the node of interest to other nodes in the network. Finally, we calculated the betweenness, the proportion of shortest paths passing through the node (Treml et al., 2008). The shortest paths represents the shortest route between any node and every other node within a network, and can be used as a proxy for most probably used dispersal pathways (Treml et al., 2008). Thus, points with high betweenness values may help to locate important dispersal pathways and highlight key locations for stepping stone movements (Treml et al., 2008).

Genotyping and population genomics

We obtained published GBS genomic data for *Symphodus* species (Torrado et al., 2020) and 2b-RAD genomic data for *Diplodus puntazzo* (Torrado, Raventós, et al., n.d.). We used the haplotype loci files for the population genomic analyses. For each species, F_{ST} pairwise values and their significance (computed by 999 permutations) were obtained by R software (R Core Team, 2015) 'hierfstat' package vs 0.04-2 (Goudet & Jombart, 2015). In order to compare genomic and net-based clusters we calculate a Rand Index (Rand, 1971) between dissimilarity matrixes whit a customize script.

Results

We found that the preferred habitats of our species are common in the West Mediterranean Sea, allowing a high habitat suitability map for all three species (Figure 1). Nevertheless, we find some zones with wide gaps, mainly in the African Coast (probably due to the lack of studies in the area) and near the rivers' mouths and extended sandy zones along the European coastline.

Based on connectivity information extracted from larval dispersal simulations, we identified the main connectivity clusters in the West Mediterranean for the three studied species (Figure 2). Starting from the eastern part, we see that Corsica and Sardinia (Figures 1 and 2) usually form one or two overlapping clusters. For *S. tinca*, we found only a big cluster, also including the easternmost parts of the Liguro-Provençal Basin and the Algerian Coast. In all species, Sardinia is connected with the eastern part of the Algerian coast. This Algerian Coast is always divided in two well connected clusters, the eastern and the western part, with highly variable connectivity between species with the

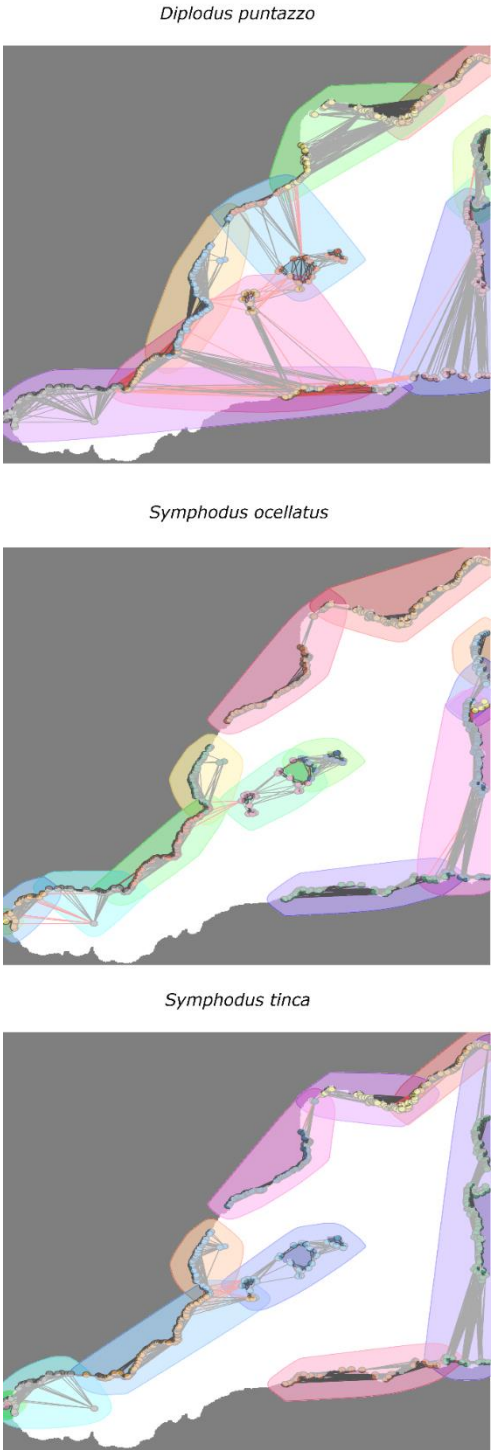


Figure 2 Clusters of connectivity based on the biophysical model.

West Algerian Basin and the Balearic Islands: very high in *D. puntazzo*, weak in *S. ocellatus* and negligible in *S. tinca*.

The Balearic Islands (Figures 1 and 2) are divided in two overlapping clusters in all the species, with the bigger eastern islands in one and the smaller western islands in the other. Nevertheless, the composition of these clusters change among species. In *S. ocellatus,* the eastern cluster only includes the two islands, whereas in *D. puntazzo* the cluster also includes the southern part of the Catalan Coast. West Algerian Basin is a highly variable part. It stands alone in *S. ocellatus*, but forms a cluster with Balearic Islands in *S. tinca* and with the islands and the Algerian Coast in *D. puntazzo*.

Alboran Sea is highly complex (Figures 1 and 2), with a variable number of clusters and subclusters in the different species. Nevertheless, the eastern part always forms part of a cluster including the West Algerian Basin. Only in *D. puntazzo* the Alboran sea is clustered together with the western part of the Algerian Coast.

We also located the nodes with highest indegree (number of paths coming into a node), outdegree (number of paths leaving a node) and/or betweenness (proportion of shortest paths passing through that node) in the study area. We can see nodes with high indegree grouping along the coast, which could be considered as sink zones, mainly in the west Alboran sea, Liguro-Provençal Basin, Algerian coast and Gulf of Valencia (Figure 3). We can also see groupings of nodes with high outdegree, (source zones) for example in the Liguro-Provençal Basin, Sardinia and the West Algerian Basin (Figure 4). On the other hand, nodes with high betweenness are scarce, with only four zones common to all species (Figure 5). Nevertheless, they are very important, as they can act as stepping-stone nodes, vital to the connection of some parts of the network.

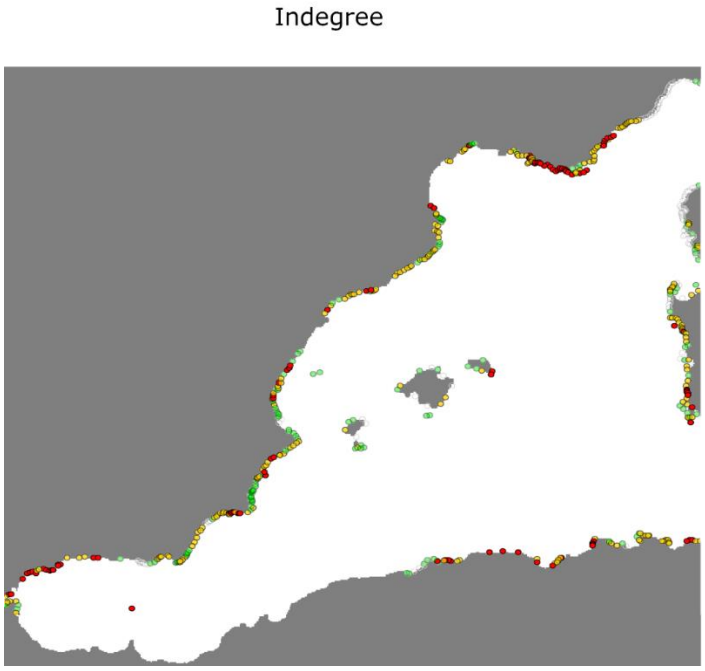


Figure 3 Nodes with Indegree values higher than mean in one (green), two (yellow) or three (red) species.

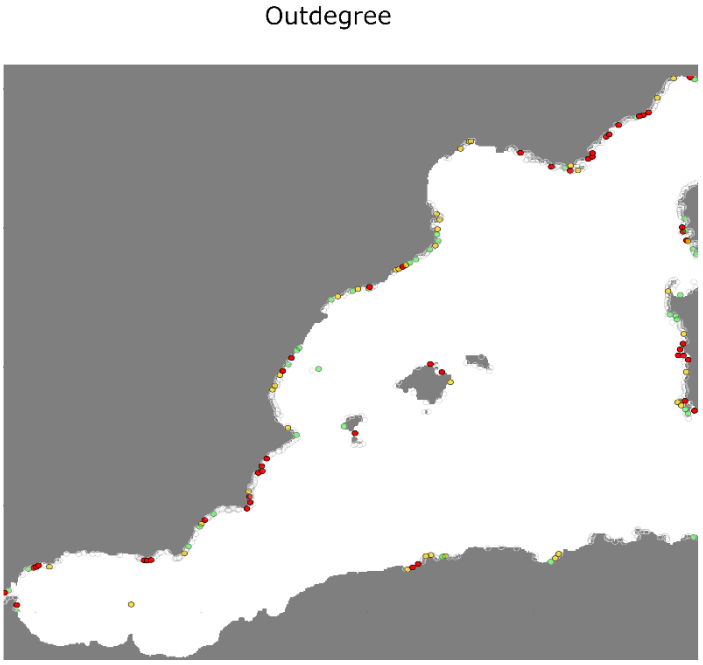


Figure 4 Nodes with Outdegree values higher than mean (sources) in one (green), two (yellow) or three (red) species.

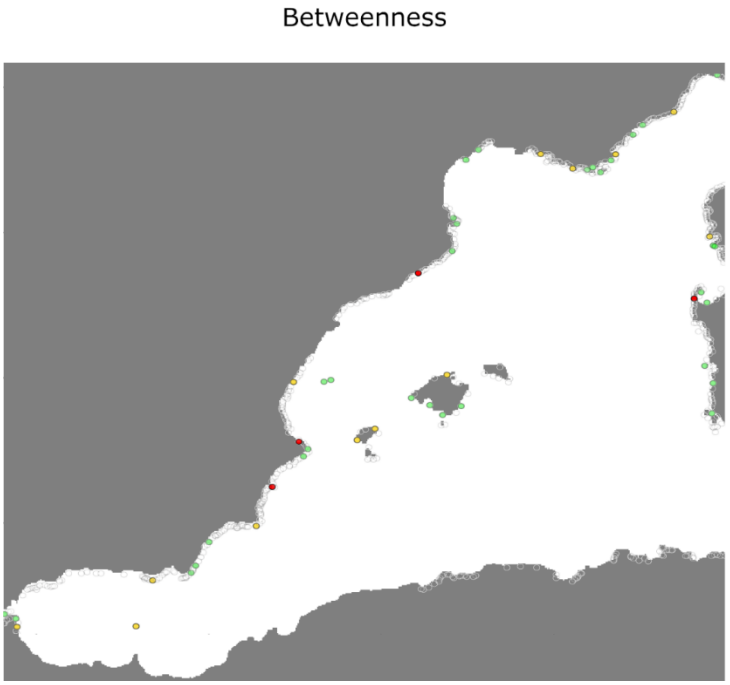


Figure 5 Nodes with Betweenness values higher than mean in one (green), two (yellow) or three (red) species.

In order to contrast our clustering information with knowing genomic structuring in this species, we used genomic data from literature (see methods). Clusters for *D. puntazzo* were identical with both methods and the Rand index was 1 (Table 2). In *S. ocellatus*, the only difference was Aguamarga locality, being clustered with both Xabia/Palos group and Herradura group in net clustering, while only with the first with genomic data (Table 3). Thus, clustering obtained from genomic data and from the biophysical model was very similar with a Rand index of 0.93. The case of *S. tinca* was the less similar. Xabia, Palos and Santa Pola were part of a different cluster that those formed by Colera/Blanes/Ametlla in the biophysical model, but all except Santa Pola with Blanes and Ametlla presented lack of genetic differentiation (Table 4). Thus, clustering obtained from genomic data and from the biophysical model showed higher discrepancies, with a Rand index of 0.6.

Table 2. Fst values (lower diagonal) and network clustering (upper diagonal, 1 same cluster, 0 different cluster) for *Diplodus puntazzo*. For F_{ST}, bold indicates significant differences at α=0.05. Values in bold are considered as different cluster (0) and otherwise considered as belonging to the same cluster (1).

	Blanes	Xabia	Aguamarga
Blanes		0	0
Xabia	**0.008**		1
Aguamarga	**0.012**	-0.002	

Table 3. Fst values (lower diagonal) and network clustering (upper diagonal, 1 same cluster, 0 different cluster) for *Symphodus ocellatus*. For F_{ST}, bold indicates significant differences at α=0.05. Values in bold are considered as different cluster (0) and otherwise considered as belonging to the same cluster (1)

	Colera	Blanes	Xabia	Palos	Aguamarga	Herradura
Colera		1	0	0	0	0
Blanes	-0.001		0	0	0	0
Xabia	**0.029**	**0.028**		1	1	0
Palos	**0.041**	**0.041**	0.003		1	0
Aguamarga	**0.038**	**0.038**	0.004	0.001		1
Herradura	**0.022**	**0.023**	**0.036**	**0.049**	**0.042**	

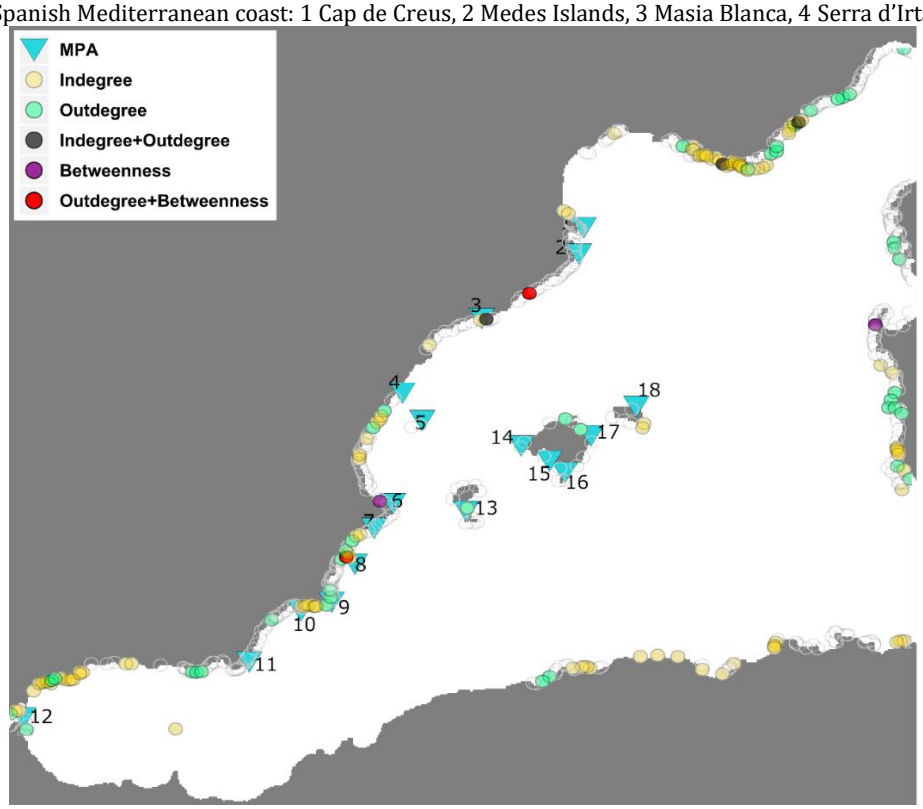
Table 4. Fst values (lower diagonal) and network clustering (upper diagonal, 1 same cluster, 0 different cluster) for *Symphodus tinca*. For F_{ST}, bold indicates significant differences at α=0.05.

	Colera	Blanes	Ametlla	Xabia	Santa Pola
Colera		1	1	0	0
Blanes	0.001		1	0	0
Ametlla	0.001	0.000		0	0
Xabia	0.000	0.001	0.001		1
Santa Pola	0.000	**0.002**	**0.002**	0.000	

We selected exclusively the marine reserves and natural parks with zones of integral protection located in the Spanish Mediterranean Coast to analyse their overlap with areas shown to be important for connection in the three species (high betweenness) and/or source areas (high outdegree. We observed some coincidences among MPAs and places with high betweenness and/or outdegrees (Figure 6) in the Spanish coast. Only four areas common to the three species with high outdegree and potential source areas in the network are protected (Figure 6): Masia Blanca (3), Nova Tabarca Island (8), Cabo de Palos (9), Ibiza-Formentera (13) and Cala Rajada (17). On

the other hand, two of the three high betweenness nodes in the Spanish coast are protected (Cap de Sant Antoni [6] and Nova Tabarca [8]). The unprotected point is located in the Catalan Coast, which is also a source zone, remarking its importance. Finally, there's not any point with high value for all tree parameters in the area. Neverheless. there is a node with high outdegree and indegree, located at Masia Blanca (3). Furthermore, near the MPA of Nova Tabarca (8) we observed a zone with high outdegree and betweenness and a close node with high indegree. Thus, these zones could also be interesting to create bigger protected areas including different categories of nodes.

Figure 6 Marine protected areas (marine reserves and natural parcs with zones of integral protection) in the Spanish Mediterranean coast: 1 Cap de Creus, 2 Medes Islands, 3 Masia Blanca, 4 Serra d'Irta, 5



Columbretes Islands, 6 Cap de Sant Antoni, 7 Serra Gelada, 8 Nova Tabarca Island, 9 Cabo de Palos, 10 Cabo Tiñoso, 11 Cabo de Gata, 12 Gibraltar Strait, 13 Ibiza-Formentera, 14 Sa Dragonera, 15 Palma Bay, 16 Mallorca Migjorn, 17 Cala Rajada, 18 Menorca. Source: MAPAMED, the database on Sites of interest for the conservation of marine environment in the Mediterranean Sea. MedPAN, UNEP/MAP/SPA-RAC. November 2017 release.

Discussion

Here, we assessed the connectivity structure of three common littoral fishes in the Western Mediterranean through modelling and network reconstruction. The same structure was observed in two of the three species, suggesting that historical and other biological processes can determine in some species the present day genetic structure. The network reconstruction with the biophysical model provided additional spatial characteristics of the suitable habitats to define potential management units for these species. With this information, we located the main nodes of their connectivity networks: source, sink and stepping-stone nodes.

The discrepancies between the biophysical model and the genomic data in *S. tinca* can be related with the low number of sampling localities. However, previous studies using microsatellites (Galarza et al., 2009), showed an absence of genetic structuring along the western Mediterranean, with a break at the Almeria-Oran front. Although the early life traits of both *Symphodus* species are quite similar (benthic eggs and short larval duration), as Galarza et al., (2009) pointed out, the mating and settlement behaviour of the *S. tinca* might explain this apparently anomalous result. Spawning of *Symphodus* takes place in nests built with branching algae that are often destroyed and transported by waves (Raventos, 2006). As such algal clumps are also suitable habitat for settlers (Raventos & MacPherson, 2005), passive transportation through drifting algae could help homogenize gene pools along the coast and islands. This potential transport is more common in spring (spawning time of *S. tinca*) than in summer (spawning time of *S. ocellatus*) as the frequency of storms in our study area is much higher the first season (Font et al., 1995; Izaguirre et al., 2011). Obviously, this hypothesis, needs further research.

The importance of MPAs in conservation strategies and protecting the spawning stocks is well known, and includes benefits as increasing biomass and/or abundances, adult spillover, population structure restoration and exportation of early-life individuals (Claudet et al., 2010; Crec'Hriou et al., 2010; Cudney-Bueno et al., 2009; Garcia-Rubies et al., 2017; Guidetti, 2006; Guidetti & Sala, 2007; Harmelin-Vivien et al.,

2008; Russ & Alcala, 2004). However, these benefits are maximised only if the MPAs are well connected forming a network (Boero et al., 2016). For these reasons, source, sink and betweenness areas are important to optimise conservation policies. In this sense, the protection of areas having a high betweenness (the ones with higher number of shortest paths passing through them) is crucial to maintain a high traffic originated by their position in the network (Treml et al., 2008; Urban et al., 2009). On the other hand, the source areas could be also considered as hatching areas and they are crucial to provide new recruits to populations, as they export the highest number of larvae to other areas (Crowder et al., 2000). Finally, the sink areas, where the settlers arrive in larger quantities, could be also considered as the main nursery areas and, thus, essential for conservation strategies. Therefore, all the areas with high connectivity and other relevant network properties should be taken into account in the design of MPA networks. In the Western Mediterranean, along the Spanish coast, we don´t see any MPA where the source, sink and betweenness overlap, but there is one zone where the high betweenness and source overlap (darkgrey point in Figure 6), but is not part of a MPA. Nonetheless, if the limits of the areas were extended, some of them could protect a combination of different nearly nodes with diverse proprieties. Thus, our study helps identifying relevant areas for protection to ensure the resilience of species by maintaining a sustainable network of MPAs.

Acknowledgments

This work was supported by the Spanish Government project 'PopCOmics' (CTM2017-88080) (MCIU, AEI/FEDER, UE). The authors CC and MP are members of the research group SGR2017-1120 and EM of SGR2017-378 (Catalan Government). HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities (FPU15/02390).

References

Almany, G. R., Berumen, M. L., Thorrold, S. R., Planes, S., & Jones, G. P. (2007). Local replenishment of coral reef fish populations in a marine reserve. *Science*,

316(5825), 742–744. https://doi.org/10.1126/science.1140597

Andrello, M., Mouillot, D., Beuvier, J., Albouy, C., Thuiller, W., & Manel, S. (2013). Low Connectivity between Mediterranean Marine Protected Areas: A Biophysical Modeling Approach for the Dusky Grouper Epinephelus marginatus. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068564

Aspillaga, E., Bartumeus, F., Linares, C., Starr, R. M., López-Sanz, À., Díaz, D., Zabala, M., & Hereu, B. (2016). Ordinary and extraordinary movement behaviour of small resident fish within a mediterranean marine protected area. *PLoS ONE*, *11*(7), 1–19. https://doi.org/10.1371/journal.pone.0159813

Berumen, M. L., Almany, G. R., Planes, S., Jones, G. P., Saenz-Agudelo, P., & Thorrold, S. R. (2012). Persistence of self-recruitment and patterns of larval connectivity in a marine protected area network. *Ecology and Evolution*, *2*(2), 444–452. https://doi.org/10.1002/ece3.208

Boero, F., Foglini, F., Fraschetti, S., Goriup, P., Macpherson, E., Planes, S., Soukissian, T., & Consortium, C. (2016). CoCoNet: towards coast to coast networks of marine protected areas (from the shore to the high and deep sea), coupled with sea-based wind energy potential. *SCIRES-IT-SCIentific RESearch and Information Technology*, *6*, 1–95. https://doi.org/10.2423/i22394303v6SpI

Calò, A., Lett, C., Mourre, B., Pérez-Ruzafa, Á., & García-Charton, J. A. (2018). Use of Lagrangian simulations to hindcast the geographical position of propagule release zones in a Mediterranean coastal fish. *Marine Environmental Research*, *134*, 16–27. https://doi.org/10.1016/j.marenvres.2017.12.011

Carreras-Carbonell, J., Macpherson, E., & Pascual, M. (2007). High self-recruitment levels in a Mediterranean littoral fish population revealed by microsatellite markers. *Marine Biology*, *151*(2), 719–727. https://doi.org/10.1007/s00227-006-0513-z

Carreras, C., García-Cisneros, A., Wangensteen, O. S., Ordóñez, V., Palacín, C., Pascual, M., & Turon, X. (2020). East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). *Diversity and Distributions*, *26*(3), 382–398. https://doi.org/10.1111/ddi.13016

Claudet, J., Osenberg, C. W., Domenici, P., Badalamenti, F., Milazzo, M., Falcón, J. M., Bertocci, I., Benedetti-Cecchi, L., García-Charton, J. A., Goni, R., Borg, J. A., Forcada, A., De Lucia, G. D., Pérez-Ruzafa, Á., Afonso, P., Brito, A., Guala, I., Diréach, L. Le, Sanchez-Jerez, P., … Planes, S. (2010). Marine reserves: Fish life history and ecological traits matter. *Ecological Applications*, *20*(3), 830–839. https://doi.org/10.1890/08-2131.1

Coleman, M. A., Chambers, J., Knott, N. A., Malcolm, H. A., Harasti, D., Jordan, A., & Kelaher, B. P. (2011). Connectivity within and among a network of temperate marine reserves. *PLoS ONE*, *6*(5). https://doi.org/10.1371/journal.pone.0020168

Cowen, R. K., Paris, C. B., & Srinivasan, A. (2006). Scaling of connectivity in marine populations. *Science*, *311*(5760), 522–527. https://doi.org/10.1126/science.1122039

Cowen, R. K., & Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annual Review of Marine Science*, *1*, 443–466. https://doi.org/10.1146/annurev.marine.010908.163757

Crandall, E. D., Treml, E. A., Liggins, L., Gleeson, L., Yasuda, N., Barber, P. H., Wörheide, G., & Riginos, C. (2014). Return of the ghosts of dispersal past: Historical spread and contemporary gene flow in the blue sea star Linckia laevigata. *Bulletin of Marine Science*, *90*(1), 399–425. https://doi.org/10.5343/bms.2013.1052

Crec'Hriou, R., Alemany, F., Roussel, E., Chassanite, A., Marinaro, J. Y., Mader, J., Rochel, E., & Planes, S. (2010). Fisheries replenishment of early life taxa: Potential export of fish eggs and larvae from a temperate marine protected area. *Fisheries Oceanography*, *19*(2), 135–150. https://doi.org/10.1111/j.1365-2419.2010.00533.x

Crowder, L. B., Lyman, S. J., Figueira, W. F., & Priddy, J. (2000). Source-sink population dynamics and the problem of siting marine reserves. *Bulletin of Marine Science*, *66*(3), 799–820.

Csardi, G., Nepusz, T., & others. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1–9.

Cudney-Bueno, R., Lavín, M. F., Marinone, S. G., Raimondi, P. T., & Shaw, W. W. (2009). Rapid effects of marine reserves via larval dispersal. *PLoS ONE*, *4*(1). https://doi.org/10.1371/journal.pone.0004140

Cumming, G. S., Bodin, Ö., Ernstson, H., & Elmqvist, T. (2010). Network analysis in conservation biogeography: Challenges and opportunities. *Diversity and Distributions*, *16*(3), 414–425. https://doi.org/10.1111/j.1472-4642.2010.00651.x

Di Franco, A., Gillanders, B. M., de Benedetto, G., Pennetta, A., de Leo, G. A., & Guidetti, P. (2012). Dispersal patterns of coastal fish: Implications for designing networks of marine protected areas. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0031681

Font, J., Garcialadona, E., & Gorriz, E. (1995). The seasonality of mesoscale motion in the Northern Current of the western Mediterranean: several years of evidence. *Oceanologica Acta*, *18*(2), 207–219.

Galarza, J. A., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G. F., & Rico, C. (2009). The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proceedings of the National Academy of Sciences*, *106*(5), 1473–1478. https://doi.org/10.1073/pnas.0806804106

García-Merchán, V. H., Robainas-Barcia, A., Abelló, P., Macpherson, E., Palero, F., García-Rodríguez, M., Gil de Sola, L., & Pascual, M. (2012). Phylogeographic patterns of decapod crustaceans at the Atlantic-Mediterranean transition. *Molecular Phylogenetics and Evolution*, *62*(2), 664–672. https://doi.org/10.1016/j.ympev.2011.11.009

Garcia-Rubies, A., Cebrian, E., Schembri, P. J., Evans, J., & Macpherson, E. (2017). Ecological Effects and Benefits of Mediterranean Marine Protected Areas: Management Implications. *Management of Marine Protected Areas: A Network Perspective*, 21–47. https://doi.org/10.1002/9781119075806.ch2

García-Rubies, A., & Macpherson, E. (1995). Substrate use and temporal pattern of recruitment in juvenile fishes of the Mediterranean littoral. *Marine Biology*, *124*(1), 35–42. https://doi.org/10.1007/BF00349144

Goudet, J., & Jombart, T. (2015). *hierfstat: estimation and tests of hierarchical F-*

statistics. R package version 0.04-22. https://cran.r-project.org/package=hierfstat

Guidetti, P. (2006). Marine reserves reestablish lost predatory interactions and cause community changes in rocky reefs. *Ecological Applications*, *16*(3), 963–976. https://doi.org/10.1890/1051-0761(2006)016[0963:MRRLPI]2.0.CO;2

Guidetti, P., & Sala, E. (2007). Community-wide effects of marine reserves in the Mediterranean Sea. *Marine Ecology Progress Series*, *335*, 43–56. https://doi.org/10.3354/meps335043

Harmelin-Vivien, M., Le Diréach, L., Bayle-Sempere, J., Charbonnel, E., García-Charton, J. A., Ody, D., Pérez-Ruzafa, A., Reñones, O., Sánchez-Jerez, P., & Valle, C. (2008). Gradients of abundance and biomass across reserve boundaries in six Mediterranean marine protected areas: Evidence of fish spillover? *Biological Conservation*, *141*(7), 1829–1839. https://doi.org/10.1016/j.biocon.2008.04.029

Heller, N. E., & Zavaleta, E. S. (2009). Biodiversity management in the face of climate change: A review of 22 years of recommendations. *Biological Conservation*, *142*(1), 14–32. https://doi.org/10.1016/j.biocon.2008.10.006

Izaguirre, C., Méndez, F. J., Menéndez, M., & Losada, I. J. (2011). Global extreme wave height variability based on satellite data. *Geophysical Research Letters*, *38*(10), 1–6. https://doi.org/10.1029/2011GL047302

Juza, M., Mourre, B., Renault, L., Gómara, S., Sebastian, K., López, S. L., Borrueco, B. F., Beltran, J. P., Troupin, C., Tomás, M. T., & others. (2016). Operational SOCIB forecasting system and multi-platform validation in the Western Mediterranean. *J. Oper. Oceanogr*, *9*, 9231.

Kinlan, B. P., Gaines, S. D., & Lester, S. E. (2005). Propagule dispersal and the scales of marine community process. *Diversity and Distributions*, *11*(2), 139–148. https://doi.org/10.1111/j.1366-9516.2005.00158.x

Kool, J. T., & Nichol, S. L. (2015). Four-dimensional connectivity modelling with application to Australia's north and northwest marine environments. *Environmental Modelling and Software*, *65*, 67–78. https://doi.org/10.1016/j.envsoft.2014.11.022

Kough, A. S., & Paris, C. B. (2015). The influence of spawning periodicity on population connectivity. *Coral Reefs*, *34*(3), 753–757. https://doi.org/10.1007/s00338-015-1311-1

Lejeune, P. (1985). Le comportement social des Labridés méditerranéens. *Cahier d'Ethologie Appliqué*, *5*, 1–208.

Millot, C. (1999). Circulation in the Western Mediterranean Sea. *Journal of Marine Systems*, *20*(1–4), 423–442. https://doi.org/10.1016/S0924-7963(98)00078-5

Millot, C., & Taupier-Letage, I. (2005). Circulation in the Mediterranean Sea. In A. Saliot (Ed.), *The Mediterranean Sea* (pp. 29–66). Springer Berlin Heidelberg. https://doi.org/10.1007/b107143

Myers, N., Mittermeler, R. A., Mittermeler, C. G., Da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Pascual, M., Palero, F., García-Merchán, V. H., Macpherson, E., Robainas-Barcia, A., Mestres, F., Roda, T., & Abelló, P. (2016). Temporal and spatial genetic differentiation in the crab *Liocarcinus depurator* across the Atlantic-Mediterranean transition. *Scientific Reports*, *6*(1), 1–10.

https://doi.org/10.1038/srep29892

Planes, S., Jones, G. P., & Thorrold, S. R. (2009). Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5693–5697. https://doi.org/10.1073/pnas.0808007106

R Core Team. (2015). *A language and environment for statistical computing.* Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 846–850. https://doi.org/10.1080/01621459.1971.10482356

Raventos, N. (2006). Nest site characteristics and nesting success of the five- spotted wrasse *Symphodus roissali* in the north-western Mediterranean Sea. *Journal of Fish Biology*, *68*, 305–309. https://doi.org/10.1111/j.1095-8649.2005.00885.x

Raventos, N., & MacPherson, E. (2005). Environmental influences on temporal patterns of settlement in two littoral labrid fishes in the Mediterranean Sea. *Estuarine, Coastal and Shelf Science*, *63*(4), 479–487. https://doi.org/10.1016/j.ecss.2004.11.018

Raventós, N., Torrado, H., Arthur, R., Alcoverro, T., & Macpherson, E. (n.d.). *Temperature reduces fish dispersal as larvae grow faster to their settlement size*.

Roberts, K. E., Cook, C. N., Beher, J., & Treml, E. A. (2020). Assessing the current state of ecological connectivity in a large marine protected area system. *Conservation Biology*, *00*(0), 1–12. https://doi.org/10.1111/cobi.13580

Russ, G. R., & Alcala, A. C. (2004). Marine reserves: Long-term protection is required for full recovery of predatory fish populations. *Oecologia*, *138*(4), 622–627. https://doi.org/10.1007/s00442-003-1456-4

Sala, E., Ballesteros, E., Dendrinos, P., Di Franco, A., Ferretti, F., Foley, D., Fraschetti, S., Friedlander, A., Garrabou, J., Güçlüsoy, H., Guidetti, P., Halpern, B. S., Hereu, B., Karamanlidis, A. A., Kizilkaya, Z., Macpherson, E., Mangialajo, L., Mariani, S., Micheli, F., … Zabala, M. (2012). The structure of mediterranean rocky reef ecosystems across environmental and human gradients, and conservation implications. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0032742

Scheltema, R. S. (1986). Long-Distance Dispersal By Planktonic Larvae of Shoal-Water Benthic Invertebrates Among Central Pacific Islands. *Bulletin of Marine Science*, *39*(2), 241–256.

Schunter, C., Carreras-Carbonell, J., Macpherson, E., Tintoré, J., Vidal-Vijande, E., Pascual, A., Guidetti, P., & Pascual, M. (2011). Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Molecular Ecology*, *20*(24), 5167–5181. https://doi.org/10.1111/j.1365-294X.2011.05355.x

Schunter, C., Pascual, M., Garza, J. C., Raventos, N., & Macpherson, E. (2014). Kinship analyses identify fish dispersal events on a temperate coastline. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1785). https://doi.org/10.1098/rspb.2014.0556

Torrado, H., Carreras, C., Raventos, N., Macpherson, E., & Pascual, M. (2020). Individual-based population genomics reveal different drivers of adaptation in sympatric fish. In *Scientific Reports* (Vol. 10, Issue 1). https://doi.org/10.1038/s41598-020-

69160-2

Torrado, H., Mourre, B., Raventós, N., Carreras, C., Tintoré, J., Pascual, M., & Macpherson, E. (n.d.). *Impact of individual early life traits in larval dispersal: a multispecies approach using backtracking models*.

Torrado, H., Raventós, N., Carreras, C., Macpherson, E., & Pascual, M. (n.d.). *Born late, die late: Genomic basis for early-life selective mortality in <i>Diplodus puntazzo<\i>.*

Treml, E. A., & Halpin, P. N. (2012). Marine population connectivity identifies ecological neighbors for conservation planning in the Coral Triangle. *Conservation Letters*, *5*(6), 441–449. https://doi.org/10.1111/j.1755-263X.2012.00260.x

Treml, E. A., Halpin, P. N., Urban, D. L., & Pratson, L. F. (2008). Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. *Landscape Ecology*, *23*(SUPPL. 1), 19–36. https://doi.org/10.1007/s10980-007-9138-y

Treml, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., & Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integrative and Comparative Biology*, *52*(4), 525–537. https://doi.org/10.1093/icb/ics101

Urban, D. L., Minor, E. S., Treml, E. A., & Schick, R. S. (2009). Graph models of habitat mosaics. *Ecology Letters*, *12*(3), 260–273. https://doi.org/10.1111/j.1461-0248.2008.01271.x

Discusión general

Discusión general

El Mediterráneo occidental y, particularmente, la costa española, proporciona un marco de trabajo ideal para el estudio de la variabilidad poblacional y la adaptación local. Ello es debido, por un lado, a la presencia de tres frentes oceanográficos en la zona y, por otro, al gradiente ambiental (temperatura y productividad, por ejemplo), que se desarrolla a lo largo de dicha costa. En esta tesis hemos abordado, desde varias aproximaciones metodológicas complementarias, cómo estas características están moldeando la estructura poblacional de los peces litorales de la zona. Asimismo, analizamos qué efectos tienen sobre los fenotipos larvarios y cómo se refleja en su información genética. Gracias a ello hemos encontrado patrones de dispersión comunes a todas las especies, sin dejar de lado una amplia variabilidad consecuencia de sus características biológicas.

Efectos de la variabilidad ambiental sobre los fenotipos larvales.

Encontramos un claro efecto del aumento de temperatura sobre algunos rasgos larvales, reduciendo la duración de la fase larvaria (PLD) y aumentando la tasa de crecimiento de la larva, tendencia que se ha observado en otras especies marinas (O'Connor et al., 2007; Sponaugle et al., 2006). Estos patrones fueron comunes en casi todas las especies, aunque más marcados en las especies cuya época reproductora tiene lugar en invierno. La excepción está representada por *Coris julis*, especie con la capacidad de retrasar su metamorfosis hasta encontrar un hábitat apropiado (Laurel et al., 2015).

Debido a estas relaciones con la temperatura, y conociendo que las predicciones para el Mediterráneo auguran un aumento de entre 3 y 4ºC (Adloff et al., 2015), las larvas tendrán considerablemente menos tiempo para dispersarse (hasta un 30% de su PLD máximo en algunas especies). Ello se debe a que crecerán más rápido hasta un

tamaño de asentamiento que se muestra en gran parte invariable, como vimos en el estudio del capítulo 1. A esto se le añade que la mayor temperatura y su consecuente aceleración del desarrollo pueden mejorar la capacidad de natación y de retención local de las larvas (Leis, 2006). Por tanto, el efecto que el cambio climático tendrá sobre las capacidades de dispersión de estas especies podría tener consecuencias en su conectividad e incluso en la composición de los ecosistemas litorales mediterráneos.

Modelización de la dispersión larval.

Con el fin de conocer la capacidad de dispersión de las larvas de las diferentes especies se realizaron una serie de simulaciones oceanográficas, basadas en los datos individuales de cada larva. En estos datos se incluyeron la fecha de nacimiento, la duración de la fase larvaria, y el lugar y fecha del asentamiento. Estos datos, combinados con unas simulaciones retrospectivas en el tiempo, nos permitieron definir los lugares de origen potencial de cada individuo.

Las características larvarias empleadas en este modelo (PLD y fecha de nacimiento y asentamiento) tuvieron efectos significativos sobre los resultados finales de las simulaciones, afectando tanto a la distancia a de dispersión como a la orientación de ésta. Nuestros resultados demuestran que las especies con PLD más largos se dispersan a mayores distancias, como también se ha visto en estudios previos (por ejemplo, Pascual et al., 2017; Siegel et al., 2003). También influyó en las distancias de dispersión la fecha de eclosión de las larvas, lo que podría estar relacionado con el efecto de la temperatura en el PLD. No obstante, el impacto de la fecha de eclosión fue nulo para las dos especies con mayor PLD (*D. vulgaris* y *C. julis*) y para *S. salpa*. Además, el signo del coeficiente de regresión fue variable entre las especies. Las que se reproducen en primavera mostraron correlaciones negativas con la distancia. En éstas, los individuos que nacen antes experimentan temperaturas más frías, y muestran PLD más largos que los que nacen más tarde en la temporada y, en consecuencia, se dispersan más. Por otra parte, las especies que se reproducen en verano (*D. puntazzo*, *O. melanura*, *C. chromis* y *S. ocellatus*) tuvieron correlaciones positivas, lo que indica que los individuos que eclosionan más tarde tienen mayores distancias de dispersión.

Dado que en este caso la temperatura aumenta con el tiempo, produciendo una disminución del PLD, el efecto de la fecha de eclosión en la distancia de dispersión probablemente se deba a la interacción de otros factores desconocidos.

Con estas simulaciones pudimos definir tres unidades hidrodinámicas separadas por dos de los principales frentes oceanográficos de la zona: el canal de Ibiza y el frente Almería-Orán. En ellas encontramos altas tasas de autorreclutamiento y una comunicación limitada entre unidades hidrodinámicas. Este bajo nivel de intercambio larvario, observado en todas las especies, sugiere que el autorreclutamiento es el mecanismo común de reposición larval en los peces costeros del Mediterráneo. Esta afirmación se ve respaldada por estudios genéticos realizados en el Mediterráneo occidental (Carreras-Carbonell et al., 2007; Schunter, Carreras-Carbonell, Macpherson, et al., 2011).

Por otra parte, las barreras oceanográficas que separan estas tres unidades no son completamente impermeables ni constantes a lo largo del tiempo. Esto puede ser explicado por los cambios temporales en las corrientes y en la intensidad del frente. Este parece el caso de *D. vulgaris*, que se reproduce en invierno, y para el que el frente Almería-Orán permitió un gran intercambio larvario en ambas direcciones. Esto concuerda con la conocida variabilidad de este frente, disminuyendo su fuerza o incluso desapareciendo en invierno (Fernández et al., 2005b; Renault et al., 2012; Tintore et al., 1988). Además, se han observado cambios en el flujo génico en esta zona en diversas especies (Calderón et al., 2012; Pascual et al., 2016; Pérez-Portela et al., 2019). En el caso del canal de Ibiza también se ha descrito variabilidad temporal en el transporte (Balbín et al., 2014; Heslop et al., 2012; Pinot et al., 2002). En nuestro caso observamos un pequeño transporte de larvas en *Symphodus ocellatus*, mientras que en *S. tinca*, del mismo género y con PLD similar, la proporción fue mayor. Y dado que estas dos especies se reproducen en diferentes estaciones, la estacionalidad temporal del frente puede explicar estas diferencias. Esta variabilidad en el canal de Ibiza queda también reflejada en la diversidad de patrones de diferenciación genética en distintas especies y con diferentes periodos de reproducción (Pascual et al., 2016; Schunter,

Carreras-Carbonell, Macpherson, et al., 2011; Schunter, Carreras-Carbonell, Planes, et al., 2011; Torrado et al., 2020).

Estudio poblacional comparativo de especies congenéricas y simpátricas.

Por otra parte, las dos especies congenéricas (*Symphodus ocellatus* y *S. tinca*), ambas con un corto PLD (7-11 días), presentaron patrones de estructuración genética muy dispares, a pesar de su gran similitud ecológica y cercanía taxonómica. En *S. ocellatus* encontramos tres poblaciones genéticas diferenciadas debido a un claro efecto del canal de Ibiza y el frente Almería-Orán, coincidiendo con lo visto en una amplia variedad de taxones con diferentes capacidades de dispersión (Pascual et al., 2017b). Sin embargo, *S. tinca* mostró una baja diferenciación genética y nulo efecto del canal de Ibiza, reforzando los resultados de un estudio previo utilizando microsatélites (Galarza et al., 2009). El efecto algo debilitado de dicho frente se observó también en las simulaciones del capítulo 1.2, en las que *S. tinca* presentaba mayor intercambio de partículas que *S. ocellatus*. Sin embargo, la dispersión observada en estas simulaciones no acaba de explicar la baja estructuración poblacional que presenta esta especie.

Ambas especies muestran cuidado parental por parte del macho, construyendo nidos con algas en el que la puesta se deposita y guarda hasta la eclosión de las larvas (Raventos, 2006; Warner & Lejeune, 1985). Además, presentan rasgos similares en la primera fase del ciclo vital, como el corto PLD, el tamaño al asentamiento y la tasa de crecimiento (Raventos & Macpherson, 2001). Sin embargo, difieren en la duración de la vida adulta, talla máxima y en la época de reproducción, presentando *S. tinca* adultos de mayor tamaño (35cm frente a 12cm) y más longevos (más de 7 años frente a unos 5), y comenzando la reproducción un mes antes que *S. ocellatus*, si bien la época de reproducción de éste último es significativamente más extensa (de mayo a junio en *S. tinca* y de junio a septiembre en *S. ocellatus*) (Bauchot, 1987; Froese & Pauly, 2016; Lejeune, 1985). Estas contadas diferencias podrían explicar, al menos en parte, las discrepancias encontradas en la estructuración poblacional de ambas especies. Sus fases adultas exhiben comportamientos altamente sedentarios (Lejeune, 1985), por lo

que la dispersión recae completamente en la fase larvaria. No obstante, nada impide que, esporádicamente, algún individuo adulto recorra distancias mayores de lo habitual, por ejemplo a causa de factores externos como tormentas, como se ha observado en *Diplodus sargus* (Aspillaga et al., 2016), especie con una sedentarismo semejante en la fase adulta. Si algunos adultos se trasladan a áreas vecinas, especies con una vida más larga como *S. tinca* podrían presentar poblaciones más conectadas que su congénere *S. ocellatus*, de vida más corta (Gordoa et al., 2000). Sin embargo, es indiscutible que los adultos son, como mínimo, mayoritariamente sedentarios, y permanecen cerca de su área de asentamiento. Por tanto, se debe considerar también el efecto que otras variables, como el momento de reproducción, pudieran tener sobre la estructura de la población. Otra posibilidad consistiría en el transporte arribazones de algas (Galarza et al., 2009), ya que estas son un hábitat apto para los reclutas (Raventos & MacPherson, 2005) o de nidos, que son frecuentemente destruidos y transportados por el oleaje (Raventos, 2006). Este transporte fuera del periodo larval es debido principalmente a tormentas, más comunes en el área de estudio en primavera (Font et al., 1995; Izaguirre et al., 2011), el periodo de reproducción de *S. tinca*. Serán necesarios futuros estudios en este tema para poder confirmar o desmentir estas hipótesis.

La variación en las fechas de nacimiento de los individuos de la misma especie y localidad puede influir en el origen potencial de las larvas, como se vio en el capítulo 1.2. Además, en estas dos especies se vio un paso de larvas distinto a través del canal de Ibiza (mayor en *S. tinca*). Por ello, no sería atrevido pensar que la diferencia en la época reproductiva de ambas especies conlleve algunos cambios en la intensidad y direccionalidad de las corrientes que afecten a sus capacidades dispersivas y al efecto de los frentes oceanográficos sobre la conectividad, aunque quizás no los suficientes como para explicar las diferencias en la estructuración de ambas. Por ejemplo, ya vimos que el frente situado en el canal de Ibiza se vuelve más permeable (aunque con variaciones interanuales) con más frecuencia en primavera que en verano (Balbín et al., 2014; Fernández et al., 2005a), lo que favorece la conectividad en las especies que se reproducen en esta época (Schunter, Carreras-Carbonell, Macpherson, et al., 2011).

En el caso concreto de estas dos especies, pudimos observar que las tasas de autoreclutamiento en las unidades hidrodinámicas eran mayores en *S. ocellatus*, mientras que la comunicación entre el mar Balear y la Cuenca Argelina Occidental era mayor en *S. tinca* (Capítulo 1-1 de esta tesis). Ello apoya la hipótesis de que, al menos en parte, la época de reproducción está influyendo en la comunicación entre cuencas y, por tanto, en la estructuración poblacional de estas especies.

Bases genómicas de la variabilidad de las características larvarias y correlación con los cambios ambientales.

A pesar de las diferencias encontradas en la estructuración poblacional, el uso combinado de varias metodologías para evaluar señales de selección reveló múltiples loci asociados a variables ambientales y fenotípicas en ambas especies. Para ello, fue de gran importancia el estudio integrado de datos a nivel individual tanto genómicos como ambientales y fenotípicos.

Estos loci parecen indicar ciertas interacciones genotipo-fenotipo-ambiente en las especies de estudio. En el caso de *Symphodus tinca*, identificamos dos loci asociados a la fecha de eclosión y también a variables ambientales, uno de ellos relacionado con la turbulencia y el otro con la temperatura. En *S. ocellatus*, 22 loci se relacionaron con el tamaño del asentamiento o la segunda variable fenotípica compuesta, muy influenciados por la PLD y el tamaño del asentamiento. Además, el 70% de ellos también estuvieron asociados a la temperatura. Asimismo, una alta proporción (50%) de esos loci se asociaron a la productividad y a la primera variable ambiental compuesta, también muy influenciada por la productividad.

En las dos especies de *Symphodus* las funciones biológicas de los genes asociados a los loci candidatos a estar sometidos a selección sugieren diferentes mecanismos de adaptación en respuesta a los factores ambientales. No obstante, ambas especies mostraron una fuerte influencia de la temperatura y la productividad. La importancia de los factores ambientales en los organismos marinos ha sido ampliamente estudiada y es bien sabido que la temperatura o la salinidad, por ejemplo,

pueden actuar como presiones selectivas para los organismos marinos (Carreras et al., 2020; Franch-Gras et al., 2018; Lamichhaney et al., 2012). La temperatura es un factor crucial para muchos organismos marinos ya que, al ser la mayoría de ellos ectotermos, puede tener importantes efectos el metabolismo, la supervivencia y el desarrollo de las larvas, entre otros (Sanford & Kelly, 2010). En el caso de la turbulencia, resultó más importante en *S. tinca* que en *S. ocellatus*. Aunque los valores de turbulencia pueden ser similares durante el período de reproducción de ambas especies, la frecuencia de tormentas en nuestra zona de estudio es mucho mayor en primavera que en verano (Font et al., 1995; Izaguirre et al., 2011). Esta inestabilidad puede explicar la considerable diferencia en la cantidad de loci candidatos asociados a esta variable, revelando su importancia en el reproductor primaveral *S. tinca*.

Optimización del protocolo 2b-RAD para estudios con especies no modelo.

La aplicación de la técnica 2b-RAD a nuestras dos especies no modelo (*Diplodus puntazo* y *Caretta caretta*) dio buenos resultados en ambas especies, si bien en el caso de *C. caretta* una de las enzimas (AlfI) tuvo un mayor rendimiento. Además, AlfI proporcionó un mayor número de loci que CspCI en ambas especies. Ello era de esperar dado que la secuencia de reconocimiento de AlfI tiene seis nucleótidos fijos, mientras que CspCI tiene siete nucleótidos fijos. Por ello, se espera que AlfI tenga una mayor densidad de sitios de restricción en cualquier genoma que CspCI, dé como resultado un mayor número de loci como se ha observado en otras especies como *Rhodnius ecuadoriensis* (Hernandez-Castro et al., 2017) e *in-silico* en diferentes especies de *Drosophila* (Seetharam & Stuart, 2013). Asimismo, las bandas de amplificación mostraron claras diferencias entre aquellos individuos con mayores y menores números de loci tras la secuenciación, por lo que recomendamos tener en cuenta estas bandas y descartar aquellas más débiles para optimizar los costes de secuenciación.

Cabe destacar los buenos resultados de esta técnica a la hora de procesar ADN degradado, haciéndola realmente útil para trabajar con especies que presentan dificultad de muestreo o fácil degradación del material genético. Ello se debe a la corta

longitud del ADN que necesitan las enzimas de tipo IIB (pues dan un fragmento digerido de 32-34 pb), lo que reduce la probabilidad de que falten loci incluso en muestras muy degradadas. Esta es una característica muy valiosa, ya que no todos los estudios pueden acceder fácilmente a muestras de alta calidad. Por ejemplo, los estudios genéticos de tortugas marinas generalmente se basan exclusivamente en el muestreo de individuos varados (Clusa et al., 2016) o embriones muertos encontrados en nidos después de la excavación (Clusa et al., 2018), dada la escasez de individuos y las dificultades para encontrarlos. En casos como estos, un protocolo genómico capaz de proporcionar resultados óptimos con muestras degradadas es extremadamente útil.

Por otra parte, el uso de adaptadores con bases selectivas demostró ser una herramienta útil para especies con tamaños genómicos grandes, dado que permite reducir la cantidad de loci obtenidos de cada muestra para optimizar el número de muestras por cada carril del secuenciador. Esto se debe a que el número de lecturas proporcionado por los secuenciadores es limitado por lo que (a igual cobertura por individuo) obtener más loci reduce su profundidad de cobertura, llevando a una pérdida de precisión en la posterior reconstrucción de genotipos y con ello a una mayor cantidad de datos faltantes (Casso et al., 2019). Así, y dado que la construcción y secuenciación de bibliotecas produce un número variable de lecturas por locus, asegurar una profundidad media de 20x garantizaría un mínimo lecturas por genotipo en la mayoría de los loci para cada muestra, resultando en una menor pérdida de genotipos y, por lo tanto, más loci retenidos. Nuestras simulaciones sobre análisis de remuestreo, permitieron la construcción de una curva de acumulación que relaciona el número de lecturas por muestra y el número de loci resultante, así como la correlación lineal entre la profundidad media por locus y el número de lecturas por individuo. Basándose en la combinación de estas dos funciones, el número de individuos a secuenciar en un carril se puede calcular fácilmente, simplificando la toma de decisiones y el diseño de análisis para optimizar el coste de los estudios de genómica de poblaciones. Con todo ello, proporcionamos unas pautas generalizables para la aplicación de técnicas de *RAD-seq* a especies no modelos que pretenden facilitar los

pasos iniciales de futuros proyectos en este campo en cualquier especie eucariota (Barbanti et al., 2020).

Estudio de mortalidad selectiva en *Diplodus puntazzo*.

Encontramos claras señales de mortalidad selectiva entre reclutas recientemente asentados de *Diplodus puntazzo* y los juveniles supervivientes tras seis meses. Esta mortalidad estuvo correlacionada sobretodo con tres características larvales: La fecha de eclosión, la tasa de crecimiento y la duración de la vida pelágica larval. Estas correlaciones solo se mantuvieron en el mismo sentido en las diferentes localidades para la fecha de nacimiento, con fechas posteriores en los supervivientes mostrando la variabilidad especial a la que se pueden ver sometidos estos patrones, variabilidad, junto a la temporal, ampliamente descrita en la literatura (Grorud-Colvert & Sponaugle, 2011; Rankin & Sponaugle, 2011; Searcy & Sponaugle, 2001).

Además, esta variabilidad fenotípica estuvo respaldada por las asociaciones fenotipo-genotipo encontradas por el análisis de redundancia (RDA). Esta asociación sugiere una base genética para las diferencias entre reclutas y supervivientes. Además, se han observado diferencias genéticas entre grupos de edad relacionadas con presiones sobre el crecimiento empleando distintos marcadores. Por ejemplo, selección alélica en alozimas relacionadas con el crecimiento en *Diplodus sargus* (Planes & Romans, 2004), estructura genética significativa entre los reclutas y supervivientes de tres meses como resultado de cambios en la frecuencia de los haplotipos de ADNmt en *Neopomacentrus filamentosus* (Vigliola et al., 2007), o la influencia de la exposición a depredadores sobre la composición genética de microsatélites en *Dascyllus aruanus*, relacionada con una depredación selectiva por tamaño (Ciotti & Planes, 2019).

Centrándonos en los dos rasgos con mayores diferencias fenotípicas y asociaciones más claras, la literatura nos muestra algunos casos similares a los encontrados en nuestro trabajo, como el retraso en en las fechas de desove a lo largo de los años, como se ve en *Theragra chalcogramma* (Yoklavich & Bailey, 1990). Sin

embargo, este patrón no es constante entre especies, y algunos estudios han mostrado un patrón opuesto en especies como *Lepomis macrochirus* (Cargnelli & Gross, 1996) o *Forsterygion lapillum* (Moginie & Shima, 2018). Esto sugiere que la presión no es constante y podría deberse a diferentes factores entre especies o poblaciones de la misma especie Además, se ha observado en otros peces, como *Symphodus tinca*, cierta asociación de este rasgo con diferencias genotípicas (Torrado et al., 2020). Por otra parte, las diferencias de selección en contra de una tasa de crecimiento larval mayor se han detectado en diversas especies como *Oncorhynchus mykiss* (Biro et al., 2004) o *Thalassoma bifasciatum* (Sponaugle et al., 2011). No obstante, también en este rasgo se encuentran estudios discrepantes que muestran el patrón contrario en otras especies, como en *Engraulis japonicus* (Takasuka et al., 2003) o *Spratelloides gracilis* (Meekan et al., 2006). Además, otros estudios han encontrado diferencias genéticas asociadas a diferencias en la tasa de crecimiento, como en *Diplodus sargus* (Planes & Romans, 2004).

Los genes asociados a la fecha de eclosión en nuestro estudio no tienen una función general claramente relacionada con este rasgo y serían necesarios datos adicionales, incluida la información de los padres, para explicar por qué los supervivientes nacen y se asientan más tarde que los reclutas que perecen. El caso de la tasa de crecimiento fue más claro, ya que se encontraron genes relacionados con el desarrollo, algunos con el del sistema nervioso, que podrían estar indicando qué presiones selectivas podrían actuar en esta especie.

Redes de conectividad del Mediterráneo Occidental.

El estudio se realizó en tres especies: *Diplodus puntazo*, *Symphodus ocellatus* y *S. tinca*. Se encontró una clara estructuración, congruente con el modelo biofísico y los datos genómicos, en *D. puntazzo* y *S. ocellatus*. Sin embargo, en el caso de *S. tinca* hubo considerables discrepancias, si bien aún se mantuvo una coincidencia del 60% entre ambas metodologías. Si bien estas discrepancias podrían estar relacionadas con el bajo número de localidades de muestreo en *S. tinca*, estudios previos utilizando microsatélites (Galarza et al., 2009) también mostraron una ausencia de estructuración

genética en la zona estudiada. Otras posibilidades, como se comentó en apartados anteriores, estarían relacionadas con movimientos fuera del periodo larval, ya sea debido al desplazamiento de adultos o a la translocación de nidos o reclutas por tormentas.

Con la información derivada de este modelo biofísico y respaldada con la información genómica, definimos las posibles unidades de gestión espacial para estas especies y localizamos los principales nodos de sus redes de conectividad: nodos muy interrelacionados, nodos fuente y nodos sumidero. Además, comparamos la localización de estos nodos con las áreas marinas protegidas de la costa española. Estas áreas tienen un gran valor en las estrategias de conservación, ya que incluyen beneficios como el aumento de la biomasa y/o de la abundancia, migración de adultos hacia áreas cercanas, restauración de la estructura de la poblacional y exportación de larvas (Claudet et al., 2010; Crec'Hriou et al., 2010; Cudney-Bueno et al., 2009; Garcia-Rubies et al., 2017; Guidetti, 2006; Guidetti & Sala, 2007; Harmelin-Vivien et al., 2008; Russ & Alcala, 2004). Sin embargo, para maximizar estos beneficios es necesario que las áreas marinas protegidas estén bien conectadas formando una red (Boero et al., 2016). Dentro de esas redes, las zonas clave son las de estos tres tipos de nodos que identificamos previamente. En primer lugar, la protección de las áreas que tienen una alta intermediación, que son aquellas que son atravesadas por un mayor número de caminos más cortos entre el resto de nodos de la red, es crucial para mantener la conectividad, pues son zonas de paso (Treml et al., 2008; Urban et al., 2009). Por otro lado, las áreas fuente, debido a su mayor exportación de larvas, son cruciales para proporcionar nuevos reclutas a otras poblaciones con las que estén conectadas (Crowder et al., 2000). Finalmente, las áreas sumidero son aquellas donde llegan mayor cantidad de larvas. Si bien el reclutamiento no depende sólo de la llegada sino también de la disponibilidad del microhábitat propicio para los reclutas (García-Rubies & Macpherson, 1995), estos sumideros pueden considerarse como las principales áreas de crianza, siendo por tanto esenciales para las estrategias de conservación.

Por tanto, todos estos nodos son de gran valor para las estrategias de conservación. En el Mediterráneo occidental, a lo largo de la costa española, sólo existe

un punto en que los tres tipos de áreas se superpongan, en la costa de Alicante. Por otro lado, hay una zona donde se superponen áreas de intermediación alta y de fuente, en la costa de Cataluña. Finalmente, hay cuatro zonas donde se superponen fuentes y sumideros, situadas en la costa de Cataluña, el golfo de Valencia, la costa de Alicante y la parte occidental del mar de Alborán. Únicamente una de estas áreas, la que tiene valores altos en todos los parámetros, situada en la isla alicantina de Nueva Tabarca, está protegida. Si observamos cada parámetro por separado, el 66% de las áreas de intermediación están protegidas, aproximadamente el 40% de las fuentes y alrededor del 30% de los sumideros. Por tanto, existe un alto número de zonas importantes desprotegidas, que deben ser consideradas para la creación de nuevas áreas marinas protegidas o la ampliación de las existentes.

Importancia del uso de datos individuales.

Si bien algunas de las metodologías empleadas a lo largo de los capítulos de esta tesis pueden aplicarse a nivel poblacional (*e.g.* Calò et al., 2018; Carreras et al., 2020), el uso de datos individuales ha demostrado ser muy útil e indispensable para analizar con detalle variaciones espaciales y temporales. Su uso en los estudios de genómica poblacional y asociación genotipo-fenotipo-ambiente ha sido crucial para obtener resultados claros sobre las interacciones de estos factores, que hubiesen sido muy complicados de obtener, si no imposibles, trabajando en su lugar con medias ambientales o fenotípicas.

Por otra parte, el estudio de su impacto en las simulaciones oceanográficas de dispersión de partículas demostró un claro efecto de los datos larvales individuales, tanto en la dirección como en la distancia viajada por estas larvas. Por ello, las trayectorias observadas hubieran sido significativamente diferentes de no disponer de datos individuales. Ello hubiera dado lugar a estimaciones considerablemente menos precisas. Gracias a todo ello, las conclusiones que se pueden alcanzar en este tipo de estudios son más sólidas.

Conclusiones

1. El incremento de temperatura aumenta la tasa de crecimiento larval y disminuye la duración de dicha fase, reduciendo las capacidades dispersivas de las especies.

2. La fecha de nacimiento y la duración de la vida larvaria tienen un claro efecto en las simulaciones oceanográficas, afectando a la distancia y orientación de la dispersión.

3. Los frentes oceanográficos actúan como barreras a la dispersión en las 9 especies estudiadas, generando tres unidades hidrodinámicas con elevado autoreclutamiento.

4. *Symphodus ocellatus* y *S. tinca* presentan estructuraciones genómicas diferentes a pesar de sus similitudes biológicas. No obstante, ambas especies muestran claras asociaciones genómicas con sus fenotipos larvales y condiciones ambientales.

5. La técnica 2b-RAD mostró buenos resultados incluso partiendo de ADN parcialmente degradado. Además, nuestros resultados facilitarán el diseño experimental y procesamiento de futuros estudios de genómica poblacional con esta y otras técnicas similares.

6. Existe mortalidad selectiva en los reclutas de *D. puntazzo,* relacionada con sus características larvales, especialmente con el día de nacimiento. Las variaciones fenotípicas quedan respaldadas por un considerable número de SNPs asociados a éstas.

7. Las redes de conectividad mostraron agrupaciones concordantes entre los datos genómicos y el modelo biofísico en *S. ocellatus* y *D. puntazzo*, y en menor medida en *S. tinca*.

8. Las redes permitieron localizar las zonas fuente, sumidero y nodos clave en la conectividad de estas tres especies. Sólo una pequeña proporción de estas zonas se halla protegida, por lo que se deben considerar en el diseño de futuras áreas marinas protegidas en la Mediterráneo Occidental.

Bibliografía

Adloff, F., Somot, S., Sevault, F., Jordà, G., Aznar, R., Déqué, M., Herrmann, M., Marcos, M., Dubois, C., Padorno, E., Alvarez-Fanjul, E., Gomis, D., 2015. Mediterranean Sea response to climate change in an ensemble of twenty first century scenarios. Clim. Dyn. 45, 2775–2802. https://doi.org/10.1007/s00382-015-2507-3

Ahrens, C.W., Rymer, P.D., Stow, A., Bragg, J., Dillon, S., Umbers, K.D.L., Dudaniec, R.Y., 2018. The search for loci under selection: trends, biases and progress. Mol. Ecol. 27, 1342–1356. https://doi.org/10.1111/mec.14549

Allain, G., Petitgas, P., Lazure, P., Grellier, P., 2007. Biophysical modelling of larval drift, growth and survival for the prediction of anchovy (*Engraulis encrasicolus*) recruitment in the Bay of Biscay (NE Atlantic). Fish. Oceanogr. 16, 489–505. https://doi.org/10.1111/j.1365-2419.2007.00443.x

Almany, G.R., Berumen, M.L., Thorrold, S.R., Planes, S., Jones, G.P., 2007. Local replenishment of coral reef fish populations in a marine reserve. Science (80-.). 316, 742–744. https://doi.org/10.1126/science.1140597

Álvarez-Romero, J.G., Munguía-Vega, A., Beger, M., del Mar Mancha-Cisneros, M., Suárez-Castillo, A.N., Gurney, G.G., Pressey, R.L., Gerber, L.R., Morzaria-Luna, H.N., Reyes-Bonilla, H., Adams, V.M., Kolb, M., Graham, E.M., VanDerWal, J., Castillo-López, A., Hinojosa-Arango, G., Petatán-Ramírez, D., Moreno-Baez, M., Godínez-Reyes, C.R., Torre, J., 2018. Designing connected marine reserves in the face of global warming. Glob. Chang. Biol. 24, e671–e691. https://doi.org/10.1111/gcb.13989

Andrello, M., Guilhaumon, F., Albouy, C., Parravicini, V., Scholtens, J., Verley, P., Barange, M., Sumaila, U.R., Manel, S., Mouillot, D., 2017. Global mismatch between fishing dependency and larval supply from marine reserves. Nat. Commun. 8, 1–9. https://doi.org/10.1038/ncomms16039

Andrello, M., Mouillot, D., Beuvier, J., Albouy, C., Thuiller, W., Manel, S., 2013. Low Connectivity between Mediterranean Marine Protected Areas: A Biophysical Modeling Approach for the Dusky Grouper Epinephelus marginatus. PLoS One 8. https://doi.org/10.1371/journal.pone.0068564

Andrello, M., Mouillot, D., Somot, S., Thuiller, W., Manel, S., 2015. Additive effects of climate change on connectivity between marine protected areas and larval supply to fished areas. Divers. Distrib. 21, 139–150. https://doi.org/10.1111/ddi.12250

Aspillaga, E., Bartumeus, F., Linares, C., Starr, R.M., López-Sanz, À., Díaz, D., Zabala, M., Hereu, B., 2016. Ordinary and extraordinary movement behaviour of small resident fish within a mediterranean marine protected area. PLoS One 11, 1–19. https://doi.org/10.1371/journal.pone.0159813

Astraldi, M., Balopoulos, S., Candela, J., Font, J., Gacic, M., Gasparini, G.P., Manca, B., Theocharis, A., Tintoré, J., 1999. The role of straits and channels in understanding the characteristics of Mediterranean circulation, in: Progress in Oceanography. Pergamon, pp. 65–108. https://doi.org/10.1016/S0079-6611(99)00021-X

Babbucci, M., Ferraresso, S., Pauletto, M., Franch, R., Papetti, C., Patarnello, T., Carnier, P., Bargelloni, L., 2016. An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*). Sci. Rep. 6, 1–11. https://doi.org/10.1038/srep38673

Balbín, R., López-Jurado, J.L., Flexas, M.M., Reglero, P., Vélez-Velchí, P., González-Pola,

C., Rodríguez, J.M., García, A., Alemany, F., 2014. Interannual variability of the early summer circulation around the Balearic Islands: Driving factors and potential effects on the marine ecosystem. J. Mar. Syst. 138, 70–81. https://doi.org/10.1016/j.jmarsys.2013.07.004

Barbanti, A., Torrado, H., Macpherson, E., Bargelloni, L., Franch, R., Carreras, C., Pascual, M., 2020. Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. Mol. Ecol. Resour. 1–12. https://doi.org/10.1111/1755-0998.13144

Barbut, L., Groot Crego, C., Delerue-Ricard, S., Vandamme, S., Volckaert, F.A.M., Lacroix, G., 2019. How larval traits of six flatfish species impact connectivity. Limnol. Oceanogr. 64, 1150–1171. https://doi.org/10.1002/lno.11104

Bauchot, M.L., 1987. Poissons osseux, Fiches FAO d'identification pour les besoins de la pêche.(rev. 1). Méditerranée et mer Noire. Zone de pêche.

Benestan, L., Quinn, B.K., Maaroufi, H., Laporte, M., Clark, F.K., Greenwood, S.J., Rochette, R., Bernatchez, L., 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). Mol. Ecol. 25, 5073–5092. https://doi.org/10.1111/mec.13811

Berumen, M.L., Almany, G.R., Planes, S., Jones, G.P., Saenz-Agudelo, P., Thorrold, S.R., 2012. Persistence of self-recruitment and patterns of larval connectivity in a marine protected area network. Ecol. Evol. 2, 444–452. https://doi.org/10.1002/ece3.208

Biro, P.A., Abrahams, M. V., Post, J.R., Parkinson, E.A., 2004. Predators select against high growth rates and risk-taking behaviour in domestic trout populations. Proc. R. Soc. B Biol. Sci. 271, 2233–2237. https://doi.org/10.1098/rspb.2004.2861

Boero, F., Foglini, F., Fraschetti, S., Goriup, P., Macpherson, E., Planes, S., Soukissian, T., Consortium, C., 2016. CoCoNet: towards coast to coast networks of marine protected areas (from the shore to the high and deep sea), coupled with sea-based wind energy potential. SCIRES-IT-SCIentific Res. Inf. Technol. 6, 1–95. https://doi.org/10.2423/i22394303v6SpI

Bottesch, M., Gerlach, G., Halbach, M., Bally, A., Kingsford, M.J., Mouritsen, H., 2016. A magnetic compass that might help coral reef fish larvae return to their natal reef. Curr. Biol. https://doi.org/10.1016/j.cub.2016.10.051

Calderón, I., Pita, L., Brusciotti, S., Palacín, C., Turon, X., 2012. Time and space: Genetic structure of the cohorts of the common sea urchin *Paracentrotus lividus* in Western Mediterranean. Mar. Biol. 159, 187–197. https://doi.org/10.1007/s00227-011-1799-z

Calò, A., Lett, C., Mourre, B., Pérez-Ruzafa, Á., García-Charton, J.A., 2018. Use of Lagrangian simulations to hindcast the geographical position of propagule release zones in a Mediterranean coastal fish. Mar. Environ. Res. 134, 16–27. https://doi.org/10.1016/j.marenvres.2017.12.011

Cargnelli, L.M., Gross, M.R., 1996. The temporal dimension in fish recruitment: birth date, body size, and size-dependent survival in a sunfish (bluegill: *Lepomis macrochirus*). Can. J. Fish. Aquat. Sci. 53, 360–367.

Carreras-Carbonell, J., Macpherson, E., Pascual, M., 2007. High self-recruitment levels in a Mediterranean littoral fish population revealed by microsatellite markers.

Mar. Biol. 151, 719–727. https://doi.org/10.1007/s00227-006-0513-z

Carreras, C., García-Cisneros, A., Wangensteen, O.S., Ordóñez, V., Palacín, C., Pascual, M., Turon, X., 2020. East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). Divers. Distrib. 26, 382–398. https://doi.org/10.1111/ddi.13016

Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E., Pascual, M., 2017. Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. Sci. Rep. 7, 43417. https://doi.org/10.1038/srep43417

Casso, M., Turon, X., Pascual, M., 2019. Single zooids, multiple loci: independent colonisations revealed by population genomics of a global invader. Biol. Invasions 21, 3575–3592. https://doi.org/10.1007/s10530-019-02069-8

Christensen, A., Daewel, U., Jensen, H., Mosegaard, H., St John, M., Schrum, C., 2007. Hydrodynamic backtracking of fish larvae by individual-based modelling. Mar. Ecol. Prog. Ser. 347, 221–232. https://doi.org/10.3354/meps06980

Ciotti, B.J., Planes, S., 2019. Within-generation consequences of postsettlement mortality for trait composition in wild populations: An experimental test. Ecol. Evol. 9, 2550–2561. https://doi.org/10.1002/ece3.4911

Claudet, J., Osenberg, C.W., Domenici, P., Badalamenti, F., Milazzo, M., Falcón, J.M., Bertocci, I., Benedetti-Cecchi, L., García-Charton, J.A., Goni, R., Borg, J.A., Forcada, A., De Lucia, G.D., Pérez-Ruzafa, Á., Afonso, P., Brito, A., Guala, I., Diréach, L. Le, Sanchez-Jerez, P., Somerfield, P.J., Planes, S., 2010. Marine reserves: Fish life history and ecological traits matter. Ecol. Appl. 20, 830–839. https://doi.org/10.1890/08-2131.1

Clusa, M., Carreras, C., Cardona, L., Demetropoulos, A., Margaritoulis, D., Rees, A.F., Hamza, A.A., Khalil, M., Levy, Y., Turkozan, O., Aguilar, A., Pascual, M., 2018. Philopatry in loggerhead turtles *Caretta caretta*: Beyond the gender paradigm. Mar. Ecol. Prog. Ser. 588, 201–213. https://doi.org/10.3354/meps12448

Clusa, M., Carreras, C., Pascual, M., Gaughran, S.J., Piovano, S., Avolio, D., Ollano, G., Fernández, G., Tomás, J., Raga, J.A., Aguilar, A., Cardona, L., 2016. Potential bycatch impact on distinct sea turtle populations is dependent on fishing ground rather than gear type in the Mediterranean Sea. Mar. Biol. 163, 1–10. https://doi.org/10.1007/s00227-016-2875-1

Cowen, R.K., Sponaugle, S., 2009. Larval dispersal and marine population connectivity. Ann. Rev. Mar. Sci. 1, 443–466. https://doi.org/10.1146/annurev.marine.010908.163757

Crandall, E.D., Treml, E.A., Liggins, L., Gleeson, L., Yasuda, N., Barber, P.H., Wörheide, G., Riginos, C., 2014. Return of the ghosts of dispersal past: Historical spread and contemporary gene flow in the blue sea star Linckia laevigata. Bull. Mar. Sci. 90, 399–425. https://doi.org/10.5343/bms.2013.1052

Crec'Hriou, R., Alemany, F., Roussel, E., Chassanite, A., Marinaro, J.Y., Mader, J., Rochel, E., Planes, S., 2010. Fisheries replenishment of early life taxa: Potential export of fish eggs and larvae from a temperate marine protected area. Fish. Oceanogr. 19, 135–150. https://doi.org/10.1111/j.1365-2419.2010.00533.x

Crowder, L.B., Lyman, S.J., Figueira, W.F., Priddy, J., 2000. Source-sink population

dynamics and the problem of siting marine reserves. Bull. Mar. Sci. 66, 799–820.
Cudney-Bueno, R., Lavín, M.F., Marinone, S.G., Raimondi, P.T., Shaw, W.W., 2009. Rapid effects of marine reserves via larval dispersal. PLoS One 4. https://doi.org/10.1371/journal.pone.0004140
Cumming, G.S., Bodin, Ö., Ernstson, H., Elmqvist, T., 2010. Network analysis in conservation biogeography: Challenges and opportunities. Divers. Distrib. 16, 414–425. https://doi.org/10.1111/j.1472-4642.2010.00651.x
Di Franco, A., Gillanders, B.M., de Benedetto, G., Pennetta, A., de Leo, G.A., Guidetti, P., 2012. Dispersal patterns of coastal fish: Implications for designing networks of marine protected areas. PLoS One 7. https://doi.org/10.1371/journal.pone.0031681
Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6, e19379. https://doi.org/10.1371/journal.pone.0019379
Faillettaz, R., Blandin, A., Paris, C.B., Koubbi, P., Irisson, J.O., 2015. Sun-compass orientation in mediterranean fish larvae. PLoS One 10, e0135213. https://doi.org/10.1371/journal.pone.0135213
Faillettaz, R., Durand, E., Paris, C.B., Koubbi, P., Irisson, J.O., 2018a. Swimming speeds of Mediterranean settlement-stage fish larvae nuance Hjort's aberrant drift hypothesis. Limnol. Oceanogr. 63, 509–523. https://doi.org/10.1002/lno.10643
Faillettaz, R., Paris, C.B., Irisson, J.O., 2018b. Larval fish swimming behavior alters dispersal patterns from marine protected areas in the North-Western Mediterranean Sea. Front. Mar. Sci. 5, 1–12. https://doi.org/10.3389/fmars.2018.00097
Fernández, V., Dietrich, D.E., Haney, R.L., Tintoré, J., 2005a. Mesoscale, seasonal and interannual variability in the Mediterranean Sea using a numerical ocean model. Prog. Oceanogr. 66, 321–340. https://doi.org/10.1016/j.pocean.2004.07.010
Fernández, V., Dietrich, D.E., Haney, R.L., Tintoré, J., 2005b. Mesoscale, seasonal and interannual variability in the Mediterranean Sea using a numerical ocean model. Prog. Oceanogr. 66, 321–340. https://doi.org/10.1016/j.pocean.2004.07.010
Font, J., Garcialadona, E., Gorriz, E., 1995. The seasonality of mesoscale motion in the Northern Current of the western Mediterranean: several years of evidence. Oceanol. Acta 18, 207–219.
Fraker, M.E., Anderson, E.J., Brodnik, R.M., Carreon-Martinez, L., DeVanna, K.M., Fryer, B.J., Heath, D.D., Reichert, J.M., Ludsin, S.A., 2015. Particle backtracking improves breeding subpopulation discrimination and natal-source identification in mixed populations. PLoS One 10. https://doi.org/10.1371/journal.pone.0120752
Franch-Gras, L., Hahn, C., García-Roger, E.M., Carmona, M.J., Serra, M., Gómez, A., 2018. Genomic signatures of local adaptation to the degree of environmental predictability in rotifers. Sci. Rep. 8, 16051. https://doi.org/10.1038/s41598-018-34188-y
Froese, R., Pauly, D., 2019. FishBase [WWW Document]. World Wide Web Electron. Publ.
Gagliano, M., McCormick, M.I., Meekan, M.G., 2007. Survival against the odds: ontogenetic changes in selective pressure mediate growth-mortality trade-offs in

a marine fish. Proc. R. Soc. B Biol. Sci. 274, 1575–1582.
Galarza, J.A., Carreras-Carbonell, J., Macpherson, E., Pascual, M., Roques, S., Turner, G.F., Rico, C., 2009. The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. Proc. Natl. Acad. Sci. 106, 1473–1478. https://doi.org/10.1073/pnas.0806804106
Garcia-Rubies, A., Cebrian, E., Schembri, P.J., Evans, J., Macpherson, E., 2017. Ecological Effects and Benefits of Mediterranean Marine Protected Areas: Management Implications. Manag. Mar. Prot. Areas A Netw. Perspect. 21–47. https://doi.org/10.1002/9781119075806.ch2
García-Rubies, A., Macpherson, E., 1995. Substrate use and temporal pattern of recruitment in juvenile fishes of the Mediterranean littoral. Mar. Biol. 124, 35–42. https://doi.org/10.1007/BF00349144
Garcia Lafuente, J., Lopez Jurado, J.L., Cano Lucaya, N., Vargas Yanez, M., Aguiar Garcia, J., 1995. Circulation of water masses through the Ibiza Channel. Oceanol. Acta 18, 245–254.
Garrido, S., Ben-Hamadou, R., Santos, A.M.P., Ferreira, S., Teodósio, M.A., Cotano, U., Irigoien, X., Peck, M.A., Saiz, E., Ré, P., 2015. Born small, die young: Intrinsic, size-selective mortality in marine larval fish. Sci. Rep. 5, 17065.
Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B.M., Brett Kenney, P., Vallejo, R.L., Palti, Y., Leeds, T.D., 2016. Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). Front. Genet. 7. https://doi.org/10.3389/fgene.2016.00203
Gordoa, A., Molí, B., Raventos, N., 2000. Growth performance of four wrasse species on the north-western Mediterranean coast. Fish. Res. 45, 43–50.
Grorud-Colvert, K., Sponaugle, S., 2011. Variability in water temperature affects trait-mediated survival of a newly settled coral reef fish. Oecologia 165, 675–686. https://doi.org/10.1007/s00442-010-1748-4
Grummer, J.A., Beheregaray, L.B., Bernatchez, L., Hand, B.K., Luikart, G., Narum, S.R., Taylor, E.B., 2019. Aquatic Landscape Genomics and Environmental Effects on Genetic Variation. Trends Ecol. Evol. 34, 641–654. https://doi.org/10.1016/j.tree.2019.02.013
Guidetti, P., 2006. Marine reserves reestablish lost predatory interactions and cause community changes in rocky reefs. Ecol. Appl. 16, 963–976. https://doi.org/10.1890/1051-0761(2006)016[0963:MRRLPI]2.0.CO;2
Guidetti, P., Sala, E., 2007. Community-wide effects of marine reserves in the Mediterranean Sea. Mar. Ecol. Prog. Ser. 335, 43–56. https://doi.org/10.3354/meps335043
Hamilton, S.L., Regetz, J., Warner, R.R., 2008. Postsettlement survival linked to larval life in a marine fish. Proc. Natl. Acad. Sci. U. S. A. 105, 1561–1566. https://doi.org/10.1073/pnas.0707676105
Hansen, M.M., Olivieri, I., Waller, D.M., Nielsen, E.E., 2012. Monitoring adaptive genetic responses to environmental change. Mol. Ecol. 21, 1311–1329. https://doi.org/10.1111/j.1365-294X.2011.05463.x
Harmelin-Vivien, M., Le Diréach, L., Bayle-Sempere, J., Charbonnel, E., García-Charton, J.A., Ody, D., Pérez-Ruzafa, A., Reñones, O., Sánchez-Jerez, P., Valle, C., 2008.

Gradients of abundance and biomass across reserve boundaries in six Mediterranean marine protected areas: Evidence of fish spillover? Biol. Conserv. 141, 1829–1839. https://doi.org/10.1016/j.biocon.2008.04.029

Hasan, M.M., DeFaveri, J., Kuure, S., Dash, S.N., Lehtonen, S., Merilä, J., McCairns, R.J.S., 2017. Sticklebacks adapted to divergent osmotic environments show differences in plasticity for kidney morphology and candidate gene expression. J. Exp. Biol. 220, 2175–2186. https://doi.org/10.1242/jeb.146027

Hernandez-Castro, L.E., Paterno, M., Villacís, A.G., Andersson, B., Costales, J.A., De Noia, M., Ocaña-Mayorga, S., Yumiseva, C.A., Grijalva, M.J., Llewellyn, M.S., 2017. 2b-RAD genotyping for population genomic studies of Chagas disease vectors: *Rhodnius ecuadoriensis* in Ecuador. PLoS Negl. Trop. Dis. 11, 1–21. https://doi.org/10.1371/journal.pntd.0005710

Heslop, E.E., Ruiz, S., Allen, J., López-Jurado, J.L., Renault, L., Tintoré, J., 2012. Autonomous underwater gliders monitoring variability at choke points in our ocean system: A case study in the Western Mediterranean Sea. Geophys. Res. Lett. 39, 1–6. https://doi.org/10.1029/2012GL053717

Holliday, D., Beckley, L., Millar, N., Olivar, M., Slawinski, D., Feng, M., Thompson, P., 2012. Larval fish assemblages and particle back-tracking define latitudinal and cross-shelf variability in an eastern Indian Ocean boundary current. Mar. Ecol. Prog. Ser. 460, 127–144. https://doi.org/10.3354/meps09730

Izaguirre, C., Méndez, F.J., Menéndez, M., Losada, I.J., 2011. Global extreme wave height variability based on satellite data. Geophys. Res. Lett. 38, 1–6. https://doi.org/10.1029/2011GL047302

Johnson, D.W., Christie, M.R., Stallings, C.D., Pusack, T.J., Hixon, M.A., 2015. Using post-settlement demography to estimate larval survivorship: a coral reef fish example. Oecologia 179, 729–739. https://doi.org/10.1007/s00442-015-3368-5

Kendall, M.S., Poti, M., Karnauskas, K.B., 2016. Climate change and larval transport in the ocean: Fractional effects from physical and physiological factors. Glob. Chang. Biol. 22, 1532–1547. https://doi.org/10.1111/gcb.13159

Kleypas, J.A., Thompson, D.M., Castruccio, F.S., Curchitser, E.N., Pinsky, M., Watson, J.R., 2016. Larval connectivity across temperature gradients and its potential effect on heat tolerance in coral populations. Glob. Chang. Biol. 22, 3539–3549. https://doi.org/10.1111/gcb.13347

Kool, J.T., Nichol, S.L., 2015. Four-dimensional connectivity modelling with application to Australia's north and northwest marine environments. Environ. Model. Softw. 65, 67–78. https://doi.org/10.1016/j.envsoft.2014.11.022

Kough, A.S., Paris, C.B., 2015. The influence of spawning periodicity on population connectivity. Coral Reefs 34, 753–757. https://doi.org/10.1007/s00338-015-1311-1

Lamichhaney, S., Barrio, A.M., Rafati, N., Sundstrom, G., Rubin, C.-J.C.-J., Gilbert, E.R., Berglund, J., Wetterbom, A., Laikre, L., Webster, M.T., Grabherr, M., Ryman, N., Andersson, L., Martinez Barrio, A., Rafati, N., Sundström, G., Rubin, C.-J.C.-J., Gilbert, E.R., Berglund, J., Wetterbom, A., Laikre, L., Webster, M.T., Grabherr, M., Ryman, N., Andersson, L., 2012. Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. Proc. Natl. Acad. Sci. U. S. A. 109, 19345–50. https://doi.org/10.1073/pnas.1216128109

Laurel, B.J., Basilio, A.J., Danley, C., Ryer, C.H., Spencer, M., 2015. Substrate preference and delayed settlement in northern rock sole larvae *Lepidopsetta polyxystra* . Mar. Ecol. Prog. Ser. 519, 183–193. https://doi.org/10.3354/meps11090

Leis, J.M., 2006. Are Larvae of Demersal Fishes Plankton or Nekton? Adv. Mar. Biol. 51, 57–141. https://doi.org/10.1016/S0065-2881(06)51002-8

Leis, J.M., Paris, C.B., Irisson, J.O., Yerman, M.N., Siebeck, U.E., 2014. Orientation of fish larvae in situ is consistent among locations, years and methods, but varies with time of day. Mar. Ecol. Prog. Ser. 505, 193–208. https://doi.org/10.3354/meps10792

Lejeune, P., 1985. Le comportement social des Labridés méditerranéens. Cah. d'Ethologie Appliqué 5, 1–208.

Lenfant, P., Planes, S., 2002. Temporal genetic changes between cohorts in a natural population of a marine fish, *Diplodus sargus*. Biol. J. Linn. Soc. 76, 9–20. https://doi.org/10.1046/j.1095-8312.2002.00041.x

Macpherson, E., 2002. Large-scale species-richness gradients in the Atlantic Ocean. Proc. R. Soc. B Biol. Sci. 269, 1715–1720. https://doi.org/10.1098/rspb.2002.2091

Maldonado, M., Young, C.M., 1999. Effects of the duration of larval life on postlarval stages of the demosponge *Sigmadocia caerulea*. J. Exp. Mar. Bio. Ecol. 232, 9–21. https://doi.org/10.1016/S0022-0981(98)00076-8

Manuzzi, A., Zane, L., Muñoz-Merida, A., Griffiths, A.M., Veríssimo, A., 2019. Population genomics and phylogeography of a benthic coastal shark (*Scyliorhinus canicula*) using 2b-RAD single nucleotide polymorphisms. Biol. J. Linn. Soc. 126, 289–303. https://doi.org/10.1093/biolinnean/bly185

McLeod, I.M., McCormick, M.I., Munday, P.L., Clark, T.D., Wenger, A.S., Brooker, R.M., Takahashi, M., Jones, G.P., 2015. Latitudinal variation in larval development of coral reef fishes: Implications of a warming ocean. Mar. Ecol. Prog. Ser. 521, 129–141. https://doi.org/10.3354/meps11136

Meekan, M.G., Vigliola, L., Hansen, A., Doherty, P.J., Halford, A., Carleton, J.H., 2006. Bigger is better: Size-selective mortality throughout the life history of a fast-growing clupeid, Spratelloides gracilis. Mar. Ecol. Prog. Ser. 317, 237–244. https://doi.org/10.3354/meps317237

Millot, C., 1999. Circulation in the Western Mediterranean Sea. J. Mar. Syst. 20, 423–442. https://doi.org/10.1016/S0924-7963(98)00078-5

Millot, C., Taupier-Letage, I., 2005. Circulation in the Mediterranean Sea, in: Saliot, A. (Ed.), The Mediterranean Sea. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 29–66. https://doi.org/10.1007/b107143

Moginie, B.F., Shima, J.S., 2018. Hatch date and growth rate drives reproductive success in nest-guarding males of a temperate reef fish. Mar. Ecol. Prog. Ser. 592, 197–206. https://doi.org/10.3354/meps12506

Mouritsen, H., Atema, J., Kingsford, M.J., Gerlach, G., 2013. Sun Compass Orientation Helps Coral Reef Fish Larvae Return to Their Natal Reef. PLoS One 8. https://doi.org/10.1371/journal.pone.0066039

Myers, N., Mittermeler, R.A., Mittermeler, C.G., Da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858. https://doi.org/10.1038/35002501

Nanninga, G.B., Manica, A., 2018. Larval swimming capacities affect genetic differentiation and range size in demersal marine fishes. Mar. Ecol. Prog. Ser. 589, 1–12.

O'Connor, J., Muheim, R., 2017. Pre-settlement coral-reef fish larvae respond to magnetic field changes during the day. J. Exp. Biol. 220, 2874–2877. https://doi.org/10.1242/jeb.159491

O'Connor, M.I., Bruno, J.F., Gaines, S.D., Halpern, B.S., Lester, S.E., Kinlan, B.P., Weiss, J.M., 2007. Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation. Proc. Natl. Acad. Sci. U. S. A. 104, 1266–1271. https://doi.org/10.1073/pnas.0603422104

Ospina-Alvarez, A., Catalán, I.A., Bernal, M., Roos, D., Palomera, I., 2015. From egg production to recruits: Connectivity and inter-annual variability in the recruitment patterns of European anchovy in the northwestern Mediterranean. Prog. Oceanogr. 138, 431–447. https://doi.org/10.1016/j.pocean.2015.01.011

Paris, C.B., Atema, J., Irisson, J.O., Kingsford, M., Gerlach, G., Guigand, C.M., 2013. Reef Odor: A Wake Up Call for Navigation in Reef Fish Larvae. PLoS One 8, e72808. https://doi.org/10.1371/journal.pone.0072808

Pascual, M., Palero, F., García-Merchán, V.H., Macpherson, E., Robainas-Barcia, A., Mestres, F., Roda, T., Abelló, P., 2016. Temporal and spatial genetic differentiation in the crab *Liocarcinus depurator* across the Atlantic-Mediterranean transition. Sci. Rep. 6, 1–10. https://doi.org/10.1038/srep29892

Pascual, M., Rives, B., Schunter, C., Macpherson, E., 2017a. Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. PLoS One 12, 1–20. https://doi.org/10.1371/journal.pone.0176419

Pascual, M., Rives, B., Schunter, C., Macpherson, E., 2017b. Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. PLoS One 12, 1–20. https://doi.org/10.1371/journal.pone.0176419

Payne, M.R., Ross, S.D., Clausen, L.W., Munk, P., Mosegaard, H., Nash, R.D.M., 2013. Recruitment decline in North Sea herring is accompanied by reduced larval growth rates. Mar. Ecol. Prog. Ser. 489, 197–211. https://doi.org/10.3354/meps10392

Pérez-Portela, R., Wangensteen, O.S., Garcia-Cisneros, A., Valero-Jiménez, C., Palacín, C., Turon, X., 2019. Spatio-temporal patterns of genetic variation in Arbacia lixula, a thermophilous sea urchin in expansion in the Mediterranean. Heredity (Edinb). 122, 244–259. https://doi.org/10.1038/s41437-018-0098-6

Pinot, J.M., López-Jurado, J.L., Riera, M., 2002. The CANALES experiment (1996-1998). Interannual, seasonal, and mesoscale variability of the circulation in the Balearic Channels. Prog. Oceanogr. https://doi.org/10.1016/S0079-6611(02)00139-8

Planes, S., Jones, G.P., Thorrold, S.R., 2009. Larval dispersal connects fish populations in a network of marine protected areas. Proc. Natl. Acad. Sci. U. S. A. 106, 5693–5697. https://doi.org/10.1073/pnas.0808007106

Planes, S., Romans, P., 2004. Evidence of genetic selection for growth in new recruits of a marine fish. Mol. Ecol. 13, 2049–2060. https://doi.org/10.1111/j.1365-294X.2004.02202.x

Prado, P., Tomas, F., Alcoverro, T., Romero, J., 2007. Extensive direct measurements of Posidonia oceanica defoliation confirm the importance of herbivory in temperate seagrass meadows. Mar. Ecol. Prog. Ser. 340, 63–71. https://doi.org/10.3354/meps340063

Rankin, T.L., Sponaugle, S., 2011. Temperature influences selective mortality during the early life stages of a coral reef fish. PLoS One 6. https://doi.org/10.1371/journal.pone.0016814

Raventos, N., 2006. Nest site characteristics and nesting success of the five- spotted wrasse *Symphodus roissali* in the north-western Mediterranean Sea. J. Fish Biol. 68, 305–309. https://doi.org/10.1111/j.1095-8649.2005.00885.x

Raventos, N., Macpherson, E., 2005. Effect of pelagic larval growth and size-at-hatching on post-settlement survivorship in two temperate labrid fish of the genus *Symphodus*. Mar. Ecol. Prog. Ser. 285, 205–211. https://doi.org/10.3354/meps285205

Raventos, N., Macpherson, E., 2001. Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. Mar. Biol. 138, 1115–1120. https://doi.org/10.1007/s002270000535

Raventos, N., MacPherson, E., 2005. Environmental influences on temporal patterns of settlement in two littoral labrid fishes in the Mediterranean Sea. Estuar. Coast. Shelf Sci. 63, 479–487. https://doi.org/10.1016/j.ecss.2004.11.018

Renault, L., Oguz, T., Pascual, A., Vizoso, G., Tintore, J., 2012. Surface circulation in the Alboran Sea (western Mediterranean) inferred from remotely sensed data. J. Geophys. Res. Ocean. 117. https://doi.org/10.1029/2011JC007659

Roberts, K.E., Cook, C.N., Beher, J., Treml, E.A., 2020. Assessing the current state of ecological connectivity in a large marine protected area system. Conserv. Biol. 00, 1–12. https://doi.org/10.1111/cobi.13580

Ross, S.D., Payne, M.R., Worsøe Clausen, L., Munk, P., Mosegaard, H., Nash, R.D.M., 2012. Coupling otolith microstructure analysis and hydrographic backtracking suggests a mechanism for the 2000s North Sea herring recruitment failure. ICES C. 2012/J 14, 33.

Rossi, V., Ser-Giacomi, E., Lõpez, C., Hernández-García, E., 2014. Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves. Geophys. Res. Lett. 41, 2883–2891. https://doi.org/10.1002/2014GL059540

Russ, G.R., Alcala, A.C., 2004. Marine reserves: Long-term protection is required for full recovery of predatory fish populations. Oecologia 138, 622–627. https://doi.org/10.1007/s00442-003-1456-4

Sabatés, A., Olivar, M.P., 1996. Variation of larval fish distributions associated with variability in the location of a shelf-slope front. Mar. Ecol. Prog. Ser. 135, 11–20. https://doi.org/10.3354/meps135011

Sabates, A., Zabala, M., García-Rubies, A., 2003. Larval fish communities in the Medes Islands Marine Reserve (North-west Mediterranean). J. Plankton Res. 25, 1035–1046. https://doi.org/10.1093/plankt/25.9.1035

Sala, E., Ballesteros, E., Dendrinos, P., Di Franco, A., Ferretti, F., Foley, D., Fraschetti, S., Friedlander, A., Garrabou, J., Güçlüsoy, H., Guidetti, P., Halpern, B.S., Hereu, B., Karamanlidis, A.A., Kizilkaya, Z., Macpherson, E., Mangialajo, L., Mariani, S.,

Micheli, F., Pais, A., Riser, K., Rosenberg, A.A., Sales, M., Selkoe, K.A., Starr, R., Tomas, F., Zabala, M., 2012. The Structure of Mediterranean Rocky Reef Ecosystems across Environmental and Human Gradients, and Conservation Implications. PLoS One 7, 1–13. https://doi.org/10.1371/journal.pone.0032742

Salat, J., 1996. Review of hydrographic environmental factors that may influence anchovy habitats in northwestern Mediterranean. Sci. Mar. 60, 21–32.

Sanford, E., Kelly, M.W., 2010. Local Adaptation in Marine Invertebrates. Ann. Rev. Mar. Sci. 3, 509–535. https://doi.org/10.1146/annurev-marine-120709-142756

Schunter, C., Carreras-Carbonell, J., Macpherson, E., Tintoré, J., Vidal-Vijande, E., Pascual, A., Guidetti, P., Pascual, M., 2011a. Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. Mol. Ecol. 20, 5167–5181. https://doi.org/10.1111/j.1365-294X.2011.05355.x

Schunter, C., Carreras-Carbonell, J., Planes, S., Sala, E., Ballesteros, E., Zabala, M., Harmelin, J.G., Harmelin-Vivien, M., Macpherson, E., Pascual, M., 2011b. Genetic connectivity patterns in an endangered species: The dusky grouper (Epinephelus marginatus). J. Exp. Mar. Bio. Ecol. 401, 126–133. https://doi.org/10.1016/j.jembe.2011.01.021

Schunter, C., Pascual, M., Garza, J.C., Raventos, N., Macpherson, E., 2014. Kinship analyses identify fish dispersal events on a temperate coastline. Proc. R. Soc. B Biol. Sci. 281. https://doi.org/10.1098/rspb.2014.0556

Schweizer, R.M., VonHoldt, B.M., Harrigan, R., Knowles, J.C., Musiani, M., Coltman, D., Novembre, J., Wayne, R.K., 2016. Genetic subdivision and candidate genes under selection in North American grey wolves. Mol. Ecol. 25, 380–402. https://doi.org/10.1111/mec.13364

Searcy, S.P., Sponaugle, S., 2001. Selective mortality during the larval-juvenile transition in two coral reef fishes. Ecology 82, 2452–2470. https://doi.org/10.1890/0012-9658(2001)082[2452:SMDTLJ]2.0.CO;2

Seetharam, A.S., Stuart, G.W., 2013. Whole genome phylogeny for 21 *Drosophila* species using predicted 2b-RAD fragments. PeerJ 2013, 1–15. https://doi.org/10.7717/peerj.226

Selkoe, K.A., Toonen, R.J., 2011. Marine connectivity: A new look at pelagic larval duration and genetic metrics of dispersal. Mar. Ecol. Prog. Ser. 436, 291–305. https://doi.org/10.3354/meps09238

Shima, J.S., Swearer, S.E., 2016. Evidence and population consequences of shared larval dispersal histories in a marine fish. Ecology 97, 25–31. https://doi.org/10.1890/14-2298.1

Siegel, D.A., Kinlan, B.P., Gaylord, B., Gaines, S.D., 2003. Lagrangian descriptions of marine larval dispersion. Mar. Ecol. Prog. Ser. 260, 83–96. https://doi.org/10.3354/meps260083

Sponaugle, S., Boulay, J., Rankin, T., 2011. Growth- and size-selective mortality in pelagic -larvae of a common reef fish. Aquat. Biol. 13, 263–273. https://doi.org/10.3354/ab00370

Sponaugle, S., Cowen, R.K., 1996. Larval supply and patterns of recruitment for two caribbean reef fishes, Stegastes partitus and Acanthurus bahianus, in: Marine and Freshwater Research. CSIRO, pp. 433–447. https://doi.org/10.1071/MF9960433

Sponaugle, S., Grorud-Colvert, K., Pinkard, D., 2006. Temperature-mediated variation

in early life history traits and recruitment success of the coral reef fish Thalassoma bifasciatum in the Florida Keys. Mar. Ecol. Prog. Ser. 308, 1–15. https://doi.org/10.3354/meps308001

Staaterman, E., Paris, C.B., Helgers, J., 2012. Orientation behavior in fish larvae: A missing piece to Hjort's critical period hypothesis. J. Theor. Biol. 304, 188–196. https://doi.org/10.1016/j.jtbi.2012.03.016

Stapley, J., Reger, J., Feulner, P.G.D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A.D., Beckerman, A.P., Slate, J., 2010. Adaptation genomics: The next generation. Trends Ecol. Evol. https://doi.org/10.1016/j.tree.2010.09.002

Swearer, S.E., Shima, J.S., Hellberg, M.E., Thorrold, S.R., Jones, G.P., Robertson, D.R., Morgan, S.G., Selkoe, K.A., Ruiz, G.M., Warner, R.R., 2002. Evidence of self-recruitment in demersal marine populations. Bull. Mar. Sci. 70, 251–271.

Takasuka, A., Aoki, I., Mitani, I., 2003. Evidence of growth-selective predation on larval Japanese anchovy Engraulis japonicus in Sagami Bay. Mar. Ecol. Prog. Ser. 252, 223–238. https://doi.org/10.3354/meps252223

Tintore, J., La Violette, P.E., Blade, I., Cruzado, A., 1988. A Study of an Intense Density Front in the Eastern Alboran Sea: The Almeria–Oran Front. J. Phys. Oceanogr. 18, 1384–1397. https://doi.org/10.1175/1520-0485(1988)018<1384:asoaid>2.0.co;2

Torrado, H., Carreras, C., Raventos, N., Macpherson, E., Pascual, M., 2020. Individual-based population genomics reveal different drivers of adaptation in sympatric fish. Sci. Rep. https://doi.org/10.1038/s41598-020-69160-2

Treml, E.A., Halpin, P.N., 2012. Marine population connectivity identifies ecological neighbors for conservation planning in the Coral Triangle. Conserv. Lett. 5, 441–449. https://doi.org/10.1111/j.1755-263X.2012.00260.x

Treml, E.A., Halpin, P.N., Urban, D.L., Pratson, L.F., 2008. Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. Landsc. Ecol. 23, 19–36. https://doi.org/10.1007/s10980-007-9138-y

Treml, E.A., Roberts, J.J., Chao, Y., Halpin, P.N., Possingham, H.P., Riginos, C., 2012. Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations, in: Integrative and Comparative Biology. Narnia, pp. 525–537. https://doi.org/10.1093/icb/ics101

Urban, D.L., Minor, E.S., Treml, E.A., Schick, R.S., 2009. Graph models of habitat mosaics. Ecol. Lett. 12, 260–273. https://doi.org/10.1111/j.1461-0248.2008.01271.x

Vigliola, L., Doherty, P.J., Meekan, M.G., Drown, D.M., Jones, M.E., Barber, P.H., 2007. Genetic identity determines risk of post-settlement mortality of a marine fish. Ecology 88, 1263–1277. https://doi.org/10.1890/06-0066

Wang, S., Meyer, E., Mckay, J.K., Matz, M. V., 2012. 2b-RAD: A simple and flexible method for genome-wide genotyping. Nat. Methods 9, 808–810. https://doi.org/10.1038/nmeth.2023

Warner, R.R., Lejeune, P., 1985. Sex change limited by paternal care: a test using four Mediterranean labrid fishes, genus *Symphodus*. Mar. Biol. 87, 89–99.

Whitehead, P.J.P., Bauchot, M.-L., Hureau, J.-C., Nielsen, J., Tortonese, E., 1984. Fishes of the North-Eastern Atlantic and the Mediterranean. Vol. 2, Copeia. Unesco. https://doi.org/10.2307/1444931

Wilson, McCormick, M.I., 1999. Microstructure of settlement-marks in the otoliths of tropical reef fishes. Mar. Biol. 134, 29–41. https://doi.org/10.1007/s002270050522

Xuereb, A., Kimber, C.M., Curtis, J.M.R., Bernatchez, L., Fortin, M.J., 2018. Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. Mol. Ecol. 27, 5035–5048. https://doi.org/10.1111/mec.14942

Yoklavich, M., Bailey, K., 1990. Hatching period, growth and survival of young walleye pollock Theragra chalcogramma as determined from otolith analysis. Mar. Ecol. Prog. Ser. 64, 13–23. https://doi.org/10.3354/meps064013

Zhong, X., Wang, X., Zhou, T., Jin, Y., Tan, S., Jiang, C., Geng, X., Li, N., Shi, H., Zeng, Q., Yang, Y., Yuan, Z., Bao, L., Liu, S., Tian, C., Peatman, E., Li, Q., Liu, Z., 2017. Genome-Wide Association Study Reveals Multiple Novel QTL Associated with Low Oxygen Tolerance in Hybrid Catfish. Mar. Biotechnol. 19, 379–390. https://doi.org/10.1007/s10126-017-9757-5

Apéndice: Publicaciones

SCIENTIFIC REPORTS
nature research

Check for updates

OPEN

Individual-based population genomics reveal different drivers of adaptation in sympatric fish

Héctor Torrado[1,2]✉, Carlos Carreras[2], Núria Raventos[1], Enrique Macpherson[1,3] & Marta Pascual[2,3]

Connectivity and local adaptation are two contrasting evolutionary forces highly influencing population structure. To evaluate the impact of early-life traits and environmental conditions on genetic structuring and adaptation, we studied two sympatric fish species in the Western Mediterranean Sea: *Symphodus tinca* and *S. ocellatus*. We followed an individual-based approach and measured early-life history traits from otolith readings, gathered information on environmental variables and obtained genome-wide markers from genotyping-by-sequencing (GBS). The two species presented contrasting population structure across the same geographic gradient, with high and significant population differentiation in *S. ocellatus*, mostly determined by oceanographic fronts, and low differentiation and no front effect in *S. tinca*. Despite their different levels of genetic differentiation, we identified in both species candidate regions for local adaptation by combining outlier analysis with environmental and phenotypic association analyses. Most candidate loci were associated to temperature and productivity in *S. ocellatus* and to temperature and turbulence in *S. tinca* suggesting that different drivers may determine genomic diversity and differentiation in each species. Globally, our study highlights that individual-based approach combining genomic, environmental and phenotypic information is key to identify signals of selection and the processes mediating them.

Population structure is highly influenced by the level of connectivity among localities, driven by dispersal potential and barriers to gene flow, but also shaped by local adaptation. In this context, environmental conditions can impose spatially differential genomic selective pressures on phenotypic traits and their combined study can allow to forecast the species adaptive capacity[1]. The use of environmental information coupled with the genetic structure of species can be essential for understanding how they are adapted to their habitats[2]. Some studies have investigated the relationship between local adaptation and environmental conditions in several marine species with temperature, salinity and productivity values being important factors affecting different biological processes and genetic differentiation at the population level[3–6]. For instance, thermal gradients observed at sea can explain the distributional patterns of numerous species, their diversification processes and the potential influence of climate changes[7].

Population genomic studies are fundamental to identify patterns of genetic structure and the processes determining them. Genotyping-by-sequencing (GBS)[8], provides a cost-effective approach for population genomic studies on non-model species. This methodology provides high density of markers that may allow identifying non-neutral genomic signatures related to adaptation processes, with the potential for annotation[6,9]. The assessment of genomic signals of adaptation on non-model organisms is usually performed in two different ways, outlier analyses (OAs) and genotype-environment association analyses (EAAs)[2]. On one hand, OAs are based on the search of genome-wide markers showing outlier values of genetic differentiation (F_{ST}) among the studied populations[10]. On the other hand, EAAs using population genomic data combined with environmental data have been successfully used to discover polymorphisms involved in adaptation to environmental conditions[3,6]. However, individual-based information could be used to further refine those associations. Individual-based

[1]Centre D'Estudis Avançats de Blanes (CEAB-CSIC), Car. Acc. Cala St. Francesc 14, 17300 Blanes, Girona, Spain. [2]Department de Genètica, Microbiologia i Estadística and IRBio, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain. [3]These authors contributed equally: Enrique Macpherson and Marta Pascual. ✉email: h.torrado@hotmail.com

information is used in genome-wide association studies (GWAS), which can identify signals of adaptation to both environmental conditions (EAAs) and phenotypic traits through phenotype association analysis (PAAs). This methodology has proven to be a powerful tool to study interesting phenotypic traits and adaptation in model species such as humans, plants and captive fish[11–13]. While the use of individual-based information is common in model species, the application in wild populations is rare and has been only used in a few studies combined with environmental data information[14]. Up to now, there is a lack of individual-based studies in natural marine populations, preventing an accurate knowledge of the relationships between genomic data, environment and early life phenotypes in determining population structure in environmental gradients. This approach can be undertaken in species were individual based information during dispersal phases can be obtained, such as benthic fishes.

In numerous Mediterranean benthic fishes, the pelagic larval stage allows dispersal and maintains connectivity between populations, whereas adults are territorial and present a high site fidelity[15–18]. The length of the pelagic larval duration (PLD) can not only affect dispersal but also other fitness characteristics such as larval growth, size at settlement and mortality rate, modulating the year-class strength in numerous species[19,20]. Moreover, environmental conditions can have major effects during the larval and settlement periods, where genetic and phenotypic polymorphism can be affected by selection and/or stochastic factors[15]. Fortunately, in most fishes, individual-based information on both phenotypic and environmental early-life variables can be obtained from otoliths (bones in the inner part of the ear) and has been successfully combined with genomic data to infer fine-scale dispersal[21]. Therefore, the study of individually based genomic variation associated to early-life history traits across geographic gradients in sympatric fish can provide insights into the adaptive potential of species while considering both connectivity and selection.

Here we use an individual-based approach to evaluate the phenotypic and environmental drivers shaping population genomic structure in two congeneric fish, *Symphodus tinca* and *S. ocellatus,* that are sympatric across a geographic/environmental gradient in the Western Mediterranean Sea. We selected two closely related species to ensure that any difference found between the two species on the drivers shaping population structure is not the result of a phylogenetic signal of divergent lineages, but a species-specific differential adaptation signal. On the other hand, adaptation signals on marine organisms may change depending on the environmental gradient present on the sampling sites and the scale of analysis[6]. For this reason, a sympatric distribution is key to check for a differential adaptation strategy, as both species are going to be under the same environmental gradient and face the same oceanographic fronts. We analysed 303 individuals from these two species with genotyping by sequencing, measured from otolith readings some individualized key early-life history traits (e.g. date of hatching, pelagic larval duration, size at settlement, growth rate) and gathered information of environmental variables during their pelagic larval duration (e.g. surface temperature, turbulence, and productivity). In order to test if phenotypic and environmental variables differentially influence the population structure in these two species, we (1) gathered individual-based information on early-life phenotypic and environmental variables, (2) assessed population differentiation, (3) identified loci putatively under selection using different methodological approaches and (4) evaluated the effect of phenotypic and environmental factors in structuring populations across geographic gradients. We expected to find similar associations and genomic structuring for both species due to their similar distribution and phylogenetic closeness. Nevertheless, considering their differences in reproduction times, temporal environmental variation and behaviour could change population connectivity and the drivers of population structure.

Results

Phenotypic and environmental analyses. We obtained reliable information on early-life history traits from otolith reading (Figure S1) for 93% and 89% of the juvenile individuals from most of the localities in *S. ocellatus* and *S. tinca*, respectively (Fig. 1, Table 1). With these data we acquired environmental and phenotypic individually based values (Table 2). Significant differences among sampling locations were detected with PERMANOVA for all phenotypic and environmental variables with three exceptions: PLD growth rate and Hatching date in *Symphodus ocellatus* and PLD in *S. tinca* (Table S1). The temperature during the larval stage was lower in northern localities for both species, even considering a delay on the hatching date in some of these localities to encounter warmer dates (Table 2). For *S. ocellatus,* individuals from the colder areas (Fig. 1), the Alboran Sea and the two northernmost localities, delayed hatching date and thus experienced milder temperatures during PLD than otherwise expected (Table 2). In the case of *S. tinca*, the individuals of colder localities (Fig. 1) did not always delay their hatching date. For instance, in Blanes, individuals hatched early in the season and thus experienced low temperatures and had longer PLD (Table 2). The information of multiple variables was synthesized with principal component analysis (PCA). The PCA yielded several components for the phenotypic (PC) and environmental (EC) variables, with different contribution from each variable, and overall explaining a large proportion of the variance (Table S2).

Population genomics. We analysed 303 individuals from the two sympatric *Symphodus* species with GBS (Table 1, Fig. 1). We obtained a mean (±SD) of 2.6 ± 0.8 and 2.6 ± 0.9 million reads per individual for *S. ocellatus* and *S. tinca,* respectively. After filtering, we found 3978 polymorphic haplotype loci and 5123 SNP loci for *S. ocellatus* and 5276 polymorphic haplotype loci and 6833 SNP loci for *S. tinca*. The mean depth per locus was 19.5 ± 11.4 for *S. ocellatus* and 19.5 ± 11.3 for *S. tinca*. The maximum number of alleles per haplotype locus was 17 for *S. ocellatus* and 15 for *S. tinca*. Observed and expected heterozygosities per locality were similar in the two species (Table 1).

For *S. ocellatus*, pairwise F_{ST} values (Table S3) as well as the DAPC plot (Fig. 2) and k-means clustering suggest the same structure identifying three groups. One group was formed by the northern localities (Colera – Blanes), another included the localities between the IC and AOF oceanographic discontinuities (Xabia, Palos

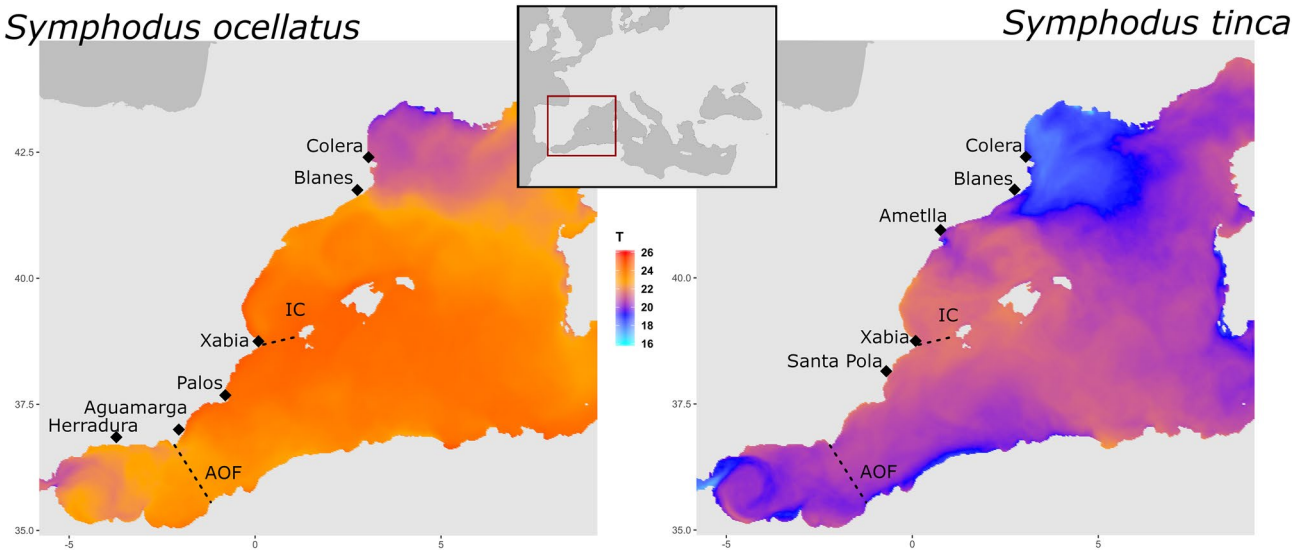


Figure 1. Map of the sampled locations in the Western Mediterranean for *Symphodus ocellatus* and *S. tinca*. Dashed lines indicate oceanographic discontinuities: Ibiza channel (IC) and Almeria-Oran front (AOF). The colour gradient of the sea identifies the mean sea surface temperature (°C) during the larval period of each species. The reduced grey map shows the location of our study area within the Mediterranean Sea. Data obtained from Western Mediterranean OPerational forecasting system (WMOP)[63].Maps were plotted using the R[60] packages 'ncdf4'[61] and 'ggplot2'[62].

and Aguamarga) as a central group, and the last one was constituted by the southernmost locality (Herradura) sampled in the Alboran Sea (Fig. 1). It is worthwhile noticing that Herradura clusters with the localities of the central group in the first axis. However, we did not find isolation by distance with a Mantel test (r = 0.45, *p* value = 0.08). For *S. tinca* no clear structuring was detected (Fig. 2, Table S4), and only two significant F_{ST} values were found, when Santa Pola was compared to Blanes and to Ametlla. When comparing only the three localities sampled in both species, we observed similar groupings in the DAPC analysis as when considering all sampled localities, with a clear effect of the IC discontinuity genetically differentiating Xabia from the northern localities in *S. ocellatus* but not in *S. tinca* (Fig. 2). As for *S. ocellatus* we did not find isolation by distance with a Mantel test (r = −0.38, *p* value = 1).

In the outlier analysis with Bayescan after correcting for multiple comparisons, we identified 95 outlier haplotype loci potentially under positive selection for *S. ocellatus* and 5 for *S. tinca* (Tables S5 and S6).

Genomic associations to phenotypic and environmental variables. The variance partition analysis using the SNP loci datafile showed a small contribution of environmental and phenotypic variables in explaining the whole genomic variance in both species, although higher for environmental variables (Table S7). The redundancy analysis (RDA) for *S. ocellatus*, using the individual environmental values and the genotypes from the SNP loci datafile, identified two clusters of individuals along the first axis, explaining 58.0% of the variation (Fig. 3). All individuals of the central group (Xabia, Palos and Aguamarga) clustered together and were positively associated with temperature and negatively associated with productivity while those from the other three localities (Colera, Blanes and Herradura) clustered also together but presented the opposite trend. Overall, 155 haplotype loci had SNPs significantly correlated with environmental predictors (Table S8) with only one locus (20878) being associated to more than one predictor, although with different SNPs (Table S5).

For *S. tinca*, the two axes of the RDA explained 67.4% of the variation and separated the individuals by locality (Fig. 3). The first axis was associated mainly to productivity while the second axis was associated to temperature and turbulence with a different sign. Consequently, Ametlla and Xabia were associated to higher temperature and low turbulence, while Xabia and Blanes seemed to have less productivity. Overall, 166 haplotype loci had SNPs significantly correlated with environmental predictors (Table S8) with only three haplotype loci (12921, 24436, 9609) associated to more than one predictor, although in all cases with different SNPs (Table S6) as in *S. ocellatus*.

The GWAS using as covariates the two first components of the MDS analyses (GWAS-Cov) identified for *S. ocellatus* only one SNP associated to the first environmental component (Table S8). However, when performing the analysis without covariates 207 haplotype loci presented SNPs significantly associated to environmental and phenotypic variables. Most of the SNPs were associated to environmental variables, particularly productivity and temperature and 63 of them were related to more than one variable (Table S5). For phenotype-genotype associations, 21 SNPs were associated to the composite phenotypic variable PC2, positively related to PLD and settlement size (Table S2). In the case of *S. tinca*, GWAS-Cov found six SNPs with associations to phenotypic and environmental variables (Table S8). The same SNPs were associated to the same variables when covariates were not used in the GWAS, with the exception of SNP 12307_17 where the association to EC3 was not detected when using covariates (Table S6).

Species	Location	Sampling date	N	Ar	Ho	He	F_{IS}
Symphodus ocellatus	Colera 42°24′26.9″N 3°09′23.8″E	June–August 2014	30 (30)	2.283	0.302	0.362	0.164
	Blanes 41°40′44.5″N 2°48′28.8″E	June–August 2014	37 (37)	2.284	0.296	0.360	0.177
	Xabia 38°47′46.1″N 0°11′01.4″E	May–July 2014	28 (26)	2.255	0.281	0.354	0.207
	Palos (Cabo de Palos) 37°38′05.4″N 0°41′31.6″W	May–August 2014	26 (24)	2.229	0.285	0.350	0.185
	Aguamarga 36°56′20.8″N 1°55′53.1″W	May–August 2014	28 (27)	2.242	0.296	0.352	0.158
	Herradura (La Herradura) 36°43′40.0″N 3°44′07.9″W	July–August 2016	13 (6)	2.224	0.284	0.352	0.193
Symphodus tinca	Colera 42°24′26.9″N 3°09′23.8″E	May–June 2015	30 (24)	2.600	0.298	0.356	0.164
	Blanes 41°40′44.5″N 2°48′28.8″E	May 2014	31 (27)	2.588	0.293	0.355	0.175
	Ametlla (L'Ametlla de Mar) 40°52′27.1″N 0°47′42.3″E	May–June 2015	26 (23)	2.597	0.301	0.356	0.155
	Xabia 38°47′46.1″N 0°11′01.4″E	May–June 2015	28 (28)	2.601	0.293	0.356	0.178
	Santa Pola 38°11′27.9″N 0°33′57.2″W	2015	26 (0)	2.592	0.293	0.355	0.173

Table 1. Sampling information and population diversity values for the locations of *Symphodus ocellatus* and *S. tinca*. Sampling date, Number of genotyped individuals (N) and in parentheses the number of juvenile individuals with otoliths information, Allelic richness (Ar), Mean observed (H_o) and expected (H_e) heterozygosities, Inbreeding coefficient (F_{IS}). All F_{IS} values are significant. The names of localities in parentheses indicate the toponymal names of the corresponding localities. Diversity values were obtained from the haplotype loci datafiles.

Species	Location	Planktonic Larval duration (PLD) (days)	Settlement size (µm)	PLD Growth rate (µm/day)	Hatching date (day of the year)	Surface temperature (°C)	Productivity (ChlA mg/m^3)	Turbulence (m)
Symphodus ocellatus	Colera	10.1 ± 1.0	45.4 ± 4.6	4.5 ± 0.6	192.8 ± 10.8	21.8 ± 0.4	0.211 ± 0.005	1.40 ± 0.24
	Blanes	9.3 ± 1.0	48.7 ± 6.2	5.3 ± 0.6	193.4 ± 13.5	22.5 ± 0.9	0.164 ± 0.009	0.67 ± 0.18
	Xabia	9.3 ± 1.2	43.3 ± 7.1	4.7 ± 0.6	186.0 ± 15.8	24.1 ± 1.2	0.171 ± 0.004	0.87 ± 0.08
	Palos	8.6 ± 0.8	42.0 ± 3.8	4.9 ± 0.6	184.2 ± 10.2	23.8 ± 0.9	0.163 ± 0.016	0.97 ± 0.08
	Aguamarga	9.3 ± 1.3	43.8 ± 4.8	4.8 ± 0.6	188.6 ± 19.0	23.1 ± 1.9	0.158 ± 0.003	0.83 ± 0.11
	Herradura	9.3 ± 0.8	46.0 ± 5.6	5.0 ± 0.6	229.7 ± 11.7	23.9 ± 0.3	0.261 ± 0.030	0.90 ± 0.22
Symphodus tinca	Colera	9.3 ± 0.9	33.0 ± 2.4	3.6 ± 0.4	159.9 ± 4.9	20.1 ± 0.9	0.306 ± 0.039	1.02 ± 0.23
	Blanes	12.4 ± 1.4	40.9 ± 4.0	3.3 ± 0.2	135.7 ± 5.7	16.5 ± 0.1	0.335 ± 0.011	0.72 ± 0.11
	Ametlla	9.0 ± 1.0	35.8 ± 4.0	4.0 ± 0.4	157.0 ± 5.2	22.0 ± 1.3	0.333 ± 0.066	0.52 ± 0.14
	Xabia	9.9 ± 1.1	42.6 ± 4.6	4.3 ± 0.4	149.0 ± 3.6	21.0 ± 0.6	0.201 ± 0.019	0.59 ± 0.10

Table 2. Mean values and standard deviation of individually based environmental and phenotypic variable values by location. No environmental and phenotypic variable is available for Santa Pola in *Symphodus tinca*, since juveniles with otolith information were not sampled in this locality.

Combining all methodologies (RDA, GWAS-Cov, GWAS and Outliers), a total of 292 and 168 haplotype loci were identified as potentially candidates to selection in *S. ocellatus* and *S. tinca*, respectively (Fig. 4). A low number of loci were shared by all methodologies being larger in *S. ocellatus* which presented higher population differentiation. The percentage of loci with SNPs associated to environmental predictors differed between the two species. Similar numbers of loci were associated to each environmental variable for *S. tinca*, although

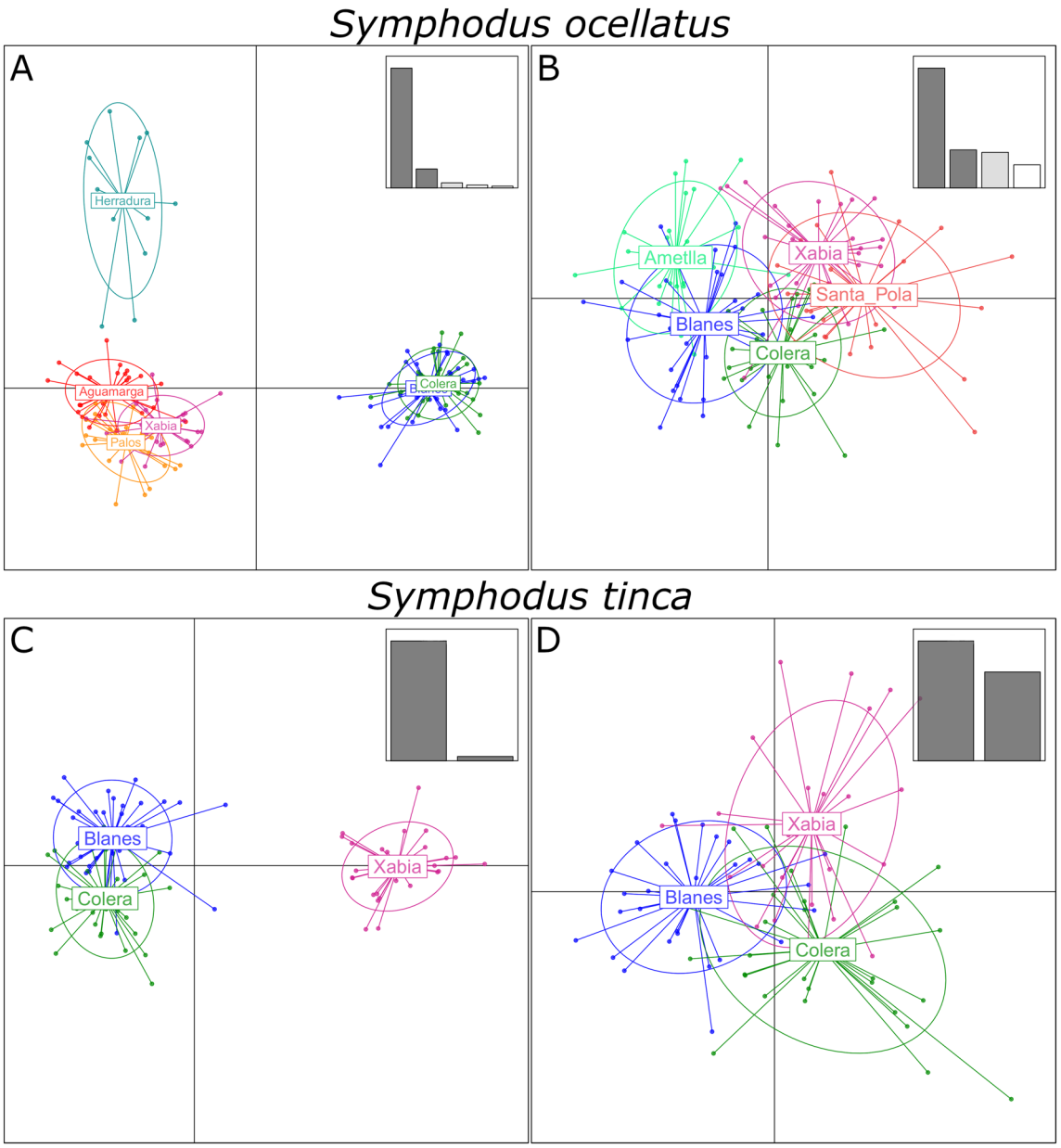


Figure 2. Discriminant Analysis of Principal Components (DAPC). DAPC plots with DA eigenvalues (top right corner) using all analysed localities for each species (**A**,**B**) or only the localities sampled in both species (**C**,**D**).

slightly larger to turbulence (Figure S2). However, for *S. ocellatus* most loci were associated to temperature and productivity. We used all candidate loci to cluster localities based on their observed heterozygosities and the frequency of their major allele from the haplotype loci dataset. In *S. ocellatus* both heatmaps clustered Colera and Blanes, the two northernmost localities, with Herradura, the southernmost locality (Fig. 5). Similar groupings were also observed when only loci related to temperature and productivity were used but differed for loci associated to turbulence (Figure S3). In *S. tinca* different population clustering was observed depending on the variable used and the SNPs associated to different environmental predictors (Fig. 5, Figure S4). Nonetheless, in both species, contrasting frequencies of heterozygotes and major alleles were found among localities (Fig. 5).

The Blast of all loci potentially under selection against the Ensemble database provided 172 hits for *S. ocellatus*, 93 of them inside genes (Table S5) and for *S. tinca* 103 loci had a blast hit, 43 of them inside genes (Table S6). The compilation of the biological functions using the GO terms in REVIGO showed that the genes identified in *S. ocellatus* and *S. tinca* had different functions (Fig. 6). On one hand, *S. ocellatus* genes were associated to embryonic caudal fin morphogenesis, receptor localisation to synapse, response to stimuli, protein stability, autophagosome assembly and cell adhesion (Fig. 6). On the other hand, *S. tinca* genes were associated to intracellular protein transport, heart contraction, thigmotaxis and swimming behaviour, regulation of GTPase activity, dephosphorylation, nucleobase catabolism and cytoskeleton organisation (Fig. 6). Furthermore, within

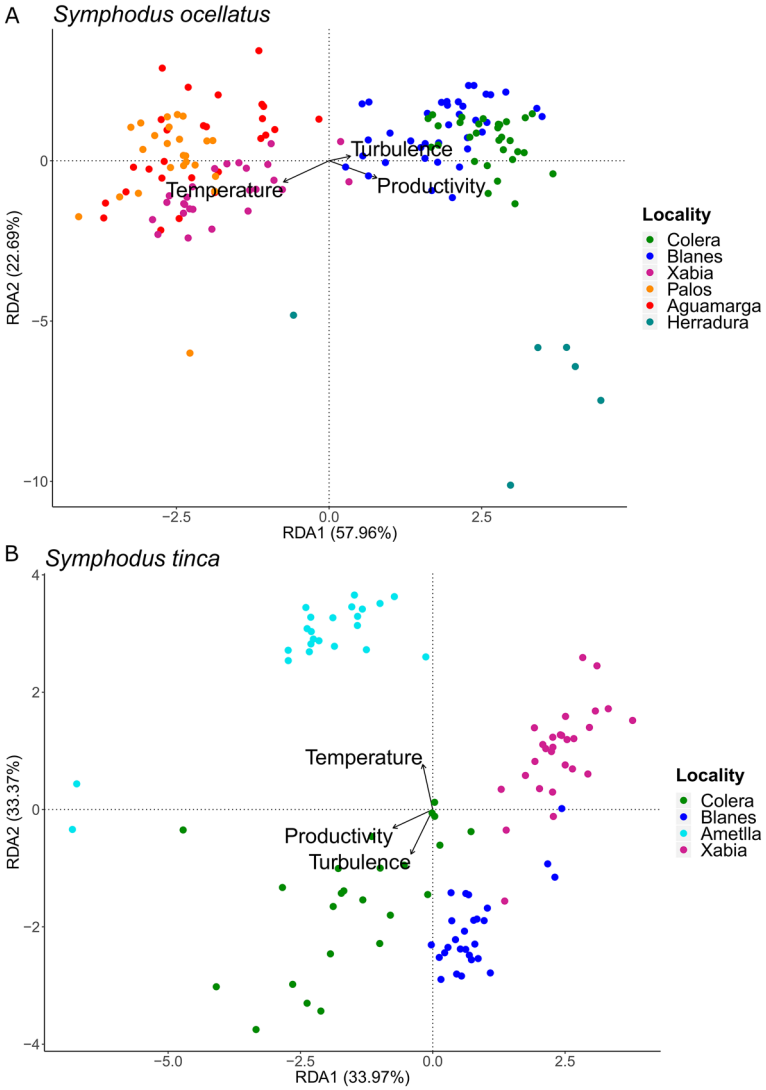


Figure 3. Redundancy analysis (RDA). RDA performed with individual environmental data and SNPs at all loci. Coloured points represent individual fish coded by locality. Vectors indicate environmental predictors of the two first RDA components. (**A**) *Symphodus ocellatus* and (**B**) *S. tinca*.

each species the genes associated to each of the environmental predictors showed different biological functions (Figure S5 and Figure S6).

Discussion

The interaction of different evolutionary forces, such as dispersal and selection, defines the actual distribution of the genomic diversity within species[4,22]. We suggest that differences in environmental and phenotypic variables at early-life stages can produce contrasting genomic response that can be revealed by comparative studies on closely related species. Our two studied species presented very contrasting patterns of genetic structure, nonetheless the use of several methodologies for assessing signals of selection revealed multiple loci associated to environmental and phenotypic variables in both of them. Our individual level approach integrating genomic data with environmental and phenotypic information was fundamental for identifying the factors and processes affecting population structuring.

Population structure and timing of reproduction. Our genome-wide analysis of the two *Symphodus* species, *S. tinca* and *S. ocellatus*, showed different levels of connectivity across the same environmental gradient. Two contrasting patterns are found despite their potentially similar dispersal capabilities, as both species build nests with algae[23,24] and present similar early-life traits such as planktonic larval duration (PLD), size at settlement and growth rate[25]. However, they have different life span and reproductive season[26,27]. If some adults move to neighbouring areas, species with longer life span such as *S. tinca* could present more connected populations than its congener *S. ocellatus* that has a shorter life span[28]. Adults are considered to be sedentary and remain

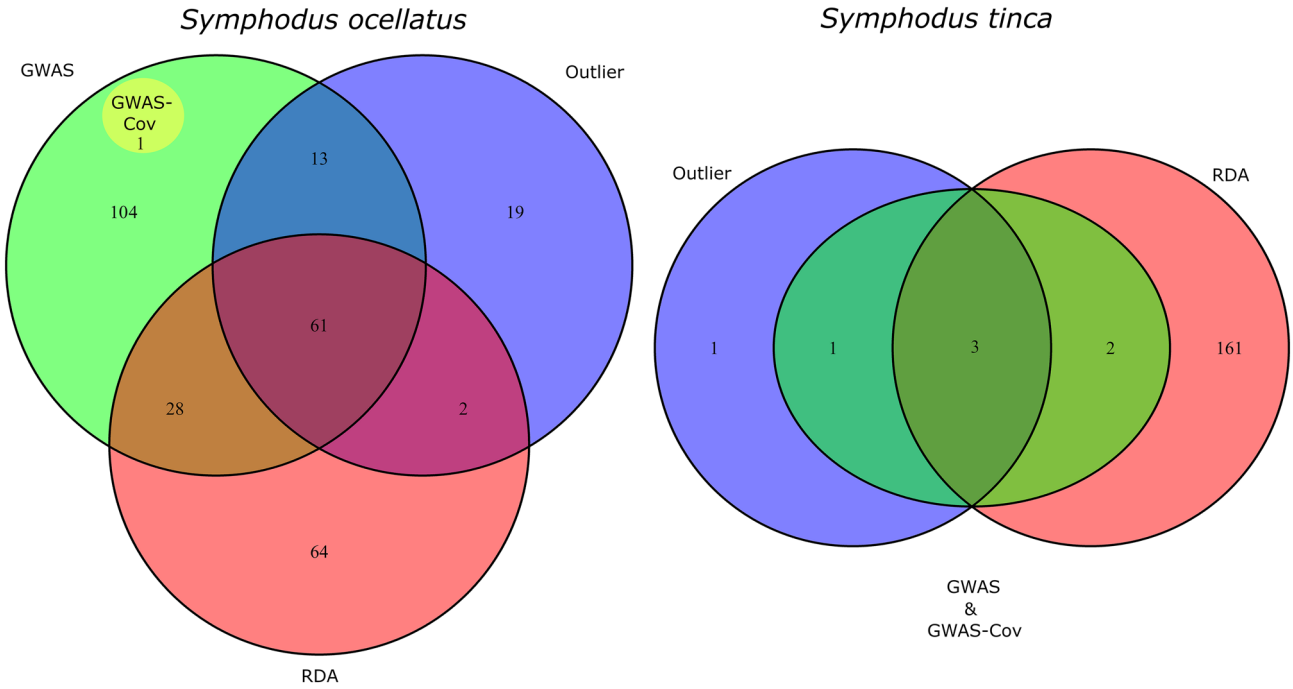


Figure 4. Venn diagram of the candidate haplotype loci showing significant values with different methodologies. Haplotype loci were considered to have a significant value for RDA and GWAS if containing SNPs with significant values for these analyses. Blue: outlier loci; Green: loci associated to environmental and/or phenotypic variables according to GWAS; Yellow: GWAS loci with the two first MDS components as covariates (GWAS-Cov); Pink: loci associated to environmental variables by RDA.

close to their settling area[26] and thus other variables such as the time of reproduction could influence population structure.

In *Symphodus ocellatus*, we found a strong population structure and a clear effect of the two main oceanographic barriers in the area, the Ibiza channel and Almeria-Oran front, genetically clustering our sampling locations in three groups. These two oceanographic barriers seem to reduce gene flow across a great variety of taxa with different dispersal capabilities[29]. On the contrary, we observed very low genetic differentiation for *S. tinca* and a null effect of the Ibiza channel, in agreement with a previous study using microsatellite loci[30]. Seasonal variation in the reproductive and larval periods could probably influence the effect of oceanographic fronts on connectivity. For instance, the Ibiza channel although presenting inter-annual variations is known to break more often in spring than in summer[31,32] thus promoting connectivity for species reproducing in this season[16,33]. In the western Mediterranean, in spring there are more extreme turbulence values and higher frequency of unpredictable energy fluxes and episodic events than in summer[34–36] which could enhance connectivity. Contrarily to the results we found in the western Mediterranean, a significant population genomic structure was unveiled for *S. tinca* in the Adriatic Sea[10]. Interestingly, during the reproductive season of *S. tinca*, lower turbulence is observed in the Adriatic Sea in comparison to the western Mediterranean[34].

Some fronts, such as the AOF, seem to reduce connectivity for many marine species across all taxa, although many other species seem not to be influenced[29,37]. This could be attributed to temporal differences. Annual variation in the current pattern across the oceanographic fronts seems to account for genetic differentiation in neighbouring locations in crabs and sea urchins[33,38,39]. Seasonal variation in the hydrodynamic conditions across the fronts[31,40] could also differentially affect population structure in different species according to their reproductive season, as found in the present study. Therefore, we hypothesize that more stable oceanographic conditions could reduce genetic connectivity, facilitate long term local adaptation and leave more detectable genomic footprints of selection. Further population genomic studies across the same barriers in a wider number of species considering the reproductive season could allow testing this hypothesis.

Genomic associations to phenotypic and environmental variables. Local adaptation might be shaping population structure along with connectivity reduction mediated by oceanographic barriers[41]. Different genotypes can be under diverse selective pressures among localities, rendering individual-based information highly valuable. The use of several methodologies was fundamental to capture the signatures of adaptation in species with contrasting degree of population structuring. We found, in both species, small contribution of our studied variables to genomic structuring with the variation partitioning of the SNPs datafile at the individual level. These results confirm that the effect of external factors in genomic structuring is small across the whole genome, but does not preclude to be important in some specific loci. For instance, in *S. ocellatus* the higher population differentiation resulted in a higher number of outlier loci and potential for detecting local adaptation in comparison to *S. tinca*. Nonetheless, for both species of *Symphodus,* the combination of the results of GWAS

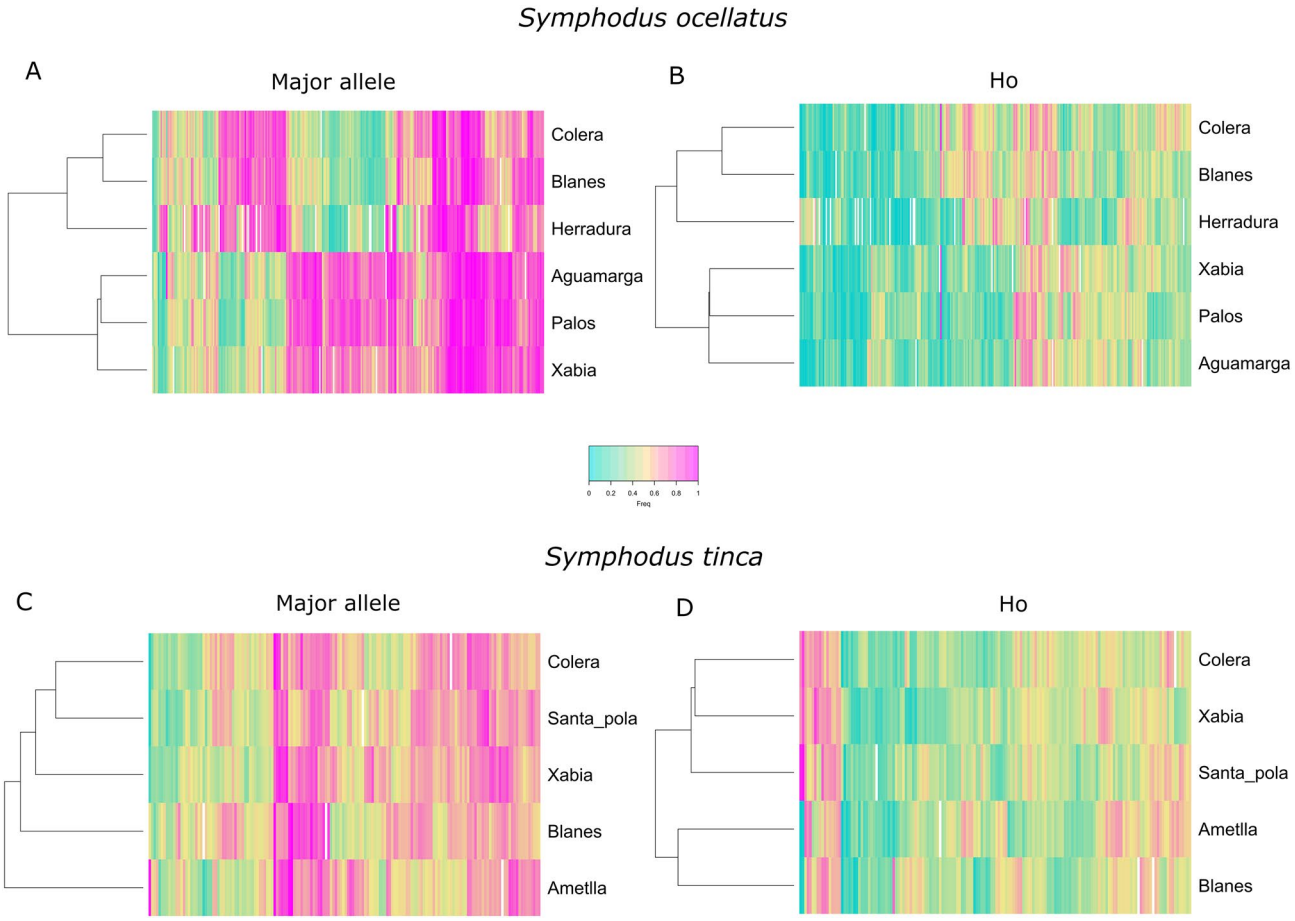


Figure 5. Haplotype loci values showing significant associations. Dendrogram and heatmap grouping the localities based on all candidate loci showing significant values with the different methodologies and using the values from the haplotype loci dataset: *Symphodus ocellatus* (**A**) frequency of the major allele and (**B**) observed heterozygosity. *Symphodus tinca* (**C**) frequency of the major allele and (**D**) observed heterozygosity.

and RDA revealed a larger number of candidate loci, putatively under selection, as well as the selection drivers of most of the outlier loci.

Redundancy analyses have been successfully used to detect signals of selection by comparing population allele frequencies to their environmental variables[4,6]. As we had individually based environmental data, we assessed how each single individual was distributed along the RDA axes based on its genotype, and according to its environmental predictors[42]. In *S. ocellatus* the individuals were separated in two groups along the first RDA axis, which was explained by surface temperature and productivity. This analysis grouped geographically separated localities, that present low temperature and high productivity, thus demonstrating that population differentiation is not only influenced by oceanographic barriers but also respond to environmental selection pressures. For, *S. tinca* individuals were grouped by locality in the RDA analysis, but their clustering was less clear, with temperature, productivity and turbulence having similar impact. Thus, despite the low number of outlier loci and population structure in that species, environmental variables can nonetheless contribute to adaptation. These apparently contradictory results rely on the power of different methodologies to detect signals of selection. The analysis of outlier loci is based on the detection of loci that show divergent patterns of genetic differentiation among localities. Thus, this analysis is highly dependent on the basal levels of population structuring as indicated in previous studies[4,42]. For instance, the abundance of outlier loci in *S. tinca* found in our study area (0.1% of all loci) was lower than that obtained for the same species in the Adriatic Sea (0.5%)[10], suggesting that a more structured region can favour the power to detect local adaptation by outlier analysis. For this reason, RDA should be also used, as it can detect signals of selection even on conditions of low structuring[42]. Moreover, this approximation can be done both at the individual and population level and can thus be implemented in species where individual based information is difficult to obtain.

The analyses of phenotypic and environmental variables associated to individual genotypes are necessary to unveil candidate regions of the genome determining these traits[3,6]. However, when there is population structure, signals of selection within them can be confounded by either demographic or selective processes among populations[43]. In both *Symphodus* species we identified only a few candidate loci associated to environmental variables by using GWAS after removing the population effect using the two first MDS axis as covariates. This methodology has been suggested to control for hierarchical structure to avoid identifying spurious associations

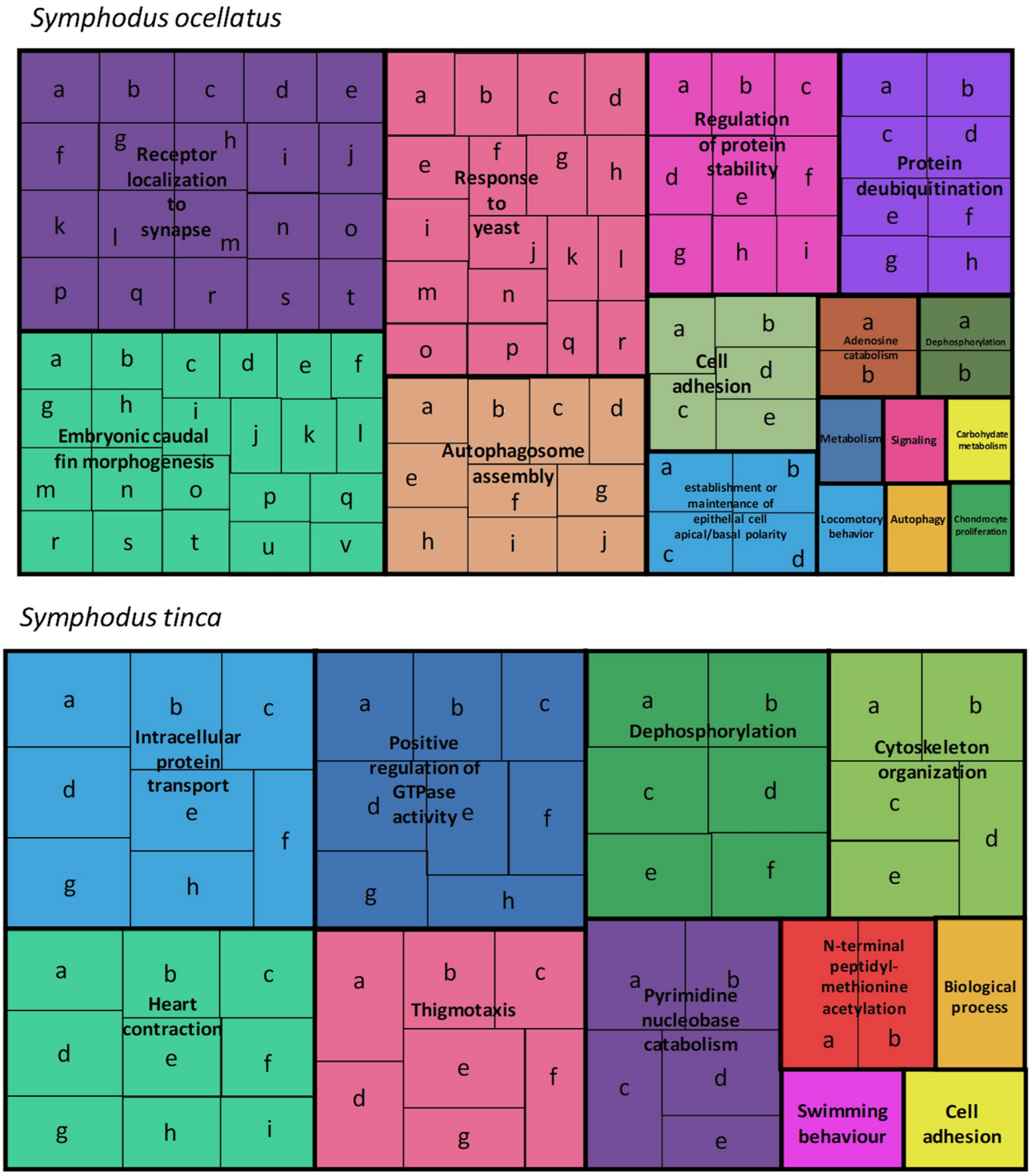


Figure 6. Gene ontology treemap of the candidate genes to selection combining all methodologies in *S. ocellatus* and *S. tinca*. The box size correlates to the –log10 p-value of the GO-term. Boxes with the same colour can be grouped together as correspond to the same upper-hierarchy GO-term which is found in the middle of each box. Description of letters and detailed information of the treemap can be found in supplementary tables S9 and S10.

due to population differentiation. However, this approach is very conservative and can underestimate phenotype-genotype associations in loci under local adaptation[4,42]. For instance, in *S. ocellatus*, the high level of population structuring undermines the power to detect selection using GWAS with covariates, while GWAS without covariates and RDA revealed a high number of candidate loci to selection. In the case of *S. tinca*, the low level of population structuring limited the resolution of the two types of GWAS while many loci under selection were

found with RDA. This highlights the need to assess selection using different methodologies as the degree of population structuring may underestimate the role of selection depending on the methodology used.

Environmental factors, such as temperature or salinity, are known to impose differential spatial signatures of selection on marine organisms[6,44,45]. The biological functions of the genes associated to the candidate loci of selection in our two *Symphodus* species suggest different mechanisms of adaptation in response to the environmental factors although both of them showed strong influence of temperature and productivity. The importance of these two variables in marine organisms is not unprecedented. Temperature is a crucial factor in many marine organisms as it may impact metabolism, larval survival and development among others[46]. Interestingly, turbulence was found to be more important in *S. tinca* than in *S. ocellatus*. Although the turbulence values can be similar during the reproduction period of both species, the frequency of storms in our study area is much higher in spring than in summer[34,47]. This temporal instability may explain the importance of this variable on the spring breeder *S. tinca* with 40% of candidate loci associated to this factor while only 11% where found to be associated in *S. ocellatus*.

In both species, we found several polymorphisms to be significantly associated with phenotypic and environmental variables, indicating a possible genotype-phenotype-environment interaction. In *Symphodus tinca*, we identified two loci associated to hatching date and also to environmental variables, being one of them related to turbulence and the other to temperature. Hatching date seems to be under selections in several marine species. For instance, in *S. tinca* reproducing females appear to prefer males with nests early in the reproductive season, and hatching success is consistently higher for eggs placed in early nests[48]. In the sunfish, *Lepomis macrochirus*, individuals born early in the season have higher survivorship than those born late[49]. In the peacock blenny, *Salaria pavo*, males select a different reproductive tactic depending of the birth date[50]. For the common triplefin, *Forsterygion lapillum*, more successful males hatch early in the hatching period[51]. Overall, these studies suggest that individuals born early could attain larger sizes at the reproductive season, rank more highly in social hierarchy and have better chances to reproduce. In *S. ocellatus,* 22 loci were related to settlement size or to PC2, highly influenced by PLD and settlement size, and 70% of them were also associated to temperature. A high proportion (50%) of those loci were also associated to productivity and EC1. Size at settlement, and other early-life traits, clearly increase post-settlement survival in some species, e.g. *Pomatomus saltatrix*[52], *Thalassoma bifasciatum*[53], *Symphodus* spp.[54] demonstrating the existence of a size-selective mortality. However, the absence of a consistent trend for selective mortality[53] and annual variation among settlement intensity, recruitment level and year class strength[55] suggest an important role of random events on survival. Therefore, natural selection could increase the frequency of alleles favouring reproduction early in the season as well as size at settlement and high stochasticity in survival maintain genetic polymorphism.

In conclusion, the results of this study show that individual-based information combining genomics, environmental and phenotypic values is key to unveil detailed patterns of evolutionary genomic signatures in living organisms and thus to glimpse the potential response to future changes. Despite the study of early life traits through otolith readings is only applicable to bone fishes, the combination of the different methodologies conducted in this paper to assess population structuring and local adaptation is useful for all species with individualized genotypic, environmental and phenotypic data. Furthermore, this approach has demonstrated to be applicable at the population level and in the absence of individual based information[4,6]. Our findings emphasize that candidate regions to local adaptation, mediated by productivity and temperature, can even be detected in species presenting low population structuring. This holistic approach should be considered in scientifically informed management actions since, in spatially and temporally variable environments, as found along the distribution ranges of most species, environmental and phenotypic factors could be imposing previously concealed selective pressures.

Materials and methods

Study species. *Symphodus ocellatus* (Linnaeus, 1758) and *Symphodus tinca* (Linnaeus, 1758) live in shallow waters (0–30 m), have similar larval duration (7–13 days) and are considered to disperse only during the larval stages, since their adults exhibit territorial behaviour[26]. These two species are representative components of the rocky, sea-grass beds and soft bottom communities of the Western Mediterranean[56,57]. They have a narrow bathymetric range (< 100 m), are found ubiquitously across the studied area and produce a clear settlement band in their otoliths[25,58]. The reproductive season of *S. tinca* extends during spring, whereas *S. ocellatus* reproduces in summer (Lejeune, 1985; Raventos & Macpherson, 2005b). *S. ocellatus* individuals live near 5 years, reach a maximum length of 12 cm and attain sexual maturity between 1 and 2 years[27]. *S. tinca* individuals live up to 15 years[28], can reach 44 cm, and attain sexual maturity in 2 or 3 years[27]. Due to the phylogenetic relatedness and shared biological traits and distribution, these species have a high value for a comparative population genomic study.

Sampling. A total of 162 individuals of *Symphodus ocellatus* (Linnaeus, 1758) and 141 of *Symphodus tinca* (Linnaeus, 1758) were sampled and analysed from six and five different locations respectively along the Spanish Mediterranean coast between 2014 and 2016 (Table 1, Fig. 1). This area contains two oceanographic discontinuities, the Almeria-Oran Front (AOF) and the Ibiza Channel (IC), that can affect connectivity among populations[29,30]. From Santa Pola, adult fin clips were provided by professional fishermen, while for the remaining localities juveniles were captured during recruitment using hand nets. The whole juvenile individuals and adult fin clips were taken and stored in 96% Ethanol. The collection of fish samples met the Spanish and European regulations. According to article 3.1 of the European Union directive (2010/63/UE) from the 22/9/2010, no approval is needed for fish sacrification with the purpose of tissue or organ analyses. Furthermore, the study species *Symphodus ocellatus* and *S. tinca* are not listed in CITES. A map of the sampling localities (Fig. 1) was plotted using the R[60] packages 'ncdf4'[61] and 'ggplot2'[62]. The colour gradient in the map represents the mean

temperature across the larval period of each species in a 2X2 Km grid obtained from Western Mediterranean OPerational forecasting system (WMOP)[63].

Phenotypic and environmental variables. At the laboratory, the otoliths (lapilli) of all juveniles were extracted and mounted in an oil droplet on a microscope slide. Age readings and the presence of settlement marks were determined using a light microscope following the standard methodology described in the literature[25,54]. In brief, we counted the age of the individual with the number of otolith marks along the longest radius, and with this information, we inferred the hatch date of each individual (Figure S1). PLD was calculated with the number of otolith marks before the settlement mark. From otolith readings, four early-life traits were considered for each individual: hatching date (expressed as the number of day of the year), pelagic larval duration (PLD, in days), growth rate during the pelagic larval duration (increment in otolith size in µm divided by PLD) and otolith settlement size (in µm). We measured the otolith width six times and used their mean to minimize measurements errors.

For each individual, the hatching date and its PLD value were used to determine the exact time of its pelagic period, over which the individual-based environmental variables were averaged. Three environmental variables were considered: two from remote sensing, surface temperature (in ºC) and productivity (measured as Chlorophyll a concentration (ChlA) in $mg{\cdot}m^{-3}$), and one from modelling, significant wave height (in meters) as a proxy for turbulence. The water surface temperature was obtained from SOCIB from a daily gridded dataset with a resolution of 0.02 degrees (Balearic Islands coastal ocean observatory and forecasting system, https://www.socib.eu/). Productivity data used in this paper were produced with the Giovanni online data system, developed and maintained by the NASA GES DISC[64], obtaining monthly mean values of ChlA with a spatial resolution of 4 km. Significant wave height was obtained from SIMAR model from the Spanish state ports agency, with hourly values and a spatial resolution of 0,25 degrees. Individual-based environmental and phenotypic variables were only available for individuals with otolith information (Table 1).

In order to evaluate the possible differences of early-life traits between locations, Permutational Multivariate Analyses of Variance (PERMANOVA) were performed for each phenotypic and environmental variable, as implemented in the R function 'adonis', from the 'vegan' package v 2.5-2[65]. Two principal components analysis (PCA) using 'dudi.pca' function from 'ade4' package v.1.7-8[66] were performed for each species after standardizing each variable, one with all phenotypic and the other with all environmental variables. We kept the minimum number of principal components necessary to explain at least 99% of the variance.

DNA extraction and genotyping. DNA was individually extracted using the QIAamp DNA Mini Kit (QIAGEN) extraction kit, following manufacturer's instructions. DNA integrity was checked by gel electrophoresis and quantified by NanoDrop. To build individual genomic libraries by GBS, 1.5–3 µg of DNA for each sample were sent to the Cornell University Biotechnology Resource Centre (BRC). DNA was digested with EcoT22I restriction enzyme and ligated to barcode and common adaptors with appropriated sticky ends as in a previous study[10]. For each plate, one blank sample and 95 samples were pooled and cleaned using QIAGEN PCR cleanup kit (following manufacturer's instructions). An amplification by PCR was applied to the resulting 96plex libraries, using generic primers matching the adaptors. The PCR was made under the following conditions: 5 min at 72ºC, 30 s at 98ºC, 18 cycles of 10 s at 98ºC, 30 s at 65ºC and 30 s at 72ºC and a final extension of 5 min at 72ºC. The PCR product was cleaned with the QIAquick PCR Purification Kit, diluted and single-end sequenced in an Illumina HiSeq 2500 platform at BRC, by using one lane per plate and the HiSeq v4 reagents kit. Individuals from the same locality were distributed in different lanes.

Raw Illumina sequences from all samples were used for genotyping with STACKS vs 1.47 software[67]. Good reads with a barcode were trimmed to 59 bp, excluding barcodes and primers. To construct a catalogue loci two mismatches were allowed between stacks within and between individuals, and a minimal stack depth of three was required. Individual genotypes were outputted both as haplotype loci and SNP loci VCF files. Additional filters were applied using VCFtools vs 1.12[68]. Individual genotypes with a depth below 5X were not considered. Loci with a missingness value higher than 30% were removed from the loci datafiles. Additionally, haplotype loci with the major allele frequency equally or higher than 0.95 (*i.e.* monomorphic at that level) were identified by the R function 'isPoly' from the package 'adegenet'[69] and removed from the vcf file by VCFtools. For SNP loci, those with a minimum allele frequency (MAF) of 0.05 were also removed from the SNP dataset by VCFtools. Finally, the loci in HW disequilibrium at more than 60% of the sampling sites[10] were eliminated of the analysis from both the haplotype and SNP loci datafiles. This filter was used step to eliminate possible paralogous loci.

Population genomic analyses. We used the haplotype loci datafiles for the population genomic analyses. For each locality, observed (Ho) and expected (He) heterozygosities, allelic richness (Ar), F_{ST} pairwise values and their significance (computed by 999 permutations) were obtained by R software[60] 'hierfstat' package vs 0.04–2[70]. A FDR correction for multiple comparisons was applied in all required analyses to estimate the threshold of differentiation[71]. Isolation by distance was evaluated between Euclidean geographic distances and F_{ST} distances with Mantel tests, as implemented by the 'mantel' function in the 'vegan' R package. Discriminant Analysis of Principal Components (DAPC) were performed retaining a number of PCAs equal to one third of the number of individuals. To detect the optimal number of genetic clusters, we performed K-means clustering from $k = 1$ to $k = 10$, using 10^5 iterations, retaining a number of PCs equal to sample size and compared the results using the Akaike Information Criterion (AIC). The analyses were performed with the R software package 'adegenet' and represented with the 'ggplot2' package[62]. To identify outlier loci (OAs), we used BayeScan vs 2.1[72], with individuals grouped by sampling locality. To minimize false positives, 100,000 simulations were run and a prior

odd of 10,000 was specified[73]. Those haplotype loci with a q-value (FDR analogue of the p-value) below 0.05 were considered as statistically significant outliers.

We used the SNP datafile for association analysis with environmental (EAAs) and phenotypic (PAAs) variables. We kept all polymorphic SNPs in each locus after filtering. Missing genotypes were imputed by replacing them with the most common genotype across all individuals. We performed a variance partition analysis for the complete matrix of SNPs, the matrix of the three environmental variables and the matrix of the four phenotypic variables using the 'varpart' function in the 'vegan' R package. We looked for SNPs strongly correlated with environmental variables by using a Redundancy Analysis (RDA) with the SNP loci datafiles following the method described by Forester and co-authors[42]. We used as response variables a genotypic matrix with the SNPs of all loci and the three environmental variables as predictors. All environmental variables were retained because they were not highly correlated ($r < 0.75$).

A genome-wide association analysis (GWAS) was performed with the linear model implemented in the software PLINK vs 1.9[74]. For each individual, we used its genotype from the SNP loci datafile and its environmental and phenotypic variables as well as the composite PCA variables previously detailed. Moreover, we performed a Multidimensional Scaling analysis (MDS) with the SNP datafile per species with the same software. The effect of genetic differentiation among localities is often removed in associated studies with the use of covariates. Thus, we carried out two association studies with PLINK: one using as covariates the two first components of the MDS (GWAS-Cov), and the other without covariates (GWAS).

For all candidate loci potentially under selection, identified as population outlier loci and/or including SNPs associated to environmental or phenotypic variables, we computed the frequency of the major allele and their observed heterozygosity for the haplotype dataset in each sampling location. Both values were represented in heatmaps and hierarchical dendrograms, using the heatmap.2 function of the package 'gplots' vs 3.0.1[75].

BLAST searches were conducted with the 59 bp sequences of all candidate loci against the *Labrus bergylta* Ascanius, 1767 genome[76], available in BLASTn search tool of Ensembl website (www.ensembl.org). Sensitivity was set to "Distant homologies" to maximise matches' length considering the phylogenetic distance between species. A maximum E-value of 10^{-3} was allowed and only matches of at least half of the locus sequence were considered. When sequences matched within a gene, the function of that gene was searched at the UniProt database (www.uniprot.org). For the list of annotated genes their GO-terms were introduced in REVIGO to produce treemaps of the biological functions[77].

Data availability

Raw read data from all individuals is available in the NCBI SRA Bioprojects PRJNA646056 (*S. ocellatus*) and PRJNA646057 (*S. tinca*). A list of genotypes of all individuals, and their sampling location, environmental and phenotypic variable values are available as Supplementary Dataset.

Received: 17 January 2020; Accepted: 3 July 2020
Published online: 29 July 2020

References

1. Grummer, J. A. *et al.* Aquatic landscape genomics and environmental effects on genetic variation. *Trends Ecol. Evol.* **34**, 641–654 (2019).
2. Ahrens, C. W. *et al.* The search for loci under selection: trends, biases and progress. *Mol. Ecol.* **27**, 1342–1356 (2018).
3. Benestan, L. *et al.* Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Mol. Ecol.* **25**, 5073–5092 (2016).
4. Xuereb, A., Kimber, C. M., Curtis, J. M. R., Bernatchez, L. & Fortin, M. J. Putatively adaptive genetic variation in the giant California sea cucumber (*Parastichopus californicus*) as revealed by environmental association analysis of restriction-site associated DNA sequencing data. *Mol. Ecol.* **27**, 5035–5048 (2018).
5. Hasan, M. M. *et al.* Sticklebacks adapted to divergent osmotic environments show differences in plasticity for kidney morphology and candidate gene expression. *J. Exp. Biol.* **220**, 2175–2186 (2017).
6. Carreras, C. *et al.* East is East and West is West: Population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). *Divers. Distrib.* **26**, 382–398 (2020).
7. Hansen, M. M., Olivieri, I., Waller, D. M. & Nielsen, E. E. Monitoring adaptive genetic responses to environmental change. *Mol. Ecol.* **21**, 1311–1329 (2012).
8. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379 (2011).
9. Stapley, J. *et al.* Adaptation genomics: the next generation. *Trends Ecol. Evol.* **25**, 705–712 (2010).
10. Carreras, C. *et al.* Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Sci. Rep.* **7**, 43417 (2017).
11. Babbucci, M. *et al.* An integrated genomic approach for the study of mandibular prognathism in the European seabass (*Dicentrarchus labrax*). *Sci. Rep.* **6**, 1–11 (2016).
12. Gonzalez-Pena, D. *et al.* Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Front. Genet.* **7**, 203 (2016).
13. Zhong, X. *et al.* Genome-wide association study reveals multiple novel QTL associated with low oxygen tolerance in hybrid catfish. *Mar. Biotechnol.* **19**, 379–390 (2017).
14. Schweizer, R. M. *et al.* Genetic subdivision and candidate genes under selection in North American grey wolves. *Mol. Ecol.* **25**, 380–402 (2016).
15. Planes, S. & Lenfant, P. Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. *Mol. Ecol.* **11**, 1515–1524 (2002).
16. Schunter, C. *et al.* Matching genetics with oceanography: Directional gene flow in a Mediterranean fish species. *Mol. Ecol.* **20**, 5167–5181 (2011).
17. Treml, E. A. *et al.* Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integr. Comp. Biol.* **52**, 525–537 (2012).

18. Aspillaga, E. *et al.* Ordinary and extraordinary movement behaviour of small resident fish within a mediterranean marine protected area. *PLoS ONE* **11**, 1–19 (2016).
19. Garrido, S. *et al.* Born small, die young: Intrinsic, size-selective mortality in marine larval fish. *Sci. Rep.* **5**, 17065 (2015).
20. Raventos, N., Macpherson, E. & García-Rubiés, A. Effect of brine discharge from a desalination plant on macrobenthic communities in the NW Mediterranean. *Mar. Environ. Res.* **62**, 1–14 (2006).
21. Schunter, C. *et al.* A novel integrative approach elucidates fine-scale dispersal patchiness in marine populations. *Sci. Rep.* **9**, 10796 (2019).
22. Warner, R. R. Evolutionary ecology: how to reconcile pelagic dispersal with local adaptation. *Coral Reefs* **16**, S115–S120 (1997).
23. Warner, R. R. & Lejeune, P. Sex change limited by paternal care: a test using four Mediterranean labrid fishes, genus *Symphodus*. *Mar. Biol.* **87**, 89–99 (1985).
24. Raventos, N. Nest site characteristics and nesting success of the five- spotted wrasse *Symphodus roissali* in the north-western Mediterranean Sea. *J. Fish Biol.* **68**, 305–309 (2006).
25. Raventos, N. & Macpherson, E. Planktonic larval duration and settlement marks on the otoliths of Mediterranean littoral fishes. *Mar. Biol.* **138**, 1115–1120 (2001).
26. Lejeune, P. L. comportement social des Labridés méditerranéens. *Cah. d'Ethologie Appliqué* **5**, 1–208 (1985).
27. Froese, R. & Pauly, D. FishBase. *World Wide Web electronic publication.* www.fishbase.org, version (10/2016) (2016).
28. Gordoa, A., Molí, B. & Raventos, N. Growth performance of four wrasse species on the north-western Mediterranean coast. *Fish. Res.* **45**, 43–50 (2000).
29. Pascual, M., Rives, B., Schunter, C. & Macpherson, E. Impact of life history traits on gene flow: a multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PLoS ONE* **12**, 1–20 (2017).
30. Galarza, J. A. *et al.* The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. *Proc. Natl. Acad. Sci.* **106**, 1473–1478 (2009).
31. Balbín, R. *et al.* Interannual variability of the early summer circulation around the Balearic Islands: driving factors and potential effects on the marine ecosystem. *J. Mar. Syst.* **138**, 70–81 (2014).
32. Fernández, V., Dietrich, D. E., Haney, R. L. & Tintoré, J. Mesoscale, seasonal and interannual variability in the Mediterranean Sea using a numerical ocean model. *Prog. Oceanogr.* **66**, 321–340 (2005).
33. Pascual, M. *et al.* Temporal and spatial genetic differentiation in the crab *Liocarcinus depurator* across the Atlantic-Mediterranean transition. *Sci. Rep.* **6**, 1–10 (2016).
34. Izaguirre, C., Méndez, F. J., Menéndez, M. & Losada, I. J. Global extreme wave height variability based on satellite data. *Geophys. Res. Lett.* **38**, 1–6 (2011).
35. Álvarez, I. *et al.* Drivers of larval fish assemblage shift during the spring-summer transition in the coastal Mediterranean. *Estuar. Coast. Shelf Sci.* **97**, 127–135 (2012).
36. Guillén, J. *et al.* Coastal oceanographic signatures of heat waves and extreme events of dense water formation during the period 2002–2012 (Barcelona, NW Mediterranean). *Sci. Mar.* **82**, 189–206 (2019).
37. Patarnello, T., Volckaert, F. & Castilho, R. Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break?. *Mol. Ecol.* **16**, 4426–4444 (2007).
38. Pérez-Portela, R. *et al.* Spatio-temporal patterns of genetic variation in *Arbacia lixula*, a thermophilous sea urchin in expansion in the Mediterranean. *Heredity (Edinb).* **122**, 244–259 (2019).
39. Calderón, I., Pita, L., Brusciotti, S., Palacín, C. & Turon, X. Time and space: genetic structure of the cohorts of the common sea urchin *Paracentrotus lividus* in Western Mediterranean. *Mar. Biol.* **159**, 187–197 (2012).
40. Renault, L., Oguz, T., Pascual, A., Vizoso, G. & Tintore, J. Surface circulation in the Alborn Sea (western Mediterranean) inferred from remotely sensed data. *J. Geophys. Res. Ocean.* **117**, 1–12 (2012).
41. Bradbury, I. R. *et al.* Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proc. R. Soc. Lond. B Biol. Sci.* **277**, 3725–34 (2010).
42. Forester, B. R., Lasky, J. R., Wagner, H. H. & Urban, D. L. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.* **27**, 2215–2233 (2018).
43. Excoffier, L., Hofer, T. & Foll, M. Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb).* **103**, 285–298 (2009).
44. Franch-Gras, L. *et al.* Genomic signatures of local adaptation to the degree of environmental predictability in rotifers. *Sci. Rep.* **8**, 16051 (2018).
45. Lamichhaney, S. *et al.* Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proc. Natl. Acad. Sci. USA* **109**, 19345–19350 (2012).
46. Sanford, E. & Kelly, M. W. Local adaptation in marine invertebrates. *Ann. Rev. Mar. Sci.* **3**, 509–535 (2010).
47. Font, J., Garcialadona, E. & Gorriz, E. The seasonality of mesoscale motion in the Northern Current of the western Mediterranean: several years of evidence. *Oceanol. Acta* **18**, 207–219 (1995).
48. Warner, R. R., Wernerus, F., Lejeune, P. & Van Den Berghe, E. Dynamics of female choice for parental care in a fish species where care is facultative. *Behav. Ecol.* **6**, 73–81 (1995).
49. Cargnelli, L. M. & Gross, M. R. The temporal dimension in fish recruitment: birth date, body size, and size-dependent survival in a sunfish (bluegill: *Lepomis macrochirus*). *Can. J. Fish. Aquat. Sci.* **53**, 360–367 (1996).
50. Fagundes, T. *et al.* Birth date predicts alternative life-history pathways in a fish with sequential reproductive tactics. *Funct. Ecol.* **29**, 1533–1542 (2015).
51. Moginie, B. F. & Shima, J. S. Hatch date and growth rate drives reproductive success in nest-guarding males of a temperate reef fish. *Mar. Ecol. Prog. Ser.* **592**, 197–206 (2018).
52. Hare, J. A. & Cowen, R. K. Size, growth, development, and survival of the planktonic larvae of *Pomatomus saltatrix* (Pisces: Pomatomidae). *Ecology* **78**, 2415–2431 (1997).
53. Searcy, S. & Sponaugle, S. Selective mortality during the larval-juvenile transition in two coral reef fishes. *Ecology* **82**, 2452–2470 (2001).
54. Raventos, N. & Macpherson, E. Effect of pelagic larval growth and size-at-hatching on post-settlement survivorship in two temperate labrid fish of the genus *Symphodus*. *Mar. Ecol. Prog. Ser.* **285**, 205–211 (2005).
55. Raventos, N. Relationships between adult population size, recruitment, and year-class strength in a labrid fish in the Mediterranean Sea. *Estuar. Coast. Shelf Sci.* **85**, 167–172 (2009).
56. García Rubies, A. & Zabala, M. Effects of total fishing prohibition on the rocky fish assemblages of Medes Islands marine reserve (NW Mediterranean). *Scientia Marina* **54**, 317–328 (1990).
57. Macpherson, E., Gordoa, A. & García-Rubies, A. Biomass size spectra in littoral fishes in protected and unprotected areas in the NW mediterranean. *Estuar. Coast. Shelf Sci.* **55**, 777–788 (2002).
58. Wilson, D. T. & McCormick, M. I. Microstructure of settlement-marks in the otoliths of tropical reef fishes. *Mar. Biol.* **134**, 29–41 (1999).
59. Raventos, N. & Macpherson, E. Environmental influences on temporal patterns of settlement in two littoral labrid fishes in the Mediterranean Sea. *Estuar. Coast. Shelf Sci.* **63**, 479–487 (2005).
60. R Core Team. *R: A Language and Environment for Statistical Computing.* http://www.r-project.org. (R Foundation for Statistical Computing, 2020).

61. Pierce, D. ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. R package version 1.16.1. (2019).
62. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, New York, 2016).
63. Juza, M. *et al.* Operational SOCIB forecasting system and multi-platform validation in the Western Mediterranean. *J. Oper. Ocean.* **9**, 9231 (2016).
64. Acker, J. G. & Leptoukh, G. Online analysis enhances use of NASA Earth Science Data. *Eos, Trans. Am. Geophys. Union.* **88**, 14–17 (2007).
65. Oksanen, J. *et al. vegan: Community Ecology Package*. R package version 2.5–2. https://CRAN.R-project.org/package=vegan (2018).
66. Dray, S. & Dufour, A. B. The ade4 Package: Implementing the Duality Diagram for Ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).
67. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: An analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
68. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
69. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
70. Goudet, J. & Jombart, T. hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04–22. (2015).
71. Narum, S. R. Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conserv. Genet.* **7**, 783–787 (2006).
72. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993 (2008).
73. Lotterhos, K. E. & Whitlock, M. C. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Mol. Ecol.* **23**, 2178–2192 (2014).
74. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
75. Warnes, M. G. R., Bolker, B., Bonebakker, L. & Gentleman, R. gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. (2016).
76. Lie, K. K. *et al.* Loss of stomach, loss of appetite? Sequencing of the ballan wrasse (*Labrus bergylta*) genome and intestinal transcriptomic profiling illuminate the evolution of loss of stomach function in fish. *BMC Genomics* **19**, 186 (2018).
77. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* **6**, e21800 (2011).

Acknowledgements

This work was supported by the Spanish Government projects 'ChallenGen' (CTM2013-48163) and 'PopCOmics' (CTM2017-88080) (MCIU, AEI/FEDER, UE) and by the European FP7 CoCoNet project (Ocean 2011-4, Grant Agreement #287844). The authors CC and MP are members of the research group SGR2017-1120 and EM of SGR2017-378 (Catalan Government). HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities (FPU15/02390).

Author contributions

H.T., C.C., E.M. and M.P. conceived and designed the study. N.R. and E.M. sampled all juveniles. N.R. did the otolith readings. H.T. conducted the laboratory and data analysis with inputs from C.C., E.M. and M.P.. H.T., C.C., E.M. and M.P. wrote the manuscript.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-69160-2.

Correspondence and requests for materials should be addressed to H.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 30 October 2019 | Revised: 4 February 2020 | Accepted: 10 February 2020

DOI: 10.1111/1755-0998.13144

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species

Anna Barbanti[1] | Hector Torrado[1,2] | Enrique Macpherson[2] | Luca Bargelloni[3] | Rafaella Franch[3] | Carlos Carreras[1] | Marta Pascual[1]

[1]Department of Genetics, Microbiology and Statistics and IRBio, University of Barcelona, Barcelona, Spain

[2]Center for Advanced Studies of Blanes (CEAB-CSIC), Blanes, Girona, Spain

[3]Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro, Italy

Correspondence
Carlos Carreras, Department of Genetics, Microbiology and Statistics and IRBio, University of Barcelona, Barcelona, Spain.
Email: carreras@ub.edu

Funding information
Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER), Grant/Award Number: CTM2017-88080 ; Generalitat de Catalunya, Grant/Award Number: SGR2017-1120 and SGR2017-378 ; Agència de Gestió d'Ajuts Universitaris i de Recerca, Grant/Award Number: FI_B 00997; Ministerio de Ciencia, Innovación y Universidades, Grant/Award Number: FPU15/02390

Abstract

High-throughput sequencing has revolutionized population and conservation genetics. RAD sequencing methods, such as 2b-RAD, can be used on species lacking a reference genome. However, transferring protocols across taxa can potentially lead to poor results. We tested two different IIB enzymes (AlfI and CspCI) on two species with different genome sizes (the loggerhead turtle *Caretta caretta* and the sharpsnout seabream *Diplodus puntazzo*) to build a set of guidelines to improve 2b-RAD protocols on non-model organisms while optimising costs. Good results were obtained even with degraded samples, showing the value of 2b-RAD in studies with poor DNA quality. However, library quality was found to be a critical parameter on the number of reads and loci obtained for genotyping. Resampling analyses with different number of reads per individual showed a trade-off between number of loci and number of reads per sample. The resulting accumulation curves can be used as a tool to calculate the number of sequences per individual needed to reach a mean depth ≥20 reads to acquire good genotyping results. Finally, we demonstrated that selective-base ligation does not affect genomic differentiation between individuals, indicating that this technique can be used in species with large genome sizes to adjust the number of loci to the study scope, to reduce sequencing costs and to maintain suitable sequencing depth for a reliable genotyping without compromising the results. Here, we provide a set of guidelines to improve 2b-RAD protocols on non-model organisms with different genome sizes, helping decision-making for a reliable and cost-effective genotyping.

KEYWORDS

Caretta caretta, conservation genomics, *Diplodus puntazzo*, genotyping-by-sequencing, high-throughput sequencing, sequencing simulations

Anna Barbanti and Hector Torrado contributed equally to the manuscript.

Carlos Carreras and Marta Pascual authors are considered senior authors.

1 | INTRODUCTION

High-throughput sequencing technologies have revolutionized the fields of population and conservation genetics during the last 10 years by providing easy access to genomic data from virtually any taxonomic group (Andrews & Luikart, 2014; Bellin et al., 2009; Davey & Blaxter, 2011). Many studies have explored the potential of genomic analysis to address a variety of questions, such as population structuring (Girault, Blouin, Vergnaud, & Derzelle, 2014), inbreeding depression (Hoffman et al., 2014), local adaptation (Savolainen, Lascoux, & Merilä, 2013) or hybridization (Hohenlohe, Amish, Catchen, Allendorf, & Luikart, 2011). Restriction site associated techniques (RAD) are based on massive sequencing after enzymatically reducing the fraction of the genome being analysed and can identify and score thousands of genetic markers, randomly distributed across the genome in many individuals simultaneously (Davey & Blaxter, 2011; Pecoraro et al., 2016). The advantage of these methodologies is that they can be carried out with no or limited previous sequence knowledge, since RAD tags can be analysed using pipelines for de novo loci identification if a reference genome is not available (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; Davey & Blaxter, 2011; Hapke & Thiele, 2016; Lu, Glaubitz, Harriman, Casstevens, & Elshire, 2012). These methods allow parallel and multiplexed sample sequencing of tag libraries, with a rapid and very cost-effective procedures resulting in high genome coverage (Baird et al., 2008; Pecoraro et al., 2016). The RAD marker approach has the flexibility to assay different number of markers depending on the restriction enzyme of choice (Baird et al., 2008).

Many studies focusing on population structure in non-model organisms have implemented different RAD technologies, such as RADseq (Lim et al., 2017; Xu et al., 2014), ddRAD (Lavretsky, DaCosta, Sorenson, McCracken, & Peters, 2019; Portnoy et al., 2015), GBS (Carreras et al., 2017; Hess et al., 2015), and 2b-RAD (Boscari et al., 2019; Galaska, Sands, Santos, Mahon, & Halanych, 2017). By shifting the realms of genomics from laboratory-based studies of model species towards studies of natural populations of ecologically well-characterized organisms, researchers can now start to address important ecological and evolutionary questions on a scale and precision that, only a few years ago, was unrealistic (Ekblom & Galindo, 2011). As for all genotyping-by-sequencing methodologies, the mean number of reads per locus (mean depth of coverage) is crucial to consider reliable the quality of markers and their genotypes (Sims, Sudbery, Ilott, Heger, & Ponting, 2014). Some recent population studies prioritised the number of sequenced individuals over depth of coverage or used improved bioinformatics pipelines to extract information from low coverage data (Buerkle & Gompert, 2013; Maruki & Lynch, 2014). However, when depth is generally low, statistical uncertainty of individual sequence data is high and calling of genotypes is difficult (Maruki & Lynch, 2017). Although probabilistic genotyping methods are thought to overcome shortcomings of low-depth sequencing data, they may behave unpredictably when compared to high-depth data (Hendricks et al., 2018). Thus, any analysis involving individual genotypes is going to be limited by coverage (Chow, Anderson, & Shedlock, 2019). For this reason, RAD sequencing techniques and laboratory protocols should be adjusted to target enough sequencing depth to obtain reliable genotypes while optimising sequencing costs.

2b-RAD is a RAD methodology that uses IIB restriction endonucleases, which cleave genomic DNA upstream and downstream of the target sites producing 32–34 bp fragments (Wang, Meyer, McKay, & Matz, 2012). This method is simple and provides a cost-effective alternative to existing reduced representation genotyping methods, allowing its use in routine experimental laboratory (Baird et al., 2008; Luo et al., 2017; Wang et al., 2012). One of the most interesting features of 2b-RAD is that the number of loci/marker density can be adjusted by using selective adaptors (Wang et al., 2012) to reduce the number of expected markers and increase the coverage per locus for a given sequencing effort. This RAD sequencing technique has been used to identify candidate genes associated with specific traits (Luo et al., 2017), to construct ultra-high density genetic maps (Fu, Liu, Yu, & Tong, 2016), to identify genomic regions under selection in population genetic studies (Pecoraro et al., 2016), and to perform genomic prediction for relevant traits in agricultural species (Palaiokostas, Ferraresso, Franch, Houston, & Bargelloni, 2016). It has also been extended to microbial ecology (Pauletto et al., 2016).

In this paper, we provide a protocol for laboratory and bioinformatic analyses to optimise studies using 2b-RAD sequencing on different non-model organisms. We focused our study on the sharpsnout seabream *Diplodus puntazzo* Walbaum, 1792 and the loggerhead turtle *Caretta caretta* Linnaeus, 1758 characterized by very different genome sizes. This study aims to unveil key elements to adapt library building of non-model organisms to best profit from this genomic method. Specifically, we focused our analyses on five main objectives. (a) Assess the effect of initial DNA quality and concentration on sequencing results. (b) Evaluate the performance of different IIB enzymes (i.e., AlfI and CspCI) on genomic library construction in the two species. (c) Calculate the optimum number of raw sequences needed per each combination of species and enzyme in order to achieve the maximum number of loci with an optimum depth per locus for a correct genotyping. (d) Assess if selective base ligation protocols have an impact on genetic differentiation among individuals. (e) Set guidelines for new population genomic studies on non-model organisms. Our study provides useful information for future studies on non-model species with different genome sizes, helping decision-making to obtain a reliable and cost-effective genotyping.

2 | MATERIALS AND METHODS

2.1 | Samples

We analysed two species with approximately three-fold different genome sizes. We consider the sharpsnout seabream (*D. puntazzo*) genome size to be similar to that of *Diplodus anularis* (0.9 Gb), its

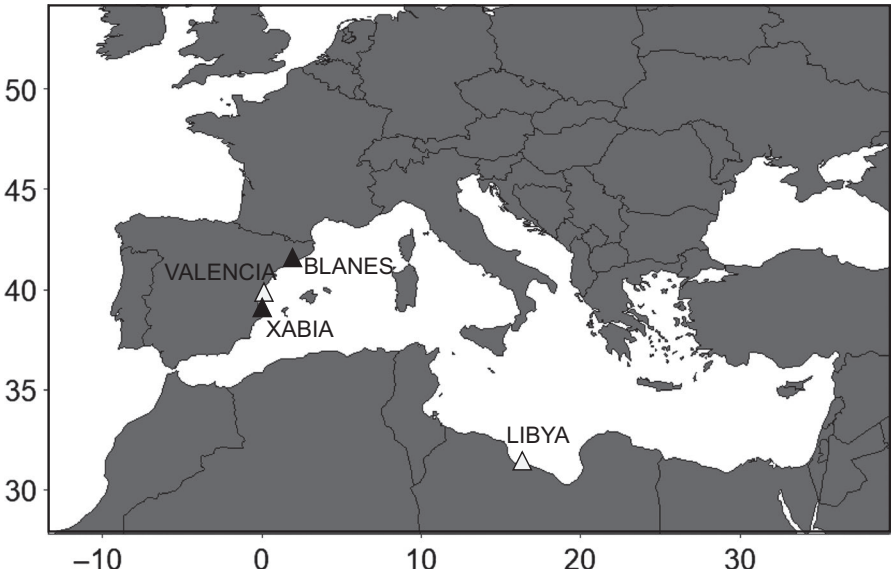


FIGURE 1 Sampling sites. White triangles show sampling sites for *Caretta caretta*, Libya is a nesting ground while Valencia is a foraging ground. Black triangles show sampling sites for *Diplodus puntazzo*

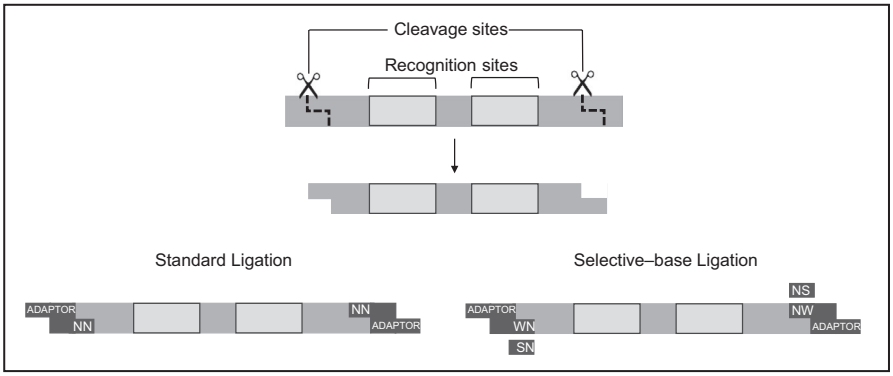


FIGURE 2 Selective-base ligation. In 2b-RAD protocol, after IIB enzyme digestion, specific fragments can be selected to reduce the density of markers to be amplified by designing customised adaptors with one fully degenerated base (N) and one partially degenerated base (S = G and C bases, W = A and T bases)

closest relative's sequenced genome (www.genomesize.com). The loggerhead turtle (*C. caretta*) genome size was considered to be similar to the genome of *Chelonia mydas* (Wang et al., 2013), which measures 2.24 Gb.

Juveniles of *D. puntazzo* were collected in Blanes (*N* = 12) and Xabia (*N* = 12) (Spain) during recruitment using hand nets (Figure 1). Samples of *C. caretta* were taken from bycaught juveniles at the foraging ground off Valencia (Spain) (*N* = 9; Figure 1) and from dead hatchlings at the nesting ground west of Sirte (Lybia) (*N* = 14; Clusa et al., 2018). We also added a sample collected from a live female turtle nesting in Pulpí (Spain) as positive control (Carreras et al., 2018). Consequently, our study included 24 samples per species. All samples were stored in 96% ethanol.

2.2 | DNA extraction and library construction

Genomic DNA was extracted using Qiagen Qiamp blood and tissue kit following the manufacturer's protocol. DNA concentration was measured with Nanodrop or Qubit, and DNA degradation assessed in 1% agarose gels. This information was recorded to be used in further statistical analysis. We coded the level of degradation as high if the DNA was mostly located at the bottom of the run in the agarose gel or the smear intensity increased in direction top-to-bottom, and as low if the DNA was mostly located at the top of the gel or the smear intensity faded in direction top-to-bottom. When possible we included samples that presented degraded DNA or low DNA concentration to test 2b-RAD efficiency for population genomics, as DNA degradation is a common issue when sampling non-model organisms (e.g., marine turtle studies targeting stranded individuals or dead embryos found after excavation of nests). A total of 24 individual libraries were constructed with each enzyme per species. Individual libraries were prepared adjusting the protocol from Wang et al. (2012). In brief, the construction of 2b-RAD libraries consisted of four main steps (for detailed protocol, see Annex S1). (a) Genomic DNA was digested by a IIB restriction enzyme providing short (32–34 bp) sequences. Each individual sample was digested with *AlfI* and *CspcI* enzymes separately. (b) During ligation, adaptors were attached to the sticky ends of the digested sequences. This step is crucial in the library preparation process because at this point, adaptors can be customised to attach to any sticky end or to attach only to sticky ends with specific sequences, based on the last two bases of the adaptor. For this study we used degenerated bases (5′-NN-3′) for our adaptors (Figure 2). (c) In the amplification step, barcodes and Illumina primers were attached to the adaptors and sequences were amplified by PCR. At the end of this step the resulting fragment is expected to measure

~165 bp. Library products were run through a 1.8% agarose gel to check amplification success. The library DNA quality of each sample was coded as 'good' when the band of the agarose gel was bright or 'bad' when it was faint (Appendix S1: Figure S1). (d) Purification was performed using magnetic beads to remove primers and sequences longer and shorter than 165 bp. At the end of this step, 2b-RAD libraries were ready to be sequenced. The DNA concentration of purified libraries was quantified using a Real Time PCR. The 48 libraries of each species were pooled for SR50 single read sequencing (one species per lane) with a HiSeq 4000 Illumina at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center.

2.3 | Genotyping

Sequences were processed using customized scripts (Annex S2). First, raw sequences were trimmed to eliminate ligation adaptors and then cut down to the same length (i.e., 32 bp for CspCI and 34 bp for AlfI). Processed sequences were used for genotyping using the STACKS versus 1.47 pipeline (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen et al., 2013). To construct a loci catalogue we used Stacks function *denovo_map.pl* setting the following parameters: a minimum depth of three reads to consider a stack within an individual (m = 3), up to three mismatches allowed between stacks (putative alleles) to merge them into a putative locus within an individual (*M* = 3), and two mismatches allowed between stacks (putative loci) during construction of the catalogue (*n* = 2). Individual genotypes were outputted as haplotype loci VCF files. We used five main filters to process loci found in our samples. We removed individual genotypes based on less than five reads, loci present in <70% of individuals, loci with outlier values of mean depth across all individuals (those above the upper whisker of the R boxplot, corresponding to 1.5 times the interquartile range from the data), loci with a major allele frequency higher than 99% and loci out of Hardy-Weinberg equilibrium (HWE) in at least one of the populations. In the case of *C. caretta* HWE was considered only for Libya, since Valencia is a feeding aggregation of individuals from different populations, and thus deviations from HWE are expected (Clusa et al., 2014). Filtering was performed with VCFTOOLS vs 1.12 (Danecek et al., 2011), with the exception of loci with a major allele frequency higher than 99%, which were identified by the function isPoly from the package ADEGENET (Jombart, 2008) and the assessment of HWE, computed with the function hw.test from the package pegas (Paradis, 2010) in R (R Core Team, 2018). We performed linear regression and Wilcoxon-Mann-Whitney test in R to assess whether initial and library DNA concentrations, initial DNA degradation and library quality influenced the number of total sequences and the final number of loci of each sample.

2.4 | Resampling analysis

We used bioinformatic simulations for each species and enzyme to obtain several sample sets, each one presenting a different number of reads per individual. We used a customised script to create new sample sets with different number of reads per sample by performing a random selection with replacement of the real data up to different target numbers of raw reads per sample (Annex S2). We performed 10 iterations for each target number. Target numbers varied for each species to accommodate the data points to the expected accumulation curve results for the different genome sizes. For *D. puntazzo* we simulated 0.5, 1, 2, 4, 8 and 10 million raw reads per sample for CspCI and AlfI enzymes. For *C. caretta* we simulated 4, 8, 12, 16 and 20 million raw reads per sample for each enzyme. Each resampled set underwent the same process of loci identification and filtering as explained above with the exception of the filter removing loci out of Hardy-Weinberg equilibrium. This filter was not applied because loci genotyping could be biased in the low depth data sets, artificially creating loci out of H-W equilibrium, since resampling was done with replacement. For this reason, this technique should not be used to artificially increase locus depth for a proper genotyping, as these genotyping errors are going to persist in the extended data sets. We calculated the formula that best fitted the accumulation curve for each species and enzyme and plotted the curve with R package 'GGPLOT2' (Wickham, 2016). We calculated the number of reads per individual needed to obtain a mean depth of coverage of 20×, since this threshold of quality is used in 2b.RAD studies (Resh, Galaska, & Mahon, 2018; Whelan et al., 2019). We also estimated values for a coverage of 25× (Warmuth & Ellegren, 2019) to evaluate if with higher coverage we can detect an improvement in the number of total loci.

2.5 | Selective-base ligation simulation

We assessed the potential impact of reducing the number of loci by selective-base-ligation in population genomic analyses. We bioinformatically selected trimmed reads of the corresponding combination of nucleotides to simulate the use of customised adaptors for selective-base ligation on each combination of species and enzyme (Annex S2). This type of ligation is usually performed in the laboratory by designing adaptors that will attach only to reads having the target base at both sticky ends (Figure 2). The simulation of a selective-base-ligation aims to test whether the processing of a proportionally lower number of loci per individual results in the same genetic differentiation as for the whole sample set. We removed from this analysis all samples that had a final mean depth per locus <10 to eliminate errors given by low depth of coverage. For *D. puntazzo* no samples were removed, while for *C. caretta* 5 samples were removed from the AlfI sample set and 7 form CspCI sample set. We used a customized script simulating the effects of building libraries with adaptors ending in 5′-WN-3′ (W = A and T) or 5′-SN-3′ (S = G and C) instead of 5′-NN-3′. These simulations aimed to select trimmed sequences by their first and last base and allocate them in separate folders. These selected sequences were then analysed with Stacks and loci were filtered with the same process as explained above for the whole data set. We calculated the genetic differentiation between pairs of

individuals using Prevosti distance with the R function prevosti.dist from the package poppr 2.8.0 (Kamvar, Brooks, & Grünwald, 2015; Kamvar, Tabima, & Grünwald, 2014) for the data set containing all combinations (NN) and for the two simulated selective-base-ligation data sets. The pairwise genetic distance matrixes among individuals for each selective-base-ligation subset were compared to the original NN matrix with a Mantel test using GENALEX v6.503 (Peakall & Smouse, 2012), then for each matrix we ran a Principal Coordinate Analysis (PCoA) to evaluate whether individuals maintained the same clustering pattern among subsets, using the same program. To detect the eventual decrease of heterozygosity in the subsets compared to their original set of loci we calculated individual observed heterozygosities for the three data sets with VCFTOOLS and used R to perform a Kruskal-Wallis test for each species and enzyme.

3 | RESULTS

3.1 | Library construction and loci identification in *Caretta caretta*

In *C. caretta* extracted DNA ranged from 17.3 to 133.5 ng/μl, and showed high level of degradation in 38% of the samples probably due to the bad condition of the tissue used (Appendix S1: Table S1). After adaptor ligation and amplification by PCR we observed generally good results with AlfI but much lower amplification success with CspCI with 46% of faint bands, as assessed with gel electrophoresis (Appendix S1: Table S1). After purification, library DNA concentration was similar for the two enzymes ranging between 6.7 and 52.3 ng/μl. The mean number of reads per sample was higher for AlfI digested samples, 7.6×10^6 reads per sample (max 10.1×10^6, min 4.0×10^6), than for CspCI digested samples, 6.6×10^6 reads per sample (max 10.7×10^6, min 2.6×10^6) mostly because some samples had low number of reads (Appendix S1: Table S1). The trimming process discarded all the sequences that were shorter than 34 bp for AlfI and 32 bp for CspCI or missed the chosen restriction site, with an average (±*SE*) lower loss per sample in AlfI (19.2 ± 2.1%) than in CspCI (41.9 ± 4.7%) (Table 1). After the loci calling, *C. caretta* showed higher total number of loci with AlfI (66,907 loci) than CspCI (25,416 loci). The mean number of loci retained after all filtering steps were slightly higher for AlfI (72.9 ± 0.4%) than for CspCI (69.4 ± 0.9%), although their final mean depth was smaller (Table 1).

3.2 | Library construction and loci identification in *Diplodus puntazzo*

In *D. puntazzo* starting concentrations ranged from 22.3 to 43.1 ng/μl and none of the samples was degraded. Adaptor ligation and amplification yielded successful amplifications with both enzymes although 17% of the samples digested with CspCI had faint bands (Appendix S1: Table S2). After purification, library DNA concentration was slightly higher for AlfI ranging between 13.6 and 109.63 ng/μl. As for *C. caretta* the sequencing of AlfI in *D. puntazzo* resulted in slightly higher mean number of reads per sample than for CspCI (Table 1). After the loci calling and filtering higher number of loci were also found for *D. puntazzo* for AlfI (84,382 loci) than for CspCI (31,111 loci). The mean number of loci retained after all filtering steps was similar for AlfI (90.6 ± 0.1%) than for CspCI (90.8 ± 0.1%), although their final mean depth was almost double in the latter (Table 1).

3.3 | Quality predictors of sequencing success

In the two species analysed and for both restriction enzymes the number of raw reads was significantly correlated to the final number of loci (Table 2). For *D. puntazzo*, initial DNA concentration, DNA degradation and library DNA quality had no significant effect in the number of raw reads or number of loci. However, for CspCI in *C. caretta*, the initial DNA concentration showed a significant impact on number of reads and loci, and on library concentration (Table 2). The library DNA concentration explained sequencing success in both species since the regression between library DNA concentration and the number of reads and loci was significant in most cases, with the exception of AlfI in *C. caretta* and the number of loci with CspCI in *D. puntazzo* (Table 2). The impact of DNA degradation on sequencing success was only assessed in *C. caretta* since in *D. puntazzo* DNA had initial good quality (Appendix S1: Tables S1 and S2). Interestingly, initial DNA degradation

TABLE 1 Summary of sequencing outcome

	Caretta caretta		*Diplodus puntazzo*	
Species enzyme	AlfI	CspCI	AlfI	CspCI
TR	$7.6 \pm 0.3 \times 10^6$	$6.6 \pm 0.4 \times 10^6$	$7.1 \pm 0.3 \times 10^6$	$6.5 \pm 0.3 \times 10^6$
TMR	$6.2 \pm 0.4 \times 10^6$	$4.2 \pm 0.5 \times 10^6$	$5.3 \pm 0.2 \times 10^6$	$4.3 \pm 0.2 \times 10^6$
IL	48,740 ± 1,489	17,811 ± 1,010	75,971 ± 130	27,989 ± 40
FL	35,576 ± 1,124	12,455 ± 732	68,978 ± 115	25,421 ± 27
RL	72.9 ± 0.4%	69.4 ± 0.9%	90.6 ± 0.1%	90.8 ± 0.1%
FMD	11.5 ± 0.7	19.3 ± 2.4	29.2 ± 1.4	52.2 ± 2.3

Note: Mean (±*SE*) values per individual are given for FMD, final mean depth of coverage per locus; FL, final number of loci after filtering; IL, initial number of loci; RL, percentage of loci retained after filtering; TR, total number of reads; TMR, number of trimmed reads.

TABLE 2 Statistical analyses of potential quality predictors

Explanatory variable	Response variable	Test	*Caretta caretta*				*Diplodus puntazzo*			
			CspCI		*AlfI*		*CspCI*		*AlfI*	
			F or *W*	*p*-Value	*F* or *W*	*p*-Value	*F* or *W*	*p*-Value	*F* or *W*	*p*-Value
Raw reads	Final loci	Linear regression	17.7	**.000**	30.4	**.000**	4.7	**.041**	34.4	**.000**
Initial DNA concentration	Raw reads	Linear regression	15.4	**.001**	2.7	.115	0.5	.469	0.4	.522
	Final loci	Linear regression	5.2	**.032**	2.1	.159	0.0	.959	1.5	.236
	Library DNA concentration	Linear regression	15.8	**.001**	0.0	.986	3.2	.087	0.3	.611
Initial DNA degradation	Raw reads	Wilcoxon-Mann-Whitney	60.0	.682	61.0	.726	na	na	na	na
	Final loci	Wilcoxon-Mann-Whitney	44.0	.170	45.0	.194	na	na	na	na
	Library DNA concentration	Wilcoxon-Mann-Whitney	34.0	**.048**	44.0	.174	na	na	na	na
Library DNA concentration	Raw reads	Linear regression	14.2	**.001**	2.0	.174	22.6	**.000**	6.3	**.020**
	Final loci	Linear regression	20.3	**.000**	3.2	.086	0.4	.559	6.2	**.021**
Library DNA quality	Raw reads	Wilcoxon-Mann-Whitney	19.0	**.002**	13.0	**.037**	na	na	25.0	.261
	Final loci	Wilcoxon-Mann-Whitney	12.5	**.001**	6.0	**.005**	na	na	9.0	**.018**

Note: Significant *p*-values after FDR correction are shown in bold. na, tests not available due to insufficient samples with bad initial DNA quality or low library DNA quality.

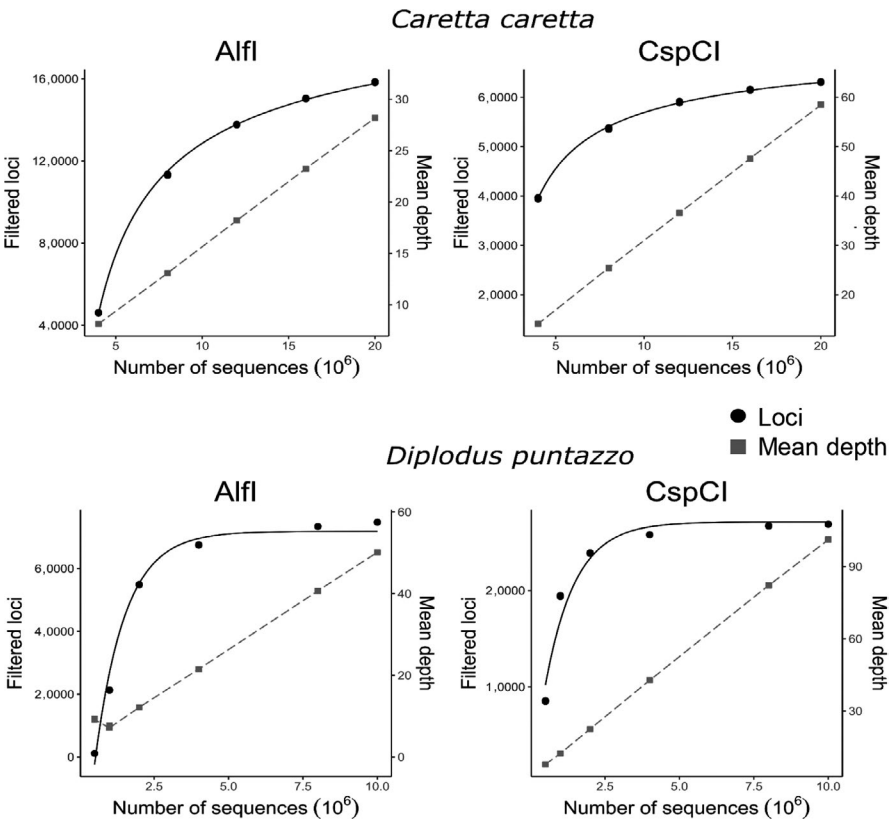


FIGURE 3 Accumulation curves resulting from the resampling analysis. The graphs show the number of final loci (circles) and the mean depth per locus (squares) obtained after filtering, for *Caretta caretta* (top) and *Diplodus puntazzo* (bottom)

was not a good predictor of neither the number of reads nor loci (Table 2). However, library DNA quality and thus amplification success assessed in an agarose gel significantly increased the number of raw reads and final number of loci (Table 2).

3.4 | Resampling analysis

We simulated the sequencing of different target number of reads per sample set and we obtained the total number of loci and mean

TABLE 3 Estimated number of loci and reads needed to obtain different mean depth per locus as derived from the accumulation curve

		Caretta caretta		*Diplodus puntazzo*	
		AlfI	CspCI	AlfI	CspCI
20×	Reads (10^6)	13.5	6.1	3.5	1.7
	Loci	142,910	49,588	68,079	22,225
25×	Reads (10^6)	17.4	7.9	4.6	2.2
	Loci	152,998	53,842	70,571	24,173

Note: The table shows the number of reads per individual and the total number of loci per set corresponding to a mean depth of coverage of 20× and 25× for each species and enzyme.

depth for each simulation (Figure 3; Appendix S1: Table S3). In all simulations, the mean depth of coverage was highly correlated to the number of reads per individual with an $R^2 > .99$. Based on the accumulation curve (Figure 3) we estimated the mean number of reads per individual and the corresponding number of loci for two mean depth of coverage, 20× and 25× (Table 3). For both species, AlfI needed a much higher number of reads per individual than CspCI to reach the desired coverage of 20×, due to the higher number of loci obtained with this enzyme. We found that, using a coverage of 25×, the total number of final loci improved in AlfI by 4% and by 7% for *D. puntazzo* and *C. caretta*, respectively, and by 9% in CspCI for both species.

3.5 | Selective-base ligation simulation

The selective-base ligation subsets obtained from *C. caretta* retained between 22.2% and 31.5% of the total loci from their original sample sets (Appendix S1: Table S4). In *D. puntazzo* the amount of loci retained was more variable between the two tested subsets (Appendix S1: Table S4), ranging from 19.8% to 43.4%. In this species we also found that for CspCI enzyme the subsets presented lower coverage than the original set, which could be a consequence of the base composition of the regions where this enzyme is cutting and related with the characteristics of the genomes that make the results species specific (Seetharam & Stuart, 2013). Mantel tests in both species showed high correlation between the pairwise genetic distances among individuals assessed with all loci and assessed with a selective base ligation, for both CspCI and AlfI enzymes (Figure 4). This was also reflected in the PCoA, as *C. caretta* samples do not have the exact same pattern among subsets whereas *D. puntazzo* patterns match perfectly despite the lower number of loci retained in the different data sets (Appendix S1: Figure S2). The Kruskal-Wallis test showed no significant differences in observed heterozygosity among any of the subsets and the original set of loci for both species and enzymes (Appendix S1: Table S5).

3.6 | Protocol optimization

We used the results obtained from these simulations to refine the laboratory protocol for 2b-RAD libraries preparation and

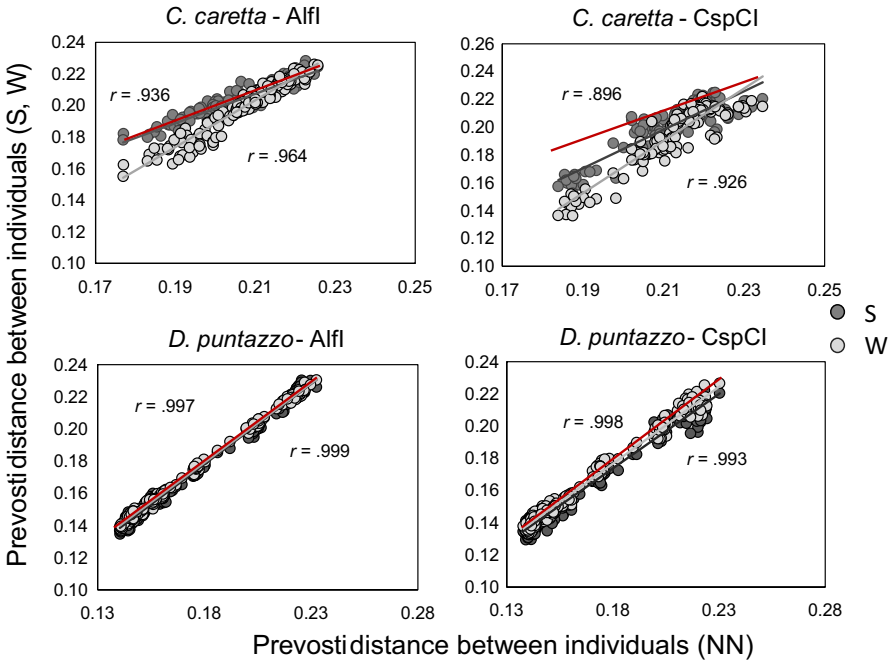


FIGURE 4 Mantel test of genetic differentiation between selective-base subsets and original sets. *X*-axes show Prevosti distance between pairs of individuals for each one of the four original sample sets (with fully degenerated bases –NN-). *Y*-axis show Prevosti distance between the same pairs of individuals for subsets obtained from bioinformatic simulations of selective base ligation (either –SN- or –WN-) for each species and enzyme. Dark grey shows genetic differentiation for S (G and C bases) subsets and light grey for W (A and T) subsets. Correlation coefficient (*r*) is given for each test above the lines for S and below for W. The red line represents the expected correlation function when no deviation in genetic distances is found in the selective-base subsets compared to NN

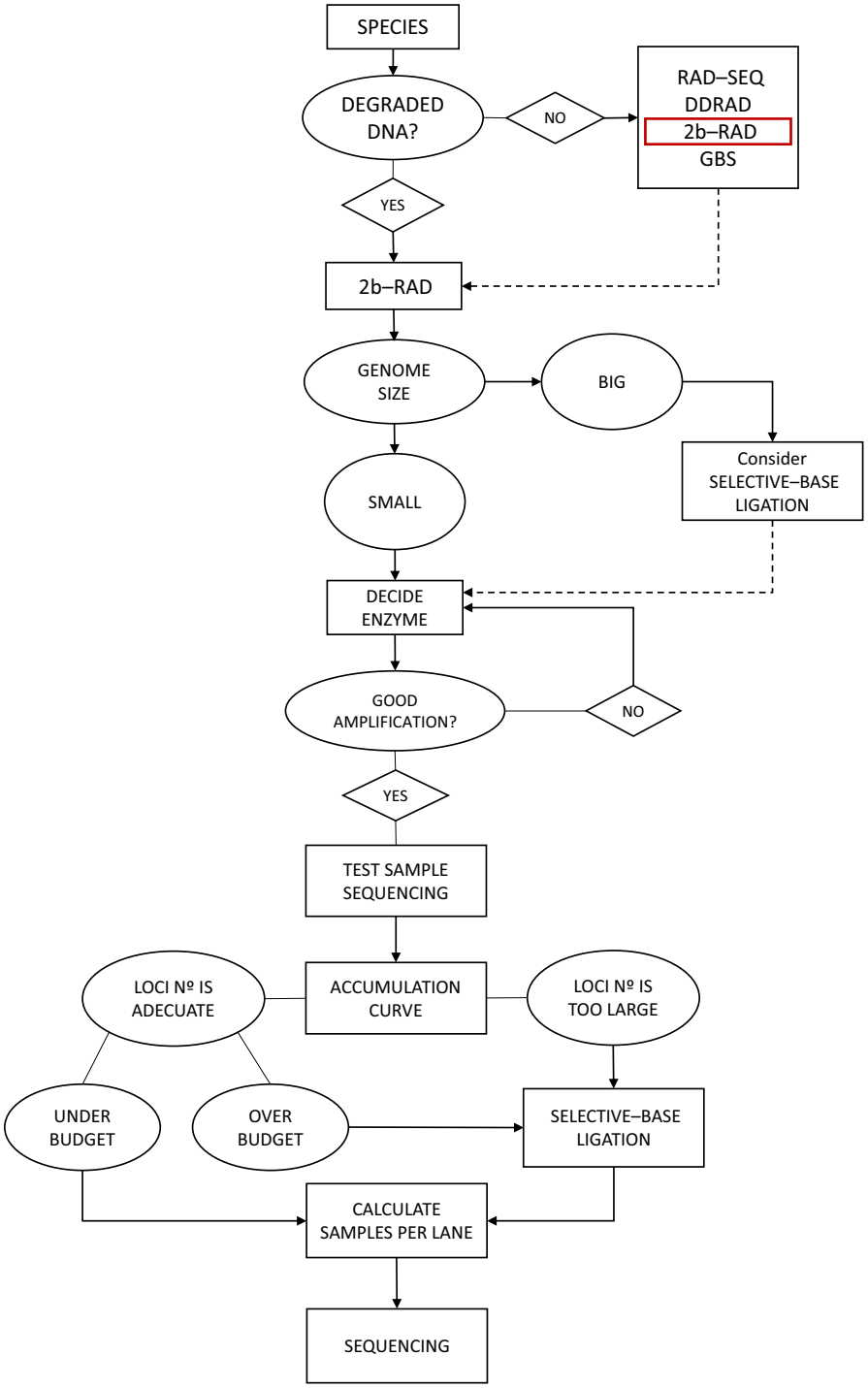


FIGURE 5 Flowchart for 2b-RAD laboratory protocol. This flowchart is meant to aid decision making for 2b-RAD laboratory protocols when studying non-model species. Together with the guidelines listed above this chart aims to make 2b-RAD studies not only easier but also more cost-effective

sequencing. In fact, given the mean value of depth of coverage, the optimum number of loci and the size of the studied species genome, we can calculate the number of samples to be sequenced in one lane to optimize costs without compromising the results. To facilitate the decision-making process, based on our results, we constructed a flowchart (Figure 5) and a set of guidelines (Box 1) to help future studies design the most efficient and cost effective protocol to reach their goals.

4 | DISCUSSION

In this study, we have shown that 2b-RAD protocol provides efficient results even with degraded samples and we demonstrated how this protocol can be optimised for population genomics of non-model species with different genome sizes. To prove this point, we analysed the sharpsnout seabream *D. puntazzo* and the loggerhead turtle *C. caretta* with two different enzymes, AlfI and CspCI, and performed

BOX 1. Guidelines for the optimisation of a 2b-RAD protocol with non-model species.

- Use 2b-RAD instead of other RAD sequencing techniques if you have degraded samples.
- If the target species has a big genome size, consider performing a selective-base ligation to retain 20%–40% of total loci.
- If the species genome is small, proceed without selective base-ligation.
- Test different IIB enzymes with the target species.
- Use library quality and concentration as predictors of sequencing success.
- Sequence the test samples with conservative conditions to obtain good coverage.
- Calculate an accumulation curve in a preliminary analysis with the test samples to identify the number of reads needed per individual and the total number of loci corresponding to a coverage ≥20×.
- If the total number of loci is adequate for the selected type of study, proceed to sequence the rest of your samples to obtain the mean number of reads needed according to the curve.
- If the total number of loci is too high for the selected study, use a selective base ligation for library building to reduce the amount of loci.
- The number of samples to be sequenced in the same lane is a trade-off between the number of reads per individual, the number of reads provided per lane and available budget.
- If the total number of loci is adequate but the cost of sequencing is over budget, use a selective base ligation for further 2b-RAD library building to reduce the amount of reads needed per sample and therefore fit more samples in one lane.

bioinformatic simulations. Our simulations allow estimating the mean number of reads needed per individual to obtain a reliable genotyping and the corresponding expected number of loci. Moreover, our results indicate that selective-base ligation can be used without compromising pairwise genetic distances among individuals.

In the case of the loggerhead turtle, where several samples had highly degraded DNA, we found that the quality of the initial DNA did not affect the number of raw reads nor the final number of loci, for both enzymes. In fact, the DNA short length for proper IIB enzyme functioning (i.e.,32–34 bp digested fragment) reduces the probability of missing loci even in highly degraded samples. This is a highly valuable characteristic of 2b-RAD methodology, since not all studies can easily access high quality samples. For instance, marine turtle genetic studies usually rely on sampling of stranded individuals (Clusa et al., 2016) or dead embryos found at nests after excavation (Clusa et al., 2018), due to the complexity of their behaviours and the paucity of individuals. In such cases, a genomic protocol capable of providing optimal results with degraded samples is invaluable.

The library quality after adaptor ligation and amplification was a good predictor of sequencing success. The electrophoresis gel after the library amplification of the loggerhead turtle clearly showed that AlfI resulted in a better amplification than CspCI, which failed to yield a clear band in 46% of individuals. Moreover, the sequencing success was poor for samples with faint amplification bands, which resulted in lower number of reads per individual and thus lower number of loci. We thus suggest discarding samples with poor library DNA quality to help optimising sequencing costs. In the case of the sharpsnout seabream, both enzymes showed good results after the amplification, although a few individuals yielded poorer amplification that resulted in significantly lower number of loci, as observed also in the loggerhead turtle were the difference in library quality with the two enzymes was even greater. Moreover, AlfI provided higher number of loci than CspCI in both species as expected, since AlfI recognition sequence has six fixed nucleotides, while CspCI has seven fixed nucleotides. Therefore, AlfI is expected to have a greater density of restriction sites across any genome than CspCI, and potentially yield more loci as observed in the kissing bug *Rhodnius ecuadoriensis* (Hernandez-Castro et al., 2017).

Obtaining more loci, though, reduces depth of coverage per locus for the same mean number of reads per individual. As expected, when using CspCI enzyme our sample sets showed higher values of mean depth then when using AlfI in both species, despite poorer amplification success for CspCI in the loggerhead turtle. A low mean depth per locus leads to less accurate genotype calling and thus higher percentage of missing data across loci (Casso, Turon, & Pascual, 2019; Chow et al., 2019; Maruki & Lynch, 2017), and for this reason a good depth coverage is important to consider data reliable. Since library construction and sequencing produces a variable number of reads per locus, a mean depth of 20× would guarantee that the minimum of five reads per genotype is consistently achieved across most loci for each sample. This would result in fewer genotypes lost and thus more loci retained over all samples. Our simulations on resampling analyses, allowed the construction of the accumulation curve relating the number of reads per sample and the resulting number of loci as well as the linear correlation between the mean depth per locus and the number of reads per individual. Based on the combination of these two functions the number of individuals to be sequenced in one lane can be calculated easily, simplifying decision-making and analysis design for optimizing population genomic studies at the lowest cost. The amount of reads per individual required by the sharpsnout seabream would allow including a fair number of individuals per lane for each enzymes, since both yielded good library DNA quality across samples. However, in the case of the loggerhead turtle, only AlfI enzyme should be used according to library DNA quality. In this case, the amount of reads needed to achieve an adequate

coverage would be very large and the number of loci obtained very high, due to the size of the genome. Under these circumstances, the number of individuals of loggerhead turtle to be included in one sequencing lane would be too small and not affordable by most research groups.

The difference between the two species is mostly related to the crucial role played by the genome size. Species with large genomes will probably produce more loci (due to a greater number of regions yielding the enzyme recognition site) and would need a greater sequencing effort to reach the suitable number of reads per sample for an adequate genotyping. Using a selective-base ligation the number of individuals can be adjusted to the needs of the study considering the number of loci projected by the accumulation curve. Our simulations of customized adaptors with selective base ligation, which extremities would end in –WN or –SN, proved that this type of reduction in the number of loci does not affect genetic differentiation between pairs of individuals. Therefore, the use of a selection of sequences for each sample instead of the whole set, would allow reducing costs by fitting more samples in one lane without compromising overall genetic differentiation. In both species we found that the subsets from the simulated selective-base ligation had a proportionally similar lower number of raw sequences and final loci than the original sets (~25%). However, some differences were observed according to the base and enzyme used in each species suggesting that the species' genome base composition may affect the outcome. Nonetheless, the high levels of correlation that we found between the subsets and the original sets, regardless of the number of loci retained, indicate that they are reliable sources of information. In fact, the slightly lower correlation in genetic distances of *C. caretta* and its differences in PCoAs patterns among subsets were probably a consequence of the bigger genome size of the species, resulting in a lower coverage. This type of selective ligation would be particularly interesting in the case of species with large genomes such as *C. caretta*. Considering the size of this species genome (2.24 Gb) and referring to our resampling simulation, we would need 13.5–17.4 million reads per sample to achieve 20×–25× of coverage, therefore only 20–25 samples could be sequenced in the same lane of a platform providing 340 million reads per run as in the present study. A selective-base ligation would allow reducing the costs of sequencing while ensuring good loci coverage, without influencing the outcome. In fact, since the selective-base ligated set would need only ~ 25% of the original set, between 3.4 and 4.4 million reads per sample are expected to reach the adequate coverage (Warmuth & Ellegren, 2019). Therefore, as much as 78–100 samples could fit in the same Illumina lane, greatly reducing costs without compromising genetic differentiation between individuals. Nevertheless, the number of loci required for a study depends on the scope, the type of analysis performed, and the target species. For instance, selective-base ligation would be less powerful for studies aiming to identify adaptation, since the probability of finding candidate genes can decrease when analysing only a small fraction of the genome (Ahrens et al., 2018).

Finally, we show that 2b-RAD methodologies can be reliable even for degraded DNA samples. Following our set of guidelines, researchers can optimize effort, time, and sequencing cost of 2b-RAD library building for non-model species while maintaining good sequencing depth for a proper genotyping (Box 1, Figure 5).

In conclusion, genomic population studies are increasing in species without reference genomes that rely on restriction-site associated DNA sequencing techniques, although some protocols require good quality DNA. Moreover, transferring protocols across taxa can potentially lead to poor results, such as low number of recovered markers or inadequate genotyping due to differential genomic features. Researchers working with species with large genome sizes or needing lower number of markers can adjust the number of loci by performing selective-base ligation, allowing the sequencing of a larger number of samples, without altering genomic differentiation between individuals as observed by our simulations. The optimal number of samples per lane can, therefore, be adjusted as a trade off with the desired target number of loci and the species genome size for an adequate mean depth of coverage for a correct genotyping. Our results and guidelines aim to improve 2b-RAD protocols on non-model organisms with different genome sizes, helping initial decision-making for a reliable, faster and cost-effective genotyping for population genomic studies.

ACKNOWLEDGEMENTS

M. Babucci (University of Padova) for helping with the initial trimming and preliminary analysis of sequenced data and J. Abril (University of Barcelona) for helping with script building for the simulation analysis with replacement. L. Cardona (University of Barcelona), J. Tomás (University of Valencia) and A.A. Hamza (Alfateh University and University Malaysia Terengganu) for providing tissue samples of *Caretta caretta* from the Valencian feeding ground and the Libyan rookery. This work was supported by the project CTM2017-88080 from the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER). CC and MP are part of the research group SGR2017-1120, and EM of SGR2017-378 (Catalan Government). HT was supported by a PhD scholarship funded by the Spanish Ministry of Science, Innovation and Universities (FPU15/02390). AB was supported by grant 2017 FI_B 00997 of the Catalan Government-AGAUR.

AUTHOR CONTRIBUTIONS

A.B., H.T., E.M., C.C., and M.P. conceived and designed the study. A.B., and H.T. did the laboratory analysis with inputs from L.B., and R.F. A.B., and H.T. conducted the data analysis. A.B., and H.T. wrote the manuscript with input from all authors.

DATA AVAILABILITY STATEMENT

Raw reads from all individuals, including information of location of all samples, can be found in the SRA Bioproject PRJNA604507.

All customised scripts (.sh files) can be found in the Supporting Information customised_scripts.zip.

ORCID

Carlos Carreras https://orcid.org/0000-0002-2478-6445
Marta Pascual https://orcid.org/0000-0002-6189-0612

REFERENCES

Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D., & Dudaniec, R. Y. (2018). The search for loci under selection: Trends, biases and progress. *Molecular Ecology*, *27*(6), 1342–1356.

Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. *Molecular Ecology*, *23*(7), 1661–1667.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, *3*(10), e3376.

Bellin, D., Ferrarini, A., Chimento, A., Kaiser, O., Levenkova, N., Bouffard, P., & Delledonne, M. (2009). Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species. *BMC Genomics*, *10*(1), 555.

Boscari, E., Abbiati, M., Badalamenti, F., Bavestrello, G., Benedetti-Cecchi, L., Cannas, R., … Fraschetti, S. (2019). A population genomics insight by 2b-RAD reveals populations' uniqueness along the Italian coastline in *Leptopsammia pruvoti* (Scleractinia, Dendrophylliidae). *Diversity and Distributions*, *25*, 1101–1117.

Buerkle, C. A., & Gompert, Z. (2013). Population genomics based on low coverage sequencing: How low should we go? *Molecular Ecology*, *22*(11), 3028–3035.

Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., Macpherson, E., & Pascual, M. (2017). Population genomics of an endemic Mediterranean fish: Differentiation by fine scale dispersal and adaptation. *Scientific Reports*, *7*, 43417.

Carreras, C., Pascual, M., Tomás, J., Marco, A., Hochscheid, S., Castillo, J. J., … Cardona, L. (2018). Sporadic nesting reveals long distance colonisation in the philopatric loggerhead sea turtle (*Caretta caretta*). *Scientific Reports*, *8*(1), 1435.

Casso, M., Turon, X., & Pascual, M. (2019). Single zooids, multiple loci: Independent colonisations revealed by population genomics of a global invader. *Biological Invasions*, *21*(12), 3575–3592. https://doi.org/10.1007/s10530-019-02069-8

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*(3), 171–182.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.

Chow, J. C., Anderson, P. E., & Shedlock, A. M. (2019). Sea turtle population genomic discovery: Global and locus-specific signatures of polymorphism, selection, and adaptive potential. *Genome Biology and Evolution*, *11*(10), 2797–2806.

Clusa, M., Carreras, C., Cardona, L., Demetropoulos, A., Margaritoulis, D., Rees, A. F., … Pascual, M. (2018). Philopatry in loggerhead turtles *Caretta caretta*: Beyond the gender paradigm. *Marine Ecology Progress Series*, *588*, 201–213.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Avolio, D., … Cardona, L. (2016). Potential bycatch impact on distinct sea turtle populations is dependent on fishing ground rather than gear type in the Mediterranean Sea. *Marine Biology*, *163*(5), 122.

Clusa, M., Carreras, C., Pascual, M., Gaughran, S. J., Piovano, S., Giacoma, C., … Cardona, L. (2014). Fine-scale distribution of juvenile Atlantic and Mediterranean loggerhead turtles (*Caretta caretta*) in the Mediterranean Sea. *Marine Biology*, *161*(3), 509–519.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFTOOLS. *Bioinformatics*, *27*(15), 2156–2158.

Davey, J., & Blaxter, M. L. (2011). RADSeq: Next-generation population genetics. *Briefings in Functional Genomics*, *9*, 108.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1.

Fu, B., Liu, H., Yu, X., & Tong, J. (2016). A high-density genetic map and growth related QTL mapping in bighead carp (*Hypophthalmichthys nobilis*). *Scientific Reports*, *6*, 28679.

Galaska, M. P., Sands, C. J., Santos, S. R., Mahon, A. R., & Halanych, K. M. (2017). Geographic structure in the Southern Ocean circumpolar brittle star *Ophionotus victoriae* (Ophiuridae) revealed from mt DNA and single-nucleotide polymorphism data. *Ecology and Evolution*, *7*(2), 475–485.

Girault, G., Blouin, Y., Vergnaud, G., & Derzelle, S. (2014). High-throughput sequencing of *Bacillus anthracis* in France: Investigating genome diversity and population structure using whole-genome SNP discovery. *BMC Genomics*, *15*(1), 288.

Hapke, A., & Thiele, D. (2016). GIBPSS: A toolkit for fast and accurate analyses of genotyping-by-sequencing data without a reference genome. *Molecular Ecology Resources*, *16*(4), 979–990.

Hendricks, S., Anderson, E. C., Antao, T., Bernatchez, L., Forester, B. R., Garner, B., … Luikart, G. (2018). Recent advances in conservation and population genomics data analysis. *Evolutionary Applications*, *11*(8), 1197–1211.

Hernandez-Castro, L. E., Paterno, M., Villacís, A. G., Andersson, B., Costales, J. A., De Noia, M., … Llewellyn, M. S. (2017). 2b-RAD genotyping for population genomic studies of Chagas disease vectors: *Rhodnius ecuadoriensis* in Ecuador. *PLOS Neglected Tropical Diseases*, *11*(7), e0005710.

Hess, J. E., Campbell, N. R., Docker, M. F., Baker, C., Jackson, A., Lampman, R., … Narum, S. R. (2015). Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, *15*(1), 187–202.

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. S., … Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences*, *111*, 3775–3780.

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, *11*, 117–122.

Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405.

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, *6*, 208.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). POPPR: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Lavretsky, P., DaCosta, J. M., Sorenson, M. D., McCracken, K. G., & Peters, J. L. (2019). ddRAD-seq data reveal significant genome-wide population structure and divergent genomic regions that distinguish the mallard and close relatives in North America. *Molecular Ecology*, *28*, 2594–2609.

Lim, H. C., Gawin, D. F., Shakya, S. B., Harvey, M. G., Rahman, M. A., & Sheldon, F. H. (2017). Sundaland's east–west rain forest population structure: Variable manifestations in four polytypic bird species examined using RAD-Seq and plumage analyses. *Journal of Biogeography*, *44*(10), 2259–2271.

Lu, F., Glaubitz, J., Harriman, J., Casstevens, T., & Elshire, R. (2012). TASSEL 3.0 Universal Network Enabled Analysis Kit (UNEAK) pipeline documentation. *White Paper*, *2012*, 1–12.

Luo, X., Shi, X., Yuan, C., Ai, M., Ge, C., Hu, M., … Yang, X. (2017). Genome-wide SNP analysis using 2b-RAD sequencing identifies the candidate genes putatively associated with resistance to ivermectin in *Haemonchus contortus*. *Parasites & Vectors*, *10*(1), 31.

Maruki, T., & Lynch, M. (2014). Genome-wide estimation of linkage disequilibrium from population-level high-throughput sequencing data. *Genetics*, *197*(4), 1303–1313.

Maruki, T., & Lynch, M. (2017). Genotype calling from population-genomic sequencing data. *G3: Genes, Genomes, Genetics*, *7*(5), 1393–1404.

Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R. D., & Bargelloni, L. (2016). Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3 (Bethesda)*, *6*(11), 3693–3700. https://doi.org/10.1534/g3.116.035220

Paradis, E. (2010). PEGAS: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420.

Pauletto, M., Carraro, L., Babbucci, M., Lucchini, R., Bargelloni, L., & Cardazzo, B. (2016). Extending RAD tag analysis to microbial ecology: A comparison between MultiLocus Sequence Typing and 2b-RAD to investigate Listeria monocytogenes genetic structure. *Molecular Ecology Resources*, *16*(3), 823–835. https://doi.org/10.1111/1755-0998.12495

Peakall, R., & Smouse, P. E. (2012). GENALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, *28*, 2537–2539.

Pecoraro, C., Babbucci, M., Villamor, A., Franch, R., Papetti, C., Leroy, B., … Cariani, A. (2016). Methodological assessment of 2b-RAD genotyping technique for population structure inferences in yellowfin tuna (*Thunnus albacares*). *Marine Genomics*, *25*, 43–48.

Portnoy, D. S., Puritz, J. B., Hollenbeck, C. M., Gelsleichter, J., Chapman, D., & Gold, J. R. (2015). Selection and sex-biased dispersal in a coastal shark: The influence of philopatry on adaptive variation. *Molecular Ecology*, *24*(23), 5877–5885.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Resh, C. A., Galaska, M. P., & Mahon, A. R. (2018). Genomic analyses of Northern snakehead (*Channa argus*) populations in North America. *PeerJ*, *6*, e4581.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807.

Seetharam, A. S., & Stuart, G. W. (2013). Whole genome phylogeny for 21 Drosophila species using predicted 2b-RAD fragments. *PeerJ*, *1*, e226. https://doi.org/10.7717/peerj.226

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121.

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012). 2b-RAD: A simple and flexible method for genome-wide genotyping. *Nature Methods*, *9*(8), 808.

Warmuth, V. M., & Ellegren, H. (2019). Genotype-free estimation of allele frequencies reduces bias and improves demographic inference from RADSeq data. *Molecular Ecology Resources*, *19*, 586–596.

Whelan, N. V., Galaska, M. P., Sipley, B. N., Weber, J. M., Johnson, P. D., Halanych, K. M., & Helms, B. S. (2019). Riverscape genetic variation, migration patterns, and morphological variation of the threatened Round Rocksnail, *Leptoxis Ampla*. *Molecular Ecology*, *28*(7), 1593–1610.

Wickham, H. (2016). GGPLOT2: *Elegant graphics for data analysis*. Berlin, Germany: Springer.

Xu, P., Xu, S., Wu, X., Tao, Y. E., Wang, B., Wang, S., … Li, G. (2014). Population genomic analyses from low-coverage RAD-Seq data: A case study on the non-model cucurbit bottle gourd. *The Plant Journal*, *77*(3), 430–442.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Barbanti A, Torrado H, Macpherson E, et al. Helping decision making for reliable and cost-effective 2b-RAD sequencing and genotyping analyses in non-model species. *Mol Ecol Resour.* 2020;00:1–12. https://doi.org/10.1111/1755-0998.13144