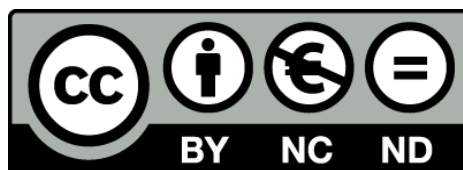


Nous marcadors moleculars per resoldre les espècies del gènere *Aeromonas*

Ariadna Sanglas Baulenas



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial – SenseObraDerivada 3.0. Espanya de Creative Commons.**

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial – SinObraDerivada 3.0. España de Creative Commons.**

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial-NoDerivs 3.0. Spain License.**



Facultat de Farmàcia

**Nous marcadors moleculars per resoldre
les espècies del gènere *Aeromonas***

**Ariadna Sanglas Baulenas
2015**



Facultat de Farmàcia
Departament de Microbiologia i Parasitologia Sanitàries
Programa de Doctorat: Biotecnologia

Nous marcadors moleculars per resoldre les espècies del gènere *Aeromonas*

Memòria presentada per **Ariadna Sanglas Baulenas** per optar al títol de Doctora per la
Universitat de Barcelona

Directores:

Dra. M^è Carme Fusté Munné

Dra. M. Isabel Farfán Sellarés

Doctoranda:

Ariadna Sanglas Baulenas
2015

Dra. M^a Carme Fusté Munné, Professora Titular d'Universitat de Microbiologia, i la **Dra. M. Isabel Farfán Sellarés**, Professora Titular Interina d'Escola Universitària de Microbiologia, del Departament de Microbiologia i Parasitologia Sanitàries de la Universitat de Barcelona,

INFORMA:

Que la memòria titulada "Nous marcadors moleculars per resoldre les espècies del gènere *Aeromonas*", presentada per ARIADNA SANGLAS BAULENAS per optar al títol de Doctora per la Universitat de Barcelona, ha estat realitzat sota la nostra direcció al Departament de Microbiologia i Parasitologia Sanitàries i considerant-la finalitzada, autoritzem la seva presentació per ser jutjada pel tribunal corresponent.

I, perquè així consti, firmem la present a Barcelona, el dia 22 de Juny de 2015,

Dra. M^a Carme Fusté Munné

Dra. M. Isabel Farfán Sellarés

AGRAÏMENTS

Tot i que ja fa alguns anys que aquest projecte de tesi va començar, no ha estat fins fa poc temps que he començat a ser realment conscient de tot el que això suposa i de la gran quantitat de persones que han estat al meu costat i que, molt probablement, sense elles i el seu suport hauria estat possible. Després de donar-hi moltes voltes, per fi ha arribat el moment donar-los les gràcies per tot el que han suposat per mi.

Voldria començar agraint a les meves directores de tesi, la Dra. Carme Fusté i la Dra. Maribel Farfán, per haver-me donat la oportunitat de realitzar la tesi en el seu grup d'investigació, per tota la dedicació i tot el que m'han ensenyat al llarg d'aquests anys i per estar en tot moment disposades a ajudar-me. També al Dr. J. Gaspar Lorén per la gran quantitat de coneixements aportats i les moltes hores de dedicació.

A la Vicen, sempre disposada a donar-me un cop de mà i per totes les hores que hem passat juntes, no tan sols al laboratori sinó també dinant, seminaris, els nostres primers congressos, etc.

A la Carla, al Guille, l'Ornella, la Mònica, al Nacho i totes aquelles persones que al llarg de tots aquests anys han passat pel laboratori amb les que també he compartit una infinitat de dinars i amb els que he arribat a riure molt.

A la Dra. M^a Jesús Montes i la Carme Guillen, sempre disposades a ajudar i resoldre dubtes, i amb les que juntament amb la Lucía Muñoz, hem fet les nostres xerradetes. Al Dr. Francisco Congregado, pel seu interès i les seves visites al laboratori d'aliments. També, a la Dra. Anna Marquès, la Dra. Merche Berlanga, el Dr. David Miñana-Galbis i a totes les persones que fan del departament de Microbiologia un magnífic lloc de treball.

A tots els amics del Voltreganès i de Sabadell, la majoria dels quals portem tota una vida junts, que formen una part important de la meva vida fora del laboratori i amb els que he passat tantes hores, gaudint de grans moments i que comparteixo magnífics i inoblidables records.

Tot i que no existeixen suficients paraules per expressar tot el meu agraïment m'agradaria donar les gràcies als meus pares, Maido i Joan, per tot el que m'han donat en aquesta vida i el seu suport incondicional, sempre disposats a donar-me ànims per tirar endavant quan ha estat necessari. A tota la meva família, tiets, cosins, iaia i a la família d'en Jordi, que des de fa molt de temps considero meva, en especial a la Montse i en Jaume. I també, molt especialment als meus avis que malauradament no són aquí per compartir amb mi aquest moment tan important de la meva vida. Sempre els tinc en el meu record.

Finalment, a tu Jordi, per estar sempre al meu costat, per creure en mi en els moments en que ni jo mateixa ho he fet, per buscar sempre la vesant positiva de totes les situacions, per ajudar-me a millorar personal i professionalment,..... Per tot, mil gràcies.

RESUM

En aquesta tesi doctoral s'ha analitzat la utilitat dels gens malat deshidrogenasa (*mdh*) i recombinasa A (*recA*) com a marcadors moleculars per a la construcció de filogènies fiables del gènere *Aeromonas*. Les dades obtingudes han permès l'obtenció, mitjançant mètodes de màxima versemblança (ML) i d'inferència bayesiana (IB), de filogènies amb un alt suport estadístic en la majoria dels nodes dels arbres obtinguts, indicant que els arbres construïts a partir de les seqüències d'aquests gens són robusts. Com a resultat d'aquesta anàlisi, s'han reclassificat algunes soques mal identificades, s'ha detectat un fragment recombinant en 4 soques d'*A.bestiarum*, del qual s'ha pogut identificar l'inici i final, així com el parental recombinant del que molt probablement deriva aquest segment, i s'ha determinat el possible efecte d'aquest esdeveniment de recombinació a l'estructura i funció de la proteïna RecA.

A partir de les filogènies obtingudes i utilitzant un punt de calibratge extern, s'ha determinat l'origen de l'ancestre comú d'*Aeromonas* i dels diferents llinatges d'aquest gènere bacterià. A més, s'ha pogut establir a partir de diferents aproximacions, el model de diversificació seguit al llarg dels anys en que *Aeromonas* ha estat evolucionant, així com la seva taxa de diversificació (nombre d'espècies/milió d'anys).

Finalment, s'ha comparat l'evolució seguida per *Aeromonas*, des del seu origen fins a l'actualitat, amb la de totes les espècies animals. Els resultats suggereixen, que des de la gran extinció del Permià-Triàsic, *Aeromonas* hauria evolucionat paral·lelament a l'augment del nombre dels gèneres animals. L'explosiva proliferació d'organismes multicel·lulars i la seva decisiva influència en l'estructura i funció dels ecosistemes va proporcionar un nou univers de nínxols ecològics potencials pels llinatges bacterians presents. La bona correlació observada entre el nombre de gèneres animals i la diversificació d'*Aeromonas* en els darrers 250 Ma, podria ser explicada per la colonització d'aquests nous nínxols.

ÍNDEX

Resum

Índex

Índex de figures

Índex de taules

Abreviatures

1. INTRODUCCIÓ	1
1.1 Reconstrucció d'arbres filogenètics	5
1.2 Incongruències a les filogènies	8
1.3 Mètodes de datació	9
1.4 Procés de diversificació de les espècies al llarg del temps	11
1.5 El gènere <i>Aeromonas</i>	12
1.5.1 Característiques generals	12
1.5.2 Antecedents històrics i situació actual	13
1.6 Gens seleccionats: malat deshidrogenasa i recombinasa A	18
2. OBJECTIUS	21
3. MATERIAL I MÈTODES	25
3.1 Selecció de soques	27
3.2 Recuperació de soques	28
3.3 Verificació de soques	28
3.4 Conservació de soques	32
3.5 Selecció de gens	33
3.6 Disseny d'encebadors	34
3.7 Extracció de DNA genòmic i quantificació	35
3.8 Reacció d'amplificació per PCR	35
3.9 Electroforesi en gel d'agarosa	37
3.10 Purificació del producte de l'amplificació per PCR	38
3.11 Seqüenciació directa	38
3.12 Obtenció de les seqüències nucleotídiques dels gens	40
3.13 Anàlisi de les seqüències de cada gen	41
3.14 Anàlisi filogenètica	42
3.14.1 Test del rellotge molecular	42
3.14.2 Elecció d'un model de substitució	42
3.14.3 Mètodes de reconstrucció	43

3.15 Anàlisi de la datació	45
3.16 Anàlisi de la diversificació	46
3.16.1 LTT plot	46
3.16.2 Models de diversificació	47
3.16.2.1 Determinació de l'estadístic γ de Pybus i Harvey	47
3.16.2.2 Anàlisi del model de diversificació	48
3.16.2.3 Anàlisi d'heterogeneïtats de taxes de diversificació	50
3.16.3 Càlcul teòric de l'origen de diversificació i el nombre mig d'espècies	51
3.17 Detecció de recombinació	51
3.17.1 Mètodes computacionals	52
3.17.2 Mètodes tradicionals	54
3.18 Predicció i construcció del model tridimensional d'una proteïna	54
3.18.1 Modelatge per homologia	55
3.18.2 Predicció dels llocs d'unió dels lligands	58
3.18.3 Perfil d'hidrofobicitat	58
4. RESULTAS: Estudi del gen malat deshidrogenasa d'<i>Aeromonas</i>	59
4.1 Obtenció de les seqüències	61
4.1.1 Disseny d'encebadors	61
4.1.2 Optimització del procés d'amplificació per PCR	62
4.1.3 Optimització del procés de seqüenciació	64
4.2 Anàlisi de les seqüències	66
4.2.1 Caracterització del gen <i>mdh</i>	66
4.2.2 Tipus de selecció	68
4.2.3 Distàncies genètiques intra- i inter-específiques	70
4.3 Filogènies	73
4.3.1 Anàlisi filogenètica del gènere <i>Aeromonas</i>	73
4.3.2 Anàlisi fenotípic de les soques amb incongruències filogenètiques	76
4.4 Datació	78
4.4.1 Datació a partir de les 150 seqüències	78
4.4.2 Datació a partir de les seqüències consens	80
4.4.3 Datació a partir de les seqüències de les soques tipus	81
4.4.4 Datació a partir de l'arbre d'espècies	83
4.5 Anàlisi de la diversificació	84
4.5.1 A partir de les seqüències consens	85
4.5.2 A partir de les seqüències de les soques tipus	89
4.5.3 A partir de l'arbre d'espècies	92

5. RESULTATS: Estudi del gen recombinasa A d'<i>Aeromonas</i>	97
5.1 Obtenció de les seqüències	99
5.1.1 Disseny d'encebadors	99
5.1.2 Optimització del procés d'amplificació per PCR	101
5.1.3 Optimització del procés de seqüenciament	103
5.2 Anàlisi de les seqüències	104
5.2.1 Caracterització del gen <i>recA</i>	104
5.2.2 Tipus de selecció	106
5.2.3 Distàncies genètiques intra- i inter-específiques	109
5.3 Filogènies	111
5.3.1 Anàlisi filogenètica del gènere <i>Aeromonas</i>	111
5.4 Recombinació	115
5.4.1 Anàlisi de la recombinació del gen <i>recA</i>	115
5.4.1.1 Anàlisi exploratòria de les seqüències	115
5.4.1.2 Mètodes de detecció amb programes bioinformàtics	117
5.4.1.2.1 Programa RAT	117
5.4.1.2.2 Programa RDP4	117
5.4.1.3 Mètodes de detecció tradicionals	119
5.4.1.3.1 Incongruències filogenètiques	119
5.4.1.3.2 Contingut en guanina i citosina (G+C)	120
5.4.1.3.3 Composició d'aminoàcids i ús de codons	121
5.4.2 Anàlisi de recombinació dels gens <i>recA+recX</i>	122
5.4.2.1 Anàlisi exploratòria de les seqüències	122
5.4.2.2 Mètodes de detecció amb programes bioinformàtics	123
5.4.2.2.1 Programa RAT	123
5.4.2.2.2 Programa RDP4	123
5.4.2.2.3 Programa SplitsTree4	125
5.4.2.3 Mètodes de detecció tradicionals	125
5.4.2.3.1 Contingut en guanina i citosina (G+C)	125
5.4.2.3.2 Composició d'aminoàcids i ús de codons	126
5.4.2.3.3 Incongruències filogenètiques	126
5.4.3 Caracterització del fragment recombinant	127
5.5 Predicció de la proteïna RecA d' <i>Aeromonas</i>	130
5.5.1 Modelatge molecular per homologia	130
5.5.1.1 Reconeixement del motlle	130
5.5.1.2 Alineament	131
5.5.1.3 Construcció del model	132
5.5.1.3.1 Predicció de l'estructura secundària	133
5.5.1.3.2 Predicció de l'estructura terciària	135
5.5.1.3.2.1 Obtenció dels models tridimensionals	135

5.5.1.3.2.2 Model tridimensional de RecA de la soca recombinant <i>A.bestiarum</i> CECT5741	138
5.5.1.4 Validació del model	143
5.5.1.5 Llocs d'unió dels lligands i perfil d'hidrofobicitat	143
5.6 Datació	144
5.6.1 A partir de les 150 seqüències	144
5.6.2 A partir de les seqüències consens	146
5.6.3 A partir de les seqüències de les soques tipus	147
5.6.4 A partir de l'arbre d'espècies	148
5.7 Anàlisi de la diversificació	150
5.7.1 A partir de les seqüències consens	150
5.7.2 A partir de les seqüències de les soques tipus	154
5.7.3 A partir de l'arbre d'espècies	157
6. RESULTATS: Estudi conjunt dels dos gens	161
6.1 Filogènies	163
6.1.1 Anàlisi filogenètica del gènere <i>Aeromonas</i>	163
6.2 Datació	166
6.2.1 A partir del concatenat de les 150 seqüències	166
6.2.2 A partir del concatenat de les seqüències consens	168
6.2.3 A partir del concatenat de les seqüències de les soques tipus	169
6.2.4 A partir de l'arbre d'espècies dels dos gens	170
6.3 Anàlisi de la diversificació	172
6.3.1 A partir del concatenat de les seqüències consens	172
6.3.2 A partir del concatenat de les seqüències de les soques tipus	175
6.3.3 A partir de l'arbre d'espècies dels dos gens	178
7. DISCUSSIÓ	183
7.1 Anàlisi dels marcadors moleculars	185
7.2 Filogènies d' <i>Aeromonas</i>	186
7.3 Anàlisi de la recombinació en el gen <i>recA</i>	188
7.4 Modelatge de la proteïna RecA	189
7.5 Datació de l'origen d' <i>Aeromonas</i>	192
7.6 Model i taxa de diversificació d' <i>Aeromonas</i>	194
8. CONCLUSIONS	201
9. ANNEXES	205
I. Soques utilitzades en l'estudi	207
II. Medis de cultiu	213

III. Solucions, rectius i colorants	218
IV. Electroforesi en gel d'agarosa	220
V. Composició d'aminoàcids del gen <i>recA</i>	221
VI. Ús de codons del gen <i>recA</i>	222
VII. Composició d'aminoàcids del gen <i>recX</i>	223
VIII. Ús de codons del gen <i>recX</i>	224
IX. Alineament de les seqüències de la proteïna RecA de diferents espècies bacterianes	225
10. BIBLIOGRAFIA	227
11. PUBLICACIONS	249
11.1 Population genetics of the " <i>Aeromonas hydrophila</i> complex"	251
11.2 The effect of recombination in <i>Aeromonas</i>	267
11.3 Reclassification of <i>Aeromonas hydrophila</i> subspecies <i>anaerogenes</i>	283
11.4 Draft genome sequence of <i>Aeromonas diversa</i> type strain	287
11.5 Draft genome sequence of <i>Aeromonas molluscorum</i> strain 848T ^T , isolated from bivalve molluscs	289

ÍNDIX DE FIGURES

Figura 3.1	Representació gràfica de la proporció de soques de cada espècie del gènere <i>Aeromonas</i> de l'estudi.	27
Figura 3.2	Esquema de la localització i l'orientació dels gens <i>recA</i> i <i>mdh</i> en el genoma d' <i>A. hydrophila</i> ATCC7966 ^T .	33
Figura 3.3	Representació gràfica dels tipus de diversificació al llarg del temps (LTT plots).	47
Figura 3.4	Representació gràfica dels models amb taxa constant.	48
Figura 3.5	Representació gràfica del model DDL.	49
Figura 3.6	Representació gràfica del model DDX.	49
Figura 3.7	Representació gràfica del model de Yule-2-rate.	50
Figura 3.8	Esquema de treball seguit per obtenir un model tridimensional per homologia d'una proteïna.	55
Figura 4.1	Posició dels encebadors utilitzats per a l'amplificació per PCR i la seqüenciació del gen <i>mdh</i> .	61
Figura 4.2	Comprovació dels amplificats del gen <i>mdh</i> obtinguts per PCR (5µL) mitjançant l'electroforesi en gel d'agarosa a l'1%.	64
Figura 4.3	Representació gràfica del grau de saturació del gen <i>mdh</i> .	67
Figura 4.4	Distribució de les substitucions sinònimes i no sinònimes per codons al llarg de les seqüències del gen <i>mdh</i> .	68
Figura 4.5	Relació entre les substitucions no sinònimes (<i>dN</i>) i sinònimes (<i>dS</i>) a nivell inter-específic del gen <i>mdh</i> .	70
Figura 4.6	Representacions gràfiques dels valors de distància genètica de les seqüències del gen <i>mdh</i> .	72
Figura 4.7	Influència del nombre de soques respecte les distàncies genètiques intra- i inter-específiques del gen <i>mdh</i> .	72
Figura 4.8	Arbre de màxima versemblança (ML) del gen <i>mdh</i> .	74
Figura 4.9	Arbre bayesià (IB) del gen <i>mdh</i> .	75
Figura 4.10	Cronograma del gènere <i>Aeromonas</i> basat en les 150 seqüències del gen <i>mdh</i> .	79
Figura 4.11	Cronograma del gènere <i>Aeromonas</i> basat en les seqüències consens del gen <i>mdh</i> de les diferents espècies.	81
Figura 4.12	Cronograma del gènere <i>Aeromonas</i> basat en les seqüències del gen <i>mdh</i> de les soques tipus.	82
Figura 4.13	Cronograma de l'arbre d'espècies del gènere <i>Aeromonas</i> obtingut a partir de les seqüències del gen <i>mdh</i> .	83
Figura 4.14	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en les seqüències consens del gen <i>mdh</i> .	86

Figura 4.15	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en les seqüències consens del gen <i>mdh</i> .	87
Figura 4.16	Resultats obtinguts amb el programa BMM a partir de les seqüències consens del gen <i>mdh</i> .	88
Figura 4.17	Resultats teòrics obtinguts a partir del model de Yule amb les dades de les seqüències consens del gen <i>mdh</i> .	88
Figura 4.18	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en les seqüències del gen <i>mdh</i> de les soques tipus.	89
Figura 4.19	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en les seqüències del gen <i>mdh</i> de les soques tipus.	90
Figura 4.20	Resultats obtinguts amb el programa BMM a partir de les seqüències del gen <i>mdh</i> de les soques tipus.	91
Figura 4.21	Resultats teòrics obtinguts a partir del model de Yule amb les seqüències del gen <i>mdh</i> de les soques tipus.	92
Figura 4.22	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en l'arbre d'espècies del gen <i>mdh</i> .	93
Figura 4.23	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en l'arbre d'espècies del gen <i>mdh</i> .	94
Figura 4.24	Resultats obtinguts amb el programa BMM a partir de les seqüències del gen <i>mdh</i> de l'arbre d'espècies.	95
Figura 4.25	Resultats teòrics obtinguts a partir del model de Yule amb l'arbre d'espècies del gen <i>mdh</i> .	95
Figura 5.1	Posició dels encebadors utilitzats per a l'amplificació per PCR i la seqüenciació del gen <i>recA</i> .	99
Figura 5.2	Comprovació dels amplificats del gen <i>recA</i> obtinguts per PCR (5µL) mitjançant l'electroforesi en gel d'agarosa a l'1%.	103
Figura 5.3	Representació gràfica del grau de saturació del gen <i>recA</i> .	105
Figura 5.4	Distribució de les substitucions sinònimes i no sinònimes per codons al llarg de les seqüències del gen <i>recA</i> .	107
Figura 5.5	Relació entre les substitucions no sinònimes (<i>dN</i>) i sinònimes (<i>dS</i>) a nivell inter-específic del gen <i>recA</i> .	108
Figura 5.6	Representacions gràfiques dels valors de distància genètica de les seqüències del gen <i>recA</i> .	110
Figura 5.7	Influència del nombre de soques respecte les distàncies genètiques intra- i inter-específiques del gen <i>recA</i> .	111
Figura 5.8	Arbre de màxima versemblança (ML) del gen <i>recA</i> .	113
Figura 5.9	Arbre bayesià (IB) del gen <i>recA</i> .	114
Figura 5.10	Ampliació del clade de l'espècie <i>A.bestiarum</i> de les filogènies obtingudes per màxima versemblança (A) i inferència bayesiana (B) del gen <i>recA</i> complet.	115

Figura 5.11	Visualització de les seqüències del gen <i>recA</i> de les soques d' <i>A.bestiarum</i> .	116
Figura 5.12	Resultat de sortida del programa RAT del clade <i>A.bestiarum</i> del gen <i>recA</i> .	117
Figura 5.13	Anàlisi de detecció de recombinació del gen <i>recA</i> utilitzant el programa RDP4.	119
Figura 5.14	Arbres de màxima versemblança construïts a partir de seqüències intragèniques del gen <i>recA</i> dels aïllats recombinants i les soques tipus d' <i>Aeromonas</i> .	120
Figura 5.15	Diagrama de punts de les posicions polimòrfiques de les seqüències <i>recA+recX</i> .	122
Figura 5.16	Resultat de sortida del programa RAT del <i>A.bestiarum</i> per les seqüències <i>recA+recX</i> .	123
Figura 5.17	Anàlisi de detecció de recombinació de la seqüència <i>recA+recX</i> utilitzant el programa RDP4.	124
Figura 5.18	Anàlisi <i>split decomposition</i> de les seqüències <i>recA+recX</i> de les soques recombinants i espècies relacionades.	125
Figura 5.19	Arbres de màxima versemblança construïts a partir de les seqüències del gen <i>recX</i> i de regions intragèniques de les soques problema i espècies relacionades.	127
Figura 5.20	Visualització de les seqüències nucleotídiques del fragment recombinant.	128
Figura 5.21	Identificació del parental recombinant.	129
Figura 5.22	Alineament de la seqüència aminoacídica de RecA d' <i>A.bestiarum</i> CECT5741 i de l'estructura 3CMW d' <i>E.coli</i> amb el programa JPred.	132
Figura 5.23	Predicció de l'estructura secundària de la proteïna RecA de la soca tipus d' <i>A.bestiarum</i> CECT4227 ^T i la soca recombinant CECT5741 a partir de la seqüència aminoacídica.	133
Figura 5.24	Predicció de l'estructura secundària de la proteïna RecA de la soca tipus d' <i>A.bestiarum</i> CECT4227 ^T i la soca recombinant CECT5741 a partir dels motlles 2REB i 3CMW.	134
Figura 5.25	Predicció de les estructures 3D de la proteïna RecA de la soca tipus d' <i>A.bestiarum</i> CECT4227 ^T i de l'aïllat recombinant CECT5741 a partir del motlle 2REB.	136
Figura 5.26	Predicció de les estructures 3D de la proteïna RecA de la soca tipus d' <i>A.bestiarum</i> CECT4227 ^T i de l'aïllat recombinant CECT5741 a partir del motlle 3CMW.	137
Figura 5.27	Predicció de les estructures 3D de la proteïna RecA de la soca recombinant CECT5741 a partir dels motlles 2REB i 3CMW.	138

Figura 5.28	Caracterització dels dominis estructurals i funcionals de la proteïna RecA de la soca recombinant CECT5741 a partir del motlle 3CMW. Predicció de 5741_3CMW.	140
Figura 5.29	Localització d'altres residus i dominis amb importància funcional de la proteïna RecA de la soca recombinant CECT5741 a partir del motlle 3CMW.	141
Figura 5.30	Anàlisi de la seqüència de la regió C-terminal de la proteïna RecA de la soca recombinant <i>A.bestiarum</i> CECT5741.	142
Figura 5.31	Representació gràfica del perfil d'hidrofobicitat de la seqüència de la soca recombinant <i>A.bestiarum</i> CECT5741 utilitzant el programa ProtScale.	144
Figura 5.32	Cronograma del gènere <i>Aeromonas</i> basat en les 150 seqüències del gen <i>recA</i> .	145
Figura 5.33	Cronograma del gènere <i>Aeromonas</i> basat en les seqüències consens del gen <i>recA</i> de les diferents espècies.	147
Figura 5.34	Cronograma del gènere <i>Aeromonas</i> basat en les seqüències del gen <i>recA</i> de les soques tipus.	148
Figura 5.35	Cronograma de l'arbre d'espècies del gènere <i>Aeromonas</i> obtingut a partir de les seqüències del gen <i>recA</i> .	149
Figura 5.36	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en les seqüències consens del gen <i>recA</i> .	151
Figura 5.37	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en les seqüències consens del gen <i>recA</i> .	152
Figura 5.38	Resultats obtinguts amb el programa BAMM a partir de les seqüències consens del gen <i>recA</i> .	153
Figura 5.39	Resultats teòrics obtinguts a partir del model de Yule amb les seqüències consens del gen <i>recA</i> .	153
Figura 5.40	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en les seqüències del gen <i>recA</i> de les soques tipus.	154
Figura 5.41	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en les seqüències del gen <i>recA</i> de les soques tipus.	155
Figura 5.42	Resultats obtinguts amb el programa BAMM a partir de les seqüències del gen <i>recA</i> de les soques tipus.	156
Figura 5.43	Resultats teòrics obtinguts a partir del model de Yule amb les seqüències del gen <i>recA</i> de les soques tipus.	157
Figura 5.44	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en l'arbre d'espècies del gen <i>recA</i> .	158
Figura 5.45	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en l'arbre d'espècies del gen <i>recA</i> .	159

Figura 5.46	Resultats obtinguts amb el programa BAMB a partir de les seqüències del gen <i>recA</i> de l'arbre d'espècies.	159
Figura 5.47	Resultats teòrics obtinguts a partir del model de Yule amb l'arbre d'espècies del gen <i>recA</i> .	160
Figura 6.1	Arbre de màxima versemblança (ML) del concatenat de seqüències dels gens <i>mdh</i> i <i>recA</i> .	164
Figura 6.2	Arbre bayesià (IB) del concatenat de seqüències dels gens <i>mdh</i> i <i>recA</i> .	165
Figura 6.3	Cronograma del gènere <i>Aeromonas</i> basat en el concatenat de les 150 seqüències dels gens <i>mdh</i> i <i>recA</i> .	167
Figura 6.4	Cronograma del gènere <i>Aeromonas</i> basat en el concatenat de les seqüències consens dels gens <i>mdh</i> i <i>recA</i> .	168
Figura 6.5	Cronograma del gènere <i>Aeromonas</i> basat en el concatenat de les seqüències de les soques tipus dels gens <i>mdh</i> i <i>recA</i> .	170
Figura 6.6	Cronograma del gènere <i>Aeromonas</i> basat en l'arbre d'espècies dels gens <i>mdh</i> i <i>recA</i> .	171
Figura 6.7	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en el concatenat de les seqüències consens dels dos gens.	173
Figura 6.8	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en el concatenat de les seqüències consens dels dos gens.	174
Figura 6.9	Resultats obtinguts amb el programa BAMB a partir del concatenat de les seqüències consens dels dos gens.	174
Figura 6.10	Resultats teòrics obtinguts a partir del model de Yule amb el concatenat de les seqüències consens dels dos gens.	175
Figura 6.11	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en el concatenat de les seqüències de les soques tipus dels dos gens.	176
Figura 6.12	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en el concatenat de les seqüències de les soques tipus dels dos gens.	177
Figura 6.13	Resultats obtinguts amb el programa BAMB a partir del concatenat de les seqüències de les soques tipus dels dos gens.	178
Figura 6.14	Resultats teòrics obtinguts a partir del model de Yule amb el concatenat de les seqüències de les soques tipus dels dos gens.	178
Figura 6.15	Anàlisi del model de diversificació del gènere <i>Aeromonas</i> basat en l'arbre d'espècies dels dos gens.	179
Figura 6.16	Determinació del millor model de diversificació del gènere <i>Aeromonas</i> basat en l'arbre d'espècies dels dos gens.	180
Figura 6.17	Resultats obtinguts amb el programa BAMB a partir de l'arbre d'espècies dels dos gens.	181

Figura 6.18	Resultats teòrics obtinguts a partir del model de Yule amb l'arbre d'espècies dels dos gens.	181
Figura 7.1	Dominis estructurals de la proteïna RecA i homòlegs de les diferents espècies bacterianes	190
Figura 7.2	Diversificació al llarg del temps del gènere <i>Aeromonas</i> comparada amb la dels gèneres animals recopilats per Sepkoski.	199

ÍNDIX DE TAULES

Taula 3.1	Atributs de les cadenes MCMC dels estudis d'inferència bayesiana realitzats per construir filogènies i cronogrames segons diferents aproximacions.	44
Taula 3.2	Característiques dels principals mètodes computacionals de detecció de recombinació implementats en el paquet informàtic RDP4.	53
Taula 3.3	Característiques dels programes més utilitzats per la modelització de proteïnes.	57
Taula 4.1	Característiques dels encebadors utilitzats per a l'amplificació i la seqüenciació del gen <i>mdh</i> .	62
Taula 4.2	Temperatures de fusió (T_m) i d'hibridació (T_a) dels encebadors del gen <i>mdh</i> .	62
Taula 4.3	Parelles d'encebadors utilitzades per a l'amplificació per PCR del gen <i>mdh</i> .	63
Taula 4.4	Composició de la barreja de reacció de PCR optimitzada per l'estudi del gen <i>mdh</i> .	63
Taula 4.5	Condicions optimitzades per l'amplificació per PCR del gen <i>mdh</i> .	64
Taula 4.6	Barreja de reacció de seqüenciació optimitzada del gen <i>mdh</i> .	65
Taula 4.7	Condicions optimitzades per la reacció de seqüenciació del gen <i>mdh</i> .	65
Taula 4.8	Anàlisi de la variabilitat i el contingut en G+C del gen <i>mdh</i> .	67
Taula 4.9	Anàlisi intra-específica dels canvis sinònims (dS) i no sinònims (dN) i el valor ω (dN/dS) del gen <i>mdh</i> .	69
Taula 4.10	Distàncies genètiques (p -distance) a nivell intra- i inter-específic del gen <i>mdh</i> .	71
Taula 4.11	Anàlisi del test del rellotge molecular del gen <i>mdh</i> .	73
Taula 4.12	Resultats de l'estudi fenotípic de les soques amb incongruències filogenètiques.	77
Taula 4.13	Proposta de reclassificació de les soques problema.	78
Taula 4.14	Edat de divergència estimada dels principals grups del cronograma obtingut a partir de les 150 seqüències del gen <i>mdh</i> .	80
Taula 4.15	Edat de divergència estimada dels principals grups del cronograma obtingut a partir de les seqüències consens del gen <i>mdh</i> .	81
Taula 4.16	Edat de divergència estimada dels grups principals del cronograma obtingut a partir de les seqüències de les soques tipus del gen <i>mdh</i> .	82

Taula 4.17	Edat de divergència estimada dels grups principals del cronograma obtingut a partir de l'arbre d'espècies del gen <i>mdh</i> .	84
Taula 5.1	Característiques dels encebadors per a l'amplificació i la seqüenciació del gen <i>recA</i> .	100
Taula 5.2	Temperatures de fusió (T_m) i d'hibridació (T_a) dels encebadors del gen <i>recA</i> .	101
Taula 5.3	Parelles d'encebadors utilitzades per a l'amplificació per PCR del gen <i>recA</i> .	102
Taula 5.4	Composició de la barreja de reacció de PCR optimitzada per l'estudi del gen <i>recA</i> .	102
Taula 5.5	Condicions optimitzades per l'amplificació per PCR del gen <i>recA</i> .	103
Taula 5.6	Barreja de reacció de seqüenciació optimitzada del gen <i>recA</i> .	104
Taula 5.7	Condicions optimitzades per la reacció de seqüenciació del gen <i>recA</i> .	104
Taula 5.8	Anàlisi de la variabilitat i contingut en G+C del gen <i>recA</i> .	106
Taula 5.9	Anàlisi intra-específic dels canvis sinònims (dS) i no sinònims (dN) i el valor ω (dN/dS) del gen <i>recA</i> .	108
Taula 5.10	Distàncies genètiques (p -distance) a nivell intra- i inter-específic del gen <i>recA</i> .	109
Taula 5.11	Anàlisi del test del rellotge molecular pel gen <i>recA</i> .	111
Taula 5.12	Esdeveniments de recombinació detectats amb el programa RDP4 pel conjunt de seqüències del gen <i>recA</i> .	118
Taula 5.13	Anàlisi del contingut en G+C (%mol) del gen complet <i>recA</i> i de regions intragèniques de les 4 soques recombinants i espècies relacionades.	121
Taula 5.14	Esdeveniments de recombinació detectats amb el programa RDP4 pel conjunt de seqüències <i>recA+recX</i> .	124
Taula 5.15	Anàlisi del contingut en G+C (%mol) del gen <i>recX</i> i de regions intragèniques de les 4 soques recombinants i espècies relacionades.	126
Taula 5.16	Resultats de la cerca d'estructures motlles amb el programa SWISS-MODEL per la soca recombinant <i>A.bestiarum</i> CECT5741.	131
Taula 5.17	Principals característiques de les estructures tridimensionals de la proteïna RecA d' <i>E.coli</i> disponibles a la base de dades PDB.	131
Taula 5.18	Identificació dels possibles aminoàcids implicats en la unió a diferents lligands.	143
Taula 5.19	Edat de divergència estimada dels principals grups del cronograma obtingut a partir de les 150 seqüències del gen <i>recA</i> .	146
Taula 5.20	Edat de divergència estimada dels principals grups del cronograma obtingut a partir de les seqüències consens del gen <i>recA</i> .	147

Taula 5.21	Edat de divergència estimada dels grups principals del cronograma obtingut a partir de les seqüències de les soques tipus del gen <i>recA</i> .	148
Taula 5.22	Edat de divergència estimada dels grups principals de l'arbre d'espècies d' <i>Aeromonas</i> obtingut a partir de l'arbre d'espècies del gen <i>recA</i> .	150
Taula 6.1	Edat de divergència estimada dels principals grups del cronograma obtingut a partir del concatenat de les 150 seqüències dels gens <i>mdh</i> i <i>recA</i> .	166
Taula 6.2	Edat de divergència estimada dels principals grups del cronograma obtingut a partir del concatenat de les seqüències consens dels gens <i>mdh</i> i <i>recA</i> .	169
Taula 6.3	Edat de divergència estimada dels principals grups del cronograma obtingut a partir del concatenat de les seqüències de les soques tipus dels gens <i>mdh</i> i <i>recA</i> .	169
Taula 6.4	Edat de divergència estimada dels principals grups del cronograma obtingut a partir de l'arbre d'espècies dels gens <i>mdh</i> i <i>recA</i> .	171
Taula 7.1	Principals característiques dels gens <i>mdh</i> i <i>recA</i> .	186
Taula 7.2	Edats de divergència empíriques i teòriques estimades a partir de les seqüències.	193
Taula 7.3	Valors estimats de λ obtinguts a partir de tres aproximacions diferents.	196
Taula 7.4	Resultats de l'anàlisi de diversificació.	197
Taula 7.5	Nombre d'espècies teòric obtingut a partir de les diferents aproximacions realitzades amb els gens <i>mdh</i> i <i>recA</i> .	198

ABREVIATURES I SÍMBOLS

A	Adenina
Å	Àngstrom
AFLP	<i>Amplified Fragment Length Polymorphism</i>
AIC	Criteri d'Informació d'Akaike
ATCC	<i>American Type Culture Collection</i>
ATP	Adenosina-5'-trifosfat
bv.	Biovar
C	Citosina
CECT	<i>Colección Española de Cultivos Tipo</i>
CIP	<i>Collection de l'Institut Pasteur</i>
DMSO	Dimetilsulfòxid
ddNTP	Didesoxinucleòtid trifosfat
dN	Substitucions no sinònimes per lloc no sinònim
dNTP	Desoxiribonucleòtid trifosfat
DNA	Àcid desoxiribonucleic
dS	Substitucions sinònimes per lloc sinònim
dsDNA	DNA de cadena doble
ESS	<i>Effective Sample Size</i>
F	<i>Forward primer</i>
g	Gram
G	Guanina
G	Distribució gamma
G+C	Contingut en guanina i citosina
h	Hora
HG	Grup d'hibridació DNA-DNA
HPD	<i>Highest Posterior Density</i>
H ₀	Hipòtesi nul·la
I	Posicions invariants
IB	Inferència bayesiana
IC	Interval de confiança
IUPAC/IUBMB	<i>International Union of Pure and Applied Chemistry / International Union of Biochemistry and Molecular Biology</i>
JCM	<i>Japan Collection of Microorganisms</i>
kDa	Kilodalton
L	Litre
LDH	Enzim lactat deshidrogenasa
LMG	<i>Belgian Co-ordinated Collections of Micro-organisms</i>
lnL	Logaritme de la versemblança
LTR	<i>Likelihood Ratio Test</i>
LTT plot	<i>Lineages Through Time plot</i>
Ma	Milió d'anys

MCC	<i>Maximum Clade Credibility tree</i>
MCMC	<i>Markov Chain Monte Carlo</i>
<i>mdh</i>	Gen malat deshidrogenasa
MDH	Enzim malat deshidrogenasa
min	Minut
mL	Mil·lilitre
ML	Màxima versemblança
MLEE	<i>Multilocus Enzyme Electrophoresis</i>
MLSA	<i>Multilocus Sequence Analysis</i>
MLST	<i>Multilocus Sequence Typing</i>
mm	Mil·limetre
ng	Nanogram
nm	Nanòmetre
<i>P</i>	Valor de <i>P</i>
pb	Parell de bases
PCR	Reacció en cadena de la polimerasa
PDB	<i>Protein Data Bank</i>
p.ex.	Per exemple
PFGE	<i>Pulsed-Field Gel Electrophoresis</i>
p/v	Pes/volum
qsp.	Quantitat suficient per
R	<i>Reverse primer</i>
r^2	Coefficient de determinació
<i>recA</i>	Gen recombinasa A
RecA	Proteïna recombinasa A
<i>recX</i>	Gen regulador
RFLP	<i>Restriction Fragment Length Polymorphism</i>
RNA	Àcid ribonucleic
rRNA	RNA ribosomal
SD	Desviació estàndard
SE	Error estàndard
seg	Segon
sp.	Espècie
ssDNA	DNA de cadena senzilla
subsp.	Subespècie
T	Timina
T	Soca tipus
Ta	<i>Annealing temperature</i>
Ti	Transicions
Tm	<i>Melting temperature</i>
tMRCA	<i>Time to the Most Recent Common Ancestor</i>
Tv	Transversions
UV	Ultraviolat
V	Volt
α	Alfa

β	Beta
γ	Estadístic Gamma de Pybus i Harvey
λ	Taxa de diversificació
ω	Omega (dN/dS)
μ	Taxa d'extinció
μg	Microgram
μL	Microlitre
μM	Micromolar
%	Percentatge
$^{\circ}\text{C}$	Grau Celsius
∞	Infinít

1. INTRODUCCIÓ

1. INTRODUCCIÓ

Actualment, hi ha una gran quantitat d'estudis en els que s'utilitzen seqüències de diferents gens per generar les anomenades filogènies moleculars. Poques vegades es consideren les característiques que han de tenir les seqüències dels gens per poder ser utilitzats en aquest tipus d'anàlisi, tot i que és un aspecte important a tenir en compte alhora de construir una veritable filogènia. Les filogènies basades en un únic gen (arbre gènic) poden no necessàriament representar la història evolutiva verdadera dels organismes que s'està analitzant, si no que poden explicar la del gen en particular, mentre que un altre gen no distant en el genoma pot haver experimentat una història evolutiva diferent. En canvi, un arbre d'espècies és un patró dels llinatges i les seves relacions a través del temps. Els esdeveniments que han tingut lloc en la història d'un gen no es corresponen necessàriament a la història de les espècies que poden estar afectades per processos com la duplicació i/o pèrdua de gens, la transferència lateral de gens, la recombinació, etc. (Maddison 1997; Liu & Pearl 2007; Heled & Drummond 2010). Aquest factor no sempre es té en compte en molts treballs, i es per això que es recomana analitzar més d'un gen amb històries evolutives comparables per construir una filogènia fiable.

A més, els gens utilitzats han d'estar sotmesos a selecció estabilitzadora, amb evolució neutra (que acumulin mutacions per deriva genètica a l'atzar) i que codifiquin funcions de metabolisme central (*housekeeping*). La manera tradicional de determinar el tipus de selecció que opera en un gen codificador de proteïna consisteix en analitzar la proporció relativa en la que s'han fixat les mutacions que originen canvis sinònims (silenciosos) i no sinònims (que canvien l'aminoàcid) en la població, és a dir, determinant el quocient dN (número de canvis no sinònims per lloc no sinònim) i dS (número de canvis sinònims per lloc sinònim) que es simbolitza amb la lletra ω . Si el canvi és deleteri, la selecció purificadora reduirà el seu percentatge de fixació i per tant ω serà < 1 . Pel contrari, quan el canvi en l'aminoàcid proporciona una avantatge selectiva, aquest es fixarà amb una percentatge superior al d'una mutació sinònima. Finalment, si el canvi en un aminoàcid és neutre es fixarà amb la mateixa proporció amb la que ho faria una mutació sinònima, amb $\omega = 1$.

Quan s'analitzen gens que codifiquen proteïnes, la majoria dels aminoàcids estan altament conservats per tal de mantenir la seva funció i molt probablement l'evolució adaptativa afectarà només a alguns aminoàcids en algun moment de la seva evolució. L'anàlisi que s'ha comentat anteriorment només pot detectar fenòmens de selecció positiva en el casos més extrems. Actualment però, es disposa de programes que permeten comparar un gen determinat codó per codó en diferents espècies per identificar llocs seleccionats positivament o episodis de selecció positiva que s'hagin produït al llarg de l'evolució del gen en la població que s'està analitzant.

Les mutacions, són una font molt important de variació genètica ja que poden alterar un gen donant lloc a un nou al·lel o poden afectar a l'estructura o al patró d'expressió depenent de la localització de la mutació. No obstant, la influència d'una mutació no té perquè ser la mateixa en la variabilitat genètica d'una població, dependrà de si aquesta és deletèria o no i de si és

seleccionada positivament entre altres paràmetres (Lewis-Rogers 2004). Les mutacions sinònimes són neutres, per tant, no estan subjectes al fenomen de selecció i per aquest motiu són molt habituals. Les mutacions no sinònimes en canvi, poden tenir un efecte positiu o negatiu. Només una petita part de totes les mutacions són beneficioses i tenen èxit, de manera que s'incorporen a l'espècie. Degut a la importància dels processos de mutació al llarg de l'evolució, un factor a tenir en compte és la taxa de mutació. Una taxa alta de mutació implica un major potencial d'adaptació ja que a l'augmentar les variants genètiques també augmenta la probabilitat d'obtenir la variant adequada per adaptar-se a un nou ambient. Per contrapartida, una taxa de mutació elevada també està associada a més possibilitats de mutacions deletèries, augmentant així la probabilitat d'extinció (Denamur & Matic 2006).

Els gens utilitzats en les filogènies han de ser bons rellotges moleculars, amb una taxa constant d'evolució, és a dir, que presentin una relació lineal entre la quantitat de substitucions i el temps de divergència. De vegades però, el ritme de substitucions pot haver-se accelerat en una espècie determinada o en un grup de soques de la mateixa espècie per la fixació de mutacions selectivament avantatjoses. El concepte de rellotge molecular va ser definit per primera vegada per Margoliasch a través d'un estudi del citocrom C de diferents espècies. No obstant, van ser Zuckerkandl i Pauling els que van postular formalment la hipòtesi del rellotge molecular pel vols de l'any 1960 (Zuckerkandl & Pauling 1962). Posteriorment, Kimura (1985) va desenvolupar la teoria neutral de l'evolució que prediu el rellotge molecular. Actualment, es disposa de mètodes d'anàlisi que permeten construir filogènies a partir de gens que no compleixen aquest requisit, els anomenats mètodes de rellotge molecular relaxat els quals inclouen enfocaments no paramètrics (Sanderson 1996), rellotges locals (Yoder & Yang 2000) i models paramètrics bayesians (Thorne *et al.* 1998; Huelsenbeck *et al.* 2001; Kishino *et al.* 2001; Drummond *et al.* 2006; Lepage *et al.* 2007).

Un altre factor a tenir en compte en aquest tipus d'anàlisi és el grau de saturació de les substitucions sinònimes i no sinònimes del gen. Si a l'analitzar el nombre de substitucions en un gen respecte a la divergència s'obté un gràfic en el que s'observa que la seqüència ja no pot acumular més mutacions (està saturada), aquest gen s'ha de rebutjar per a la construcció de la filogènia ja que quan s'arriba a la saturació perd la senyal filogenètica de manera que les seqüències ja no són informatives del procés evolutiu seguit. En aquesta situació qualsevol estima de la distància genètica o arbres filogenètics, sigui quin sigui el mètode de reconstrucció utilitzat, no tindrà sentit ja que les seqüències tendiran a agrupar-se segons el grau de similitud de la seva composició de bases independentment de la seva veritable genealogia. Aquest problema de saturació sovint és obviat en les reconstruccions filogenètiques quan, de fet, és un problema crucial. En les seqüències codificants, les terceres posicions del codó normalment evolucionen més ràpidament que les primeres i segones, per tant és més probable que estiguin saturades especialment si es comparen taxa distants. Un sistema per evitar el problema de la saturació és excloure les terceres posicions de l'anàlisi o utilitzar la seqüència d'aminoàcids corresponent (Lemey *et al.* 2009b).

1.1 RECONSTRUCCIONS D'ARBRES FILOGENÈTICS

Els mètodes de reconstrucció filogenètics tenen com a resultat l'obtenció d'un arbre en el qual les branques i els nodes uneixen els diferents tàxons. Els arbres filogenètics es construeixen a partir d'alineaments múltiples de seqüències nucleotídiques o aminoacídiques, que en les regions més variables poden presentar substitucions i espais (*gaps*).

Un dels reptes de la reconstrucció filogenètica és que l'acumulació de diferències en els llinatges generalment no es produeix de manera uniforme, és a dir, unes branques poden experimentar molts canvis mentre d'altres poden no experimentar-ne cap. Així doncs, degut a que les taxes de divergència poden variar en funció dels escenaris, és necessari que els mètodes d'anàlisi filogenètics tinguin en compte les forces evolutives que actuen entre aquests taxa en els diferents llinatges (Barton *et al.* 2007). La distància genètica proporciona una mesura de la similitud entre les seqüències per inferir un arbre filogenètic. Un requisit important en la reconstrucció d'una filogènia és l'especificació prèvia d'un model de substitució que proporciona una estima d'aquest procés, determinant quin és el model de substitució que millor s'adapta a les seqüències analitzades. Alguns dels models de substitució més utilitzats des del més simple al més complex són: el Jukes-Cantor (1969), el Kimura 2-parameter (1980), el Tamura 3-parameter (1992), el Tamura-Nei (1993) i el GTR (1986). A alguns d'aquests models també se'ls poden afegir dos paràmetres, la distribució gamma (G) i els llocs invariants (I). La distribució gamma consisteix en que no totes les regions de la seqüència evolucionen a la mateixa velocitat, és a dir, les taxes de substitució poden variar al llarg de la seqüència, mentre que els llocs invariants corresponen a algunes posicions que no poden variar, com és el cas dels codons d'inici.

Actualment existeixen diferents mètodes de reconstrucció d'arbres, que es poden agrupar en mètodes basats en distàncies i en caràcters. Els mètodes de distàncies (Felsenstein 1988) es fonamenten en el principi que els organismes que comparteixen un avantpassat comú són, en general, més similars entre si i, per tant, hauria de ser possible inferir relacions evolutives a partir de la similitud de les seves seqüències. Aquest principi és subjacent a tots aquells mètodes de reconstrucció filogenètica basats en distàncies els quals es caracteritzen per realitzar en primer lloc una matriu de distàncies que s'obté de la comparació per parelles de cadascun dels taxa amb totes les altres. En definitiva doncs, els mètodes basats en distàncies estimen el nombre de substitucions que s'han produït a través de l'aplicació d'algoritmes que consideren supòsits diferents sobre la naturalesa dels canvis per calcular la distància entre dues seqüències (*pairwise distance*). A priori es pot pensar que es podria utilitzar un càlcul tan senzill com és el percentatge de posicions canviades entre les dues seqüències que s'estan analitzant, però cal tenir en compte que en l'alineament s'observen els canvis definitius però no els que han afectat a la mateixa posició varies vegades, la qual cosa és important a tenir en compte. Finalment, a partir de la matriu de distàncies obtinguda es construeix l'arbre filogenètic en base al model d'evolució que millor s'adapta a les seqüències estudiades. Els dos mètodes basats en distàncies més emprats són: *Unweighted Pair Group Method with Arithmetic* (UPGMA) i *Neighbor-joining* (NJ). El mètode UPGMA, utilitzat tradicionalment per construir arbres a partir de dades fenotípiques, es basa en la identificació de les parelles més semblants, calculant els valors de mitjana de les distàncies entre elles per construir l'arbre.

Aquest mètode, que actualment s'utilitza també en el cas de filogènies moleculars, genera arbres ultramètrics, per tant cal tenir en compte que els gens utilitzats han d'evolucionar com un rellotge molecular. El mètode NJ és el més utilitzat en les filogènies de seqüències. Els arbres generats no són ultramètrics i per tant, és el més recomanable si no s'està segur que les seqüències analitzades compleixen la hipòtesi de rellotge molecular. Un tercer mètode de distàncies és el *Minimal Evolution* (ME) descrit per primera vegada per Kidd i Sgaramella-Zonta (1971) i posteriorment per Rzhetsky i Nei (1992), que considera com la millor filogènia aquell arbre que minimitza més les longituds de les branques (Lemey *et al.* 2009b).

Pel que fa als mètodes basats en caràcters, utilitzen un alineament múltiple de seqüències i comparen els caràcters per a cada posició de l'alineament. Hi ha diferents mètodes com els de màxima parsimònia (*Maximum Parsimony*, MP), els de màxima versemblança (*Maximum Likelihood*, ML) i els d'inferència bayesiana (IB) (Lemey *et al.* 2009b). El mètode de MP es basa en el principi de la navalla d'Ockham, és a dir, en la filosofia que l'explicació més simple que requereix menys canvis és la correcta. S'obtenen arbres que ordenen les branques de tal manera que es minimitza el nombre de mutacions que han d'haver tingut lloc. El resultat d'aquest algoritme no té perquè ser un únic arbre ja que hi poden haver diversos arbres que continguin el mateix nombre de mutacions.

Una de les limitacions més importants dels mètodes basats en distàncies i els mètodes de parsimònia és que tot i que poden seleccionar un arbre en base a algun criteri, no és possible dir si un arbre és més plausible que un altre, en canvi, els mètodes de màxima versemblança i d'inferència bayesiana proporcionen valors estadístics que permeten determinar-ho.

El mètode de ML, proposat per Felsenstein (1981), es caracteritza perquè pretén trobar l'arbre més versemblant, és a dir, donat un conjunt de dades (seqüències) i uns paràmetres (un model d'evolució), l'objectiu és trobar la topologia d'un únic arbre en el que la longitud de les seves branques maximitzin la probabilitat de les dades. S'utilitzen mètodes heurístics per analitzar solament aquells arbres més raonables, i utilitza la informació filogenètica continguda en un alineament amb major eficiència que els mètodes anteriors. Per tant, amb el mateix alineament tendeix a generar l'arbre més pròxim a la realitat. El principal inconvenient és l'elevat cost computacional (Rizzo & Rouchka 2007).

En la construcció de filogènies, la inferència bayesiana (IB) ha experimentat un creixement explosiu en els últims 10 anys, tot i això en el camp de la microbiologia encara és relativament poc utilitzada. La IB de la filogènia es basa en el càlcul de la probabilitat posterior d'un arbre, és a dir, la probabilitat de que l'arbre (A) sigui correcte donades unes dades (D) i un model d'evolució (θ). La probabilitat posterior es calcula mitjançant el teorema de Bayes, segons el qual la probabilitat posterior d'un arbre $P(A, \theta | D)$ és proporcional a la seva probabilitat prèvia $P(A, \theta)$ multiplicada per la seva versemblança $P(D | A, \theta)$. És a dir, el mètode bayesià, també utilitza la versemblança, però maximitza la probabilitat d'un arbre i un model evolutiu a partir d'unes dades (Felsenstein 2004; Drummond & Rambaut 2007). El càlcul de la probabilitat posterior implica avaluar tots els arbres possibles, i per cada arbre, investigar totes les possibles combinacions de longitud de branques i paràmetres del model evolutiu. Aquest càlcul és computacionalment molt complex, i per això, s'utilitzen les cadenes de Markov

Monte Carlo (MCMC, *Markov Chain Monte Carlo*) per obtenir una aproximació de la distribució de la probabilitat posterior (Yang 2006). Les MCMC són algorismes de simulació estocàstica que eviten el càlcul directe de les probabilitats posteriors, i tenen la propietat de convergir cap a un estat d'equilibri, permetent obtenir una mostra de la distribució posterior (Hastings 1970). La idea central consisteix en partir d'un arbre i canviar certes posicions de manera que el nou arbre obtingut té certa probabilitat de ser acceptat o rebutjat. Si s'accepta, llavors serveix com a punt de partida per a la recerca d'arbres posteriors mentre que si es rebutja s'utilitza l'arbre anterior. D'aquesta manera es genera un registre dels arbres avaluats i es calcula la probabilitat posterior de cadascun d'ells.

A l'inici del procés, els arbres i les longituds de les branques són escollides arbitràriament, és per això que els valors de probabilitat són baixos. Al llarg de l'anàlisi però, la cadena avança cap a regions on la probabilitat és tan elevada que típicament es mesura amb una escala logarítmica. La fase inicial es coneix com *burn-in* i acostuma a ser rebutjada i eliminada de l'anàlisi ja que està fortament influenciada per al punt de partida i encara no s'ha arribat a la convergència. Per tal de determinar la qualitat de l'anàlisi és important confirmar que s'ha arribat a la convergència. Per això s'utilitzen principalment tres vies diferents: el valor d'*Effective Sample Size* (ESS) que és una mesura de l'autocorrelació de les cadenes de MCMC, la comparació de valors al llarg de la mateixa cadena, i la comparació de valors entre diferents *runs*, la més poderosa de totes. Decidir si la topologia de l'arbre ha arribat a la convergència és un dels punts més difícils d'aquest anàlisi. Una forma molt intuïtiva de saber-ho és comparar la variació d'arbres al llarg del mateix *run* i entre *runs* diferents que han estat mostrejats.

Un punt important en la construcció d'una filogènia és l'avaluació de la seva robustesa i fiabilitat. Un mètode filogenètic sempre generarà un resultat (un arbre) a partir de les dades utilitzades, però aquest no té perquè ser fiable. Els mètodes bayesians es caracteritzen perquè en cada node s'obté un valor de probabilitat posterior que identifica la confiança (Drummond & Rambaut 2007; Leaché & Rannala 2011). En els altres mètodes, cal determinar la fiabilitat dels nodes a posteriori. El mètode més utilitzat per a la seva avaluació consisteix en crear rèpliques dels alineaments a partir de l'original i aplicar posteriorment el mètode filogenètic corresponent per generar l'arbre. El pas final serà avaluar per a cada node el percentatge d'arbres en els que aquest node apareix. Aquest valor es denomina *bootstrap*. Els nodes amb un valor elevat de *bootstrap* tenen una elevada probabilitat de ser correctes mentre que aquells que tenen valors baixos poden haver-se generat a l'atzar (Barton *et al.* 2007). El valor límit per considerar un node fiable o no, ha estat objecte de discussions. Evidentment, un node que apareix en el 95% de les rèpliques pot considerar-se sòlid però un que apareix només en el 50% no sembla gaire fiable. És per aquest motiu que s'ha establert que valors superiors al 70% donen lloc a arbres robusts i fiables.

2. INCONGRUÈNCIES A LES FILOGÈNIES

Sovint, es pot observar com els patrons de ramificació dels arbres filogenètics no corresponen a les agrupacions esperades i presenten el que s'anomenen incongruències filogenètiques. Aquestes discordances poden ser explicades per diferents motius com les diferències reals en els processos d'evolució o història, degut a errors de mostreig i també poden aparèixer com a conseqüència de fenòmens com la recombinació o transferència de gens o duplicació gènica (Galtier & Daubin 2008).

Degut a que els bacteris es reproduïen asexualment, durant molt temps es va considerar que la única font de variació era la mutació puntual (Levin 1981). No obstant, fa alguns anys es va veure que aquesta premissa no era certa i la transferència horitzontal de gens mitjançant el procés de recombinació tenia un cert paper en la creació de variació. Avui en dia és totalment acceptat que una de les característiques dels microorganismes és la capacitat d'adquirir gens d'altres organismes (Ochman & Davalos 2006; Cohan & Koeppel 2008), cosa que els ha permès explorar nous ambients i respondre a noves pressions selectives. Hauria estat precisament aquesta capacitat més que la modificació seqüencial de les funcions dels gens per acumulació de mutacions la responsable d'aquestes adaptacions (Davison 1999; Ochman *et al.* 2000). La gran quantitat de genomes sencers disponibles des de fa alguns anys ha permès detectar que una part dels marcs oberts de lectura (ORFs) han estat objecte de transferència horitzontal. Un exemple clar d'aquest fenomen és l'adquisició de gens de resistència a antibiòtics per part de microorganismes patògens per tal de sobreviure i adaptar-se a un nou nínxol ecològic animal o humà tractat freqüentment amb aquests fàrmacs (Brien 2002; Wiedenbeck & Cohan 2011). Els fenòmens de recombinació poden tenir lloc entre diferents gens (recombinació intergènica) o bé es poden produir entre al·lels d'un mateix gen (recombinació intragènica). Encara que existeixen diferents mecanismes per generar recombinants, el resultat d'aquests processos són en gran part iguals (Posada *et al.* 2002).

Determinades espècies bacterianes com *Neisseria meningitidis*, *Staphylococcus aureus*, *Streptococcus pneumoniae* i *Streptococcus pyogenes* presenten taxes de recombinació elevades, en canvi d'altres com *Escherichia coli* presenten uns valors inferiors (Posada *et al.* 2002; Lewis-Rogers 2004). Per tant, la taxa de recombinació pot variar en funció del microorganisme. A més, a diferència d'organismes superiors que només poden patir fenòmens de recombinació amb espècies molt properes, els bacteris són capaços de recombinar amb totes les divisions bacterianes, i fins i tot amb els altres dominis dels organismes vius (Garcia-Vallvé *et al.* 2000; Rest & Mindell 2003; Wiedenbeck & Cohan 2011). La major part dels casos de transferència horitzontal de gens entre bacteris i animals es produeix en aquest sentit, sent molt menys freqüent la transferència d'eucariotes a bacteris. Fins al moment es creu que aquest mecanisme de transferència afectaria a animals invertebrats i seria particularment prevalent en bacteris endosimbionts i els seus respectius hostes. Una gran part dels casos de transferència horitzontal de gens entre endosimbionts i eucariotes té com a protagonista *Wolbachia*, un microorganisme endosimbiont que es troba en una àmplia varietat d'artròpodes, nematodes i d'insectes (Cordaux *et al.* 2001; Duploux *et al.* 2013; Zha *et al.* 2014).

En presència de fenòmens com recombinació, transferència de gens, duplicacions, etc., les històries evolutives d'un determinat grup de taxa no sempre estan ben representades per un arbre. La utilització d'aquest tipus de figura pressuposa que els processos evolutius subjacents són bifurcats i que una vegada dos elements s'han separat mai tornen a interactuar. Per aquest motiu, en els casos en que això no és així, les relacions entre els taxa analitzats s'haurien de representar mitjançant una xarxa (*network*). Les xarxes són, a més, molt útils en altres aspectes ja que al no pressuposar implícitament un procés evolutiu en forma bifurcada, no forcen les dades en una figura d'arbre. La seva utilització permet visualitzar si la forma d'arbre és la que s'ajusta millor o no a les seqüències analitzades (Nakhleh & Ruths; Posada & Crandall 2002; Huson & Bryant 2006; Lemey *et al.* 2009b; Martin *et al.* 2011)

Aquestes xarxes poden classificar-se en dos tipus, els *splits networks* dissenyats bàsicament per a representar incompatibilitats entre seqüències com a conseqüència dels processos anteriorment esmentats i utilitzats en l'anàlisi evolutiu de les seqüències, i els *reticulate networks* dissenyats per representar i confirmar històries evolutives en presència d'aquests fenòmens (Huson & Bryant 2013). En definitiva doncs, les xarxes filogenètiques són una alternativa que a diferència dels arbres filogenètics si que permeten visualitzar diferents històries evolutives aportades per tots aquests fenòmens (Huson & Scornavacca 2011).

Quan la recombinació es produeix entre grups estretament relacionats o es tracta d'un fenomen antic, només una de les històries és inferida. En aquests casos, la filogènia majoritària és en general la que es recuperarà. En canvi, quan aquest fenomen és recent i el punt de recombinació divideix l'alineament en dues regions de longitud similar, s'obté una filogènia diferent a les obtingudes per a cadascuna de les regions per separat (Posada & Crandall 2002). Per tant, en alguns casos la recombinació pot donar lloc a topologies errònies o bé passar desapercebuda. En qualsevol cas és important localitzar els punts on s'ha produït, per tal de determinar-ne l'abast i decidir si cal centrar les anàlisis en determinades regions o cal excloure aquestes seqüències de les anàlisis perquè distorsionen els resultats. Des del punt de vista filogenètic, els episodis de recombinació poden classificar-se en funció de l'efecte que provoquen en la topologia de l'arbre filogenètic com: fenòmens que no modifiquen la longitud de les branques, fenòmens que afecten a la longitud de les branques però no canvien la topologia de l'arbre i fenòmens que canvien la topologia de l'arbre (Wiuf *et al.* 2001; Posada 2002).

1.3 MÈTODES DE DATACIÓ

Els arbres filogenètics construïts a partir de seqüències de DNA contenen solament informació sobre el temps relatiu dels esdeveniments d'especiació. La longitud de les branques d'aquests arbres correspon al producte de la taxa de substitució i el temps transcorregut. Per estimar el temps de divergència de cadascun dels nodes, és necessari separar aquests dos paràmetres per a cadascuna de les branques. Per fer això cal saber com varia la taxa d'evolució en cada branca del arbre. Una vegada es coneix aquesta taxa, és fàcil obtenir el temps de divergència, però primer s'ha de veure si les seqüències evolucionen amb una taxa constant, és a dir, si evolucionen com un rellotge molecular o no. Si les taxes de substitució varien entre les branques cal utilitzar models que considerin variacions en la taxa de substitució en els llinatges

per estimar els temps de divergència relatius, els anomenats models de rellotge molecular relaxat. Una vegada determinat el temps de divergència relatiu, aquest pot ser transformat a temps absolut utilitzant informació externa sobre la datació d'un o més nodes de l'arbre filogenètic per tal de ser utilitzat com a punt de calibratge i així obtenir el temps de divergència absolut. Tradicionalment, els punts de calibratge s'han obtingut a partir d'esdeveniments geofísics o a través del registre fòssil, però aquest no està disponible en totes les espècies, està restringit a 600 milions d'anys i en el cas dels procariotes, és escàs i està confinat a unes poques espècies (Barracough & Nee 2001).

Un exemple de la deducció a través de l'associació amb esdeveniments geològics va ser descrit per Ochman i Wilson (Ochman & Wilson 1987) els qual van aconseguir vincular l'aparició dels llinatges bacterians a esdeveniments a gran escala que han tingut lloc en moments coneguts del passat geològic construint així una escala del temps de l'evolució bacteriana. Concretament, van calcular les taxes absolutes de divergència del 5S i 16S rRNA observant que aquestes eren bastant uniformes en les diferents espècies bacterianes i similars a les observades per a molècules homòlogues eucariotes. En base al temps de divergència del 16S rRNA (1% per cada 50 milions d'anys) van estimar que *Escherichia coli* i *Salmonella typhimurium* (actualment referida com *S. enterica* subsp. *enterica* serovar Typhimurium) haurien compartit un ancestre comú fa aproximadament entre 120-160 Ma. També cal destacar altres estudis els quals han estat capaços d'obtenir informació relacionada amb l'evolució al llarg del temps dels procariotes a través de l'anàlisi de concentracions isotòpiques i detecció de determinats biomarcadors en el registre fòssil (Summons *et al.* 1999; Hinrichs 2002; Brocks *et al.* 2003; Battistuzzi *et al.* 2004) i evidències geològiques com per exemple el fet que els organismes aeròbics haurien evolucionat després que aparegués l'oxigen en l'atmosfera terrestre (Carrine 2004; Battistuzzi *et al.* 2004).

En relació al registre fòssil, l'any 1993 Moran i col·laboradors van ser capaços de determinar el temps de divergència dels membres de l'espècie del gènere *Buchnera*, un proteobacteri, a partir del registre fòssil. Els microorganismes d'aquest gènere són organismes endosimbionts que viuen dins de les cèl·lules dels àfids. Són essencials per al creixement i la reproducció dels seus hostes i els hereten citoplasmàticament. Moran i col·laboradors van determinar que hi ha una concordança completa entre les filogènies moleculars d'aquests bacteris i els àfids, la qual cosa indica una transmissió vertical estricta que s'ha mantingut durant més de 100 Ma. Aquest fet implica una radiació sincrònica i una coespeciació dels microorganismes i els hostes durant aquest temps. Per tant, calibrant l'evolució del 16S rRNA de *Buchnera* amb les dades del registre fòssil dels àfids van arribar a la conclusió que en aquests bacteris s'ha produït una divergència d'entre 1-2% per cada 50 Ma. En aquest cas però, cal considerar que les taxes d'evolució en els organismes simbiotes són més accelerades segurament perquè la mida petita de la seva població promou la fixació de mutacions deletèries a través de la deriva gènica, essent gairebé dues vegades més ràpida que la dels microorganismes del mateix gènere de vida lliure.

Finalment, i des de fa algunes dècades, a partir de seqüències de DNA i aminoàcids, assumint que les molècules distribuïdes universalment han evolucionat amb una taxa similar en totes les formes de vida juntament amb l'extrapolació de les taxes per aquells organismes els quals

disposen de registre fòssil, ha estat possible reconstruir una escala temporal d'evolució (Ochman & Wilson 1987; Doolittle *et al.* 1996; Feng *et al.* 1997; Hedges *et al.* 2001; Sheridan *et al.* 2003)

1.4 PROCÉS DE DIVERSIFICACIÓ DE LES ESPÈCIES AL LLARG DEL TEMPS

La biologia evolutiva és la branca de la biologia que s'encarrega de l'estudi dels canvis que es produeixen com a conseqüència dels processos d'especiació i extinció en la diversificació de les espècies al llarg del temps i els mecanismes que els propicien (Paradis 2003). La quantificació de les taxes d'especiació i extinció contribueix a millorar el coneixement dels mecanismes evolutius que operen en la població (Paradis 2003; Ricklefs 2007).

L'especiació és un procés que condueix a l'acumulació de diversitat en les espècies, i per tant, entendre els patrons i els mecanismes d'especiació és fonamental per explicar la diversitat. Per això és important trobar respostes a preguntes com quines són les causes de l'especiació, com les taxes d'especiació i extinció varien al llarg dels temps i entre grups taxonòmics diferents, etc. En el cas dels microorganismes, contestar aquestes preguntes és difícil (Barracough & Nee 2001).

Des de que Nee i col·laboradors (1994) van proposar un mètode per estimar les taxes d'especiació i extinció d'un llinatge a partir de filogènies reconstruïdes amb els taxa contemporanis, s'han desenvolupat diferents mètodes majoritàriament basat en models de naixement i mort (Sanderson & Donoghue 1996; Nee 2006). En el model més senzill, les taxes de naixement i mort es mantenen constants al llarg del temps. Aquestes taxes però, poden variar durant les radiacions evolutives i entre els llinatges (Rabosky & Lovette 2008; Morlon *et al.* 2010). És per aquest motiu que alguns autors han desenvolupat mètodes que permeten estimar canvis en les taxes de diversificació al llarg del temps i entre els llinatges a partir de les dades filogenètiques de les espècies actuals (Nee *et al.* 1994; Paradis *et al.* 2004; Rabosky 2006a). Aquests mètodes són especialment útils en el cas d'organismes amb un registre fòssil escàs o inexistent, com en el cas dels procariotes.

Els arbres filogenètics derivats de les seqüències de DNA contenen únicament informació sobre el temps relatiu dels fets d'especiació reconstruïts. Si aquestes seqüències no evolucionen d'acord a un rellotge molecular, tal com sembla ser el cas de la majoria de gens, tant en procariotes com en eucariotes (Drummond *et al.* 2006), cal utilitzar, tal com s'ha comentat abans, models que tenen en compte una taxa de variació entre llinatges per obtenir millors estimes del temps de divergència, els anomenats models de rellotge molecular relaxat.

La reconstrucció filogenètica a partir de dades moleculars ha considerat dues premisses importants alhora d'utilitzar filogènies en estudis d'especiació. En primer lloc, per obtenir una visió acurada del fenomen d'especiació és important considerar totes les espècies que conformen el gènere per tal que reflecteixi la història evolutiva completa de tots els integrants i també per tal d'esbiaixar-la el mínim possible (Barracough & Nee 2001). En segon i últim lloc, cal tenir en compte la situació de les espècies incloses en l'arbre filogenètic, és a dir, si es tracta d'un grup monofilètic, parafilètic o fins i tot xarxes filogenètiques (Barracough & Nee

2001). Una limitació de la reconstrucció filogenètica basada en espècies existents és que els llinatges extints no estan representats. Aquesta limitació fa que l'estimació de les taxes d'especiació i extinció a vegades sigui problemàtica (Ricklefs 2007). A més, la datació a partir de dades moleculars també ha estat motiu de controvèrsies, sobretot quan no concorden amb altres dades com és el cas del registre fòssil (Welch & Bromham 2005) ja que com s'ha esmentat anteriorment, aquests mètodes converteixen mesures de distància genètica entre seqüències en estimacions del temps en el qual van divergir.

Un altre repte en l'estimació de les taxes d'evolució molecular i les dates de divergència és obtenir estimacions fiables del nombre real de substitucions que s'han produït en les seqüències analitzades a cadascun dels llinatges des que van divergir a partir d'un ancestre comú. Aquest nombre real de substitucions inclou tant les substitucions observades com les que han estat emmascarades per altres substitucions. Per tal d'acotar millor aquest valor és necessari utilitzar un model de substitució de nucleòtids apropiat (Arbogast *et al.* 2002) ja que la selecció del model determina les estimacions de distància genètica i és crucial per a la precisió dels resultats (Bromham & Penny 2003).

Per realitzar aquest estudi de la diversificació temporal s'han utilitzat mètodes amb plantejaments i fonaments estadístics diferents, com l'anàlisi dels gràfics semilogarítmics del nombre de llinatges al llarg del temps (LTT plot), la determinació de l'estadístic γ de Pybus i Harvey i l'anàlisi del model de cladogènesi que millor s'ajusta a les dades mitjançant una aproximació de màxima versemblança descrita per Rabosky (Nee 2001; Rabosky 2006a; Morlon *et al.* 2010, 2011; Stadler 2010a). A més, s'ha utilitzat també el programa BAMM que permet realitzar anàlisis específiques sobre el model de cladogènesi mitjançant inferència bayesiana.

1.5. EL GÈNERE *Aeromonas*

1.5.1 Característiques generals

El gènere *Aeromonas* (aer- del grec: gas; -monas, unitats; unitats productores de gas) pertany a la Classe *Gamma-Proteobacteria*, Ordre *Aeromonadales* i a la Família *Aeromonadaceae* que actualment inclou 5 gèneres: *Aeromonas*, *Oceanimonas*, *Tulomonas*, *Oceanisphaera* i *Zobellella*. Actualment aquest gènere està constituït per 27 espècies essent l'espècie tipus *Aeromonas hydrophila* subsp. *hydrophila*.

Aquest gènere agrupa bacils gramnegatius, oxidasa i catalasa positius amb un diàmetre d'entre 0,3-1,0 μm i una longitud d'entre 1,0-3,5 μm . Es poden observar aïllats, en parelles o agrupats formant petites cadenes. Normalment presenten mobilitat degut a la presència d'un únic flagel polar. Són anaerobis facultatius, quimioorganòtrofs i el seu metabolisme és respiratori i fermentatiu. Com conseqüència de la degradació de la glucosa i altres carbohidrats produeixen àcid i freqüentment gas. Normalment són arginina dihidrolasa positius i ornitina i lisina descarboxilasa negatius. El test de la ureasa i la fenilalanina desaminasa són positius mentre que els test de la DNasa i la gelatinasa són negatius. També redueixen els nitrats. La majoria d'espècies són fermentadores de carbohidrats com la maltosa, la D-galactosa i la trehalosa.

Són resistents a l'agent vibriostàtic O129 (2,4-diamino-6,7-diisopropilpteridina) i produeixen tot tipus d'enzims extracel·lulars com arilamidases, esterases, amilases, elastases, desoxiribonucleases, quitinases, peptidases i lipases (Abbott *et al.* 2003; Alavandi & Ananthan 2003; Parker & Shaw 2011; Igbiosa *et al.* 2012)

Pel que fa a la temperatura, el gènere *Aeromonas* es divideix en espècies mesòfiles les quals poden créixer fins a una temperatura de 44°C i espècies psicròfiles que es caracteritzen perquè no creixen a partir dels 30°C. La temperatura òptima de creixement es troba entre els 22-30°C, encara que moltes espècies (però no totes) poden créixer a 37°C. El rang de pH òptim es situa entre 5,5-9 però totes les espècies d'*Aeromonas* resisteixen a pH d'entre 4,5-9. Finalment la concentració òptima de clorur sòdic es troba entre el 0-4% (Isonhood & Drake 2002; Janda & Abbott 2010; Parker & Shaw 2011; Igbiosa *et al.* 2012).

La distribució d'aquests microorganismes és ubíqua, aïllant-se d'ecosistemes molt diferents. El seu hàbitat principal és el medi aquàtic, on es poden aïllar tant a partir d'aigua dolça com salada i també a partir d'aigües residuals. Els microorganismes pertanyents a aquest gènere també s'han aïllat de peixos i del tracte intestinal d'altres animals com els rèptils, amfibis, aus i, en el cas dels mamífers d'ovelles i cavalls (Janda & Abbott 2010). A més, algunes espècies s'han aïllat d'altres fonts d'origen alimentari com la llet crua, vegetals, diferents tipus de carns, menjars preparats, etc. (McMahon & Wilson 2001).

El fet que l'interès pel gènere *Aeromonas* hagi augmentat en els últims anys és degut a la seva capacitat per provocar diferents tipus d'infeccions tant en animals com en éssers humans. Les principals espècies causants de malalties en animals són *A.salmonicida*, *A.hydrophila* i *A.veronii* les quals són responsables de la furunculosis en peixos i principalment en salmònids, provocant importants danys econòmics en piscifactories (Han *et al.* 2011). Pel que fa als éssers humans, les malalties més comunament produïdes són la diarrea i la bacterièmia tot i que també s'han documentat casos d'infeccions en ferides i conjuntiva ocular, peritonitis i meningitis, en persones immuno-competents i immuno-compromeses. En aquest cas, les espècies responsables són principalment *A.caviae*, *A.veronii* bv. Sobria, *A.schubertii* (Holt *et al.* 1994; Janda & Abbott 1998, 2010; von Graevenitz 2007; Igbiosa *et al.* 2012; Chao *et al.* 2013). Aquesta capacitat infecciosa és deguda a l'adquisició de diferents factors de virulència que inclouen tant components estructurals (sideròfors, adhesines com les fimbries o pilis, flagel, lipopolisacàrid i proteïnes de membrana externa) com productes extracel·lulars com exotoxines (enterotoxines citotòxiques i citotòniques) i enzims (lipases, proteases, etc) (Janda 1991; Tomás 2012).

1.5.2 Antecedents històrics i situació actual

L'any 1891 Sanarelli va descriure un nou bacteri aïllat de sang i teixit ganglionar d'una granota infectada i que va anomenar *Bacillus hydrophilus fuscus*. Aquest va ser el primer aïllament reconegut d'*Aeromonas*. Posteriorment, l'any 1901 Chester va definir aquest mateix microorganisme com *Bacterium hydrophilum*. L'any 1894 Emmerich i Weibel van descriure el *Bacillus der forellenseuche* i els anys 1896 i 1901 Lehmann i Neumann i Chester, respectivament, van canviar el nom a *Baterium salmonicida*, actualment conegut com

Aeromonas salmonicida. Els anys posteriors, al voltant de l'any 1920, es van descriure soques similars aïllades en diferents animals que van rebre nomenclatures diferents i es van reclassificar en altres gèneres fins que l'any 1936 Kluyver i van Neil van proposar la seva assignació a un nou gènere anomenat *Aeromonas*, nomenclatura que va ser confirmada pels estudis de Stainer l'any 1943 el qual també va reclassificar l'espècie *Proteus hydrophilus* (descrita anteriorment com *Bacterium hydrophilus*) com *Aeromonas hydrophila*.

Al llarg de tots aquests anys, les classificacions dels aïllats es basaven exclusivament en criteris fenotípics. Aquesta forma de classificació va donar lloc al concepte de fenoespècie o complex fenotípic que es pot definir com una espècie o grup d'espècies fenotípicament semblants entre elles però diferents a la resta d'espècies del mateix gènere en base a característiques bioquímiques i morfològiques (Altwegg *et al.* 1990; Carnahan *et al.* 1991; Holt *et al.* 1994). Aquesta classificació va donar lloc a dos grups diferenciats: el grup mesòfil i el grup psicròfil (Saavedra *et al.* 2006; Igbinosa *et al.* 2012).

Per una banda, el grup mesòfil, representat principalment per l'espècie *Aeromonas hydrophila* (però també *A. caviae* i *A. veronii*) agrupava microorganismes mòbils amb un bon creixement a temperatures d'entre 35-37°C associats a diferents infeccions humanes (Janda & Duffey 1988) encara que també podien aïllar-se del medi ambient (aigua dolça i salada) i aliments (begudes, carn crua, peix fresc, etc.) tot i que es creu que la presència d'*Aeromonas* en aquests últims pot ser deguda a una contaminació a l'entrar en contacte amb l'aigua a les plantes industrials o durant la seva preparació. Per altra banda, el grup psicròfil, representat principalment per l'espècie *Aeromonas salmonicida*, es caracteritzava per no ser mòbils, patògens de peixos i tenir una temperatura òptima de creixement entre els 22-25°C (Janda & Abbott 2010; Parker & Shaw 2011).

Tot i que l'any 1943 es reconeix el gènere *Aeromonas*, aquest no apareix publicat com a tal en el *Bergey's Manual of Determinative Bacteriology* fins l'any 1957 (Bergey & Breed 1966). En la setena edició, el gènere *Aeromonas* va aparèixer inclòs en la família *Pseudomonadaceae* format per quatre espècies: *A. hydrophila*, *A. punctata*, *A. liquefaciens* i *A. salmonicida*, les 3 primeres mòbils i l'última immòbil. Durant la dècada dels seixanta es van realitzar nombrosos estudis taxonòmics el resultat dels quals van ser diferents propostes de classificació. L'any 1969 Schubert proposa la classificació del gènere en tres espècies que inclouen 8 subespècies: dues espècies mòbils, *A. hydrophila* formada per tres subespècies (*A. hydrophila* subsp. *hydrophila*, *A. hydrophila* subsp. *anaerogenes* i *A. hydrophila* subsp. *proteolytica*) i *A. punctata* formada per dues subespècies (*A. punctata* subsp. *punctata* i *A. punctata* subsp. *caviae*) i una espècie immòbil *A. salmonicida* formada alhora per tres subespècies més (*A. salmonicida* subsp. *salmonicida*, *A. salmonicida* subsp. *achromogenes* i *A. salmonicida* subsp. *masoucida*). Aquesta classificació va ser publicada en la vuitena edició del *Bergey's Manual of Determinative Bacteriology* però a diferència de l'anterior es va classificar dins la família *Vibrionaceae* (Bergey *et al.* 1974).

Pocs anys després van començar a aplicar-se les primeres tècniques de biologia molecular en la classificació dels bacteris, mitjançant la hibridació DNA-DNA es va demostrar que *A. hydrophila* biotip. *anaerogenes* (descrit anteriorment com una subespècie) pertanyia a

l'espècie *A.caviae* i que les tres espècies *A.hydrophila*, *A.caviae* i *A.sobria* incloïen dos o tres grups d'hibridació que no podien separar-se amb proves bioquímiques (Popoff *et al.* 1981). Aquests resultats van ser publicats per la primera edició del *Bergey's Manual of Systematic Bacteriology* en el qual es van incloure 3 espècies mòbils (*A.hydrophila*, *A.caviae* i *A.sobria*) i una immòbil *A.salmonicida* amb tres subespècies (*A.salmonicida* subsp. *salmonicida*, *A.salmonicida* subsp. *achromogenes* i *A.salmonicida* subsp. *masoucida*) amb diferents grups d'hibridació cadascun. Aquests estudis van ser ratificats per Hickman-Brenner (1987).

Durant molts anys el gènere *Aeromonas* s'ha considerat dins la família *Vibrionaceae* juntament amb els gèneres *Vibrio*, *Plesiomonas* i *Photobacterium*, ja que presenta algunes característiques taxonòmiques semblants a la família *Vibrionaceae* i *Enterobacteriaceae*. No obstant, l'any 1986 Colwell i col·laboradors van proposar, en base a les seqüències dels gens 16S rRNA i 5S rRNA i juntament amb els resultats obtinguts de les proves d'hibridació DNA-DNA, que el gènere *Aeromonas* és considerés com una nova família, la família *Aeromonadaceae*.

L'última edició del *Bergey's Manual of Systematic Bacteriology* descriu el gènere *Aeromonas* dins la Família *Aeromonadaceae* i alhora en un nou ordre anomenat *Aeromonadales* (Martin-Carnahan & Joseph 2005). En aquest nou ordre s'inclou la família *Aeromonadaceae* amb els gèneres *Aeromonas*, *Oceanimonas* i *Tolumonas*, i també la família *Succinivibrionaceae*. En aquesta edició, el gènere *Aeromonas* està constituït per 14 espècies, *A.allosaccharophila*, *A.bestiarum*, *A.caviae*, *A.encheleia*, *A.eucrenophila*, *A.hydrophila*, *A.jandaei*, *A.media*, *A.popoffii*, *A.salmonicida*, *A.schubertii*, *A.sobria*, *A.trota*, *A.veronii* (biovar *Sobria* i *Veronii*), i dos grups de hibridació, *Aeromonas* sp. HG11 i *Aeromonas* sp. HG13.

Actualment, la família *Aeromonadaceae* agrupa també els gèneres *Oceanisphaera* i *Zobellella* (Euzéby 2015). A més, en els últims anys s'ha descrit una gran quantitat d'espècies de manera que a dia d'avui *Aeromonas* consta de 27 espècies: *A.allosaccharophila*, *A.australiensis*, *A.bestiarum*, *A.bivalvium*, *A.cavernicola*, *A.caviae*, *A.dhakensis*, *A.diversa*, *A.encheleia*, *A.eucrenophila*, *A.fluvialis*, *A.hydrophila*, *A.jandaei*, *A.media*, *A.molluscorum*, *A.piscicola*, *A.popoffii*, *A.rivuli*, *A.salmonicida*, *A.sanarelli*, *A.schubertii*, *A.simiae*, *A.sobria*, *A.taiwanensis*, *A.tecta*, *A.trota* i *A.veronii* algunes de les quals contenen diferents subespècies i d'altres tenen espècies sinònimes. Aquest és el cas d'*A.culicicola* i *A.ichthiosmia* que han estat considerades espècies sinònimes d'*A.veronii* i per aquest motiu s'han inclòs dins del mateix grup (Huys *et al.* 2001, 2005), o el d'*A.trota* espècie sinònima d'*A.enteropelogenes* (Huys *et al.* 2002). L'espècie *A.dhakensis* proposada darrerament (Beaz-Hidalgo *et al.* 2013; Oren & Garrity 2014) inclouria les soques prèviament considerades com *A.aquariorum* i *A.hydrophila* subsp. *dhakensis*. Més recentment, aquest més de maig, s'han publicat tres noves espècies d'*Aeromonas* que encara no han estat validades, *A.aquatica*, *A.finlandiensis* i *A.lacus* (Beaz-Hidalgo *et al.* 2015).

L'any 1992 Abbot i col·laboradors van proposar 18 proves bioquímiques les quals permetien identificar una gran part de les espècies del gènere, determinant també que per algunes soques en particular era necessari realitzar proves complementàries. Aquestes 18 proves bàsiques van ser: oxidasa, indol, producció de gas a partir de glucosa, ornitina i lisina descarboxilases, arginina dihidrolasa, hidròlisi d'esculina, Voges-Proskauer, hemòlisi en agar

sang, producció d'elastasa, oxidació de gluconat de potassi, producció d'àcid sulfhídric en medi gelatina-cisteïna-tiosulfat i la fermentació de L-arabinosa, D-mannitol, salicina, manosa, cel·lobiosa i sacarosa.

En els darrers anys s'han utilitzats diferents mètodes moleculars amb l'objectiu de discriminar les diferents espècies d'*Aeromonas*, que en ocasions resulten difícils de separar amb una caracterització fenotípica, com són la hibridació DNA-DNA, el ribotipat 16S rRNA, RADP (*Randomly Amplified Polymorphic DNA*), PCR (*Polimerase Chain Reaction*), AFLP (*Amplified Fragment Length Polymorphism*), RFLP (*Restriction Fragment Length Polymorphism*), PFGE (*Pulsed-Field Gel Electrophoresis*) i PCR múltiplex (Sen 2005). No obstant, alguns d'aquests mètodes han quedat en desús, ja que són molt laboriosos, de baixa reproductibilitat i en la major part dels casos no són prou discriminatoris (Martinez-Murcia *et al.* 1992). Per tant, tot i que al llarg dels anys s'han utilitzat diferents tècniques per portar a terme estudis taxonòmics del gènere *Aeromonas* avui en dia les eines essencials per a la definició d'espècies bacterianes són juntament amb els caràcters fenotípics, les proves d'hibridació DNA-DNA, la seqüenciació del gen 16S rRNA i la de diferents gens conservats (*housekeeping*) (Stackebrandt *et al.* 2002).

No va ser fins als anys 1979 i 1981 que Mac-Innes i col·laboradors i Poppoff, respectivament, van portar a terme els primers experiments utilitzant la tècnica de la hibridació DNA-DNA en *Aeromonas*. Concretament, Poppoff i col·laboradors van classificar 55 soques pertanyents al gènere en diferents grups d'hibridació comprovant que els grups genètics no es corresponien amb els grups fenotípics. Poppoff va observar que les 4 espècies inicialment descrites en el *Bergey's Manual of Systematic Bacteriology* en base a característiques fenotípiques s'inclouien en diferents grups d'hibridació: *A. hydrophila* (HG 1, 2 i 3), *A. caviae* (HG 4, 5A, 5B i 6), *A. sobria* (HG 7, 8 i 9) i *A. salmonicida* (HG3). A partir d'aquest moment apareix el terme genomoespècie o espècies genòmiques que es pot definir com espècie o grup d'espècies genèticament diferents a la resta en relació a la prova d'hibridació del DNA. Es considera que dos aïllats pertanyen a la mateixa espècie quan els valors d'hibridació DNA-DNA són superiors al 70%. En canvi, si els valors obtinguts són inferiors al 70% es considera que pertanyen a espècies diferents.

La seqüenciació del gen 16S rRNA és un dels mètodes més utilitzats alhora de discriminar i determinar la relació filogenètica entre bacteris. En el cas de les espècies d'aquest gènere però, és important tenir en compte que la seqüència aminoacídica d'aquest gen presenta un elevat grau de conservació (97,8-100%) (Martin-Carnahan & Joseph 2005) a part de l'existència de múltiples còpies. Martínez Murcia i col·laboradors (1992) van realitzar un estudi en el qual van utilitzar seqüències codificadores del gen 16S rRNA per construir una filogènia d'*Aeromonas*. Aquest estudi va revelar que algunes espècies com *A. trota* i *A. caviae* diferien exclusivament en 3 nucleòtids al llarg de la seqüència analitzada encara que els valors obtinguts en la hibridació DNA-DNA eren baixos (30%) tenint en compte la distància que les relaciona (Martinez-Murcia *et al.* 1992). En canvi, en el cas d'*A. sobria* i *A. veronii* on la seqüència nucleotídica diferia en 12 bases el resultat de la hibridació DNA-DNA era de l'ordre d'entre el 60-65% indicant una relació clara. Aquestes contradiccions van portar a Sneath a suggerir una possible transferència lateral i recombinació com a possible explicació a aquestes incongruències (Sneath 1993). Malgrat això, en la resta d'espècies, els resultats obtinguts

d'anàlisis filogenètiques preliminars basades en el gen 16S rRNA es van correlacionar bastant bé amb els grups d'hibridació DNA-DNA, però el problema principal va ser la falta de resolució per diferenciar espècies properes (Martinez-Murcia *et al.* 1992). Per aquesta raó, juntament amb el descobriment de l'escassa heterogeneïtat intragenòmica del gen 16S rRNA en el gènere *Aeromonas* es va suggerir que la identificació basada únicament en les seqüències d'aquest gen no era apropiada per a la caracterització de les espècies (Morandi *et al.* 2005).

En conseqüència, es va plantejar la possibilitat d'utilitzar gens conservats per la caracterització de les espècies d'*Aeromonas*. Els gens *housekeeping* són aquells gens que codifiquen proteïnes amb funcions essencials per a la supervivència de la cèl·lula. La informació filogenètica que contenen aquest tipus de gens és superior a la del 16S rRNA ja que tenen una taxa evolutiva major i les variacions es troben distribuïdes per tot el gen. És important que els gens escollits no estiguin sotmesos a transferència horitzontal, es trobin en tots els bacteris i siguin de còpia única. Alguns dels gens *housekeeping* utilitzats en el gènere *Aeromonas* han estat el gen *gyrB* que codifica per la subunitat β de la DNA girasa (Yáñez *et al.* 2003), *rpoD* que codifica pel factor sigma 70 de la RNA polimerasa (Soler *et al.* 2004), *rpoB* que codifica per la subunitat β de la RNA polimerasa (Küpfer *et al.* 2006), *recA* que codifica per una proteïna que intervé en recombinació homòloga del DNA, reparació del DNA danyat i indueix el sistema de resposta SOS (Sepe *et al.* 2008), *dnaJ* que codifica per una xaperona de 40 kDa (Nhung *et al.* 2007) i finalment *cpn60* que codifica per una xaperona tipus I també anomenada Hsp60 o GroEL (Miñana-Galbis *et al.* 2009). Aquests gens han estat utilitzats com a marcadors moleculars amb l'objectiu de clarificar les relacions filogenètiques de les espècies d'aquest gènere. En aquest cas també s'ha observat que alguns d'aquests gens es comporten d'una forma diferent alhora de delimitar algunes espècies donant lloc a resultats ambigus.

Degut als inconvenients acabats de mencionar, posteriorment es va utilitzar una tècnica anomenada tipatge de seqüències multilocus (*Multilocus Sequence Typing, MLST*), un mètode altament discriminatori per a caracteritzar aïllats microbians que es basa en la seqüenciació de fragments interns d'una longitud d'entre 400-500 pb de diversos gens (Maiden *et al.* 1998). El tipatge de seqüències multilocus es basa en el mateix principi que la tècnica d'electroforesi d'enzims multilocus (*Multilocus Enzyme Electrophoresis, MLEE*), però a diferència d'aquesta, l'anàlisi no es basa en la numeració d'al·lels en funció de la seva migració en el gel, sinó que depèn directament de les diferències derivades de les seqüències del DNA (Maiden *et al.* 1998). Aquest tipus de tècnica es va desenvolupar inicialment per estudis epidemiològics i de vigilància mundial, però posteriorment es va portar a terme i amb èxit en l'estudi de poblacions bacterianes (Urwin & Maiden 2003; Sullivan *et al.* 2005).

La tècnica d'anàlisi de seqüències multilocus (*Multilocus Sequence Analysis, MLSA*), una variant de l'MLST, s'ha utilitzat com un mètode alternatiu en microbiologia per delimitar espècies, tan d'*Aeromonas* com d'altres gèneres bacterians. Aquest mètode es basa en analitzar les relacions filogenètiques establertes entre diferents tàxons o un conjunt de soques representatives a partir de la anàlisi de cinc o més gens *housekeeping*. A més, ofereix l'oportunitat d'incorporar un concatenat per obtenir la filogènia (Pascual *et al.* 2010). Una de les avantatges més importants d'aquest tipus d'aproximacions és l'existència de seqüències

disponibles i el fet que s'eviten problemes de manca de reproductibilitat com succeeix en altres tècniques com la hibridació DNA-DNA.

Actualment, la possibilitat de seqüenciar genomes complets degut a l'aparició de noves tecnologies de seqüenciació, està canviant radicalment la oferta de marcadors moleculars útils per a la reconstrucció de filogènies ja que permet utilitzar un nombre més elevat de gens repartits per tot el genoma per tal d'obtenir una visió global. La publicació dels genomes complets de la soca tipus d'*Aeromonas*, *A. hydrophila* ATCC7966^T (Seshadri *et al.* 2006) i la soca A449 d'*A. salmonicida* subsp. *salmonicida* (Reith *et al.* 2008) i posteriorment d'*A. veronii* (Li *et al.* 2011) i *A. media* (Chai *et al.* 2012), van suposar un gran avenç en el coneixement d'aquests microorganismes, demostrant que poden ser molt versàtils metabòlicament i amb una gran capacitat de colonització de diferents hàbitats. Actualment es disposa de genomes de la major part de les espècies que formen aquest gènere: *A. allosaccharophila* (2), *A. australiensis* (1), *A. bestiarum* (1), *A. bivalvium* (1), *A. caviae* (9), *A. dhakensis* (4), *A. diversa* (2), *A. encheleia* (1), *A. enteropelogenes* (3), *A. eucrenophila* (1), *A. fluvialis* (1), *A. jandaei* (3), *A. molluscorum* (1), *A. piscicola* (1), *A. popoffii* (1), *A. rivuli* (1), *A. sanarellii* (1), *A. schubertii* (1), *A. simiae* (1), *A. sobria* (1), *A. taiwanensis* (2), *A. tecta* (1); a més de les espècies mencionades anteriorment, *A. hydrophila* (30), *A. media* (4), *A. salmonicida* (10) i *A. veronii* (12) (Colston *et al.* 2014; *GenBank data base*, NCBI maig-2015, <http://www.ncbi.nlm.nih.gov/>) amb excepció d'*A. cavernicola*. Excepte en el cas dels quatre primers, la majoria d'aquests genomes estan incomplets (*draft genomes*) i per tant, contenen dades preliminars i molt poca informació relacionada amb gens específics implicats en virulència i metabolisme.

1.6 GENS SELECCIONATS: malat deshidrogenasa i recombinasa A

La malat deshidrogenasa (MDH) és un enzim d'entre 30-35 kDa que forma part del cicle dels àcids tricarbòxilics catalitzant la interconversió de malat a oxalacetat. S'han identificat diferents isoformes per aquest enzim, depenent de la seva localització subcel·lular (en el cas dels eucariotes) i la seva especificitat pels coenzims NAD o NADP. Són proteïnes multimèriques en els que les diferents subunitats s'agrupen en dímers o tetràmers, funcionant cada subunitat com una única unitat catalítica independent (Minárik *et al.* 2002) i es troba àmpliament distribuït ja que s'ha detectat la seva presència tant en eubacteris com arquees, fongs, plantes i mamífers, a més de mitocondris i plastis. Les isoformes d'aquest enzim d'alguns eubacteris (*E. coli* i *S. enterica* subsp. *enterica*) presenten una identitat de seqüència relativament alta amb les isoformes mitocondrials d'eucariotes, mentre que d'altres isoformes de procariotes (*Thermus* spp.) semblen estar més relacionades amb els corresponents isoenzims protoplasmàtics d'eucariotes (Goward & Nicholls 1994). A més, curiosament, la seqüència del gen malat deshidrogenasa dels arquees presenta una major identitat amb les lactat deshidrogenases (LDH), suggerint un possible punt d'unió en l'evolució entre les proteïnes MDH i LDH. Aquest fet s'ha vist reforçat per l'estudi cristal·logràfic d'aquest enzim d'un arqueà halòfil (*Haloarcula marismortui*), que ha presentat una estructura tridimensional més similar a la forma tetramèrica de l'enzim LDH que a les dimèriques d'MDH (Richard *et al.* 2000).

En la proteïna MDH es poden diferenciar dues regions: el domini N-terminal (residus 1-150), que correspon a la regió d'unió al coenzim i el domini catalític (residus 151-final). L'estructura primària d'aquestes proteïnes presenta generalment un grau de similitud baix, però hi ha alguns residus que estan absolutament conservats, els que corresponen al punt d'unió al coenzim i els llocs catalítics (Minárik *et al.* 2002).

En relació a la proteïna RecA, aquesta està implicada en els processos de reparació i manteniment del DNA, i a més, té un paper rellevant en la recombinació homòloga. És una proteïna de 38 kDa de la que es coneixen homòlegs estructurals i funcionals en tots els organismes vius, Rad51 en eucariotes i RadA en arquees (Campbell & Davis 1999; Yu *et al.* 2001; Vanloock *et al.* 2003; Wang *et al.* 2008). Aquest enzim és caracteritzat perquè presenta més d'un lloc d'unió a DNA, i per tant pot unir-se al mateix temps a DNA de cadena senzilla i de cadena doble, facilitant així l'entrecruament entre dues regions homòlogues d'aquestes molècules i la conseqüent recombinació. A més, conté un lloc addicional per la unió i la hidròlisi d'ATP, atès que és una ATPasa dependent de DNA (McGrew & Knight 2003; Bell 2005).

En la proteïna RecA d'*E.coli* s'ha descrit la presència de tres dominis estructurals: el domini N-terminal (residus 1-33), el domini central (residus 34-269) i el C-terminal (residus 270-final) (Lusetti & Cox 2002; Xing & Bell 2004a). El primer dels dominis (N-terminal) estaria relacionat amb una de les característiques més importants de RecA, el seu caràcter oligomèric (Egelman & Stasiak 1993), ja que diferents estudis han suggerit que determinades regions d'aquest domini (i també altres factors) afectarien l'auto-associació d'aquestes proteïnes (Masui *et al.* 1997). El domini central, el més gran de tots, es caracteritza perquè conté els llocs d'unió al DNA i a l'ATP (Lusetti & Cox 2002; McGrew & Knight 2003; Bell 2005). Finalment pel que fa a la regió C-terminal, alguns estudis han proposat que modularia al·lostèricament la conformació del domini central (Egler *et al.* 2003; Drees *et al.* 2004; Bell 2005) i jugaria un paper important en la unió al DNA formant part d'una porta d'entrada a través de la qual el DNA accediria al centre del filament (Kurumizaka *et al.* 1996; Lusetti *et al.* 2003a; b). També es poden diferenciar dominis funcionals (d'unió a ATP i DNA), localitzats al domini central de la proteïna. Concretament, en RecA d'*E.coli* s'han descrit tres motius d'unió a ATP, el Walker A localitzat entre els residus 47-74 (Walker *et al.* 1982; Lusetti & Cox 2002), el Walker B situat entre les posicions 140-150 (Karlin & Brocchieri 1996; McGrew & Knight 2003) i finalment el motiu MAW (de l'anglès *Make ATP Work*), que correspon al motiu més gran, localitzat entre les posicions 42-65 (Story & Steitz 1992; Roca & Cox 1997). Per altra banda, també en *E.coli*, s'han identificat dos motius d'unió al DNA, el *Loop* L1 o lloc d'unió secundari i *Loop* L2 o lloc d'unió primari situats en les posicions 157-164 i 195-209, respectivament (Bell 2005). Finalment, aquest domini central també inclouria altres residus amb funcions importants (McGrew & Knight 2003).

2. OBJECTIUS

2. OBJECTIUS

Els objectius d'aquesta tesi doctoral s'emmarquen en l'estudi de la filogènia i l'evolució d'*Aeromonas*, gènere bacterià que inclou espècies patògenes de peixos i altres animals, i que darrerament està adquirint importància com a patògen oportunista humà. Aquest gènere és un bon model per aquest tipus d'anàlisis ja que inclou espècies de vida lliure i patògens oportunistes i obligats.

Els objectius específics són els següents:

- 1.- Avaluar les característiques i la capacitat discriminatòria dels gens *mdh* i *recA* per a la construcció de filogènies fiables del gènere *Aeromonas*.
- 2.- Optimitzar els processos de PCR i de seqüenciació per l'obtenció de les seqüències dels gens *mdh* i *recA* de 150 soques representatives de totes les espècies d'*Aeromonas*, i construir filogènies mitjançant mètodes de màxima versemblança i d'inferència bayesiana.
- 3.- Generar cronogrames a partir de les filogènies obtingudes, utilitzant com a punt de calibratge per a la datació l'edat de divergència d'*E.coli* i *S.typhimurium*.
- 4.- Determinar el model de diversificació temporal del gènere *Aeromonas* juntament amb la taxa de diversificació, a partir de les seqüències consens, les seqüències de les soques tipus i l'arbre d'espècies.

3. MATERIAL I MÈTODES

3. MATERIAL I MÈTODES

3.1 Selecció de les soques

Per a realitzar aquest treball s'ha seleccionat un conjunt de 150 soques pertanyents a totes les espècies i les subespècies del gènere *Aeromonas* (Figura 3.1). L'origen d'aquestes soques és divers, un 68% són ambientals, un 30,7% d'origen clínic i la resta (1,3%) se'n desconeix l'origen. Les principals característiques de cada soca (referència, altres designacions, font i lloc geogràfic de l'aïllament) es detallen a l'Annex I.

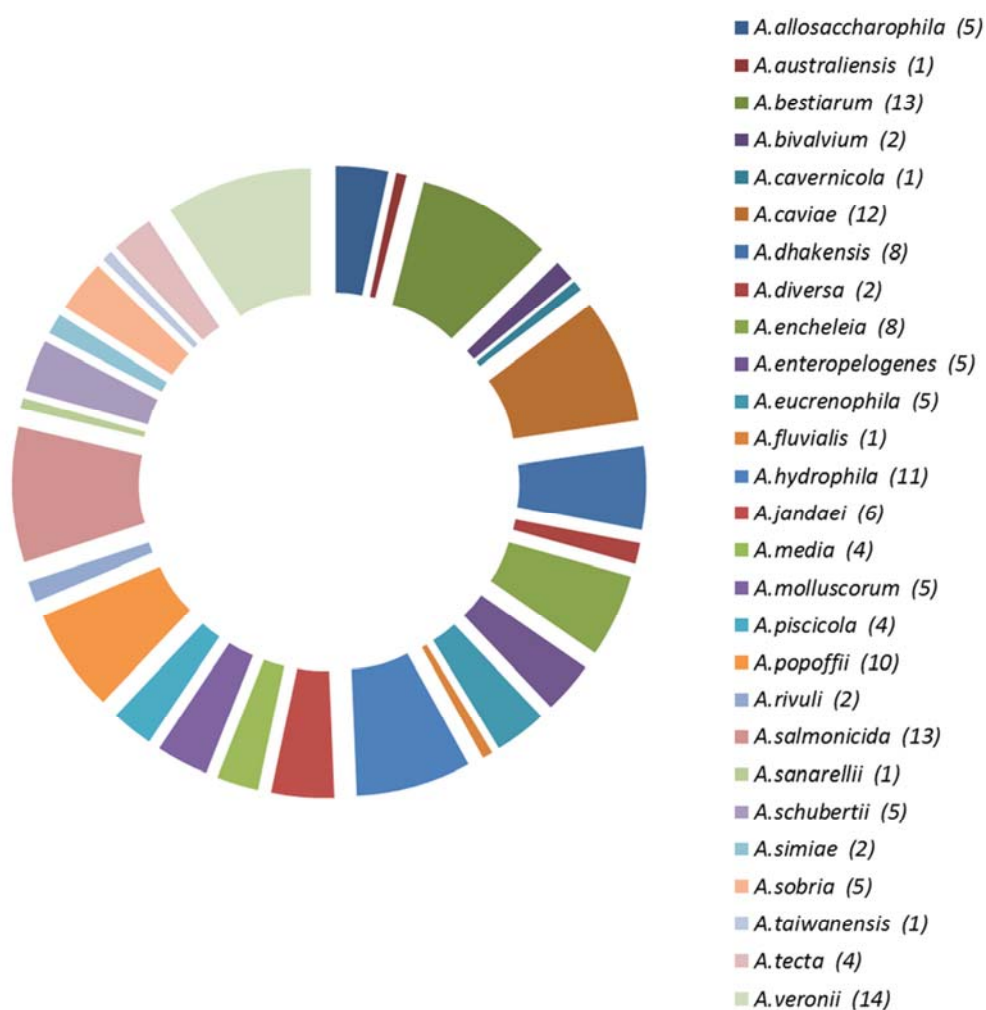


Figura 3.1 Representació gràfica de la proporció de soques de cada espècie del gènere *Aeromonas* de l'estudi. A la llegenda s'indica el nombre de soques per cada espècie entre parèntesis.

La majoria d'aquestes soques provenen de col·leccions de cultius tipus (CIP, *Collection de l'Institut Pasteur*; CECT, *Colección Española de Cultivos Tipo*; JCM, *Japan Collection of Microorganisms*; LMG, *Belgian Co-ordinated Collections of Micro-organisms*), d'altres han estat cedides per diferents investigadors o grups de recerca nacionals i internacionals (Dra. Katri

Berg, University of Helsinki, Finlàndia; Dr. Yogesh Shouche, National Center for Cell Science, Índia; Dra. Margarita Gomila, Universitat de les Illes Balears; Dr. Antonio Martínez-Murcia, Universidad Miguel Hernández; Dra. M^a José Figueras, Universitat Rovira i Virgili, i finalment algunes que han estat aïllades pel nostre propi grup (Miñana-Galbis *et al.* 2002).

3.2 Recuperació de soques

La recuperació de soques s'ha realitzat a partir de congelats en crioboles (Protect™) o en medi líquid (TSB amb un 20% de glicerol) conservats a -40°C o -80°C, i també a partir de liofilitzats. Els principals medis de cultiu que s'han utilitzat han estat el brou de triptona de soja (TSB) i l'agar de triptona de soja (TSA) amb una concentració final de clorur de sodi (NaCl) a l'1%. Per a la recuperació de les soques conservades en crioboles, s'ha agafat una criobola amb unes pinces prèviament esterilitzades, s'ha dipositat en un tub amb 5 mL de TSB, s'ha homogeneïtzat bé i s'ha incubat durant 24 hores a 25-30°C, depenent de la temperatura òptima de creixement de cada soca. Després d'aquesta incubació, s'ha sembrat en estria o per esgotament en una placa de TSA, per tal d'obtenir colònies aïllades. Per a la recuperació de les soques conservades en medi líquid, s'han inoculat 50 µL del congelat en un tub de 5 mL de TSB, i s'ha seguit el mateix procediment que en les crioboles per l'aïllament de colònies.

Les soques liofilitzades es conserven en un vial de vidre estèril tancat hermèticament, a l'interior del qual hi ha el liofilitzat i una tira de cel·lulosa impregnada d'aquest. Un vegada obert el vial en condicions asèptiques, s'han afegit 200 µL de TSB a l'interior per resuspendre el liofilitzat, i s'han inoculat 100 µL en el mateix tub de TSB, per posteriorment incubar-ho. També s'ha agafat la tira de cel·lulosa i s'ha dipositat en una placa de TSA amb ajut d'unes pinces estèrils. Després de la incubació (24h a 25-30°C), s'han resseminat cadascun dels aïllaments obtinguts en estria en medi TSA per tal d'obtenir colònies aïllades. Un cop recuperades les soques, s'han realitzat diverses proves per comprovar que les soques aïllades pertanyen al gènere *Aeromonas*.

3.3 Verificació de soques

Per a la verificació de les soques recuperades, s'han obtingut cultius purs i s'ha realitzat una tinció de Gram i les proves de la catalasa i l'oxidasa. Addicionalment, s'han realitzat altres proves bioquímiques i de sensibilitat a antibiòtics per comprovar la correcta identificació d'aquelles soques que ha presentat incongruències en les filogènies obtingudes. A continuació es detallen el fonament, el protocol i la lectura de les proves realitzades. La composició de tots els medis de cultiu, les solucions, els reactius i els colorants utilitzats s'indiquen en els Annexes II i III.

▪ Tinció de Gram

La tinció de Gram és una prova tradicional que s'utilitza per a separar els bacteris en funció de la composició de la paret cel·lular en dos grans grups: grampositius i gramnegatius. Per fer aquesta tinció es necessiten quatre solucions diferents (Annex III): un colorant bàsic (cristall violeta), un mordent (Iugol), un decolorant (etanol) i un colorant de contrast (safranina). El

mordent és una substància que augmenta i millora l'afinitat entre les cèl·lules i el colorant. La safranina, és un colorant bàsic que s'utilitza per tenyir aquelles cèl·lules que no han retingut el primer colorant després de ser decolorades per l'etanol. Les cèl·lules capaces de retenir el cristall violeta, al microscopi òptic s'observen de color lila i s'anomenen grampositives, mentre que les que s'han tenyit de color rosat són les gramnegatives. Algunes espècies no es tenyeixen de forma evident, igual que els cultius envellits degut a que l'afinitat pels colorants es pot veure alterada.

▪ Prova de la catalasa

La catalasa és un enzim present en molts tipus de cèl·lules, la funció del qual és contrarestar l'efecte tòxic del peròxid d'hidrogen (H_2O_2 , aigua oxigenada) produït com a conseqüència de diverses reaccions d'oxidació-reducció. Per realitzar aquesta prova s'ha dipositat una gota d'aigua oxigenada sobre un portaobjectes net, s'ha afegit una colònia recollida amb escuradents o nansa estèril, i s'ha barrejat. Si transcorreguts uns segons s'observen bombolles, el resultat de la prova és positiu, i el microorganisme té aquest enzim. Per contra, si no s'observen bombolles, la prova és negativa, per tant el microorganisme no presenta aquest enzim (Reddy *et al.* 2007). Com a controls s'han utilitzat *Escherichia coli* ATCC10536 (positiu) i *Streptococcus hirae* ATCC10541 (negatiu).

▪ Prova de l'oxidasa

La prova de l'oxidasa permet detectar la presència o absència de l'enzim citocrom C oxidasa. Concretament, amb aquesta prova, es determina l'oxidació o no del compost tetrametil-p-fenilendiamina. El reactiu de l'oxidasa (Annex III) generalment es prepara al moment, ja que la llum l'oxida ràpidament. Acabat de preparar, aquest reactiu és incolor o lleugerament rosat. Si no es prepara just abans de realitzar la prova, es recomana desar-lo a 4°C protegit de la llum. Per fer aquesta prova s'ha d'agafar una colònia amb una nansa de plàstic o de platí o bé un escuradents estèril, dipositar-la sobre un tros de paper de filtre i afegir-hi una gota del reactiu. Si transcorreguts uns segons s'observa un canvi de color del cultiu a blau-lilós, el resultat és positiu indicant la presència de l'enzim, mentre que si no s'observa cap canvi de color, el resultat és negatiu, i per tant el microorganisme no presenta aquest enzim (Reddy *et al.* 2007). Els controls utilitzats han estat *Aeromonas bestiarum* CECT4227^T (positiu) i *Escherichia coli* ATCC10536 (negatiu).

▪ Voges-Proskauer

La prova del Voges-Proskauer és particularment útil per a la classificació d'enterobacteris. Les peptones que conté el brou VPRM (Annex II) poden ser metabolitzades a través de diferents vies metabòliques. Quan la glucosa és fermentada per la via butanodiòlica el resultat final és un cúmul de productes neutres, com el 3-hidroxi-2-butanona, que són oxidats a diacetil quan reaccionen amb el reactiu de Barrit (format per dues solucions A i B, Annex III) donant una coloració vermellosa al llarg del tub 15 minuts després d'haver-lo afegit. Per cada mil·lilitre de cultiu cal afegir 0,6 mL d' α -naftol (solució A) i 0,2 mL d'hidròxid de potassi (solució B). Cal barrejar energèticament una vegada s'han afegit les dues solucions (Reddy *et al.* 2007). El

temps d'incubació necessari per realitzar aquesta prova és de 24h. Els controls utilitzats han estat *Enterobacter aerogenes* ATCC13048 (positiu) i *Escherichia coli* ATCC10536 (negatiu).

- **Roig de metil**

Per la prova del roig de metil s'utilitza el mateix medi de cultiu que el Voges-Proskauer, el brou VPRM (Annex II). En aquest cas, s'avalua la capacitat dels microorganismes de fermentar la glucosa present en el medi a través de la via àcid-mixta. Els productes final del metabolisme d'aquesta via són àcids orgànics com l'àcid làctic, l'acètic i el fòrmic. Aquests productes fan disminuir el pH del medi de manera que quan s'afegeix el reactiu roig de metil (Annex III), el medi vira a vermell si el resultat és positiu, mentre que manté la coloració groguenca quan el resultat és negatiu (Reddy *et al.* 2007). El temps d'incubació de la prova és de 24h. Com a controls s'han utilitzat *Enterobacter aerogenes* ATCC13048 (positiu) i *Escherichia coli* ATCC10536 (negatiu).

- **Producció de sulfhídric a partir de cisteïna**

Alguns microorganismes es caracteritzen per presentar enzims com la cisteïnasa, la desulfhidrasa i la tiosulfat reductasa capaços d'alliberar els àtoms de sofre dels aminoàcids, els quals són reduïts amb l'hidrogen dels substrats per formar àcid sulfhídric. Aquest medi es caracteritza perquè s'obté barrejant un medi A que conté agar, peptones i clorur de potassi i un medi B que conté la cisteïna i citrat fèrric amònic (Annex II). Es sembla amb nansa de picadura i el temps d'incubació és de 4 dies. La producció de sulfhídric es detecta quan el gas entra en contacte amb indicadors com el citrat fèrric, el sulfat ferrós, el tiosulfat de sodi, l'acetat de plom, etc., donant lloc a un precipitat negre al voltant de la zona de picadura (Popoff & Lallier 1984). Com a controls s'han utilitzat *Aeromonas bestiarum* CECT4227^T (positiu) i *Aeromonas bivalvium* 868E^T (negatiu).

- **Lisina i ornitina descarboxilasa**

Les descarboxilases són uns enzims que actuen sobre els grups carboxils dels aminoàcids, donant lloc a amines i alliberant diòxid de carboni. Cada descarboxilasa és específica d'un aminoàcid. En microbiologia, les proves més utilitzades són la lisina, l'ornitina i l'arginina descarboxilases. Quan el medi s'inocula amb un microorganisme capaç de fermentar la glucosa, es produeix àcid, la qual cosa disminueix el pH canviant el color del medi de lila a groc. Si el microorganisme produeix l'enzim adient, aquesta condició àcida estimula l'activitat descarboxilasa provocant la degradació de l'aminoàcid a la seva corresponent amina. La descarboxilació de la lisina produeix cadaverina, mentre que la d'ornitina dona lloc a putrescina. La presència d'aquestes amines eleva el pH del medi canviant una altre vegada el color de l'indicador que en aquesta ocasió passa de groc a lila. Si el microorganisme no produeix descarboxilasa el medi resta àcid (groc) per la fermentació de la glucosa. Per a realitzar aquesta prova s'ha utilitzat el brou base de descarboxilacions de Moeller amb un 1% de lisina monohidroclorad o ornitina monohidroclorad (Annex II) i s'ha incubat durant 4 dies (Reddy *et al.* 2007). Com a control positiu s'ha utilitzat *Serratia marcescens* 274 i com a control negatiu *Proteus vulgaris* ATCC13315. Degut al medi de cultiu emprat, en aquest cas també ha

estat necessari inocular el medi sense aminoàcid amb el control positiu i el negatiu per evitar falsos positius.

▪ Hidròlisi d'elastina

L'elastasa és un enzim hidrolític que es caracteritza per tenir la capacitat de digerir l'elastina, i que pot estar implicat en la virulència de determinats microorganismes. Per determinar la capacitat d'hidrolitzar l'elastina d'un microorganisme s'utilitza l'agar d'elastina (Annex II) que es sembra per estria amb una nansa degudament esterilitzada. El període d'incubació és de 15 dies. La presència d'un halo transparent al voltant del creixement bacterià indica la hidròlisi de l'elastina essent la prova és positiva (Popoff & Lallier 1984). Com a controls s'han emprat *Aeromonas bestiarum* CECT4227^T (positiu) i *Aeromonas bivalvium* 868E^T (negatiu).

▪ Hidròlisi d'esculina

L'esculina és un glucòsid d'origen vegetal, la hidròlisi de la qual allibera glucosa i esculetina. La prova es realitza amb el medi de cultiu agar esculina (Annex II) que conté peptones utilitzades com a font de carboni, aminoàcids, nitrogen, vitamines i altres nutrients essencials per al creixement bacterià, a més d'esculina. Habitualment també conté bilis per inhibir el creixement de determinats bacteris. Aquesta prova mesura la capacitat d'hidrolitzar l'esculina del medi, de manera que l'esculetina formada reacciona amb el ió fèrric produint un precipitat negre al voltant de la zona de creixement evidenciant la reacció positiva de la hidròlisi (Reddy *et al.* 2007). El temps d'incubació és de 72h. Els controls utilitzats han estat *Streptococcus hiraе* ATCC10541 (positiu) i *Aeromonas sobria* CECT4245^T (negatiu).

▪ Hidròlisi de midó

Alguns microorganismes poden utilitzar midó com a font de carboni i energia gràcies a que contenen amilases que els hi permeten degradar aquesta molècula complexa en monòmers de glucosa. La detecció de la presència d'amilases es realitza mitjançant l'agar midó (Annex II). Aquest medi es sembra en estria i s'incuba durant 24-48h. Per a la detecció de la hidròlisi cal inundar la placa amb una solució de lugol (Annex III) que en contacte amb el medi farà que aquest es torni blau, excepte en aquelles zones on s'hagi produït hidròlisi (Reddy *et al.* 2007). Els controls emprats han estat *Bacillus subtilis* ATCC6633 (positiu) i *Escherichia coli* ATCC10536 (negatiu).

▪ Citrat de Christensen

L'agar citrat de Christensen (Annex II) és un medi que s'utilitza per a la diferenciació d'enterobacteris en base a la capacitat d'utilitzar el citrat com a font de carboni i energia. Els microorganismes que metabolitzen el citrat com a única font de carboni, ho fan a través de l'enzim citritasa donant lloc a oxalacetat i acetat. Posteriorment, l'enzim oxalacetat descarboxilasa transformarà l'oxalacetat a piruvat alliberant diòxid de carboni, que combinat amb el sodi i l'aigua donaran lloc a carbonat sodi, que alcalinitzarà el medi provocant un canvi del color pel viratge de l'indicador roig fenol (Gerhardt *et al.* 1994). El temps d'incubació és

d'entre 1-7 dies. Com a controls s'han utilitzat *Enterobacter aerogenes* ATCC13048 (positiu) i *Escherichia coli* ATCC10536 (negatiu).

▪ Sensibilitat a antibiòtics

S'ha determinat la sensibilitat a antibiòtics mitjançant el mètode de Kirby-Bauer o de difusió en agar (Bauer *et al.* 1966) que consisteix en sembra plaques d'agar Mueller-Hinton (Annex II) amb el microorganisme problema sobre les que es dipositen disquets impregnats amb una concentració coneguda d'antibiòtic (Biomérieux). Posteriorment, després de l'incubació es mesura el diàmetre dels halos d'inhibició (mm). Per interpretar els resultats s'utilitzen unes taules estandarditzades que determinen si el microorganisme és sensible (S), intermedi (I) o resistent (R) enfront dels diferents antibiòtics. En aquest estudi s'han assajat un total de 8 antibiòtics: amikacina 30 µg (AN30), cefalotina 30 µg (CF30), cefoxitina 30 µg (FOX30), eritromicina 15 µg (E15), estreptomina 10 µg (S10), tetraciclina 30 µg (TE30), ticarcil·lina 75 µg (TIC75) i tobramicina 10 µg (NN10).

El protocol realitzat es descriu a continuació:

- Inocular una colònia de la soca a assajar en un tub que conté 5 mL de brou Mueller-Hinton (Annex II).
- Incubar *overnight* a 30°C.
- Mesurar la densitat òptica a 520 nm.
- Quan el creixement es troba entre 0,6-0,8 sembrar en una placa d'agar Mueller-Hinton (Annex II) utilitzant un hisop estèril. La sembra es realitza en tres direccions, per tal d'arribar a tota la superfície de la placa.
- Dipositar en condicions asèptiques 4 disquets d'antibiòtics en cadascuna de les plaques.
- Incubar 18h a 30°C.
- Fer la lectura dels antibiogrames.

3.4 Conservació de soques

En el procés de conservació de soques és important evitar possibles contaminacions per mantenir els cultius purs, assegurar la supervivència fent que sobrevisquin com a mínim entre el 70-80% dels microorganismes, i que es mantinguin genèticament estables. En aquest cas, s'han realitzat conservacions de soques a curt i llarg termini.

Per a la conservació a curt termini s'han sembrat plaques de TSA per esgotament o tubs inclinats de TSA. El temps de conservació en placa és inferior al de tub, el qual permet el manteniment durant 2-4 setmanes perquè la dessecació del medi és menor, i la probabilitat de contaminació també és més baixa.

Per a la conservació de soques a llarg termini, s'han utilitzat dos mètodes diferents (crioboles i TSB amb glicerol) desats a temperatura de congelació. Totes les soques s'han conservat per duplicat o triplicat a una temperatura de -40°C i -80°C. Per a la congelació d'un aïllament s'ha partit d'un creixement confluent d'un cultiu pur en TSA que s'ha incubat a una temperatura de

25-30°C depenent de la soca durant 24 hores. La conservació en crioboles és un sistema comercial (Protect® Microorganism Preservation System, Technical Service Consultants Ltd) que consisteix en un vial que a l'interior conté unes petites boles, a les quals s'adhereixen els microorganismes, immerses en una solució crioprotectora. Una vegada s'ha obtingut el cultiu pur, amb una nansa prèviament esterilitzada s'ha transferit el creixement al vial, s'ha homogeneïtzat, s'ha eliminat l'excés de líquid crioprotector amb una pipeta, i s'ha emmagatzemat a temperatura de congelació. Per a la conservació de soques en medi líquid s'ha utilitzat una barreja de TSB amb un 20% de glicerol, el qual actua com un agent crioprotector, ja que a l'incorporar-lo al medi aquest esdevé hipertònic provocant que la cèl·lula alliberi paulatinament l'aigua per equilibrar el gradient osmòtic. Prèviament s'ha al·liquotat el medi líquid (TSB amb glicerol) estèril en diferents tubs Eppendorfs, s'ha transferit mitjançant una nansa estèril una part del cultiu en placa, s'ha homogeneïtzat, i s'ha congelat fins al seu ús.

3.5 Selecció dels gens

Per a la realització d'aquest treball s'han escollit dos gens: el gen de la malat deshidrogenasa (*mdh*) i el gen de la recombinasa A (*recA*). Ambdós són gens conservats (*housekeeping*), i participen en funcions vitals de les cèl·lules ja que intervenen en processos bàsics per al manteniment de la integritat cel·lular. El gen *mdh* codifica la malat deshidrogenasa, un enzim que forma part del cicle dels àcids tricarbòxics (cicle de Krebs) que catalitza la reacció d'oxidació reversible sobre l'oxalacetat, que dóna lloc a malat. El gen *recA* és important pel procés de recombinació i de reparació del DNA danyat. A la Figura 3.2 es mostra la situació d'aquests gens en el genoma d'*A. hydrophila* ATCC7966^T (GenBank accession number: CP000462; (Seshadri *et al.* 2006)).

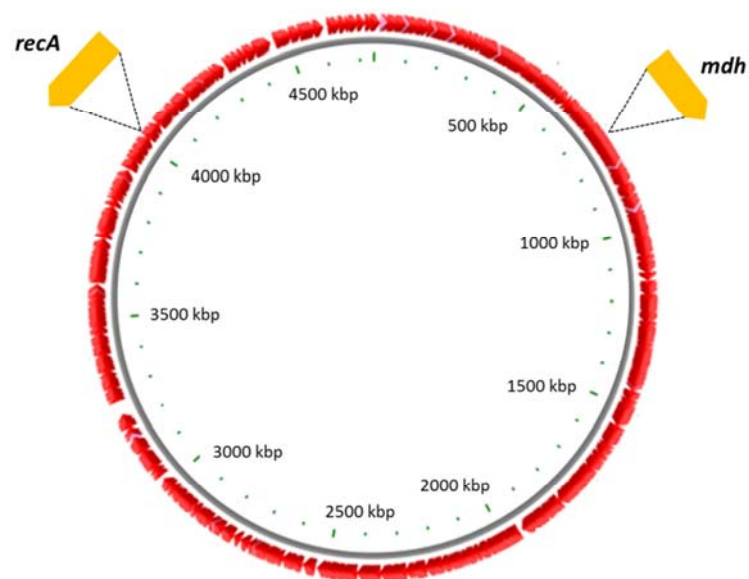


Figura 3.2 Esquema de la localització i l'orientació dels gens *recA* i *mdh* en el genoma d'*A. hydrophila* ATCC7966^T.

3.6 Disseny d'encebadors

El disseny d'encebadors o oligonucleòtids és una de les etapes més importants a tenir en compte alhora de realitzar una amplificació per PCR, ja que els encebadors mal dissenyats poden donar lloc a amplificacions inespecífiques. Per això, sempre que sigui possible cal intentar dissenyar encebadors específics, és a dir, que la complementarietat entre l'encebador i el DNA motlle sigui del 100%. En cas que no sigui possible, es poden dissenyar encebadors degenerats que són aquells que en determinades posicions permeten més d'un nucleòtid.

Algunes de les consideracions que cal tenir en compte per dissenyar bons encebadors són les següents:

- Es recomana que la longitud de l'encebador oscil·li entre 17-22 pb.
- La temperatura de fusió (T_m , *melting temperature*) dels dos encebadors (*forward* i *reverse*) ha de ser semblant, amb una diferència màxima de 8°C i no pot superar els 60°C.
- El contingut en G+C òptim ha de ser entre un 40-60%, i en cap cas pot superar el 60%.
- És recomanable que la seqüència comenci i acabi sempre que sigui possible amb 1 o 2 bases púriques.
- No hi pot haver fenòmens de complementarietat entre les bases del mateix encebador, degut a que podrien formar estructures secundàries i unir-se a llocs inespecífics.
- S'han d'evitar les cues de poli A.
- En el cas que s'hagin d'afegir seqüències addicionals, aquestes s'han d'afegir a l'extrem 5', i no s'han de tenir en compte alhora de calcular la temperatura de fusió.
- A vegades és necessari dissenyar encebadors degenerats. En aquest cas, cal tenir en compte que el risc d'amplificació inespecífica augmenta.

Actualment existeixen diferents programes informàtics que permeten dissenyar encebadors per l'amplificació per PCR. En aquest treball s'ha utilitzat el Primer3 (Untergasser *et al.* 2012, <http://primer3.sourceforge.net/>) desenvolupat per Steve Rozen i col·laboradors (*Whitehead_Institute and Howard Hughes Medical Institute*). Aquest programa permet especificar un gran nombre de variables (com per exemple, discriminar les regions de la seqüència que cal incloure, les que cal excloure, la mida del producte amplificat, els valors mínims de temperatura de fusió, el percentatge en G+C, etc.) i obtenir encebadors en funció de les necessitats.

Per optimitzar les condicions de PCR cal tenir en compte la temperatura d'hibridació (T_a , *annealing temperature*) dels encebadors al DNA motlle, la qual bé donada per la temperatura de fusió (T_m , *melting temperature*), que la proporciona el propi fabricant dels encebadors, en aquest cas InvitrogenTM (Life Technologies) i IsogenTM (LifeScience). També es pot estimar teòricament aplicant la fórmula $T_a = T_m - 5^\circ\text{C}$, on T_m es calcula amb la fórmula de Wallace (Wallace *et al.* 1979), que es defineix com $T_m = 2(A+T) + 4(G+C)$, on A, G, C i T corresponen al nombre de cada nucleòtid present en l'encebador.

Seguint les instruccions del fabricant, cada vial amb l'encebador liofilitzat s'ha resuspès en un 1mL d'aigua Milli-Q estèril i s'ha conservat a temperatura de -20°C fins a la seva utilització. També s'han preparat dues solucions de treball, una per a la barreja de PCR (10 µM) i una altre per a la barreja de seqüenciació (5 µM), ambdues amb un volum final de 500 µL.

3.7 Extracció de DNA genòmic i quantificació

Per a l'extracció de DNA s'ha utilitzat un cultiu pur en TSA (Annex II), s'ha resuspès una colònia en un tub amb 5 mL de TSB (Annex II) i s'ha incubat durant 18-24h a la temperatura òptima de creixement de cada soca. Un cop obtingut el cultiu, s'ha centrifugat una alíquota d'1,5 mL, i a partir del sediment (*pellet*) s'ha iniciat el procés de lisi i extracció del DNA. En aquest treball s'ha utilitzat el kit comercial REALPURE® Genomic DNA Extraction Kit (Durviz), i s'ha seguit el protocol per a l'extracció de DNA de bacteris gramnegatius que recomana el fabricant.

Una vegada s'ha realitzat l'extracció, s'ha determinat la concentració i la qualitat del DNA genòmic mitjançant un espectrofotòmetre (Nanodrop® 1000 Thermo Scientific), que permet analitzar quantitats molt petites de mostra (1-2 µL). Aquesta determinació s'ha fet a una longitud d'ona de 260 nm (màxim d'absorció dels àcids nucleics), ja que en aquesta longitud d'ona, l'absorció serà proporcional a la quantitat de DNA. També s'ha determinat l'absorció a 280 nm (màxim d'absorció de les proteïnes). La relació A260/A280 obtinguda permet conèixer la puresa del DNA extret, ja que la presència de proteïnes a la mostra pot afectar negativament a les reaccions d'amplificació. Es considera que la qualitat del material genètic extret és bona quan el valor obtingut és igual o pròxim a 1.

3.8 Reacció d'amplificació per PCR

Aquest procés va ser descrit per Kary Mullis l'any 1980. Es tracta d'una reacció enzimàtica *in vitro* que permet amplificar fragments específics de DNA. És una metodologia amb una elevada especificitat, sensibilitat i velocitat que s'ha anat millorant al llarg dels anys. Aquesta tècnica es basa en la utilització d'una DNA polimerasa obtinguda a partir d'un bacteri termòfil (*Thermus aquaticus*) que pot actuar a diferents temperatures, permetent d'aquesta manera utilitzar una mateixa polimerasa al llarg de tot el procés. En aquesta reacció es sintetitza una cadena de DNA complementària a una cadena motlle. L'enzim actua afegint nucleòtids al grup OH de l'extrem 3'. La necessitat d'unir-se a una seqüència inicial (encebador) per poder portar a terme la polimerització permet delimitar les regions específiques que es volen amplificar. Al final de la reacció d'amplificació s'obtenen milions de còpies que reben el nom d'amplicons.

La tècnica de la PCR es basa en 3 etapes que es repeteixen cíclicament. Però, comença amb una etapa inicial de desnaturalització i acaba amb una etapa d'extensió final que només tenen lloc una vegada. Cadascun dels 3 cicles dura pocs minuts i es repeteix entre 20-40 vegades. S'han utilitzat diversos termocicladors per realitzar les reaccions de PCR (Thermal Cyler 2720 i Veriti 96-Well Thermal Cyler, Applied Biosystems® (Life Technologies)).

Les etapes de l'amplificació per PCR es descriuen a continuació:

- Desnaturalització inicial: en aquesta etapa s'ha utilitzat una temperatura de 95°C per tal d'activar la Taq DNA polimerasa que requereix activació per calor.
- Desnaturalització: consisteix en separar les dues cadenes de DNA mitjançant un augment de la temperatura d'entre 95-100°C. Es tracta d'una etapa molt crítica, ja que és fonamental que el DNA motlle es desnaturalitzi completament. Si la mostra té un elevat contingut en G+C és recomanable augmentar el temps.
- Hibridació dels encebadors: en aquesta etapa els encebadors s'uneixen a les cadenes de DNA dissociades prèviament. Cada encebador és complementari a una de les dues cadenes de DNA. Els oligonucleòtids s'afegeixen en una concentració adequada, per tal d'afavorir la unió a les cadenes de DNA i evitar així que s'associïn entre ells. En aquesta etapa un augment de la temperatura afavoreix l'especificitat i disminueix les unions incorrectes entre els encebadors i les cadenes complementàries.
- Extensió: una vegada els encebadors s'han associat amb les cadenes de DNA corresponents, la Taq DNA polimerasa comença a afegir nucleòtids complementaris i a formar una nova cadena a partir de l'encebador en direcció 5'-3'. Aquesta etapa s'ha de realitzar a la temperatura a la qual la polimerasa té màxima activitat (72°C). El temps d'extensió depèn de la mida de l'amplicó (s'estima un minut per cada 1000 nucleòtids).
- Extensió final: normalment es realitza un cicle final d'extensió per assegurar que les polimeritzacions de tots els productes d'amplificació s'han completat.

Els components i reactius utilitzats per les optimitzacions de les PCR són:

- DNA motlle.
- Taq DNA polimerasa: s'ha utilitzat l'AmpliTaQ Gold® DNA Polymerase d'Applied Biosystems® (Life Technologies), que és inactiva a temperatura ambient, i que requereix un preescalfament a temperatura de 95°C per la reacció d'activació (*Hot-start*).
- Encebadors: específics i degenerats sintetitzats per Invitrogen™ (Life Technologies) i Isogen™ (LifeScience).
- Desoxiribonucleòtids trifosfat (dNTPs): s'ha utilitzat GeneAmp® dNTP Blend d'Applied Biosystems® (Life Technologies), una barreja equimolar que conté els 4 tipus de dNTPs (dATP, dTTP, dCTP i dGTP).
- Magnesi: en general es necessita una concentració d'entre 0,5-2,5 mM. Aquesta concentració s'ha d'optimitzar per cada assaig, ja que afecta tant a l'especificitat com al rendiment de la reacció de PCR. Elevades concentracions de magnesi poden donar lloc a

unions inespecífiques. En aquest estudi s'ha utilitzat Magnesium Chloride Solution d'Applied Biosystems® (Life Technologies).

- Solució tamponada: s'ha utilitzat la solució tamponada GeneAmp® 10X PCR Buffer I d'Applied Biosystems® (Life Technologies) que conté 500 mM KCl, 100 mM Tris-HCl, pH 8,3, 15 mM MgCl₂ i 0,01% (p/v) de gelatina.
- Aigua: s'utilitza com a solvent de la barreja de reacció, i el més recomanable és l'aigua Milli-Q estèril.
- Adjuvants: en determinades ocasions és necessari afegir-los a la barreja, per tal d'augmentar l'especificitat i el rendiment de la reacció. Els més utilitzats són el dimetilsulfòxid, Tween® 20 o Triton® X-100. En aquest treball s'ha utilitzat dimetilsulfòxid (DMSO, Sigma-Aldrich®).

3.9 Electroforesi en gel d'agarosa

S'han emprat gels d'agarosa a l'1% per comprovar la mida i la qualitat dels amplificats obtinguts per PCR mitjançant un procés d'electroforesi. Per preparar els gels (Annex IV), s'ha utilitzat agarosa (Sigma-Aldrich®) i una solució de Tris-borat-EDTA (TBEx1) com a solució tamponant. Com a revelant s'ha emprat el GelRed™ (Biotium), una alternativa al bromur d'etidi (substància irritant i fortament mutagènica). El GelRed™ actua com un agent intercalant i estructuralment està molt relacionat amb el bromur d'etidi, de manera que les seves propietats òptiques són pràcticament idèntiques. Quan s'exposa a la llum ultraviolada emet fluorescència que s'intensifica fortament després de la unió al DNA amb la qual cosa és possible observar les bandes corresponents al DNA amplificat.

Per a l'obtenció del gel d'agarosa (Annex IV), s'ha barrejat en una ampolla Pyrex de 100 mL, agarosa en pols (0,5 g) i una solució TBEx1 (50 mL) i s'ha escalfat a temperatura elevada al microones fins a dissoldre-ho totalment. Després, s'ha deixat refredar uns segons, s'ha afegit la solució de GelRed™ (3 µL) i s'ha barrejat suaument. Finalment, s'ha transferit al motlle, muntat prèviament i amb les pintes col·locades, i s'ha deixat solidificar durant una hora aproximadament. Transcorregut aquest temps s'han retirat les pintes i s'ha col·locat el gel, que està sobre un suport, en la cubeta d'electroforesi, quedant totalment cobert per la solució tamponada (TBEx1).

Per preparar les mostres de DNA que s'han carregat en el gel, s'ha barrejat de forma homogènia en un tub Eppendorf estèril el producte de PCR, el tampó de càrrega de mostres (Annex IV) i l'aigua Milli-Q estèril en una proporció 5:2:4. La presència de glicerol en el tampó de càrrega augmenta la densitat de les mostres, i facilita que es dipositin uniformement al fons de cada pou i evitant que surin en la solució tamponada de la cubeta. El color blau del tampó de càrrega (blau de bromofenol) permet seguir visualment l'evolució de l'electroforesi.

Amb l'ajut d'una micropipeta s'han carregat totes les mostres en els diferents pous, tenint en compte que el primer pou correspon al marcador de pesos moleculars (REAL® Escala N°1

RBMM1, Durviz), una solució comercial que conté un conjunt de fragments de DNA de mida coneguda (Annex IV), i l'últim un control positiu. Un cop s'han carregat totes les mostres en el gel, s'ha realitzat el procés d'electroforesi, tapant la cubeta i connectant-la a una font d'alimentació a 90 V durant 45 minuts. En acabar, s'ha realitzat la lectura del gel col·locant-lo sobre un transiluminador de llum ultraviolada (Fisher Scientific, Pittsburg, PA, USA), per tal de visualitzar les bandes de DNA. Al comparar les bandes obtingudes amb el marcador de pesos moleculars, s'ha pogut estimar de forma aproximada la longitud (pb) i la concentració (ng) dels amplificats (Annex IV).

3.10 Purificació del producte de l'amplificació per PCR

Per portar a terme el procés de seqüenciació ha estat necessari purificar prèviament el producte amplificat per eliminar les restes dels components de la barreja de PCR que no han reaccionat (encebadors, dNTPs, Taq DNA polimerasa, etc.) i que poden interferir en el procés de seqüenciació. La purificació dels amplificats s'ha realitzat mitjançant el mètode de columnes MSB® Spin PCRapace kit (Invitex, STARTEC Molecular) seguint les instruccions del fabricant. Aquest mètode utilitza una solució tamponant (*Binding Buffer*) que provoca que els fragments amplificats de DNA s'uneixin en el filtre d'una columna, i un cop units, s'elueixen amb un altre tampó (*Elution Buffer*) per obtenir el producte de PCR purificat. Es conserva a temperatura de congelació de -20°C.

3.11 Seqüenciació directa

En aquest treball s'ha realitzat la seqüenciació automàtica de DNA basat en el mètode Sanger, que consisteix en marcar *in vitro* l'amplificat de PCR purificat. En aquesta reacció s'utilitza una barreja de desoxinucleòtids (dNTPs) i didesoxinucleòtids (ddNTPs). La incorporació de ddNTPs bloqueja l'elongació de la cadena la qual cosa dóna lloc a un conjunt de fragments de DNA de longituds diferents. La identitat del nucleòtid que acaba la cadena en cada posició es pot determinar realitzant 4 reaccions d'elongació separades, cadascuna de les quals conté un ddNTP (ddATP, ddCTP, ddTTP i ddGTP) o bé en una única reacció d'elongació combinant els 4 ddNTPs marcant cadascun d'ells específicament. D'aquesta manera el conjunt resultant de molècules es separa en funció de la seva mida mitjançant una electroforesi en gel de poliacrilamida o per capil·laritat, i la seqüència s'obté correlacionant l'ordre dels fragments en l'electroforesi i el ddNTP terminal de cadascun d'ells.

En la seqüenciació automàtica, la utilització d'un marcatge fluorescent (4 fluorocroms diferents per a cadascun dels ddNTPs) permet la detecció del DNA durant el procés d'electroforesi, on un lector làser detecta la fluorescència emesa per cada base. Es seqüencien les dues cadenes complementàries de l'amplificat per separat, cadascuna amb un encebador diferent, per poder validar per comparació la seqüència resultant. La qualitat del DNA és el factor més important per obtenir bons resultats en la reacció de seqüenciació, perquè la presència de restes de components de la PCR pot interferir en el procés. La seqüenciació de fragments de mida petita pot ser problemàtica, per això es recomana que tinguin un mínim de 150 pb.

En aquest cas, el procés de seqüenciació s'ha desenvolupat en dues etapes en llocs diferents. En el laboratori, s'han realitzat totes les reaccions de seqüenciació (per amplificar i marcar fluorescentment el producte de PCR purificat) en un termociclador (Thermal Cycler 2720 i Veriti 96-Well Thermal Cycler, Applied Biosystems®). Cada reacció es caracteritza per ser un procés cíclic (25 cicles) amb les mateixes etapes que la PCR, on es fa ús d'un únic encebador i de terminadors (ddNTPs) marcats amb fluorocroms. En la seqüenciació, la temperatura d'hibridació (T_a) de l'encebador és un grau menor que en la reacció de PCR. La segona etapa d'anàlisi dels productes de la reacció de seqüenciació, s'ha realitzat a la Unitat de Genòmica dels Centres Científics i Tecnològics de la Universitat de Barcelona (CCiTUB), amb el sistema d'electroforesi per capil·laritat ABI PRISM 3700 DNA Analyzer (Applied Biosystems®).

Els components i reactius utilitzats per les optimitzacions de les reaccions de seqüenciació són:

- Producte PCR purificat.
- Encebador: per cada reacció s'utilitza un únic oligonucleòtid. La concentració de la solució de treball és inferior ($5 \mu\text{M}$) a la utilitzada en les reaccions per PCR ($10 \mu\text{M}$).
- Barreja de seqüenciació: en aquest treball s'ha emprat ABI PRISM Big Dye® Terminator Cycle Sequencing v3.1. Aquesta solució conté la DNA polimerasa per amplificar els fragments purificats i també els desoxiribonucleòtids marcats amb fluorocroms (ddNTPs).
- Solució tamponant: s'ha utilitzat Buffer Big Dye®.
- Aigua Milli-Q estèril: fins a un volum final de $20 \mu\text{L}$.

Els problemes més habituals que poden aparèixer durant el procés de seqüenciació i les seves possibles causes s'expliquen a continuació:

- No s'obté seqüència o la senyal decreix poc després de començar.
 - No hi ha DNA o molt menys del necessari.
 - En la reacció de seqüenciació no hi ha encebador, hi ha menys de l'indicat o s'ha degradat.
 - L'encebador no s'uneix a la cadena motlle.
 - Els productes de PCR amplificats no corresponen al producte d'amplificació dels encebadors.
- La senyal és baixa degut a una unió pobre entre l'encebador i el motlle.
 - La T_m de l'encebador és massa baixa perquè té una longitud massa curta o un contingut en G+C baix.
 - L'encebador forma estructures secundàries, particularment a l'extrem 3'.
 - El DNA motlle té una estructura secundària al lloc d'unió de l'encebador.
 - La qualitat o la quantitat del DNA no és l'adequada.

- La seqüència té molt soroll de fons.
 - El DNA presenta una estructura secundària a la zona d'unió a l'encebador.
 - Existeixen seqüències de menor longitud.
 - La mostra conté més d'un amplificant.
 - El producte de PCR no ha estat correctament purificat.

- S'obtenen dues seqüències superposades:
 - L'encebador s'uneix a diferents regions del DNA motlle.
 - Hi ha més d'un DNA motlle en la mostra.
 - El producte d'amplificació no s'ha purificat, o s'ha purificat malament i hi ha restes d'encebador.

- La senyal al principi és intensa però decau ràpidament:
 - Pot ser conseqüència de la presència d'algun contaminant en la mostra.
 - Les quantitats de DNA motlle i encebador no estan equilibrades.

- La seqüència obtinguda és normal, però de cop i volta la senyal decau:
 - El DNA està tallat o digerit i la reacció finalitza.
 - La pròpia seqüència produeix determinats motius que provoquen aquest efecte en la reacció. Normalment es tracta d'estructures secundàries.

- La senyal de la seqüència disminueix abans de l'esperat:
 - La mostra pot contenir masses sals o altres contaminants arrossegats durant el procés d'extracció, purificació, amplificació, etc.

- La part inicial de la seqüència presenta uns pics molt grans:
 - Pot tenir lloc quan els encebadors formen dímers.

- La seqüència es llegeix bé fins que en un determinat moment apareix una barreja de pics o un gran soroll de fons.
 - Apareix quan hi ha més d'una seqüència.
 - S'ha arribat a una zona de poli A o poli T.
 - És una zona rica en GT, GA i homopolímers de G.
 - DNA motlle difícils de seqüenciar (rics en GC o AT, DNA amb estructures secundàries, repeticions, palíndroms, etc.).

3.12 Obtenció de les seqüències nucleotídiques dels gens

Una vegada la Unitat de Genòmica (CCiTUB) ha completat el procés d'electroforesi i el posterior anàlisi, proporciona els resultats de les seqüències en format FASTA i també en forma d'electroferograma o cromatograma, que es pot visualitzar amb programes com el Chromas Lite (versió 2.1, <http://technelysium.com.au/>).

Per tal d'obtenir la seqüència completa de cada gen ha estat necessari acoblar les seqüències obtingudes de cada encebador corresponents a les cadenes *forward* i *reverse*. L'assemblatge

de cada cadena s'ha realitzat a partir dels alineaments obtinguts amb el programa ClustalX (versió 2.0.3, Thompson *et al.* 1997, <http://www.clustal.org/>) i un editor de text (Notepad ++, bloc de notes). Abans de finalitzar, s'han comprovat les bases dubtoses amb els corresponents cromatogrames, i s'ha obtingut la seqüència gènica definitiva per cada soca.

3.13 Anàlisi de les seqüències de cada gen

Per tal de caracteritzar els gens seleccionats en aquest estudi, s'han utilitzat diferents eines bioinformàtiques, com els programes DnaSP (*DNA Sequence Polymorphism*, versió 5.10, Librado & Rozas 2009, (<http://www.ub.edu/dnasp/>) i MEGA6 (*Molecular Evolutionary Genetics Analysis*, versió 6.06, Tamura *et al.* 2013, <http://www.megasoftware.net/>). També s'ha utilitzat el llenguatge i entorn de programació R (versió 3.1.3, <http://www.r-project.org/>) per generar gràfiques i realitzar càlculs estadístics.

A partir de l'alineament de seqüències s'ha determinat la longitud total, el nombre de llocs invariants i de posicions polimòrfiques (totals, llocs variables en una posició i llocs parsimòniament informatius) i el percentatge de contingut en guanina i citosina (G+C %mol) de totes les seqüències amb el programa DnaSP. El nombre de transicions (T_i , canvis entre purines o entre pirimidines), de transversions (T_v , canvis de purina per pirimidina o a l'inrevés), i la taxa T_i/T_v s'han calculat amb el programa MEGA6. Les taxes de substitució sinònima (el nombre de substitucions sinònimes per lloc sinònim, d_S), i no sinònima (el nombre de substitucions no sinònimes per lloc no sinònim, d_N) s'han determinat a nivell intra- i inter-específic amb el mètode de Nei-Gojobori (MEGA6). El grau de saturació de cada gen s'ha estimat per comparació de cada tipus de substitució (d_S o d_N) respecte la distància genètica (p -distance) entre tàxons calculada amb el programa MEGA6.

Per determinar el tipus de selecció que opera en els gens analitzats s'han realitzat varies proves. La prova Z (Z -test) del programa MEGA6 permet determinar, mitjançant un valor de P , si les seqüències analitzades es troben sota selecció positiva, neutra o purificadora. S'ha de rebutjar la hipòtesi nul·la en favor de l'alternativa amb un valor de $P < 0,05$ (Nei & Gojobori 1986). Una segona prova és el Test de Neutralitat de Tajima (Tajima 1989) del programa DnaSP que s'ha aplicat al conjunt de seqüències analitzades per determinar si segueixen un model de neutralitat, o estan afectades per algun tipus de selecció natural. S'ha calculat l'estadístic D de Tajima, i el valor obtingut si és negatiu indica la presència de mutacions deletèries (purificadora), si és positiu que hi ha alguns al·lels sota selecció positiva, i si és igual a zero que es troba en equilibri neutre. Aquest valor D de Tajima és considerat un resultat significatiu estadísticament si s'obté un valor de $P < 0,1$. Una altra mesura de la pressió selectiva que actua sobre un gen codificador de proteïna és la determinació del valor $d'\omega$ (d_N/d_S), on valors $d'\omega > 1$ indiquen selecció positiva, $\omega = 1$ selecció neutre, i $\omega < 1$ selecció purificadora (Yang & Bielawski 2000). També s'ha analitzat la distribució de substitucions sinònimes i no sinònimes codó per codó al llarg de les seqüències a partir d'un alineament traduït utilitzant el programa SNAP (*Synonymous Non-synonymous Analysis Program*, v2.1.1, Kober 2000, <http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html>).

La matriu de distàncies intra- i inter-específiques de les seqüències de cada gen s'ha generat a partir d'un alineament múltiple de seqüències amb el programa MEGA6. Els valors de distància p (*p-distance*) entre parelles de seqüències s'han calculat dividint el nombre de diferències nucleotídiques observades pel nombre total de nucleòtids comparats, sense cap correcció. Finalment, s'ha utilitzat R per representar gràficament els valors de distàncies, i obtenir el gràfic de densitats de distàncies, la correlació entre distàncies intra- i inter-específiques, i la correlació entre els valors de distància i el nombre de soques per cada espècie.

3.14 Anàlisi filogenètica

Per realitzar l'anàlisi filogenètica amb marcadors moleculars, s'han construït les filogènies d'*Aeromonas* per cada gen utilitzant dos mètodes de reconstrucció, el de màxima versemblança (ML) i el d'inferència bayesiana (IB). Abans però, ha estat necessari obtenir i seleccionar les seqüències i fer un alineament múltiple (sense *gaps*), estimar si el conjunt de seqüències evoluciona com un rellotge molecular i calcular el millor model de substitució.

3.14.1 Test del rellotge molecular

S'ha realitzat la prova *Test Molecular Clock ML* del programa MEGA6 (Tamura & Nei 1993; Tamura *et al.* 2013) a partir d'un alineament múltiple de seqüències i un arbre filogenètic de màxima versemblança. El resultat d'aquest test són dos valors de probabilitat, un per al model que segueix la hipòtesi de rellotge molecular estricte i l'altre pel que no. El logaritme dels valors de versemblança obtinguts (L) han estat comparats amb el *Likelihood Ratio Test* (LRT) o test de versemblança, el qual es calcula de la forma següent:

$$LRT = 2 (L_{\text{no rellotge}} - L_{\text{rellotge}})$$

Quan els dos models comparats estan niats, l'LRT segueix aproximadament una distribució χ^2 amb $n-2$ graus de llibertat, on n correspon al nombre de seqüències de l'alineament múltiple. Si el resultat de la prova dona un valor de $P < 0,05$ s'ha de rebutjar la hipòtesi nul·la de rellotge molecular estricte.

3.14.2 Elecció d'un model de substitució

Els models d'evolució de seqüències són utilitzats en les anàlisis filogenètiques amb la finalitat de corregir o compensar les substitucions no observades al llarg de les seqüències. Les diferències nucleotídiques per lloc que s'observen són poc informatives respecte al nombre de substitucions que realment es donen, ja que pot ser que en un mateix lloc s'hagin donat dues o més substitucions. Per tant, quantificar únicament les diferències nucleotídiques observades implica molt probablement subestimar el nombre de substitucions reals que s'han donat per lloc. Aquest fet és especialment important en el cas dels bacteris, perquè fa molt més temps que estan evolucionant en comparació amb els representants eucariotes, i com a conseqüència, acumulant molts més canvis.

Els models d'evolució de nucleòtids es defineixen matemàticament mitjançant dos classes de paràmetres que determinen el canvi, la freqüència de cada nucleòtid i els tipus de substitucions i les seves corresponents taxes de substitució. A partir de les combinacions possibles d'aquests paràmetres s'han descrit més de 200 models diferents, i de tots ells el més complex és el *General Time Reversible* o GTR. A més, per incrementar l'ajust entre el model i les dades, es poden afegir dos paràmetres, la distribució gamma (+G), per modelar l'heterogeneïtat de taxes entre llocs, i/o el nombre de llocs invariants (+I).

En aquest treball, per determinar el model de substitució que millor explica l'evolució del conjunt de seqüències s'ha fet una anàlisi *Find Best DNA/Protein Models ML* amb el programa MEGA6 (Tamura *et al.* 2013), seleccionant el tipus de substitució (nucleòtids o aminoàcids), i donant com a resultat un llistat de models classificats segons els valors de màxima versemblança obtinguts a partir dels diferents paràmetres considerats, que permet escollir el model evolutiu més apropiat per a l'anàlisi.

3.14.3 Mètodes de reconstrucció

Per construir un arbre filogenètic basat en el mètode de màxima versemblança s'ha fet servir el programa PhyML (versió 3.0, Guindon & Gascuel 2003; Guindon *et al.* 2010, <http://www.atgc-montpellier.fr/phyml/>) a partir d'un alineament múltiple de seqüències en format PHYLIP, i utilitzant el millor model de substitució nucleotídica. El suport estadístic dels nodes s'ha avaluat mitjançant una anàlisi de *bootstrap* (Felsenstein 1985) amb 1000 rèpliques. L'arbre de ML obtingut en format .txt s'ha arrelat amb el programa NJPlot (versió 2.3, Perriere & Gouy 1996, <http://doua.prabi.fr/software/njplot>) i s'ha guardat en format NEWICK.

La reconstrucció de filogènies per inferència bayesiana s'ha realitzat amb el programa BEAST (*Bayesian Evolutionary Analysis Sampling Trees*, versió 1.8.1, Drummond & Rambaut 2007; Drummond *et al.* 2012, <http://beast.bio.ed.ac.uk/>), que va acompanyat d'altres programes necessaris per a l'anàlisi, com BEAUti, LogCombiner, i TreeAnnotator. En primer lloc, amb el programa BEAUti s'ha convertit l'alineament de seqüències en format NEXUS a un arxiu .xml que conté les instruccions necessàries per realitzar l'anàlisi d'IB amb el BEAST. Aquesta anàlisi es caracteritza perquè a l'inici cal especificar unes creences prèvies sobre les dades a analitzar, i per això és necessari conèixer algunes característiques. En aquest estudi, s'ha considerat el gènere *Aeromonas* com un grup monofilètic, s'ha utilitzat el millor model de substitució prèviament obtingut, s'ha estimat la freqüència de bases, i s'ha escollit un model de rellotge molecular relaxat log-normal. També, s'han d'especificar alguns atributs de les cadenes MCMC. A la Taula 3.1 es mostren els atributs que s'han fet servir en les diferents anàlisis d'IB, indicant el nombre de generacions en milions (M), el nombre de generacions que ha de passar per fer un mostreig (arbre), i el conjunt final d'arbres que s'ha obtingut.

Un cop generat l'arxiu .xml, aquest s'ha carregat en el programa BEAST i s'ha iniciat l'anàlisi. La cadena MCMC parteix d'un arbre a l'atzar, i els algorismes MCMC modifiquen lleugerament una de les variables de l'arbre inicial (p.ex. posició o longitud d'una branca) i avalua el nou arbre. Cada vegada que es produeix una modificació es considera una generació. Després de cada generació, el canvi proposat s'accepta o es rebutja en funció del valor de la ràtio entre les

probabilitats posteriors dels estats actual i anterior. Si aquesta raó es més gran que 1 (el nou estat és més probable), el canvi s'accepta i la cadena MCMC continua des d'aquest punt. Si pel contrari, la ràtio és menor que 1, el canvi pot acceptar-se o rebutjar-se en funció del valor que pren la ràtio entre probabilitats posteriors. L'aproximació de la probabilitat posterior serà més exacta quantes més generacions s'avaluïn. Durant les generacions inicials de les cadenes MCMC, els arbres solen tenir una probabilitat posterior baixa, com a resultat d'haver començat a experimentar combinacions aleatòries de topologia, longituds de branca i valors de paràmetres. Després d'aquesta fase inicial (*burn-in*), les cadenes MCMC arriben a la convergència, a una fase estacionària on els arbres mostrejats tenen una probabilitat posterior elevada. Al finalitzar l'anàlisi, el programa dóna com a resultat varis fitxers de sortida (*.log*, *.ops*, *.(subs).tree*, *.(time).tree*).

Taula 3.1 Atributs de les cadenes MCMC dels estudis d'inferència bayesiana realitzats per construir filogènies i cronogrames segons diferents aproximacions.

BEAST				
	Subgrup	Nº Generacions	Mostreig	Conjunt d'arbres
<i>mdh</i>	152 seqüències	60 M	3.000	20.000
	Seqüències consens	10 M	1.000	10.000
	Seqüències de les soques tipus	10 M	1.000	10.000
	Arbre d'espècies	100 M	5.000	20.000
<i>recA</i>	152 seqüències	60 M	3.000	20.000
	Seqüències consens	20 M	2.000	10.000
	Seqüències de les soques tipus	20 M	2.000	10.000
	Arbre d'espècies	100 M	5.000	20.000
<i>mdh+ recA</i>	152 seqüències	50 M	5.000	10.000
	Concatenat de les seqüències consens	20 M	2.000	10.000
	Concatenat de les seqüències de les soques tipus	20 M	2.000	10.000
	Arbre d'espècies	100 M	5.000	20.000

Per determinar si s'ha arribat a la convergència, el fitxer de sortida *.log* s'ha visualitzat amb el programa Tracer, el qual assigna per cada paràmetre considerat (p.ex. posterior, *likelihood*, etc.) un valor d'ESS (*Effective Sample Size*). L'ESS és un estadístic que permet estimar la bondat dels resultats obtinguts, valors d'ESS > 200 és consideren bons resultats, mentre que si són inferiors a 100, indiquen que el *run* ha estat massa curt i no s'ha arribat a la convergència.

Una vegada s'ha comprovat que varis *runs* (2-3) han arribat a la convergència amb bons resultats, el segon fitxer *.tree* de cadascun d'aquests *runs*, s'ha carregat en el programa LogCombiner, que permet combinar el conjunt d'arbres de tots els *runs* correctes. Prèviament però, s'han descartat els arbres del *burn-in* de cada *run* (20-30% dels arbres inicials) que no han arribat a la convergència. A partir del conjunt d'arbres amb probabilitat posterior elevada de varis *runs*, i amb ajut del programa TreeAnnotator s'ha obtingut un únic arbre, l'arbre de màxima credibilitat (*maximum clade credibility tree*) en format NEXUS. Aquest arbre resultant s'ha visualitzat i editat amb el programa FigTree (versió 1.5,

<http://tree.bio.ed.ac.uk/software/figtree/>) que permet veure els valors de probabilitat posterior obtinguts en els nodes per avaluar la robustesa de la filogènia.

3.15 Anàlisi de datació

L'estimació de l'edat d'aparició de l'ancestre comú (també anomenat tMRCA, *the Most Recent Common Ancestor*) del gènere *Aeromonas* i dels seus diferents llinatges s'ha portat a terme considerant cada gen de l'estudi per separat i els dos gens conjuntament (seqüències concatenades), a partir de diferents aproximacions: A) les seqüències de totes les soques, B) les seqüències consens per a cadascuna de les espècies, C) les seqüències de les soques tipus de cada espècie del gènere i D) a partir de l'arbre d'espècies.

Els estudis de datació s'han realitzat a partir de filogènies bayesianes amb un punt extern de calibratge amb el programa BEAST (Drummond & Rambaut 2007; Drummond *et al.* 2012), obtenint els corresponents arbres ultramètrics datats o cronogrames. Els temps de divergència absoluts s'han calculat indirectament a partir del temps de divergència estimat entre *Escherichia coli* i *Salmonella typhimurium* de 120-160 milions d'anys (Ma) (Ochman & Wilson 1987). Per totes les anàlisis s'ha utilitzat el valor mig de 140 Ma com a punt de calibratge.

Per generar els cronogrames, en les condicions prèvies que poden ser especificades al programa s'ha considerat el millor model de substitució, s'ha fet una estimació de la freqüència de les bases, excepte en el cas dels arbres d'espècies en que s'han tingut en compte les dades empíriques, s'ha considerat *Aeromonas* com un grup monofilètic, s'ha escollit un model de rellotge molecular relaxat log-normal i el model de Yule com a model d'especiació. Els detalls de les cadenes MCMC per a cadascuna de les aproximacions s'especifiquen a la Taula 3.1.

Per obtenir les seqüències consens corresponents a totes les espècies, sempre i quan es disposi de més d'una soca, s'ha utilitzat el paquet seqinr (Charif & Lobry 2007, <http://cran.r-project.org/web/packages/seqinr/index.html>) d'R. La seqüència consens es genera a partir dels nucleòtids comuns més freqüents en cadascuna de les posicions de les seqüències. Per construir la filogènia, s'ha considerat el conjunt de seqüències consens i les seqüències úniques d'aquelles espècies que només tenen un representant.

S'ha utilitzat el programa DAMBE (*Data Analysis in Molecular Biology and Evolution*, versió 5.3.10, Xia 2013, <http://dambe.bio.uottawa.ca/dambe.asp>) per a l'obtenció de les seqüències concatenades a partir d'un alineament múltiple (ClustalW). Una vegada s'han concatenat les seqüències, ha estat necessari eliminar els gaps, calcular el model d'evolució que millor s'ajusta a les seqüències (MEGA6), i s'ha generat l'arbre corresponent.

En aquest estudi s'han construït arbres d'espècies datats a partir del conjunt de seqüències de cadascun dels gens analitzats i dels dos gens amb el programa BEAST. L'arbre d'espècies representa el patró de ramificació dels llinatges de les espècies mitjançant un procés d'especiació. Els esdeveniments que han tingut lloc en la història d'un gen (duplicacions, insercions, deleccions, recombinació, etc.) no s'han de correspondre necessàriament a la

història evolutiva de les espècies. La metodologia emprada coestima els arbres de gens i un arbre d'espècies comú a tots ells a partir de mostres de varis individus per espècie, i permet l'aplicació de distribucions sobre els valors a priori de gran varietat de paràmetres. Previ a l'anàlisi amb el programa BEAST, s'ha generat el corresponent fitxer .xml amb el programa BEAUti, marcant la casella *Use species tree ancestral reconstruction* (Heled & Drummond 2010). Els priors utilitzats han estat els mateixos que per la resta de cronogrames, a excepció de la freqüència de les bases que s'ha estimat empíricament com recomana el mètode (Heled *et al.* 2013), i els atributs de les cadenes MCMC que es mostren a la Taula 3.1. En el cas de la construcció de l'arbre d'espècies a partir dels alineaments de dos gens, aquests s'han carregat separatament en el programa, i s'ha indicat el millor model de substitució per a cada gen, que pot ser diferent. Un cop generat el fitxer.xml s'ha realitzat l'anàlisi amb el programa BEAST, que en acabar ha donat com a resultat varis fitxers de sortida (.log, .ops, .(subs).tree, .(time).tree, .species.tree). Amb l'últim fitxer *species.tree*, combinant els arbres obtinguts de varis runs (2-3) i eliminant els arbres inicials de *burn-in*, s'ha obtingut un únic arbre que correspon a l'arbre d'espècies.

3.16 Anàlisi de la diversificació

L'anàlisi del model temporal de diversificació del gènere *Aeromonas* s'ha realitzat a partir dels cronogrames obtinguts amb el programa BEAST de les seqüències consens, les seqüències corresponents a les soques tipus i amb l'arbre d'espècies. Cal utilitzar únicament una seqüència per espècie, ja que la inclusió de més d'una seqüència de la mateixa espècie augmentaria artificialment el nombre d'esdeveniments de diversificació a l'extrem dels arbres (Fontaneto *et al.* 2012). Per a l'anàlisi de diversificació, primerament s'ha realitzat un gràfic de l'aparició de llinatges al llarg del temps (*Lineages Through Time* o LTT plot) per tal d'observar el patró de diversificació temporal. A més, s'ha determinat també el model i la taxa de diversificació. Per fer aquestes anàlisis s'han utilitzat funcions dels paquets LASER (Rabosky 2006a, <http://cran.r-project.org/web/packages/laser/index.html>), *ape* (Paradis *et al.* 2004, <http://cran.r-project.org/web/packages/ape/index.html>), TreeSim (Stadler 2010, <http://cran.r-project.org/web/packages/TreeSim/index.html>) i MASS (Venables & Ripley 2002, <http://cran.r-project.org/web/packages/MASS/index.html>) d'R (versió 3.1.3, <http://www.r-project.org/>) i el programa BAMB (*Bayesian Analysis of Macroevolutionary Mixtures*, <http://bamm-project.org/introduction.html>).

3.16.1 LTT plot

Aquest gràfic semi-logarítmic, que utilitza un cronograma, permet observar augments o disminucions en la diversificació des de l'origen o veure si es manté constant. Per obtenir els LTT plots s'han utilitzat els paquets *ape* i LASER d'R (Paradis *et al.* 2004; Rabosky 2006a).

Si el gràfic obtingut (Figura 3.3) segueix un perfil A indica que s'ha produït una disminució en la diversificació des de l'origen. En canvi, si el gràfic s'ajusta al model C indica que s'ha produït un augment de la diversificació des de l'origen. En aquest cas, a l'inici, la taxa de diversificació és baixa i s'incrementa ràpidament en els últims temps. Finalment, si segueix un perfil B, la taxa de diversificació es manté constant al llarg del temps.

Per tal de corroborar que el patró de diversificació temporal obtingut és constant s'ha calculat la corresponent recta de regressió. El pendent de la recta dóna un valor estimat de la taxa de diversificació (λ). També s'han generat 1000 filogènies simulades amb el paquet TreeSim d'R, considerant el model d'especiació de Yule, s'han construït els corresponents LTT plots, i s'ha obtingut un gràfic a on es representen els 1000 LTT plots obtinguts a partir de les simulacions juntament amb el generat a partir de les dades empíriques. En el gràfic s'ha assenyalat amb línies discontinües els límits que inclouen el 95% dels LTT plots obtinguts.

A partir de l'LTT plot empíric obtingut, s'han ajustat les dades a un model de taxa constant, i mitjançant una prova de màxima versemblança s'han estimat els valors de taxa de diversificació (λ) i de taxa d'extinció (μ) amb el paquet LASER d'R.

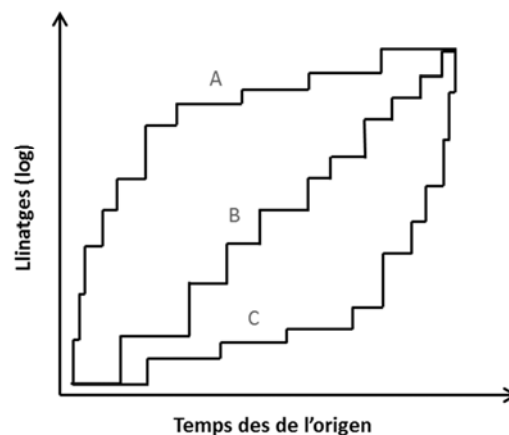


Figura 3.3 Representació gràfica dels tipus de diversificació al llarg del temps (LTT plot).

3.16.2 Models de diversificació

També s'ha determinat el model de diversificació seguit per *Aeromonas* a partir de tres aproximacions diferents: A) el càlcul de l'estadístic Υ de Pybus i Harvey (Pybus & Harvey 2000), B) una anàlisi basada en els valors de ML obtinguts per a diferents models de diversificació amb taxes d'evolució constants i variables, i C) un estudi de les heterogeneïtats en la taxa de diversificació amb el programa BAMM.

3.16.2.1 Determinació de l'estadístic Υ de Pybus i Harvey

S'ha calculat el valor empíric de la Υ de Pybus i Harvey (Pybus & Harvey 2000). Es tracta d'un mètode molt robust i àmpliament utilitzat en aquest tipus d'estudis. Aquest estadístic compara la posició relativa de cadascun dels nodes obtinguts en el cronograma empíric amb la posició esperada sota un model amb taxa de diversificació constant, en el que l'estadístic segueix una distribució normal (Pybus & Harvey 2000). El càlcul d'aquest valor s'ha realitzat a partir dels cronogrames obtinguts en l'anàlisi de datació amb el paquet LASER d'R.

La significació del valor empíric de γ s'ha determinat simulant 1.000 filogènies sota el model de Yule, amb el mateix nombre d'espècies que en l'anàlisi i el valor de λ obtingut empíricament. Aquestes simulacions s'han realitzat amb el paquet TreeSim d'R. Els resultats dels diferents valors de γ (1.000) s'han mostrat en un gràfic, en el que a més s'ha indicat l'ajust a una distribució normal, els límits dels valors del 95% i el valor empíric obtingut.

3.16.2.2 Anàlisi del model de diversificació

Per a realitzar aquesta anàlisi s'han considerat diferents models de diversificació constants i variables. S'ha utilitzat el paquet LASER d'R per fer una anàlisi de màxima versemblança, per tal de veure si la taxa de diversificació canvia al llarg del temps, contrastant els valors de màxima versemblança sota models amb taxes de diversificació constant i models en els que la taxa varia al llarg del temps. Els models analitzats han estat:

Models amb una taxa constant d'especiació (Figura 3.4)

- Model de naixement i mort:

També conegut com *Birth-Death*, es defineix com $N = N_0 \cdot e^{-\lambda t}$. Es caracteritza per tenir una taxa de diversificació o $\lambda > 0$ i una taxa de extinció o $\mu \geq 0$.

- Model de naixement pur:

També conegut com *Pure Birth* o model de Yule es caracteritza perquè la taxa d'especiació o $\lambda > 0$ i la taxa d'extinció o $\mu = 0$.

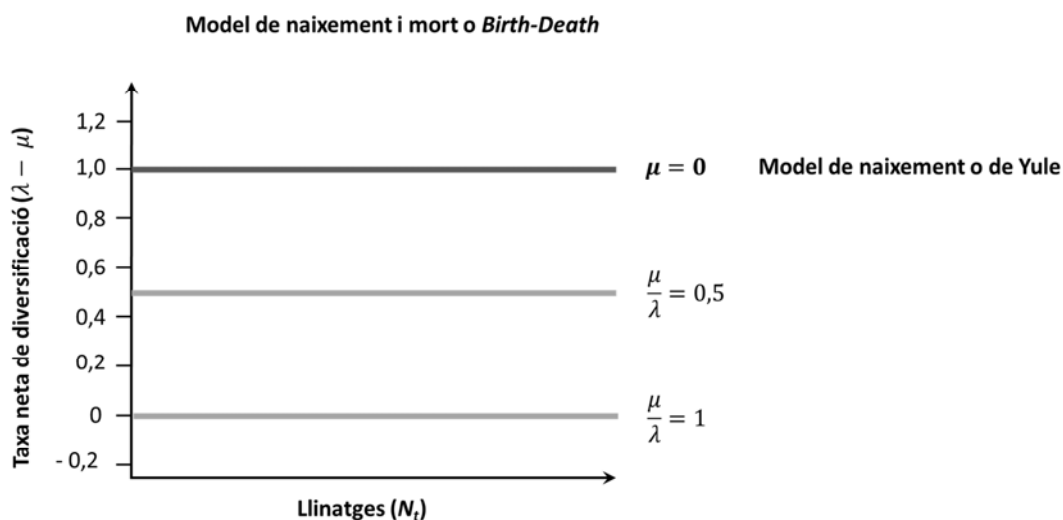


Figura 3.4 Representació gràfica dels models amb taxa constant.

Models amb una taxa de diversificació variable al llarg del temps (Figures 3.5, 3.6 i 3.7)▪ Model logístic dependent de densitat o DDL:

En el qual el valor de λ en el temps es mesura com $\lambda(t) = \lambda_0 (1 - N_t / K)$, on λ_0 és la taxa d'especiació inicial, N_t el nombre de llinatges en el temps t i K és una constant anàloga al paràmetre de capacitat de càrrega de l'ecologia de la població (Figura 3.5).

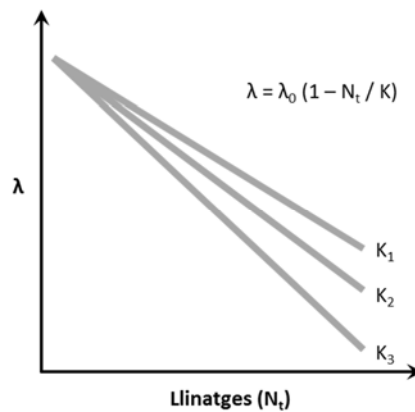


Figura 3.5 Representació gràfica del model DDL.

▪ Model exponencial dependent de densitat o DDX:

El valor de λ en el temps correspon a $\lambda(t) = \lambda_0 N_t^{-x}$, x i controla el valor de la taxa de canvi en el nombre de llinatges en qualsevol punt en el temps (Figura 3.6).

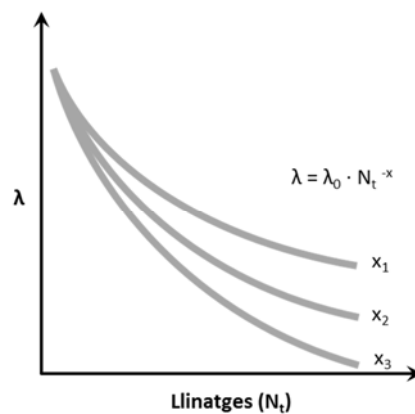


Figura 3.6 Representació gràfica del model DDX.

- Yule 2-rate:

Es caracteritza perquè la taxa d'especiació (λ) experimenta un canvi al llarg del temps de manera que s'observen un punt de ruptura, a partir del qual la taxa d'especiació varia respecte de l'anterior a aquest punt (Figura 3.7).

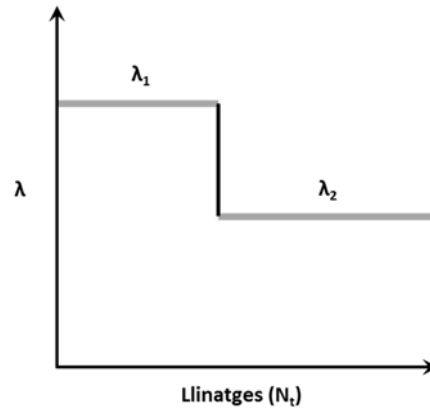


Figura 3.7 Representació gràfica del model de Yule-2-rate.

S'ha utilitzat el test ΔAIC_{RC} (Rabosky 2006a; b) per avaluar estadísticament l'ajust del patró de diversificació temporal d'*Aeromonas* a aquest conjunt de models amb taxes constants i variables. Aquesta prova es calcula com $\Delta AIC_{RC} = AIC_{RC} - AIC_{RV}$, on AIC_{RC} és el valor del criteri d'informació d'Akaike (AIC) pel model de taxa constant i AIC_{RV} el valor d'AIC pel model de taxa variable que millor s'ajusten a les dades. L'AIC és una mesura de la qualitat relativa d'un model estadístic per a un conjunt de dades, que permet calcular més de dos models simultàniament i no requereix que aquests estiguin niats.

Es calcula utilitzant la següent fórmula:

$$AIC = 2k - 2\ln L$$

on k correspon al nombre de paràmetres del model i L al valor de màxima versemblança. El model que millor s'ajusta a les dades observades és aquell que té un valor d'AIC més baix. Aquesta anàlisi de màxima versemblança s'ha realitzat amb el paquet LASER a partir d'un cronograma.

Per tal de determinar la significació estadística del valor ΔAIC_{RC} obtingut, s'ha realitzat el Test de Rabosky (Rabosky 2006b) amb els paquets LASER i TreeSim d'R, basat en la simulació d'un nombre de filogènies suficientment elevat (5.000) sota la hipòtesi nul·la de model de Yule ($\lambda \geq 0$ i $\mu = 0$), considerant també el nombre d'espècies de l'anàlisi i el valor de λ obtingut empíricament. El valor d' ΔAIC_{RC} proporciona un error de tipus I amb un $\alpha = 0,05$, corresponent al percentil 95% de la distribució d'aquests valors sota la hipòtesi de model constant.

3.16.2.3 Anàlisi d'heterogeneïtats de taxes de diversificació

S'ha utilitzat el programa BAMB (*Bayesian Analysis of Macroevolutionary Mixtures*, <http://bamb-project.org/>), que aplica la inferència bayesiana per estudiar i quantificar la dinàmica d'especiació, extinció i evolució d'arbres filogenètics. Per portar a terme una anàlisi amb aquest programa s'ha utilitzat un cronograma, a més d'un arxiu d'entrada en el qual cal especificar el model de diversificació, el nombre de generacions de MCMC i cada quan es pren la mostra, entre altres paràmetres. En aquest cas, s'han generat MCMC de 20M, 50M o 100M segons el subgrup, i després del mostreig s'han obtingut un total de 10.000 arbres. Els resultats d'aquestes anàlisis es guarden en 5 fitxers de sortida, un dels quals permet determinar si s'ha arribat a la convergència i els altres contenen la resta de resultats, que han estat analitzats i interpretats amb els paquets BAMBtools (Rabosky *et al.* 2014, <http://cran.r-project.org/web/packages/BAMBtools/index.html>) i *ape* d'R.

S'han obtingut els valors de taxa de diversificació de cadascuna de les branques de l'arbre, s'ha determinat la probabilitat de que la taxa d'especiació és mantingui constant al llarg del temps o de que presenti un o més canvis, i s'ha calculat el factor de Bayes (Kass & Raftery 1995) associat a la constància d'aquesta taxa avaluant l'evidència a favor de la hipòtesi nul·la. Els valors de referència per interpretar el valor obtingut són: 1-3,2 (lleugera), 3,2-10 (substancial), 10-32 (forta), 32-100 (molt forta) i >100 (decisiva). Finalment, s'ha representat gràficament la taxa neta de diversificació des de l'aparició de l'ancestre.

3.16.3 Càlcul teòric de l'origen de diversificació i el nombre mig d'espècies

S'han utilitzat els paquets TreeSim i MASS d'R per tal de determinar el valor teòric del tMRCA del gènere *Aeromonas* per a cadascuna de les aproximacions considerades en l'estudi a partir del nombre d'espècies analitzat i el valor de λ obtingut empíricament. A més, s'han realitzat 10.000 simulacions i s'han determinat els corresponents intervals de confiança inferior i superior del 95%.

També s'ha calculat el nombre teòric mig d'espècies que tindria *Aeromonas* des de l'aparició de l'ancestre comú fins a l'actualitat (10.000 simulacions), amb els mateixos paquets TreeSim i MASS d'R. L'estimació s'ha fet considerant que s'ajusta al model estocàstic de Yule, que segueix una distribució binomial negativa, i tenint en compte els valors d'edat de diversificació i de λ empírics. També s'ha calculat la desviació estàndard (SD) d'aquest valor mig d'espècies esperat.

3.17 Detecció de recombinació

Per tal de visualitzar en les seqüències la presència d'indels, regions repetides i/o invertides, o recombinants s'han realitzat alineaments múltiples de seqüències amb el programa ClustalW (Thompson *et al.* 1994) del paquet informàtic MEGA6 (versió 6.0.6, Tamura *et al.* 2013, <http://www.megasoftware.net/>) i diagrames de punts ó dots plots, que són representacions gràfiques que s'han generat mitjançant les funcions dels paquets *phyclust* (Wei-Chen 2011) i *ape* (Paradis *et al.* 2004) d'R (versió 3.1.3, <http://www.r-project.org/about.html>). Per obtenir el

diagrama de punts s'ha utilitzat un alineament múltiple de seqüències en format FASTA i un arbre filogenètic en format NEWICK. El dots plot és un gràfic que indica els llocs polimòrfics o segregants de cada seqüència de l'anàlisi al comparar-la amb la seqüència consens, de manera que si la base nucleotídica és la mateixa es deixa en blanc i si és diferent es marca amb un color segons la base que varia. La seqüència consens es genera a partir de totes les seqüències incloses en l'estudi i es mostra a la part superior del gràfic.

Per realitzar l'estudi de recombinació, s'han emprat mètodes computacionals, més actuals, i que en general compten amb un suport estadístic per validar els resultats obtinguts, juntament amb mètodes tradicionals, utilitzats des dels primers estudis de recombinació, que inclouen determinacions com el contingut en G+C, l'ús de codons, la detecció d'incongruències filogenètiques i d'estructures en mosaic.

3.17.1 Mètodes computacionals

Existeixen una gran varietat de mètodes computacionals desenvolupats amb l'objectiu de detectar la presència de recombinació, identificar la localització i la longitud del fragment recombinant, i determinar el seu possible origen. En l'estudi de recombinació s'han utilitzat 3 programes diferents. Un primer programa (RAT) basat en distàncies, que no està recolzat estadísticament, i que s'ha utilitzat com una eina de cribratge. Un segon programa (RDP4) format per un conjunt de mètodes per detectar recombinació, que aporta informació qualitativa i quantitativa, així com la significació estadística dels resultats. Un tercer programa (*SplitsTree*) que pot evidenciar la presència d'un possible esdeveniment de recombinació en les seqüències a l'observar l'estructura del gràfic generat, en el que el resultat obtingut està recolzat estadísticament.

El programa *Recombination Analysis Tools*, també anomenat RAT, és una aplicació de java de lliure accés (versió 1.0, Etherington *et al.* 2005, <http://cbr.jic.ac.uk/dicks/software/RAT/>) desenvolupat pel *Dicks Computational Biology Group (Department of Computational and Systems Biology, Norwich, UK)* per l'anàlisi d'alt rendiment de seqüències amb recombinació. Aquest programa permet realitzar cerques automàtiques d'esdeveniments de recombinació a partir d'alineaments múltiples de seqüències en format FASTA. Els resultats es presenten en forma d'un gràfic basat en distàncies, mostrant el traçat dels valors de distància obtinguts al llarg de tota la seqüència problema al comparar-los amb els de la resta de seqüències de l'anàlisi. En presència de recombinació, es poden identificar els punts d'inici i final d'entrecruament entre seqüències, i el possible origen del fragment recombinant (parental recombinant o *minor parent*).

El paquet informàtic *Recombination Analysis Detection Program* és de lliure accés (RDP4, versió 4, Martin *et al.* 2010, 2015, <http://web.cbio.uct.ac.za/~darren/rdp.html>) i té implementats 12 mètodes de detecció de recombinació (Taula 3.2). Aquest programa és una eina útil i altament automatitzada que permet analitzar alineaments d'un gran nombre de seqüències amb diferents programes de detecció simultàniament. De forma general, aquests mètodes es classifiquen en cinc categories: similitud, distància, filogenètic, compatibilitat i distribució de les substitucions nucleotídiques (Posada *et al.* 2002). Els mètodes de similitud es

basen en la detecció d'un nombre més elevat de substitucions sinònimes en les regions variables respecte les regions conservades (Ohta & Basten 1992). Pel que fa als mètodes de distàncies, es basen en la detecció d'inversions en els patrons de distància entre seqüències (Weiller 1998), amb la qual cosa no és necessari conèixer la filogènia, i són molt eficients computacionalment (Posada *et al.* 2002). Els mètodes filogenètics es caracteritzen per inferir la recombinació a través de genealogies amb tipologies discordants o quan s'agrupen gens ortòlegs de diferents espècies (Hein 1990; Zhou *et al.* 1997; Martin & Rybicki 2000; Husmeier & Wright 2001). Els basats en compatibilitat també detecten incongruències filogenètiques però no és necessària una filogènia de les seqüències estudiades (Jakobsen *et al.* 1997). Finalment, els mètodes basats en la distribució de les substitucions inclouen estratègies que analitzen les substitucions significatives de les seqüències o analitzen si s'adapten a una distribució estadística esperada (Gibbs *et al.* 2000; Mcvean *et al.* 2002).

Taula 3.2 Característiques dels principals mètodes computacionals de detecció de recombinació implementats en el paquet informàtic RDP4 (Martin *et al.* 2010, 2015). Les columnes A, B i C indiquen la identificació de parentals (A), la detecció i la identificació dels punts d'inici i final del fragment recombinant (B), si presenten suport estadístic (C). Els programes que s'han utilitzat en aquest treball estan marcats en gris.

Mètode	Implementació	Categoria	A	B	C	Referència
RDP	RDP4	Filogenètic	+	+	+	(Martin & Rybicki 2000)
GENECONV	GENECONV - RDP4	Substitució	+	+	+	(Padidam <i>et al.</i> 1999)
BootScan	PHYLIP - RDP4	Filogenètic	+	+	+	(Salminen <i>et al.</i> 1995)
MaxChi	RDP4	Substitució	+	+	+	(Maynard Smith 1992)
Chimaera	RDP4	Substitució	+	+	+	(Posada & Crandall 2001)
SiScan	RDP4	Substitució	+	+	+	(Gibbs <i>et al.</i> 2000)
3Seq	RDP4	Substitució	+	+	+	(Boni <i>et al.</i> 2007)
LARD	LARD - RDP4	Filogenètic	-	+	+	(Holmes <i>et al.</i> 1999)
Distance Plot	PHYLIP - RDP4	Distància	-	+	-	(Martin <i>et al.</i> 2010)
PhylPro	RDP4	Distància	+	+	-	(Weiller 1998)
DSS/TOPAL	PHYLIP & SEQ-GEN - RDP4	Filogenètic	-	+	+	(McGuire & Wright 2000)
VisRD	RDP4	Filogenètic	+	+	-	(Lemey <i>et al.</i> 2009a)

Per fer aquesta anàlisi de recombinació amb el programa RDP4 s'ha requerit d'un alineament múltiple de seqüències, i s'han aplicat els següents 7 mètodes de detecció: RDP (Martin & Rybicki 2000), GENECONV (Padidam *et al.* 1999), BootScan (Martin & Posada 2005), MaxChi (Maynard Smith 1992), Chimaera (Posada & Crandall 2001), Siscan (Gibbs *et al.* 2000) i 3Seq (Boni *et al.* 2007). Per cada mètode, els esdeveniments de recombinació detectats es mostren gràficament, i proporciona un valor de P que indica el suport estadístic dels resultats. El nivell de significació utilitzat és de 0,05, per contrastar dues hipòtesis, la hipòtesi nul·la d'absència de recombinació i l'alternativa de presència de recombinació. Per tal de descartar possibles falsos positius és necessari que un esdeveniment de recombinació sigui detectat per dos o més mètodes amb significació estadística ($P < 0,05$), i a ser possible, pertanyents a categories

diferents (Posada *et al.* 2002). A més, també permet identificar els punts d'inici i final de la regió recombinant (*breakpoints*) i els possibles parentals (*major and minor parents*).

El programa *SplitsTree* és un programa de lliure accés (versió 4.13.1, Huson & Bryant 2006, <http://www.splitsree.org/>) que permet realitzar una anàlisi filogenètica a partir d'un conjunt de tàxons representats per caràcters, distàncies, arbres filogenètics, etc. en una gran varietat de formats (NEXUS, PHYLIP, FASTA i NEWICK, entre altres) (Huson 1998). Aquest programa permet detectar fenòmens de recombinació a través de la construcció de xarxes filogenètiques (*splits networks*) i l'aplicació de la prova *Phi Test* (*Pairwise Homoplasy Index*) (Bruen *et al.* 2006) amb significació estadística. En aquest treball s'ha utilitzat el mètode *MedianJoining*, que a partir de les seqüències alineades calcula la matriu de distàncies i genera directament una xarxa filogenètica. Al visualitzar l'estructura construïda, si presenta una malla o xarxa, indica que podria ser degut a un procés de recombinació. Després, cal comprovar la significació dels resultats obtinguts amb la prova *Phi Test*, que dóna un valor de *P*. Si el resultat obtingut és inferior a 0,05, es rebutja la hipòtesi nul·la de no recombinació, i es recolza l'evidència de recombinació en les seqüències analitzades.

3.17.2 Mètodes tradicionals

El contingut en guanina i citosina (G+C) és una característica de cada genoma, i generalment s'expressa en percentatge (%mol). Aquest paràmetre s'ha calculat a partir d'un alineament múltiple de seqüències amb el programa DnaSP (versió 5.10, Librado & Rozas 2009, <http://www.ub.edu/dnasp/>). Els codons de terminació han estat assignats com a posicions no codificants.

Les determinacions de la composició d'aminoàcids i d'ús de codons s'han realitzat amb el programa MEGA6 (versió 6.0.6, Tamura *et al.* 2013, <http://www.megasoftware.net/>), a partir d'un alineament de les seqüències traduïdes. S'ha calculat la freqüència relativa de cadascun dels 20 aminoàcids per a una seqüència específica o per un conjunt de seqüències, donant el seu valor mig. També s'ha determinat l'ús de codons per una seqüència o grup de seqüències, estimant els codons més freqüentment utilitzats per codificar cada aminoàcid.

Les incongruències detectades a partir de les filogènies construïdes (p.ex. soques mal agrupades) permeten sospitar d'un possible esdeveniment de recombinació. En aquest estudi, s'han construït arbres filogenètics pel mètode de màxima versemblança (Apartat 3.14.3), considerant les seqüències d'un gen (*recA* o *recX*) o de regions intragèniques (no recombinant i recombinant), utilitzant el millor model de substitució que s'ajusta a les seqüències, i fent una anàlisi de *bootstrap* que és una mesura de la robustesa de la filogènia obtinguda.

3.18 Predicció i construcció del model tridimensional d'una proteïna

La predicció de l'estructura d'una proteïna consisteix en l'obtenció de la seva estructura tridimensional a partir de la seqüència d'aminoàcids. Existeixen dues estratègies per fer-ho: la predicció de *novo* (modelatge *ab initio*), que es basa en construir models proteics des de zero, tenint en compte, per exemple, característiques físico-químiques (Bradley *et al.* 2005; Schmidt

et al. 2013), i la predicció per comparació (modelatge per homologia), en el que es poden obtenir models tridimensionals de la proteïna problema per homologia a una proteïna coneguda motlle o patró (*template*) similar en estructura.

3.18.1 Modelatge per homologia

En aquest treball s'ha utilitzat aquest mètode per obtenir els models tridimensionals de la proteïna RecA. Aquesta tècnica utilitza estructures prèviament resoltes experimentalment com a punt de partida o motlle que es troben dipositades en bases de dades. La predicció per comparació amb aquestes estructures motlle es pot dividir en 4 etapes: el reconeixement del motlle, l'alineament, la construcció del model tridimensional i la validació (Schwede *et al.* 2003; Kopp & Schwede 2004; Arnold *et al.* 2006; Schmidt *et al.* 2013) (Figura 3.8).

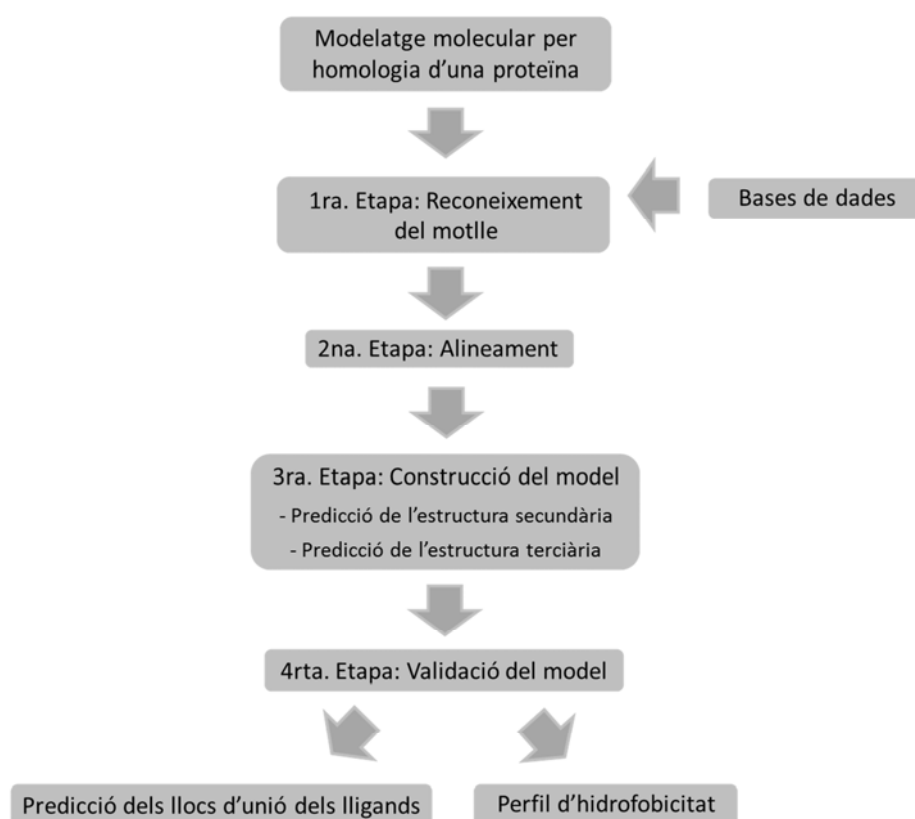


Figura 3.8 Esquema de treball seguit per obtenir un model tridimensional per homologia d'una proteïna.

La primera etapa de reconeixement del motlle té com a objectiu principal cercar seqüències amb una estructura tridimensional resolta de la proteïna que presentin un grau elevat d'homologia amb les seqüències problema de l'estudi. Aquesta cerca s'ha realitzat directament en bases de dades especialitzades com és el *Protein Data Bank* (PDB, *RCSB Protein Data Bank*, Berman *et al.* 2000, 2013, <http://www.rcsb.org/pdb/home/home.do>), i també a través dels servidors web de lliure accés Phyre2 (Protein Homology/analogY Recognition Engine v2.0, Bennett-Lovsey *et al.* 2007; Kelley & Sternberg 2009; Kelley *et al.* 2015, <http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) i SWISS-MODEL (Schwede *et al.* 2003; Arnold *et al.* 2006; Guex *et al.* 2009; Biasini *et al.* 2014), <http://swissmodel.expasy.org/>),

que connecten amb bases de dades per cercar seqüències homòlogues i proporcionar un llistat de resultats de possibles proteïnes motlles.

El SWISS-MODEL selecciona automàticament a través d'un BLAST aquells motlles a partir d'una biblioteca pròpia del programa (obtinguda a partir del PDB) en funció d'un *E-value* obtingut, de manera que s'afavorirà aquelles estructures amb propietats estereoquímiques raonables i segons l'energia del camp de força. Quan no s'identifiquen motlles o només algunes parts s'utilitzen dues aproximacions addicionals per a fer una detecció més sensible de proteïnes més distants filogenèticament. En aquest cas es porta a terme un PSI-BLAST i una cerca basada en models de Markov ocults (HMM). En el primer cas, els models es construeixen utilitzant el mètode Promod-II (Guex & Peitsch 1997) a partir de la plantilla seleccionada de manera que les possibles insercions o delecions són remodelades utilitzant una biblioteca de fragments. Les cadenes laterals són les últimes estructures en ser reconstruïdes.

La segona etapa és l'alineament entre la seqüència aminoacídica problema i les seqüències de les estructures homòlogues identificades com a possibles motlles per a la construcció del model tridimensional. Els alineaments d'estructures primàries s'han realitzat amb els programes ClustalW (MEGA6) i JPred (versió 3, Cole *et al.* 2008, <http://www.compbio.dundee.ac.uk/jpred>) i els d'estructures secundàries amb els servidors web Phyre2 i SWISS-MODEL. Les possibles proteïnes motlles es classifiquen en funció del nombre de residus alineats i la qualitat de l'alineament (la similitud de la distribució dels residus per a cada posició, la similitud de l'estructura secundària i la presència o absència d'insercions i delecions).

En la tercera etapa de construcció del model tridimensional, cal tenir en compte que cap mètode és totalment fiable, per això s'aconsella combinar les prediccions de diferent fonts per evitar errors. Actualment es disposa de varis servidors web de lliure accés per fer la predicció d'estructures de proteïnes (Taula 3.3). En aquest treball, per la predicció de les estructures secundària i terciària, i la construcció del model tridimensional de les seqüències problema s'han utilitzat els programes Phyre2 i SWISS-MODEL, a partir de diferents proteïnes motlle seleccionades. Per visualitzar els models tridimensionals obtinguts en format PDB s'han emprat els programes PyMOL (versió 0.99rc6, <https://www.pymol.org/>) i JMol (versió 13.0, <http://jmol.sourceforge.net/index.es.html>). Per fer prediccions de regions desordenades dels models, s'ha treballat en el servidor web PSIPRED (versió 2, (McGuffin *et al.* 2000; Buchan *et al.* 2013), <http://bioinf.cs.ucl.ac.uk/psipred/>) utilitzant el mètode DISOPRED (Ward *et al.* 2004) que permet determinar el grau de desordre que pot presentar una proteïna en les diferents regions, a partir de la seva seqüència aminoacídica en format FASTA.

Taula 3.3 Característiques dels programes més utilitzats per la modelització de proteïnes. Els programes emprats en aquest treball s'indiquen en gris.

Programa	Web i referència
Modelatge de proteïnes	
Phyre2	http://www.sbg.bio.ic.ac.uk/~phyre/ (Kelley & Sternberg 2009)
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/ (Zhang 2008)
SAM-T08	http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html (Karplus 2009)
HHpred	http://toolkit.tuebingen.mpg.de/hhpred (Söding <i>et al.</i> 2005)
GenThreader	http://www.chemogenomix.com/chemogenomix/GenThreader.html (Jones 1999)
PCONS	http://pcons.net/ (Wallner <i>et al.</i> 2003)
FFAS	http://ffas.burnham.org/ffas-cgi/cgi/ffas.pl (Jaroszewski <i>et al.</i> 2005)
Robetta	http://rosetta.bakerlab.org/ (Kim <i>et al.</i> 2004)
Protein Model Portal	http://www.proteinmodelportal.org/ (Haas <i>et al.</i> 2013)
ICM	http://www.molsoft.com/homology.html (Cardozo <i>et al.</i> 1995)
IntFOLD	http://www.reading.ac.uk/bioinf/IntFOLD/ (Roche & Buenavista 2011)
Modeller	http://www.salilab.org/modeller/ (Sali 1995)
SWISS-MODEL	http://swissmodel.expasy.org/ (Arnold <i>et al.</i> 2006)
Proteus2	http://www.proteus2.ca/proteus2/ (Montgomerie <i>et al.</i> 2008)
CPHModels	http://www.cbs.dtu.dk/services/CPHmodels/ (Nielsen <i>et al.</i> 2010)
Visualització d'estructures	
Chimera	https://www.cgl.ucsf.edu/chimera/
RasMol	http://www.umass.edu/microbio/rasmol/
PyMOL	http://www.pymol.org/
VMD	http://www.ks.uiuc.edu/Research/vmd/
Jmol	http://jmol.sourceforge.net/index.es.html
SwissPDB Viewer	http://spdbv.vital-it.ch/
FoldIT	http://fold.it/portal/
Avaluació d'estructures	
PROCHECK	http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/ (Laskowski <i>et al.</i> 1993)
ProsaII	https://prosa.services.came.sbg.ac.at/prosa.php (Wiederstein & Sippl 2007)
WHAT_CHECK	http://swift.cmbi.ru.nl/gv/whatcheck/ (Hooft <i>et al.</i> 1996)
ERRAT	http://nihserver.mbi.ucla.edu/ERRATv2/ (Colovos & Yeates 1993)
ProQ	http://www.sbc.su.se/~bjornw/ProQ/ProQ.cgi (Wallner & Elofsson 2003)
VERIFY_3D	http://nihserver.mbi.ucla.edu/Verify_3D/ (Lüthy <i>et al.</i> 1992)

Una vegada es disposa del model tridimensional cal validar-lo (quarta etapa). Sovint els mateixos programes de modelatge inclouen algun mètode de validació, però també hi ha programes dissenyats per avaluar estructures (Taula 3.3). En aquest estudi, la validació dels models obtinguts s'ha realitzat emprant dos servidors diferents: ProQ i SWISS-MODEL.

El programa de lliure accés ProQ (*Protein Quality Predictor*, Wallner & Elofsson 2003, <http://www.sbc.su.se/~bjornw/ProQ/ProQ.cgi>) permet predir la qualitat del model obtingut en format PDB a partir d'una sèrie de característiques estructurals. Els dos paràmetres que utilitza aquest programa per determinar el grau de qualitat són el valor de LGScore i Maxsub (Wallner

& Elofsson 2003). En funció dels resultats obtinguts, la predicció es classifica en correcta (LGscore > 1,5 i MaxSub > 0,1), bona (LGscore > 3 i MaxSub > 0,5) i molt bona (LGscore > 5 i MaxSub > 0,8).

El programa SWISS-MODEL realitza una validació del model mitjançant els paràmetres QMEAN i GMQE. El QMEAN4 (*Qualitative Model Energy ANalysis*, Benkert *et al.* 2009, 2011) és una combinació lineal de quatre paràmetres estructurals que té com a resultat una puntuació del model. El valor obtingut oscil·la entre 0 i 1, de manera que els millors models obtenen resultats propers a 1. En relació al paràmetre GMQE (*Global Model Quality Estimation*) proporciona una estimació del model a partir de la combinació de propietats de l'alineament del motlle i la predicció. També s'expressa amb un valor entre 0 i 1, de tal manera que els valors propers a 1 indiquen una major fiabilitat del model.

3.18.2 Predicció dels llocs d'unió dels lligands

Per predir els llocs d'unió dels lligands en els models estructurals obtinguts s'ha utilitzat el servidor 3DLigandSite (*Ligand Binding Site Prediction Server*, Wass *et al.* 2010, <http://www.sbg.bio.ic.ac.uk/3dligandsite/advanced.cgi>), que utilitza models resoltos per identificar els llocs d'unió a cofactors i altres paràmetres en proteïnes modelades. D'aquesta manera, els lligands units a estructures similars es superposen en el model i es realitza una predicció dels llocs d'unió. Les prediccions obtingudes es poden observar amb detall amb el programa de visualització d'estructures Jmol.

3.18.3 Perfil d'hidrofobicitat

Per estudiar determinats perfils de proteïnes a partir de la seqüència d'aminoàcids es pot utilitzar el programa ProtScale del servidor Expasy de lliure accés (Gasteiger *et al.* 2005, <http://web.expasy.org/cgi-bin/protscale/>), que presenta els resultats en forma de gràfic, on l'eix d'abscisses designa la posició dels aminoàcids al llarg de la seqüència i l'eix d'ordenades assigna un valor numèric a cadascun dels aminoàcids.

Els perfils més utilitzats són els d'hidrofobicitat, la majoria dels quals deriven d'estudis experimentals sobre la disposició dels pèptids en dissolvents polars i apolars. En aquest estudi s'ha generat el perfil d'hidrofobicitat de les seqüències aminoacídiques problema, per determinar l'accessibilitat dels aminoàcids als solvents, i així poder predir les regions de la proteïna exposades al medi. S'ha utilitzat l'algoritme Kyte & Doolittle (Kyte & Doolittle 1982) que utilitza una escala de valors pels aminoàcids entre -3,5 (àcid glutàmic) i 4,2 (valina), de manera que valors superiors a 0 corresponen a regions hidrofíliques, amb elevada probabilitat de trobar-se a la superfície, mentre que valors per sota de 0 a regions hidrofòbiques.

4. RESULTATS:
Estudi del gen malat
deshidrogenasa d'*Aeromonas*

4. RESULTATS: ESTUDI DEL GEN MALAT DESHIDROGENASA d'*Aeromonas*

4.1 OBTENCIÓ DE LES SEQÜÈNCIES

En aquest treball s'han dissenyat encebadors específics per obtenir la seqüència completa del gen *mdh* d'*Aeromonas*, i s'han estandarditzat les barreges de reacció i les condicions de PCR i de seqüenciació de les soques de l'estudi que no estaven prèviament amplificades (114 de les 150 del total). Les 36 seqüències restants han estat obtingudes també pel nostre grup i publicades en dos treballs anteriors (Farfán *et al.* 2010; Fusté *et al.* 2012). Per a l'anàlisi s'han considerat un total de 150 seqüències nucleotídiques completes d'aquest gen pertanyents a totes les espècies i subespècies del gènere *Aeromonas* descrites fins al moment, incloent totes les soques tipus i algunes soques sinònimes (Annex I).

4.1.1 Disseny d'encebadors

Per a l'amplificació per PCR i la posterior seqüenciació del gen *mdh* ha estat necessari dissenyar varis encebadors externs (*mdh*-F, *mdh*-I, *mdh*-MC, *mdh*-3, *mdh*-6) i interns (*mdh*-A, *mdh*-8), tot i que se'n disposava d'alguns (*mdh*-C2B, *mdh*-CC, *mdh*-MC2, *mdh*-1) emprats en estudis previs (Farfán *et al.* 2010; Fusté *et al.* 2012). Pel disseny d'aquests nous encebadors es van utilitzar les seqüències dels dos genomes d'*Aeromonas* disponibles en aquell moment a la base de dades del *GenBank*: *A. hydrophila* ATCC7966^T (*GenBank* accession number: CP000462, (Seshadri *et al.* 2006)) i *A. salmonicida* subsp. *salmonicida* A449 (*GenBank* accession number: CP000644, (Reith *et al.* 2008)). A la Figura 4.1 es representa de forma gràfica el gen *mdh* i els seus gens flanquejants, *artM* (extrem 5') i *motX* (extrem 3'), així com la posició i el sentit de tots els encebadors utilitzats.

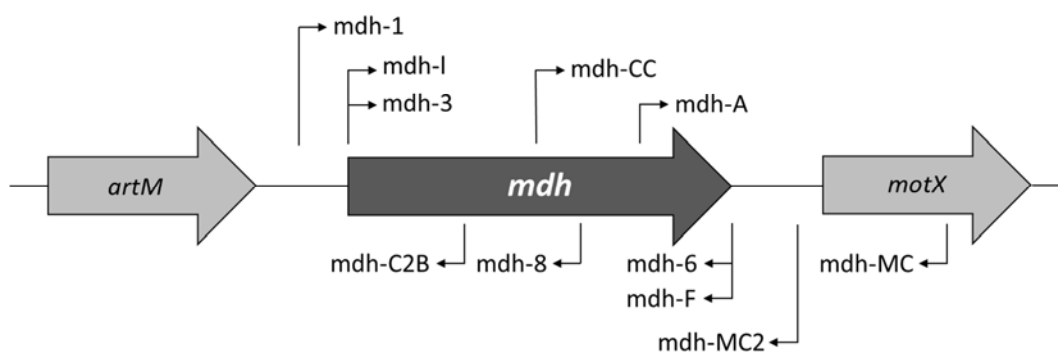


Figura 4.1 Posició dels encebadors utilitzats per a l'amplificació per PCR i la seqüenciació del gen *mdh*.

A la Taula 4.1 es mostren les seqüències i altres característiques dels encebadors emprats en l'estudi. A la Taula 4.2 es detallen les seves temperatures de fusió (T_m) calculades teòricament amb la fórmula de Wallace i les proporcionades pel fabricant, així com les temperatures d'hibridació de PCR i de seqüenciació.

Taula 4.1 Característiques dels encebadors utilitzats per a l'amplificació i la seqüenciació del gen *mdh*.

Encebador	Longitud (pb)	Seqüència 5' → 3'	Sentit ^a	Posició ^b
mdh-A	18	TTCGGTCTGTCGCTCATC	F	+700, +717
mdh-CC	17	ACCAACCCGGTNAACAC	F	+352, +368
mdh-I	18	ATGGAGTTGATCATGAAA	F	+47, +64
mdh-1	18	TTATTGGCTCACAGAACC	F	-122, -104
mdh-3	20	GAGTTGATCATGAAAGTTGC	F	-9, +11
mdh-C2B	17	GTGTTNACCGGGTTGGT	R	+368, +352
mdh-F	18	NNCTTACTTGACGAACTC	R	+939, +922
mdh-MC	20	ATCCAGTTGAGCAGCTCGTC	R	+1470, +1451
mdh-MC2	20	GTAANGGGTTTCATCATAAG	R	+1077, +1058
mdh-6	18	CTTGACGAACTCGACACC	R	+933, +916
mdh-8	18	YTCCAGRACGGTRTCGAC	R	+840, +823

^a F, *forward*; R, *reverse*

^b Posició nucleotídica de cada encebador en relació al genoma d'*A. hydrophila* ATCC7966^T

Taula 4.2 Temperatures de fusió (T_m) i d'hibridació (T_a) dels encebadors del gen *mdh*.

Encebador	T fusió (°C)		T hibridació (°C)	
	Tèorica	Fabricant	PCR	Seqüenciació
mdh-A	56	52,4	51	50
mdh-CC	51	54,0	46	45
mdh-I	48	58,0	43	42
mdh-1	52	62,0	47	46
mdh-3	56	53,2	51	50
mdh-C2B	51	54,0	46	45
mdh-F	48	51,4	43	42
mdh-MC	62	59,4	57	56
mdh-MC2	53	52,5	48	47
mdh-6	56	56,0	51	50
mdh-8	51	57,1	46	45

4.1.2 Optimització del procés d'amplificació per PCR

Per l'amplificació del gen complet *mdh*, en la majoria de casos s'ha utilitzat una sola parella d'encebadors externs. A la Taula 4.3 es mostren les diferents combinacions d'encebadors assajades en aquest estudi, la temperatura d'hibridació optimitzada i la mida de l'amplicó.

Taula 4.3 Parelles d'encebadors utilitzades per a l'amplificació per PCR del gen *mdh*.

PCR	Parelles d'encebadors	Temperatura d'hibridació (°C)	Longitud de l'amplicó (pb) ^a
PCR1	1/MC2	46	1199
PCR2	I/MC2	47	1031
PCR3	1/6	49	1055
PCR4	I/6	47	887
PCR5	I/MC	50	1424
PCR6	1/MC	52	1592
PCR7	1/F	46	1061
PCR8	3/F	47	948
PCR9	I/F	43	893

^a Basat en el genoma d'*A. hydrophila* ATCC7966^T

En relació a la composició de la barreja de reacció, s'han optimitzat les amplificacions modificant les concentracions de DNA genòmic i d'encebadors (Taula 4.4). Segons estudis previs de seqüenciació de diferents gens d'*Aeromonas* (Farfán *et al.* 2009, 2010) s'ha comprovat que les barreges de PCR requereixen una concentració addicional de MgCl₂ per obtenir millors resultats. Per això, s'ha utilitzat una solució tamponada de la Taq DNA polimerasa (Buffer I) que conté 15 mM de MgCl₂, a la que s'ha afegit una concentració addicional d'aquesta sal. A més, per algunes soques de les espècies *A. schubertii*, *A. simiae*, *A. diversa*, *A. rivuli*, *A. bivalvium*, *A. australiensis*, *A. cavernicola* i *A. eucrenophila* ha estat necessari addicionar DMSO a la barreja per obtenir més producte de PCR. La resta de components (dNTPs, solució tamponada Buffer I 10x i Taq DNA polimerasa) no s'ha variat la concentració.

Taula 4.4 Composició de la barreja de reacció de PCR optimitzada per l'estudi del gen *mdh*.

Components de la reacció de PCR	Volum (µL)
DNA genòmic	0,5 - 5
Encebador <i>forward</i> 10 µM	0,5 - 7
Encebador <i>reverse</i> 10 µM	0,5 - 7
Buffer I 10x	5
dNTPs Mix 10 mM	4
MgCl ₂ 25 mM	4
DMSO	0 - 5
Ampli Taq Gold® 5U/µL	0,25
Aigua Milli-Q estèril	qsp. 50

Les condicions i etapes d'amplificació per PCR s'han optimitzat segons mostra la Taula 4.5. Les reaccions de PCR s'han estandarditzat modificant la temperatura i el temps d'hibridació dels encebadors i el nombre de cicles a repetir el procés (35 cicles) per obtenir major quantitat de

producte amplificat. Les temperatures d'hibridació optimitzades per cada parella d'encebadors utilitzats per amplificar el gen *mdh* s'indiquen a la Taula 4.3.

Taula 4.5 Condicions optimitzades per l'amplificació per PCR del gen *mdh*.

Etapes	Temperatura (°C)	Temps	Cicles
Desnaturalització inicial	95	5 min	1
Desnaturalització	94	45 seg	
Hibridació	46 - 52	1 min	35
Extensió	72	2 min	
Extensió final	72	10 min	1
Final	4	∞	

Totes les parelles d'encebadors emprades per amplificar el gen *mdh* d'*Aeromonas* han funcionat correctament en la reacció de PCR. A l'avaluar les parelles d'encebadors amb les que s'han obtingut millors resultats, destaquen la PCR1 (39,47%) i la PCR2 (28,95%), que amplifiquen un fragment d'aproximadament 1200 i 1000 pb, respectivament (Taula 4.3, Figura 4.2). La resta de combinacions (PCR3, 18,42%; PCR4, 6,14%; PCR5, 3,50%; PCR6, 0,88%; PCR7, 0,88%; PCR8, 0,88%; PCR9, 0,88%) ha permès amplificar aïllats diferents de les soques tipus de les espècies *A.molluscorum*, *A.bivalvium*, *A.rivuli*, *A.schubertii*, *A.simiae*, *A.diversa*, i també, les descrites posteriorment, com *A.australiensis* i *A.cavernicola*.

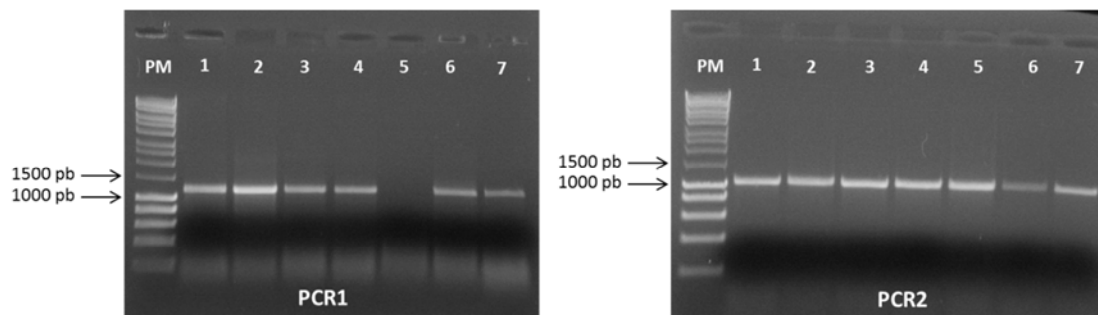


Figura 4.2 Comprovació dels amplificats del gen *mdh* obtinguts per PCR (5µL) mitjançant l'electroforesi en gel d'agarosa a l'1%. PM: marcador de pesos moleculars. 7: control positiu.

4.1.3 Optimització del procés de seqüenciació

Per tal d'obtenir la seqüència completa del gen, a partir dels productes de PCR purificats s'ha fet una seqüenciació directa utilitzant els mateixos encebadors de les PCRs, i també altres encebadors interns (*mdh-A*, *mdh-C2B*, *mdh-CC*, *mdh-8*) degut a la longitud dels amplificats (Taulas 4.1 i 4.2). En la seqüenciació, cada encebador ha permès obtenir una seqüència nucleotídica fiable d'entre 600-800 pb. Per cada amplificat s'han seqüenciat ambdues cadenes

de DNA, i han estat necessàries dues seqüenciacions amb encebadors diferents per cada cadena, per tal d'obtenir la seqüència nucleotídica completa del gen *mdh*. A la Taula 4.2 es mostren les temperatures d'hibridació optimitzades dels encebadors externs i interns utilitzats en les reaccions de seqüenciació.

A la Taula 4.6 s'indica la composició de la barreja de seqüenciació que ha donat millors resultats. Al comprovar l'amplificat obtingut per PCR per electroforesi en gel d'agarosa, la intensitat de banda observada sota llum UV ha permès conèixer de forma aproximada la concentració de DNA per comparació amb el patró de pesos moleculars. En general, per la barreja de seqüenciació s'ha utilitzat 1 µL del producte de PCR purificat i 1,5 µL d'encebador. Però, en aquells casos que s'han obtingut fragments amb bandes més intenses s'ha canviat per una relació 1:2 de DNA i encebador, i per aquelles amb intensitat més baixa per una relació 1:1 o 2:1.

Taula 4.6 Barreja de reacció de seqüenciació optimitzada del gen *mdh*.

Components de la reacció de seqüenciació	Volum (µL)
DNA purificat	1 - 2
Encebador 5µM	1 - 2
Buffer Big Dye®	3
Big Dye® Terminator v3.1	1
Aigua Milli-Q estèril	qsp. 20

Finalment, a la Taula 4.7 es mostren les condicions estandarditzades per la reacció de seqüenciació del gen *mdh*, indicant les temperatures, els temps i els cicles realitzats en cada etapa. Les temperatures d'hibridació per la seqüenciació s'han optimitzat restant un grau a les temperatures d'hibridació de PCR de cada encebador, tal com s'indica a la Taula 4.2.

Taula 4.7 Condicions optimitzades per la reacció de seqüenciació del gen *mdh*.

Etaques	Temperatura (°C)	Temps	Cicles
Desnaturalització inicial	96	1 min	1
Desnaturalització	96	10 seg	
Hibridació	42 - 56	5 seg	20
Extensió	60	4 min	
Final	4	∞	

4.2 ANÀLISI DE LES SEQÜÈNCIES

4.2.1 Caracterització del gen *mdh*

S'ha obtingut la seqüència nucleotídica completa del gen *mdh* de totes les soques de l'estudi, presentant totes elles la mateixa longitud de 936 pb. Les seqüències han estat dipositades a la base de dades del *GenBank* i s'han identificat amb els números d'accés *GenBank* que s'indiquen a l'Annex I.

Els resultats de l'anàlisi de la variabilitat nucleotídica del conjunt de seqüències es mostren a la Taula 4.8. Per aquelles espècies d'*Aeromonas* que només es disposa d'una única soca (*A.australiensis*, *A.cavernicola*, *A.fluvialis*, *A.sanarellii* i *A.taiwanensis*) no s'ha pogut realitzar aquest estudi, amb excepció de la determinació del contingut en G+C. Pel conjunt de soques s'han identificat un total de 566 llocs invariants i 370 posicions polimòrfiques, de les quals 46 han estat llocs variables en una posició (*singleton variable sites*) i 324 llocs parsimòniament informatius (*parsimony informative sites*). A partir de l'estudi de l'alineament múltiple de seqüències s'ha observat que tots els canvis detectats han estat substitucions nucleotídiques, i no s'ha trobat ni insercions ni delecions. En aquest cas s'ha observat, tal com era d'esperar, que les transicions (*Ti*, canvis entre purines o entre pirimidines) són més freqüents que les transversions (*Tv*, canvis de purina per pirimidina o a l'inrevés), amb un valor global de la taxa de *Ti/Tv* de 2,88.

S'ha determinat el contingut mig en G+C que ha estat de 60,4% pel total de seqüències analitzades. A la Taula 4.8 s'indiquen els valors de contingut en G+C per cadascuna de les espècies. Si s'observa amb més detall, *A.cavernicola* ha presentat el valor més baix (58%) i *A.diversa* el més elevat (62,4%).

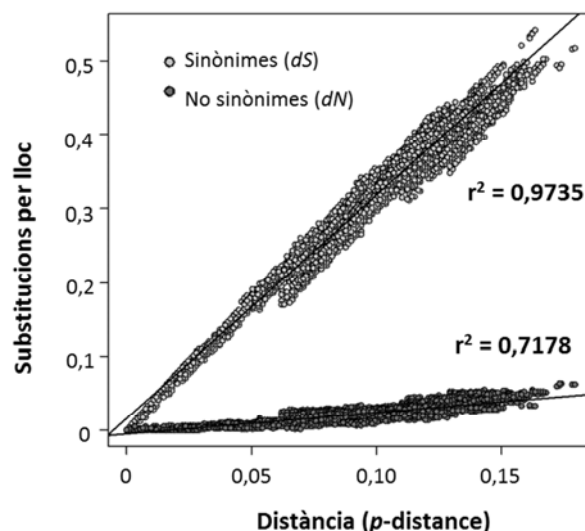


Figura 4.3 Representació gràfica del grau de saturació del gen *mdh*.

La saturació és un fenomen que afecta freqüentment a les seqüències de DNA, que es produeix quan una mateixa posició nucleotídica ha canviat més d'una vegada des de que les dues seqüències es van separar del seu ancestre comú. El grau de saturació en els canvis mutacionals de les seqüències del gen *mdh*, s'ha determinat calculant la proporció de canvis sinònims per lloc sinònim (*dS*) i canvis no sinònims per lloc no sinònim (*dN*) en relació amb la distància genètica entre tàxons. Tal com s'observa a la Figura 4.3, el nombre de substitucions sinònimes i no sinònimes en cap cas està saturat.

Taula 4.8 Anàlisi de la variabilitat i el contingut en G+C del gen *mdh*.

Espècie	Llocs invariants	Llocs variables	Llocs variables en una posició	Llocs parsimoniament informatius	G+C (%mol)
<i>A.allosaccharophila</i>	909	27	21	6	60,0
<i>A.australiensis</i>	-	-	-	-	59,9
<i>A.bestiarum</i>	904	32	21	11	59,6
<i>A.bivalvium</i>	928	8	8	0	61,3
<i>A.cavernicola</i>	-	-	-	-	58,0
<i>A.caviae</i>	895	41	29	12	61,4
<i>A.dhakensis</i>	901	35	18	17	61,4
<i>A.diversa</i>	933	3	3	0	62,4
<i>A.encheleia</i>	914	22	7	15	62,0
<i>A.enteropelogenes</i>	913	23	15	8	60,6
<i>A.eucrenophila</i>	913	23	20	3	61,6
<i>A.fluvialis</i>	-	-	-	-	60,0
<i>A.hydrophila</i>	879	57	41	16	61,0
<i>A.jandaei</i>	914	22	11	11	60,3
<i>A.media</i>	886	50	34	16	61,0
<i>A.molluscorum</i>	926	10	6	4	59,7
<i>A.piscicola</i>	908	28	15	13	59,8
<i>A.popoffii</i>	918	18	7	11	59,7
<i>A.rivuli</i>	929	7	7	0	60,7
<i>A.salmonicida</i>	886	50	16	34	58,9
<i>A.sanarellii</i>	-	-	-	-	62,0
<i>A.schubertii</i>	929	7	2	5	62,0
<i>A.simiae</i>	933	3	3	0	61,3
<i>A.sobria</i>	865	71	52	19	59,4
<i>A.taiwanensis</i>	-	-	-	-	62,2
<i>A.tecta</i>	888	48	42	6	59,8
<i>A.veronii</i>	859	77	30	47	59,7

4.2.2 Tipus de selecció

Per tal d'estudiar el tipus de selecció que actua en aquest gen s'ha realitzat una anàlisi per determinar estadísticament si hi ha selecció positiva, neutra o purificadora mitjançant la prova Z (*Z-test*). El resultat obtingut han permès rebutjar la hipòtesi de selecció positiva ($P > 0,05$), però no han permès discriminar entre selecció neutra o purificadora. Per aquest motiu s'ha portat a terme el Test de Neutralitat de Tajima (Apartat 3.13), que ha permès rebutjar la hipòtesi de neutralitat ($D = 1,11388$, $P > 0,1$). Per tant, aquestes anàlisis estadístiques indiquen que el model de selecció que opera sobre aquest gen és de tipus purificador.

A partir de les substitucions nucleotídiques detectades pel conjunt de seqüències s'ha pogut determinar un valor global de canvis nucleotídics sinònims per lloc sinònim de 0,4064 (*dS*), de no sinònims per lloc no sinònim de 0,0223 (*dN*) i un valor d' ω (*dN/dS*) de 0,0549. Cal destacar que el nombre de canvis sinònims (*dS*) ha estat d'un ordre superior a les substitucions no sinònimes (*dN*) d'aquest gen. El valor de $\omega < 1$ suggereix que els canvis detectats en el gen *mdh* es trobaven sota una forta selecció negativa o purificadora. També s'ha determinat la proporció de *dS* i *dN* codó per codó al llarg del gen (Figura 4.4), en el que s'observa un augment constant en el nombre de substitucions sinònimes, mentre que en el cas de les no sinònimes aquest augment és menor i es poden diferenciar algunes regions sense canvis no sinònims.

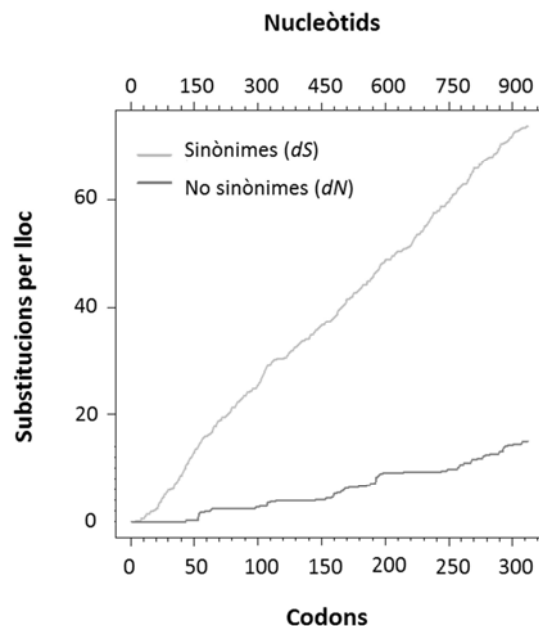


Figura 4.4 Distribució de les substitucions sinònimes i no sinònimes per codons al llarg de les seqüències del gen *mdh*.

També s'han determinat les posicions sinònimes i no sinònimes d'aquest gen entre les soques d'una mateixa espècie (Taula 4.9) i entre les diferents espècies considerades (Figura 4.5). Per aquelles espècies que només es disposa d'una soca, òbviament no s'ha pogut realitzar el càlcul a nivell intra-específic. Els resultats mostren que hi ha més canvis sinònims a nivell intra-

específic, excepte *A.simiae* en que les dues seqüències són pràcticament idèntiques, i únicament presenten tres canvis no sinònims. A més, s'han detectat 3 espècies (*A.bivalvium*, *A.enzeleia* i *A.molluscorum*) que només tenen canvis sinònims. A la Figura 4.5 es mostra la correlació que existeix entre els canvis no sinònims (dN) i sinònims (dS) entre parelles d'espècies i la recta de regressió corresponent.

Taula 4.9 Anàlisi intra-específica dels canvis sinònims (dS) i no sinònims (dN) i el valor ω (dN/dS) del gen *mdh*.

Espècie	dS	dN	ω
<i>A.allosaccharophila</i>	0,0486	0,0012	0,0240
<i>A.bestiarum</i>	0,0364	0,0018	0,0488
<i>A.bivalvium</i>	0,0238	0	0
<i>A.caviae</i>	0,0439	0,0013	0,0287
<i>A.dhakensis</i>	0,0477	0,0021	0,0446
<i>A.diversa</i>	0,0083	0,0014	0,1738
<i>A.enzeleia</i>	0,0386	0	0
<i>A.enteropelogenes</i>	0,0375	0,0026	0,0701
<i>A.eucrenophila</i>	0,0384	0,0012	0,0302
<i>A.hydrophila</i>	0,0663	0,0019	0,0292
<i>A.jandaei</i>	0,0368	0,0013	0,0344
<i>A.media</i>	0,0910	0,0049	0,0535
<i>A.molluscorum</i>	0,0202	0	0
<i>A.piscicola</i>	0,0410	0,0011	0,0266
<i>A.popoffii</i>	0,0240	0,0009	0,0371
<i>A.rivuli</i>	0,0168	0,0043	0,2592
<i>A.salmonicida</i>	0,0734	0,0025	0,0335
<i>A.schubertii</i>	0,0117	0,0015	0,1245
<i>A.simiae</i>	0	0,0043	-
<i>A.sobria</i>	0,1374	0,0032	0,0234
<i>A.tecta</i>	0,1055	0,0024	0,0229
<i>A.veronii</i>	0,0851	0,0027	0,0318

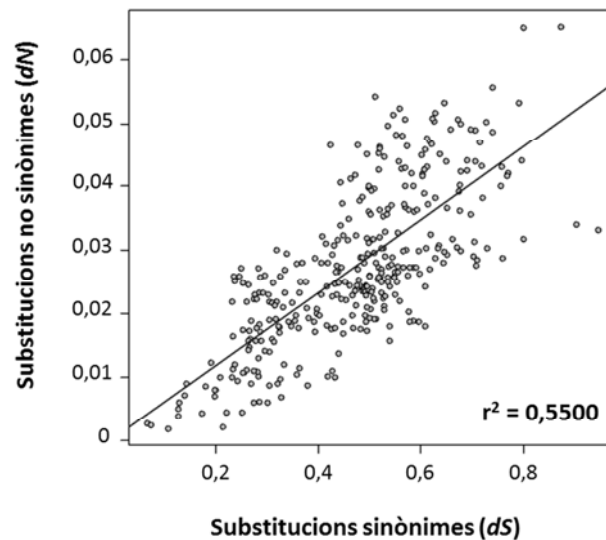


Figura 4.5 Relació entre les substitucions no sinònimes (dN) i sinònimes (dS) a nivell inter-específic del gen *mdh*.

4.2.3 Distàncies genètiques intra- i inter-específiques

A partir de l'alineament múltiple de seqüències, s'han determinat els valors de distància genètica (p -distance) entre les soques d'una mateixa espècie (intra-específica) i entre espècies (inter-específica). A la Taula 4.10 es mostren els resultats obtinguts, on també s'indica el nombre de soques de cada espècie. El valor més alt de distància intra-específica correspon a l'espècie *A.sobria* (0,0353), que inclou 5 soques, mentre que el més baix ha estat el d'*A.diversa* (0,0032), amb 2 soques. A l'analitzar les distàncies inter-específiques, *A.simiae* ha estat l'espècie que ha donat el valor més alt (0,1442) i *A.dhakensis* el valor més baix (0,0900). En aquelles espècies que només es disposa d'una soca no s'ha pogut calcular el valor de distància intra-específica.

A partir dels valors de distància obtinguts, s'ha construït un gràfic de densitats per veure la distribució d'aquests valors (Figura 4.6A). Tal com es mostra a la figura s'observa una clara separació entre la distribució dels valors de distàncies intra- i inter-específiques, sense solapament entre les dues distribucions. A més, s'ha determinat la correlació de distàncies intra- i inter-específiques de cada espècie (Figura 4.6B). Les línies discontinues corresponen a la mitjana d'aquestes dues distàncies. En el quadrant superior esquerra apareixen aquelles espècies amb una millor separació inter-específica (*A.simiae*, *A.schubertii*, *A.diversa*, *A.molluscorum*, *A.rivuli* i *A.enteropelogenes*).

Taula 4.10 Distàncies genètiques (p -distance) a nivell intra- i inter-específic del gen *mdh*.

Espècie	Distància Intra-específica	Distància Inter-específica	Nombre de soques
<i>A.allosaccharophila</i>	0,0131	0,1028	5
<i>A.australiensis</i>	-	0,1096	1
<i>A.bestiarum</i>	0,0106	0,0909	9
<i>A.bivalvium</i>	0,0061	0,1053	3
<i>A.cavernicola</i>	-	0,1431	1
<i>A.caviae</i>	0,0119	0,0975	10
<i>A.dhakensis</i>	0,0135	0,0900	8
<i>A.diversa</i>	0,0032	0,1270	2
<i>A.encheleia</i>	0,0097	0,1006	8
<i>A.enteropelogenes</i>	0,0115	0,1096	5
<i>A.eucrenophila</i>	0,0105	0,0962	5
<i>A.fluvialis</i>	-	0,1130	1
<i>A.hydrophila</i>	0,0179	0,0939	9
<i>A.jandaei</i>	0,0103	0,1055	6
<i>A.media</i>	0,0253	0,0999	5
<i>A.molluscorum</i>	0,0051	0,1238	5
<i>A.piscicola</i>	0,0112	0,0930	8
<i>A.popoffii</i>	0,0068	0,0989	9
<i>A.rivuli</i>	0,0075	0,1133	2
<i>A.salmonicida</i>	0,0196	0,1030	13
<i>A.sanarellii</i>	-	0,1003	1
<i>A.schubertii</i>	0,0041	0,1373	5
<i>A.simiae</i>	0,0032	0,1442	2
<i>A.sobria</i>	0,0353	0,1086	5
<i>A.taiwanensis</i>	-	0,0957	1
<i>A.tecta</i>	0,0271	0,1054	4
<i>A.veronii</i>	0,0230	0,1062	17

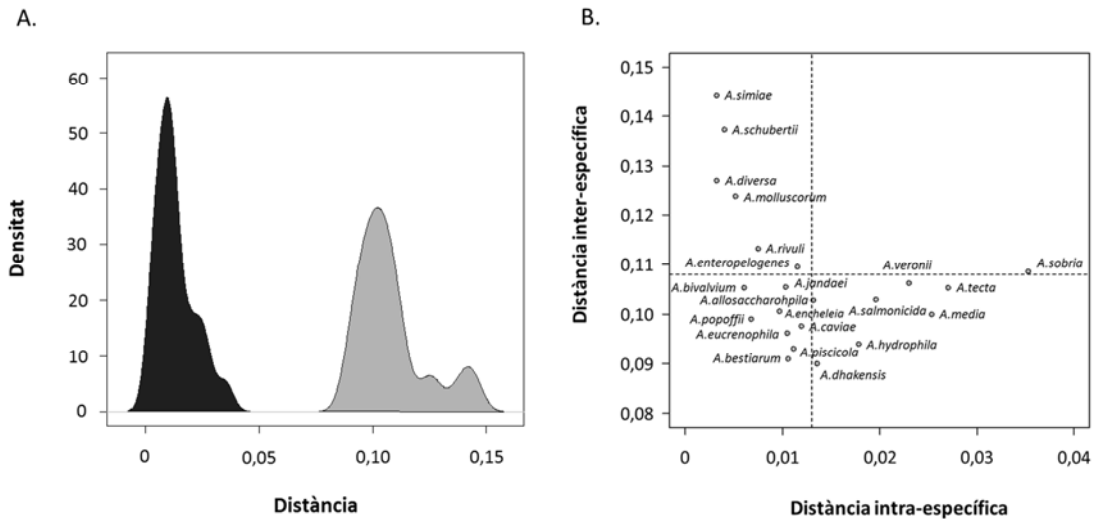


Figura 4.6 Representacions gràfiques dels valors de distància genètica de les seqüències del gen *mdh*. A) Gràfic de densitats de les distàncies intra- (negre) i inter-específiques (gris). B) Correlació de les distàncies inter- i intra-específiques de cada espècie, i les línies discontinües indiquen els valors migs de distàncies.

Per últim, tenint en compte el desigual nombre d'aïllats d'algunes espècies analitzades (Annex I), s'ha determinat la possible influència del nombre de soques en el càlcul dels valors de distància. A la Figura 4.7 es mostra la correlació entre les distàncies intra- i inter-específiques i el nombre de soques de cada espècie. Els baixos valors r^2 de les rectes de regressió indiquen una baixa correlació entre les dues variables, i per tant, que els valors de distàncies obtinguts en cap cas estan influenciats pel nombre de soques de cada espècie de l'estudi.

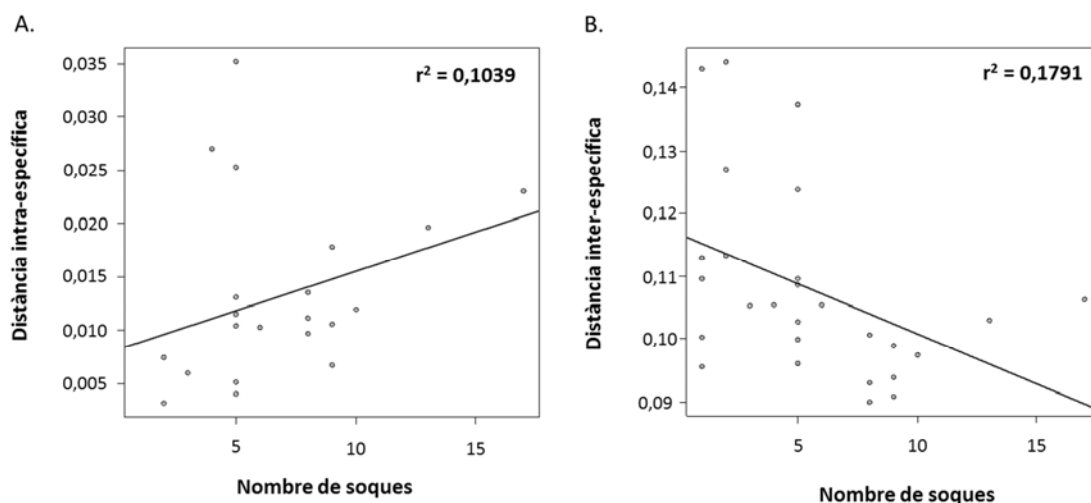


Figura 4.7 Influència del nombre de soques respecte les distàncies genètiques intra- (A) i inter-específiques (B) del gen *mdh*.

4.3 FILOGÈNIES

4.3.1 Anàlisi filogenètica del gènere *Aeromonas*

S'han realitzat reconstruccions filogenètiques a partir de les seqüències del gen *mdh* d'una mostra representativa de 150 soques que inclou totes les espècies i subespècies del gènere *Aeromonas* (Annex I). S'han utilitzat dos mètodes diferents, el mètode de màxima versemblança (ML) i el mètode d'inferència bayesiana (IB).

En primer lloc, per veure si el gen seleccionat és adequat per a la construcció d'una filogènia, s'ha determinat si evoluciona com un rellotge molecular. A la Taula 4.11 es mostren els resultats obtinguts, que indiquen que els canvis en les seqüències no són proporcionals al llarg del temps ($P \sim 0$), i per tant, aquest gen no es comportaria com un rellotge molecular estricte. Per establir les filogènies, en aquells casos en que ha estat possible (IB) s'ha assumit un mètode de rellotge molecular relaxat.

Taula 4.11 Anàlisi del test del rellotge molecular del gen *mdh*.

	lnL ^a	Model	+G ^b	+I ^c	Likelihood Ratio Test (LRT)		
					2 (L _{no rellotge} - L _{rellotge})	n-2 ^d	Valor de P
Rellotge	-1226319,30	TN93	0,64	0,52	2431726,34	148	P ~ 0
No rellotge	-10456,13		0,78	0,49			

^a lnL, logaritme de la versemblança

^b G, valor estimat del paràmetre α de la distribució gamma

^c I, proporció de posicions invariants

^d n-2, graus de llibertat

Un altre paràmetre a tenir en compte és la determinació del model de substitució nucleotídica que millor s'ajusta a les dades, que s'estableix a partir de l'alineament múltiple de seqüències. El millor model pel gen *mdh* ha estat el TN93+G+I.

A les Figures 4.8 (ML) i 4.9 (IB) es mostren les filogènies construïdes pel conjunt de soques. En ambdós casos les topologies dels arbres són pràcticament idèntiques, obtenint-se en els nodes dels grups principals valors de *bootstrap* (ML) i de probabilitat posterior (IB) molt elevats (> 70%). Aquests resultats indiquen que les filogènies obtingudes amb aquest gen són molt robustes.

Les filogènies mostren 5 grups principals: un primer grup que inclou les espècies *A.bestiarum* i *A.salmonicida*, així com *A.popoffii* i *A.piscicola*; un segon grup que agrupa les espècies *A.hydrophila* i *A.dhakensis*, juntament amb *A.caviae*, *A. eucrenophila*, *A. encheleia*, *A.tecta*, *A.media*, *A.sanarellii* i *A.taiwanensis*; un tercer grup format per *A.bivalvium*, *A.molluscorum* i *A.rivuli*; un quart grup que inclou *A.enteropelogenes*, juntament amb *A.fluvialis*, *A.jandaei*, *A.allosaccharophila*, *A.sobria*, *A.veronii* i les espècies sinònimes (*A.culicicola*, *A.ichthiosmia*), *A.cavernicola* i *A.australiensis*, i finalment el cinquè grup format per *A.simiae*, *A.schubertii* i *A.diversa* en una branca més profunda.

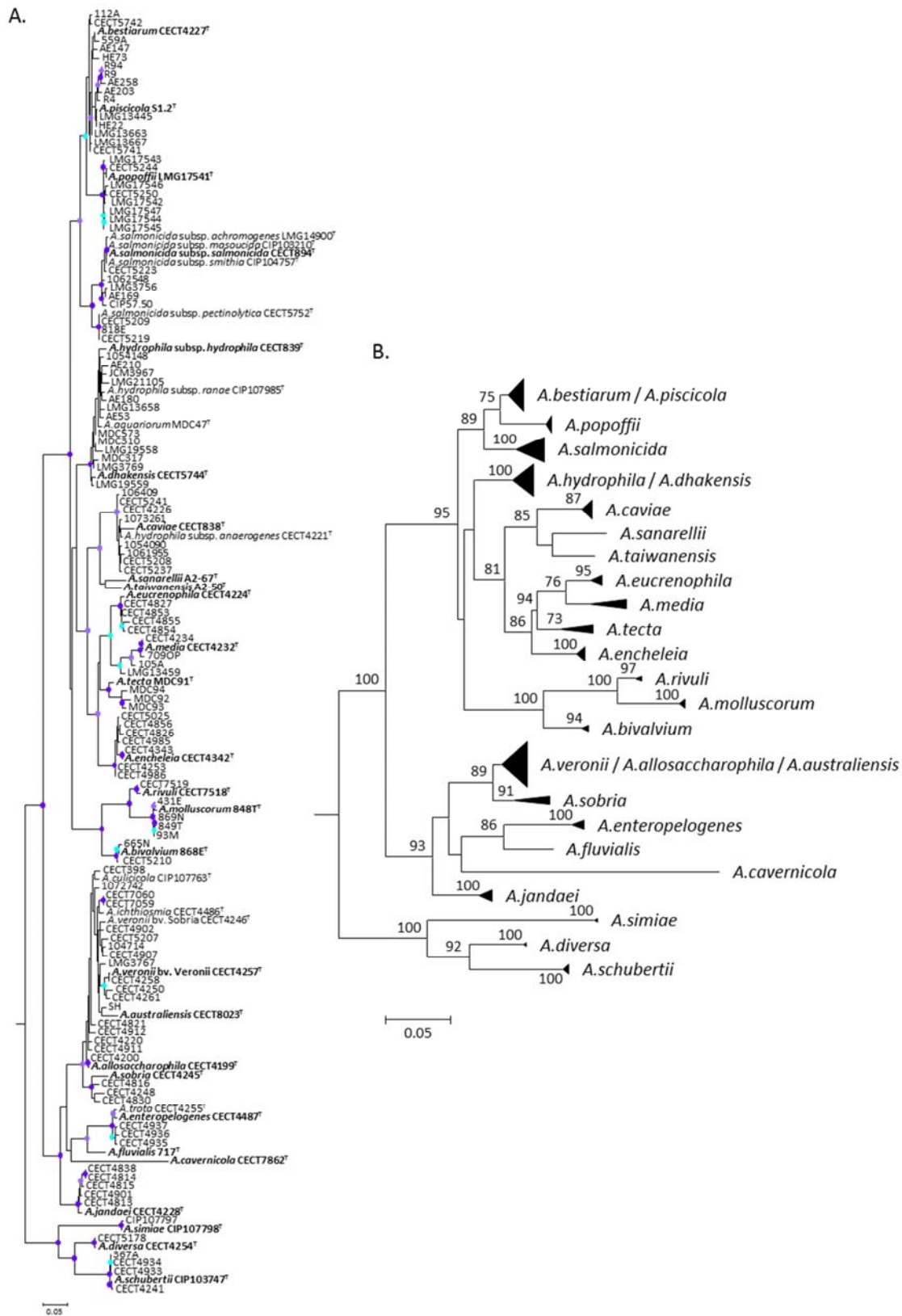


Figura 4.8 Arbre de màxima versemblança (ML) del gen *mdh*. A) Arbre amb les 150 soques. B) Arbre col·lapsat per espècies o grans grups. L'escala indica el nombre de substitucions nucleotídiques per lloc. Els nodes mostren els valors de *bootstrap* iguals o superiors al 70% (cian, 70-80%; lila clar, 80-90% i lila fosc, 90-100%).

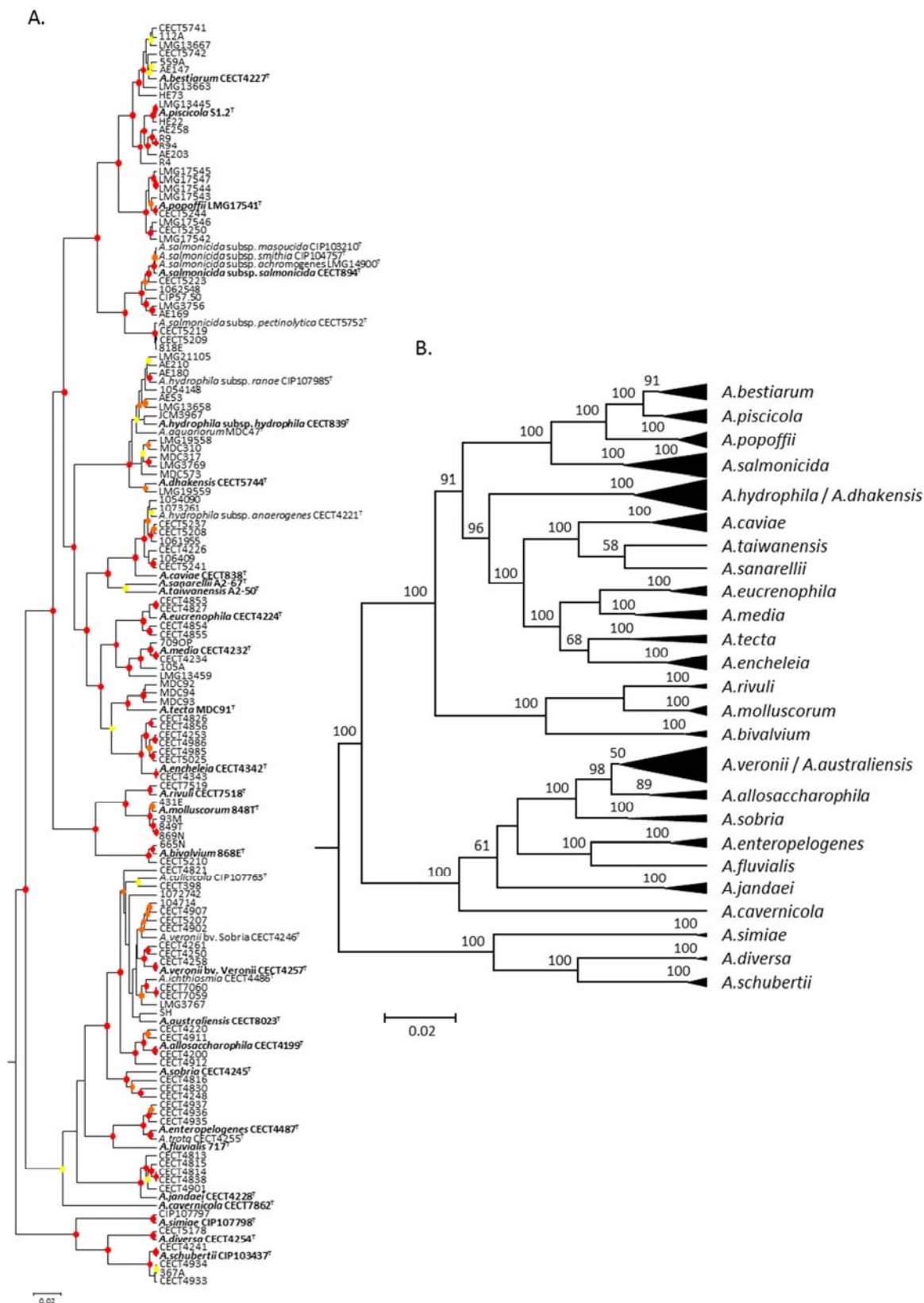


Figura 4.9 Arbre bayesià (IB) del gen *mdh*. A) Arbre amb les 150 soques. B) Arbre col·lapsat per espècies o grans grups. L'escala indica el nombre de substitucions nucleotídiques per lloc. Els nodes mostren els valors de probabilitat posterior iguals o superiors al 50% (groc, 50-70%; taronja, 70-90% i vermell, 90-100%).

Totes les soques assignades a cada espècie s'han posicionat correctament en les filogènies construïdes, amb excepció de 6 soques que s'han situat en clústers diferents dels que correspondria (Figures 4.8 i 4.9). S'han detectat 4 soques (*A.caviae* CECT5207, *A.hydrophila* CECT398, *A.hydrophila* LMG3767, *A.sobria* CECT4250) que s'han agrupat amb les espècies del complex *A.veronii*, una soca d'*A.popoffii* (CECT5210) que s'ha localitzat en el clúster d'*A.bivalvium*, i per últim, una soca d'*A.veronii* (CECT4816) que s'ha inclòs dins del grup *A.sobria*. Aquests mateixos resultats també han estat observats en les filogènies construïdes pel gen *recA* (Apartat 5.3.1).

4.3.2 Anàlisi fenotípica de les soques amb incongruències filogenètiques

Per tal de confirmar la posició taxonòmica d'aquestes soques, s'han realitzat diferents proves bioquímiques convencionals per comprovar si estan mal identificades. En aquest estudi fenotípic s'han inclòs també un conjunt de soques tipus d'*Aeromonas* corresponents tant a les espècies amb les que s'han agrupat en les filogènies com a les assignades per a cada soca, amb excepció d'*A.popoffii*, *A.hydrophila* i *A.caviae*, en que s'han utilitzat resultats prèviament publicats (Miñana-Galbis *et al.* 2004; Martin-Carnahan & Joseph 2005). Els resultats obtinguts es mostren a la Taula 4.11, assenyalant en ombrejat gris les soques problema i els resultats de les proves bioquímiques corresponents. A la part inferior de cada soca problema s'indiquen les soques tipus que s'han utilitzat i els resultats per tal de fer l'anàlisi comparativa.

Si s'observa amb detall els resultats de la Taula 4.12, es veu que per la soca *A.caviae* CECT5207 s'han fet 7 proves per discriminar entre la soca tipus *A.caviae* i espècies del complex *A.veronii*, diferint en 3 proves (Voges-Proskauer, lisina descarboxilasa i hidròlisi d'esculina) d'*A. caviae*, i presentant el mateix fenotip que *A.culicicola*, una soca sinònima d'*A.veronii*, tal com s'agrupa en les filogènies. Per les soques *A.hydrophila* CECT398 i LMG3767 s'han realitzat 6 proves discriminatòries entre les soques tipus d'*A.hydrophila* i espècies del complex *A.veronii*, comprovant que ambdues presenten el mateix fenotip que *A.culicicola*. La soca *A.popoffii* CECT5210 s'ha comparat amb les soques tipus d'*A.popoffii* i *A.bivalvium*, comprovant la seva incorrecta assignació com *A.popoffii* en base als resultats de 6 proves. Per estudiar la soca *A.sobria* CECT4250 s'han realitzat 8 proves per poder discriminar entre la soca tipus *A.sobria* i espècies del complex *A. veronii*, i en aquest cas els resultats obtinguts indiquen que pertany a l'espècie *A.culicicola* del complex *A.veronii*. Per últim, la soca *A.veronii* CECT4816 que encara que s'agrupa en les filogènies amb *A.sobria*, fenotípicament es classificaria com *A. veronii* bv. *Veronii*.

En alguns casos dubtosos, la caracterització s'ha complementat amb l'estudi de sensibilitat a diferents antibiòtics. A partir de la interpretació dels resultats dels antibiogrames, s'ha comprovat que la soca problema *A.caviae* CECT5207 junt amb la soca tipus d'*A.veronii* han estat sensibles a la cefoxitina i la soca tipus d'*A.caviae* resistent. La soca *A.sobria* CECT4250 ha estat resistent a ticarcil·lina com la soca tipus d'*A.veronii*, en canvi la soca tipus d'*A.sobria* ha estat sensible. Finalment, la soca *A.popoffii* CECT5210 també ha mostrat sensibilitat a ticarcil·lina a diferència de la soca tipus de la mateixa espècie.

Taula 4.12 Resultats de l'estudi fenotípic de les soques amb incongruències filogenètiques.

	Oxidasa	Catalasa	Voges-Proskauer	Roig de metil	H ₂ S a partir de cisteïna	Lisina descarboxilasa	Ornitina descarboxilasa	Hidrolisi d'elastina	Hidrolisi d'esculina	Hidrolisi de midó	Citrat de Christensen
<i>A.caviae</i> CECT5207	+	+	+			+	-	-	+		
<i>A.caviae</i> CECT838 ^T	+	+	-			-	-	+	+		
<i>A.veronii</i> bv. Veronii CECT4257 ^T	+	+	+			+	+	+	+		
<i>A.veronii</i> bv. Sobria CECT4246 ^T	+	+	+			+	-	-	-		
<i>A.culicicola</i> CIP107763 ^T	+	+	+			+	-	-	+		
<i>A.hydrophila</i> CECT398	+	+					-	-	-	+	
<i>A.hydrophila</i> CECT839 ^T	+	+					-	+	+	+	
<i>A.veronii</i> bv. Veronii CECT4257 ^T	+	+					+	-	+	+	
<i>A.veronii</i> bv. Sobria CECT4246 ^T	+	+					-	-	-	-	
<i>A.culicicola</i> CIP107763 ^T	+	+					-	-	-	+	
<i>A.hydrophila</i> LMG3767	+	+					-	-	-	+	
<i>A.hydrophila</i> CECT839 ^T	+	+					-	+	+	+	
<i>A.veronii</i> bv. Veronii CECT4257 ^T	+	+					+	-	+	+	
<i>A.veronii</i> bv. Sobria CECT4246 ^T	+	+					-	-	-	-	
<i>A.culicicola</i> CIP107763 ^T	+	+					-	-	-	+	
<i>A.popoffii</i> CECT5210	+	+	-	-	+				+		
<i>A.popoffii</i> LMG17541 ^T	+	+	+	+	-				-		
<i>A.bivalvium</i> 868E ^T	+	+	-	-	+				+		
<i>A.sobria</i> CECT4250	+	+	+	-			-	-	+	+	
<i>A.sobria</i> CECT4245 ^T	+	+	+	-			-	-	+	-	
<i>A.veronii</i> bv. Veronii CECT4257 ^T	+	+	+	+			+	+	+	+	
<i>A.veronii</i> bv. Sobria CECT4246 ^T	+	+	+	-			-	-	-	-	
<i>A.culicicola</i> CIP107763 ^T	+	+	+	-			-	-	+	+	
<i>A.veronii</i> CECT4816	+	+	+	+			+	+	+	+	
<i>A.veronii</i> bv. Veronii CECT4257 ^T	+	+	+	+			+	+	+	+	
<i>A.veronii</i> bv. Sobria CECT4246 ^T	+	+	+	-			-	-	-	-	
<i>A.culicicola</i> CIP107763 ^T	+	+	+	-			-	-	+	+	
<i>A.sobria</i> CECT4245 ^T	+	+	+	-			-	-	+	-	

En resum, de les 6 soques analitzades, 5 han estat reclassificades en base als resultats de les proves bioquímiques, els antibiogrames i la seva agrupació en els arbres filogenètics. A continuació, a la Taula 4.13 s'indica la designació actual de les soques que s'han detectat incongruències filogenètiques i la proposta de reclassificació.

Taula 4.13 Proposta de reclassificació de les soques problema.

Classificació actual	Reclassificació
<i>A.caviae</i> CECT5207	<i>A.culicicola</i> (complex <i>A.veronii</i>)
<i>A.hydrophila</i> CECT398	<i>A.culicicola</i> (complex <i>A.veronii</i>)
<i>A.hydrophila</i> LMG3767	<i>A.culicicola</i> (complex <i>A.veronii</i>)
<i>A.popoffii</i> CECT5210	<i>A.bivalvium</i>
<i>A.sobria</i> CECT4250	<i>A.culicicola</i> (complex <i>A.veronii</i>)
<i>A.veronii</i> CECT4816	-

4.4 DATACIÓ

A partir de les seqüències del gen *mdh* s'ha realitzat un estudi de datació per estimar l'origen del gènere *Aeromonas* i els temps de divergència dels seus llinatges. Aquesta anàlisi s'ha fet utilitzant diferents aproximacions: A) a partir de les 150 soques incloses en l'estudi, B) amb les 27 seqüències consens de totes les espècies, C) amb les 27 seqüències de les soques tipus de cada espècie, i D) a partir de l'arbre d'espècies (27 espècies). Per tal de realitzar la datació, s'han construït arbres ultramètrics datats (cronogrames) per inferència bayesiana mitjançant el programa BEAST, utilitzant el millor model d'evolució per les dades, un model de rellotge molecular relaxat, i com a punt de calibratge el temps de diversificació entre *Escherichia coli* i *Salmonella typhimurium* (140 Ma).

4.4.1 A partir de les 150 seqüències

S'ha obtingut un cronograma del gènere *Aeromonas*, considerant el model de substitució TN93+G+I que és el que millor s'ajusta a les seqüències de les 150 soques (Figura 4.10). L'arbre obtingut és robust, ja que els valors de probabilitat posterior dels nodes principals han estat elevats. A la Taula 4.14 es mostren els temps de divergència dels principals grups i els seus intervals de confiança (IC).

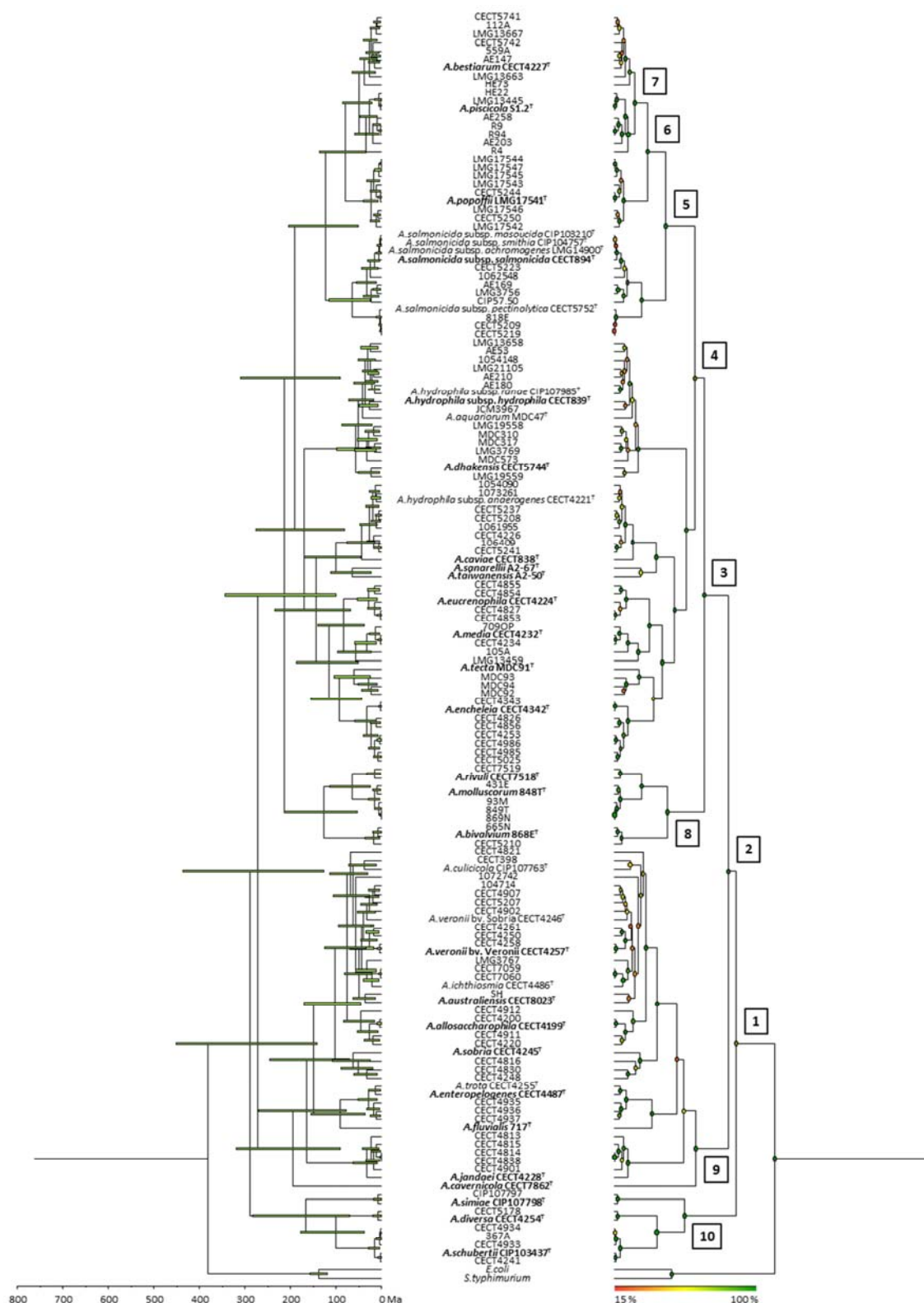


Figura 4.10 Cronograma del gènere *Aeromonas* basat en les 150 seqüències del gen *mdh*. A l'esquerra es mostra el cronograma a escala temporal (Ma), i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 4.14 Edat de divergència estimada dels principals grups del cronograma obtingut a partir de les 150 seqüències del gen *mdh*.

Clúster	Edat	IC inferior	IC superior
1	290,0	142,9	451,1
2	271,4	127,4	436,8
3	213,9	101,6	344,8
9	194,9	91,1	320,2
4	191,7	90,5	309,2
10	167,5	71,1	283,6
8	126,6	54,1	214,0
5	122,5	51,9	203,8
6	79,2	34,0	136,3
7	49,8	20,5	85,5

Segons els resultats obtinguts amb les 150 seqüències del gen *mdh*, l'origen de l'ancestre comú de totes les espècies descrites fins a l'actualitat del gènere *Aeromonas* s'estimaria en aproximadament 290 Ma. A partir d'aquest ancestre s'hauria produït la primera separació d'un llinatge que ha donat lloc a la majoria d'espècies actuals del gènere, i d'un segon grup que conté les espècies *A.diversa*, *A.schubertii* i *A.simiae*. També s'ha estimat el temps de diversificació de les espècies més recents: *A.bestiarum*-*A.piscicola* (50 Ma), *A.sanarellii*-*A.taiwanensis* (64 Ma), *A.rivuli*-*A.molluscorum* (65 Ma) i *A.allosaccharophila*-*A.veronii* (75 Ma). De forma general, la taxa evolutiva obtinguda pel gen *mdh* ha sigut de $3,8 \times 10^{-4}$ substitucions/lloc/milió d'any amb una desviació estàndard de $4,6 \times 10^{-6}$.

4.4.2 A partir de les seqüències consens

S'ha determinat l'origen del gènere *Aeromonas* a partir de les seqüències consens de totes les espècies. Per això, s'ha construït un arbre bayesià, seguint el model de substitució TN93+G+I, i utilitzant el mateix punt de calibratge que en l'anàlisi anterior. A la Figura 4.11 es mostra el cronograma obtingut, i a la Taula 4.15 l'edat dels principals nodes i els seus intervals de confiança (IC). Els valors de probabilitat posterior en la majoria dels nodes han estat propers al 100%. En base a les seqüències consens i tenint en compte el punt de calibratge considerat, aquest gènere hauria començat a divergir fa aproximadament 184 Ma. Les diversificacions d'espècies més recents van ocórrer fa uns 11 Ma per *A.bestiarum*-*A.piscicola* i *A.hydrophila*-*A.dhakensis* i 15 Ma per *A.australiensis*-*A.veronii*.

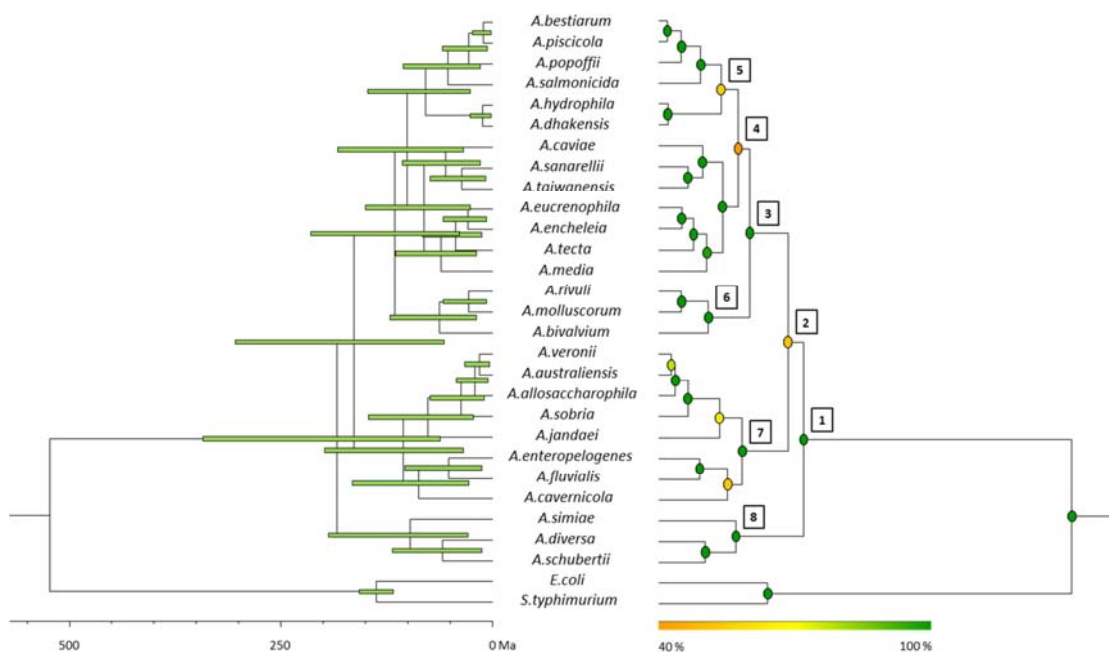


Figura 4.11 Cronograma del gènere *Aeromonas* basat en les seqüències consens del gen *mdh* de les diferents espècies. A l'esquerra es mostra el cronograma a escala temporal (Ma) i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 4.15 Edat de divergència estimada dels principals grups del cronograma obtingut a partir de les seqüències consens del gen *mdh*.

Clúster	Edat	IC inferior	IC superior
1	183,5	61,5	342,1
2	163,7	57,5	303,8
3	115,2	39,6	215,2
7	105,8	34,8	198,1
4	100,6	34,6	182,7
8	97,6	29,5	193,8
5	78,9	26,5	147,3
6	62,8	19,3	120,6

4.4.3 A partir de les seqüències de les soques tipus

S'ha estimat l'edat d'aparició de l'ancestre d'*Aeromonas* a partir de les seqüències de les soques tipus de cada espècie considerades en l'estudi. En aquest cas, s'ha construït l'arbre bayesià corresponent seguint el model de substitució nucleotídica TN93+G+I i considerant el mateix punt de calibratge (140 Ma). El cronograma obtingut mostra valors de probabilitat

posteriors elevats en els nodes principals (Figura 4.12), i a la Taula 4.16 s'indica l'edat dels grups principals i els seus intervals de confiança (IC). Segons les seqüències del gen *mdh* de les soques tipus de totes les espècies, el gènere *Aeromonas* s'hauria originat fa aproximadament 183 Ma. Les diversificacions més recents es van produir quan es van separar les espècies *A.bestiarum*-*A.piscicola* (8 Ma), *A.hydrophila*-*A.dhakensis* (25 Ma) i *A.allosaccharophila*-*A.sobria* (25 Ma).

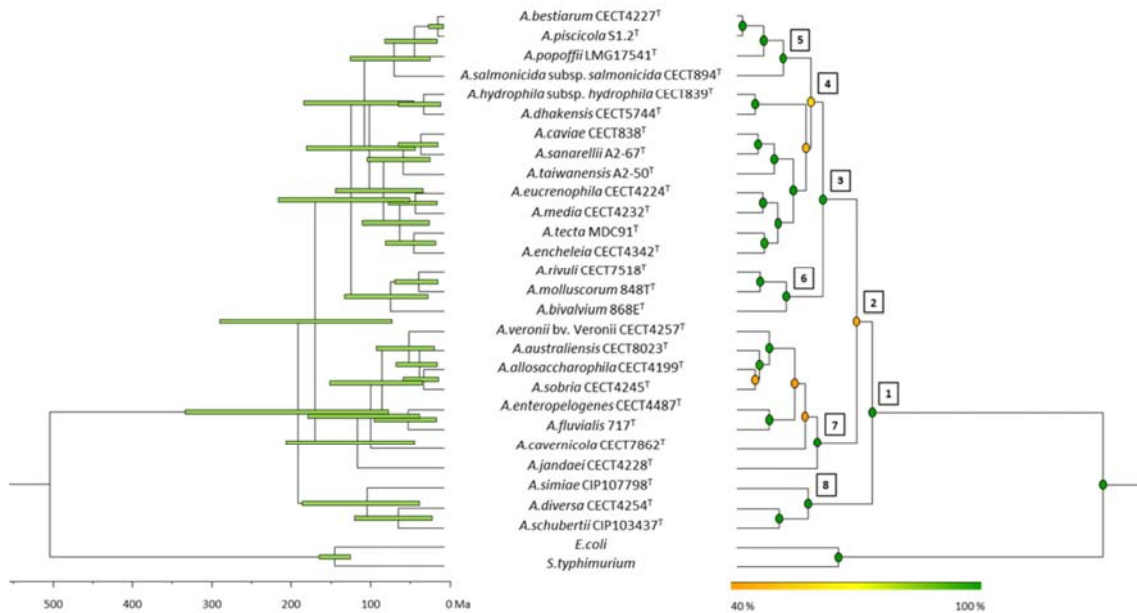


Figura 4.12 Cronograma del gènere *Aeromonas* basat en les seqüències del gen *mdh* de les soques tipus. A l'esquerra es mostra el cronograma a escala temporal (Ma) i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 4.16 Edat de divergència estimada dels grups principals del cronograma obtingut a partir de les seqüències de les soques tipus del gen *mdh*.

Clúster	Edat	IC inferior	IC superior
1	183,4	69,8	326,1
2	162,2	65,6	282,0
3	117,3	43,7	208,9
7	108,8	37,2	198,8
4	100,6	38,0	176,2
8	97,0	30,9	178,2
6	67,5	20,5	125,2
5	62,9	17,9	118,4

4.4.4 A partir de l'arbre d'espècies

S'ha construït un arbre d'espècies datat del gènere *Aeromonas* basat en les seqüències del gen *mdh*, utilitzant com a millor model de substitució el TN93+G+I, i considerant el mateix punt de calibratge que en totes les datacions anteriors. A la Figura 4.13 es mostra el cronograma obtingut i a la Taula 4.17 s'indiquen els valors estimats de datació dels grups principals i els seus intervals de confiança (IC). L'arbre d'espècies obtingut a partir de les seqüències del gen *mdh* fixaria l'origen d'aquest gènere bacterià fa aproximadament uns 190 Ma. També s'han estimat les diversificacions més recents quan es van separar les espècies *A.bestiarum*-*A.piscicola* (15 Ma), *A.hydrophila*-*A.dhakensis* (16 Ma) i *A.australiensis*-*A.veronii* (22 Ma).

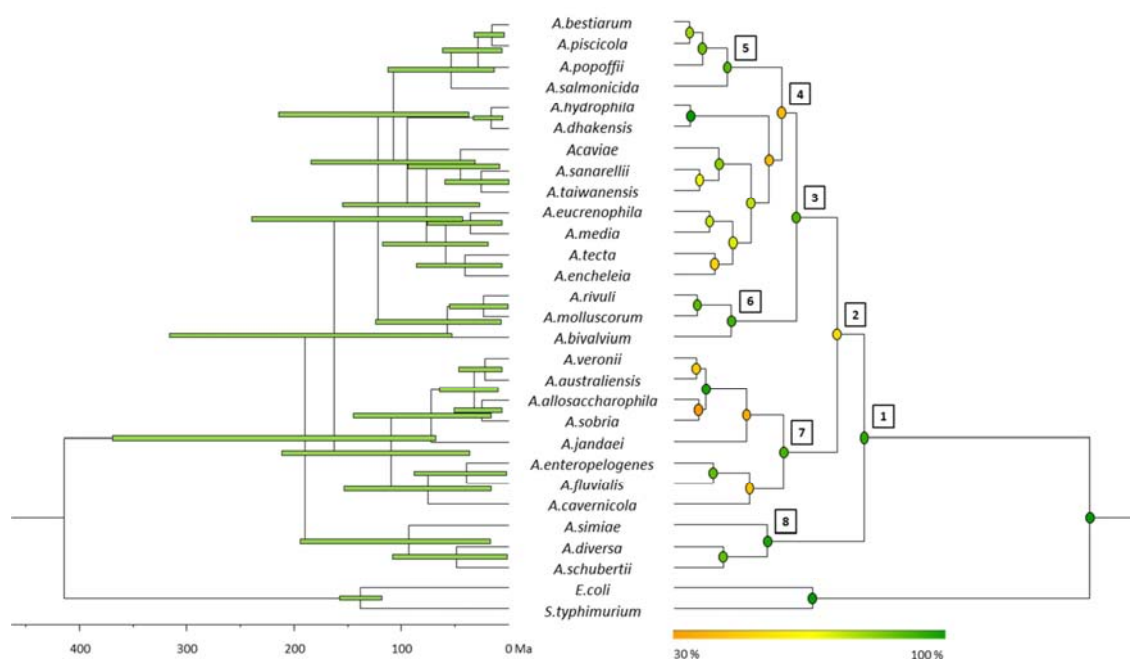


Figura 4.13 Cronograma de l'arbre d'espècies del gènere *Aeromonas* obtingut a partir de les seqüències del gen *mdh*. A l'esquerra es mostra el cronograma a escala temporal (Ma) i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 4.17 Edat de divergència estimada dels grups principals del cronograma obtingut a partir de l'arbre d'espècies del gen *mdh*.

Clúster	Edat	IC inferior	IC superior
1	189,5	67,5	369,4
2	162,2	52,7	315,7
3	121,8	42,8	239,1
7	109,3	36,5	211,1
4	107,2	36,9	214,0
8	93,0	16,8	194,1
6	57,2	6,9	123,7
5	53,1	13,2	112,1

4.5. ANÀLISI DE LA DIVERSIFICACIÓ

Tal com s'ha comentat anteriorment, per determinar el model i la taxa de diversificació d'un taxó és necessari partir d'un cronograma que contingui una sola seqüència per cada espècie. Per fer aquest estudi de diversificació del gènere *Aeromonas* s'han utilitzat 3 aproximacions diferents a partir de cronogrames obtinguts pel mètode bayesià de: A) les 27 seqüències consens, B) les 27 seqüències corresponents a les soques tipus, i C) a partir de l'arbre d'espècies (27 espècies).

En cada cas, s'ha determinat gràficament el patró de diversificació dels llinatges al llarg del temps (LTT plot), per tal de veure si segueix un model constant o bé apareixen acceleracions o desacceleracions al llarg del procés. La fiabilitat dels resultats obtinguts s'ha determinat mitjançant l'ajust a una recta de regressió i per simulació. A més, a partir del pendent de la recta de regressió s'ha obtingut un valor estimat de λ . El valor de taxa de diversificació (λ) s'ha calculat també a partir del cronograma, ajustant les dades a un model constant mitjançant una prova de màxima versemblança.

S'ha determinat l'estadístic Υ de Pybus i Harvey, el test de ΔAIC_{RC} comparant models de diversificació constants i variables que millor s'ajusten a les dades, i el test de Rabosky per validar mitjançant simulació el millor model de diversificació.

S'ha completat l'estudi realitzant una anàlisi amb el programa BMM per inferència bayesiana per detectar i quantificar heterogeneïtat en les taxes d'evolució. S'han realitzat diverses aproximacions: s'ha determinat la taxa de diversificació per cada branca de la filogènia bayesiana, s'ha representat gràficament la taxa neta de diversificació al llarg del temps, s'ha estimat la probabilitat de que es produeixi o no un o més punts de canvi (*shift*) en la taxa de diversificació, i s'ha calculat el factor de Bayes que és una mesura del suport estadístic de les taxes d'evolució estimades.

4.5.1 A partir de les seqüències consens

Per tal de visualitzar el patró de diversificació temporal s'ha realitzat un LTT plot (Figura 4.14A), que suggereix que els llinatges s'haurien originat de manera constant al llarg del temps, obtenint-se un bon ajust lineal per les dades (Figura 4.14B). El valor de λ estimat a partir del pendent de la recta de regressió ha estat de 0,0156. A partir del cronograma, s'han ajustat les dades a un model constant de naixement i mort (*Birth-Death*) ó de naixement pur (*Pure-Birth* o model de Yule) mitjançant una prova de màxima versemblança, obtenint un valor de taxa de diversificació (λ) de 0,0133 (SE = 0,0019) i una taxa d'extinció μ de 0, indicant que seguiria un model de Yule.

Per avaluar la fiabilitat dels resultats obtinguts, s'han simulat 1000 filogènies utilitzant el valor λ obtingut empíricament, el nombre d'espècies analitzades, i Yule com a model diversificació. A partir dels cronogrames generats s'han obtingut els corresponents LTT plots (Figura 4.14C). En la simulació s'observa que l'LTT plot obtingut amb les dades empíriques (línia contínua negra) està clarament dins del 95% dels LTTs obtinguts a partir de la simulació (línies discontinues).

També s'ha calculat l'estadístic Υ de Pybus i Harvey, per tal de veure si el model de diversificació s'ajusta a un model constant o no. A la Figura 4.14D es mostra la seva distribució per simulació de 1000 filogènies seguint un model constant de Yule. El valor empíric de Υ ha estat de -1,7177 (fletxa vertical), i s'ha situat entre el límit superior (1,0662) i inferior (-2,2369) de l'interval de confiança del 95%.

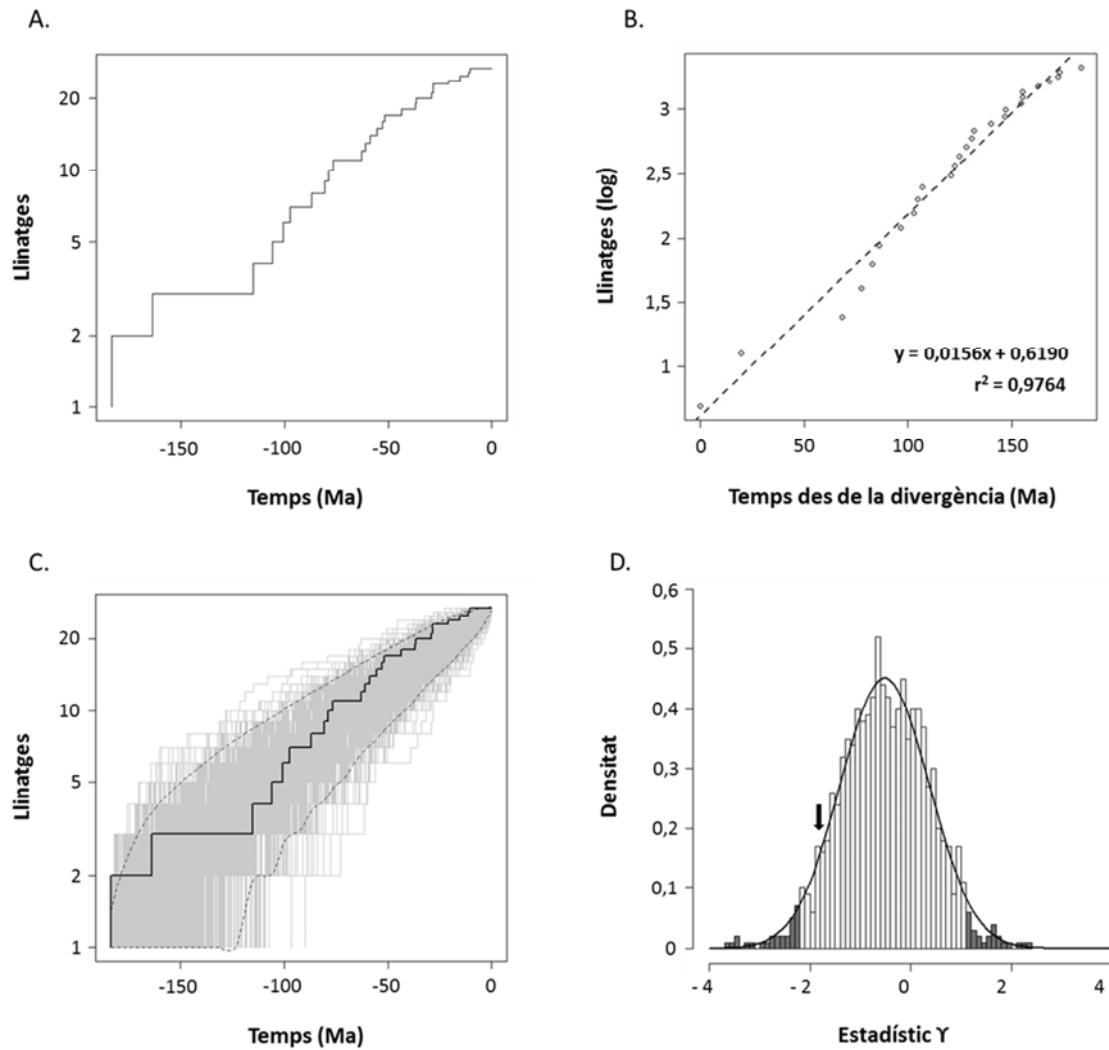


Figura 4.14 Anàlisi del model de diversificació del gènere *Aeromonas* basat en les seqüències consens del gen *mdh*. A) LTT plot obtingut a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió de l'LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic γ de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

A més, s'han analitzat diferents models de diversificació variables i constants mitjançant una prova de màxima versemblança, per veure quin és el model que millor s'ajusta a les dades (Figura 4.15A). A partir dels valors d'AIC obtinguts s'ha estimat el millor model de diversificació constant, que en aquest cas ha estat el model de Yule, i el millor variable, que ha sigut el model DDL (*Density Dependent Logistic*). Mitjançant el test ΔAIC_{RC} s'ha determinat que el millor model seria el DDL, de creixement logístic amb una taxa d'especiació depenent de la densitat de la població, amb un valor inicial de $\lambda = 0,0267$. Després, s'ha avaluat el nivell de significació del valor ΔAIC_{RC} obtingut (2,3449) realitzant el test de Rabosky (Figura 4.15B), obtenint-se un valor de P de 0,1105, clarament superior a 0,05, que indica que no es pot rebutjar la hipòtesi nul·la de model de Yule (Figura 4.15B).

A.

Selecció del model de diversificació					
Models constants			Models variables		
Model	<i>Birth-Death</i>	Yule	DDX	DDL	Yule 2-rate
AIC	147,4103	145,4103	145,6942	143,0654	143,6204

Millor model constant		Millor model variable	
Model	Yule	Model	DDL
AIC	145,4103	AIC	143,0654

Test ΔAIC_{RC}	
Model	DDL
AIC	143,0654
ΔAIC_{RC}	2,3449

B.

Test de Rabosky		
ΔAIC_{RC}	Valor de P	Decisió
2,3449	0,1105	→ $P > 0,05$ - Acceptar hipòtesi nul·la (H_0) de model constant de Yule

Figura 4.15 Determinació del millor model de diversificació del gènere *Aeromonas* basat en les seqüències consens del gen *mdh*. A) Selecció del model i resultat del test ΔAIC_{RC} . B) Test de Rabosky.

Encara que la majoria dels resultats coincideixen que seguiria un model de diversificació constant de Yule, s'ha completat l'estudi realitzant una anàlisi bayesiana amb el programa BAMM. La filogènia bayesiana obtinguda es mostra a la Figura 4.16A, i s'indiquen les taxes de diversificació per cada branca, que oscil·len entre 0,016 i 0,013. Aquesta lleugera variació també s'observa en el gràfic de la taxa neta de diversificació al llarg del temps, on el perfil presenta una lleugera disminució (Figura 4.16B). Per determinar si aquests resultats són significatius s'ha calculat la probabilitat que es produeixi un punt de canvi (1 *shift*) ó cap (0 *shift*) en la taxa de diversificació, donant valors de 0,0084 i 0,99, respectivament. Finalment, s'ha estimat el factor de Bayes (118,403), que al donar un valor superior a 100 mostra una evidència decisiva que no s'ha produït cap canvi, i reforça el resultat de model constant de Yule.

Per últim, de complir-se el model de Yule obtingut (λ empírica = 0,0133 i $\mu = 0$), s'ha calculat el valor teòric de l'edat de l'ancestre comú més recent (tMRCA) i el valor esperat del nombre mig d'espècies que s'haurien produït des de l'origen del gènere *Aeromonas* (Figura 4.17). Segons les determinacions realitzades, aquest gènere hauria aparegut fa aproximadament 217 Ma, un

valor similar a l'obtingut empíricament (184 Ma), i estaria constituït per un valor mig de 23 espècies (SD = 16).

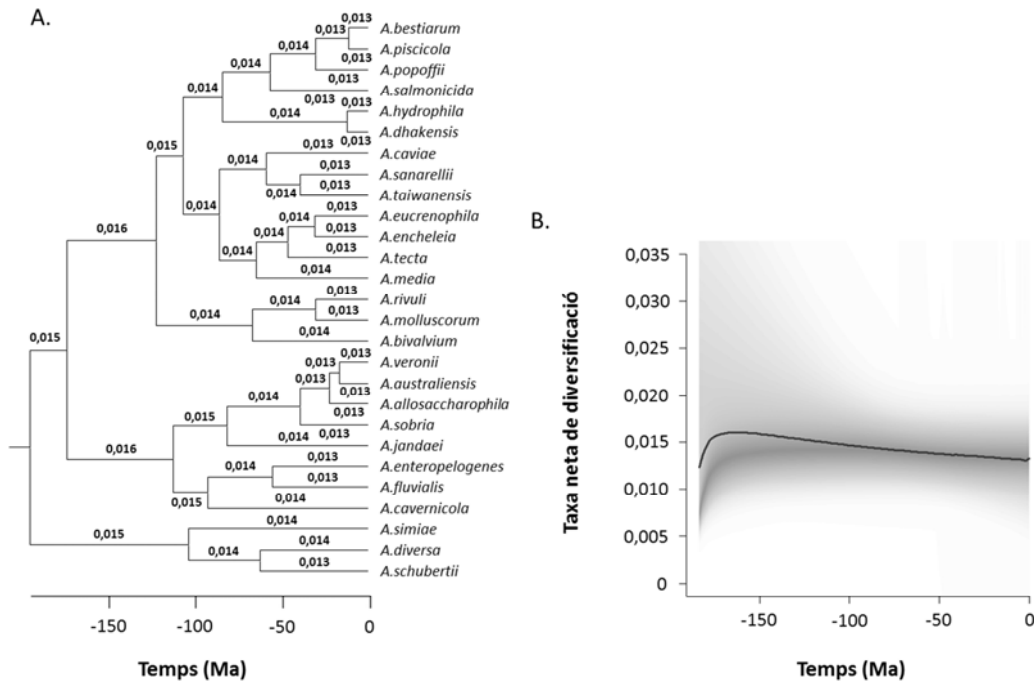


Figura 4.16 Resultats obtinguts amb el programa BAMM a partir de les seqüències consens del gen *mdh*. A) Arbre bayesià on s'indiquen les taxes de diversificació per cada branca. B) Representació gràfica de la taxa neta de diversificació al llarg del temps.

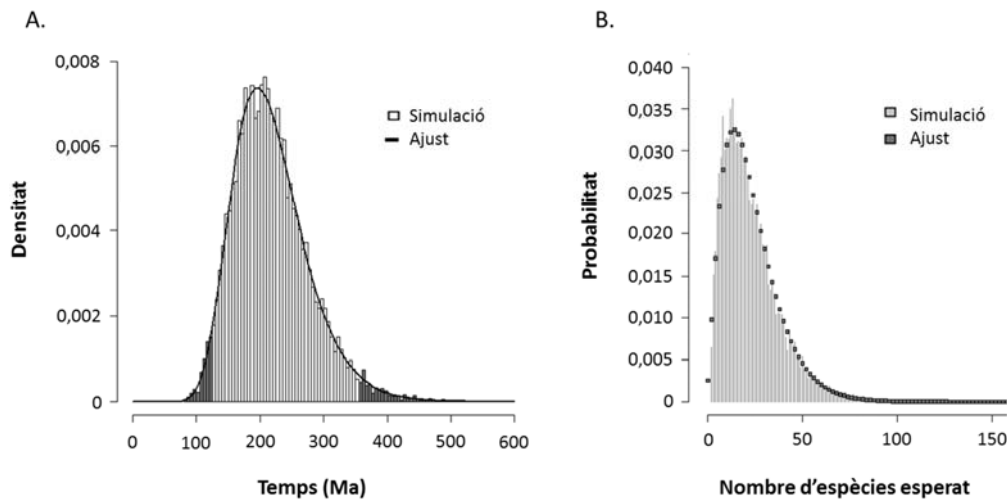


Figura 4.17 Resultats teòrics obtinguts a partir del model de Yule amb les dades de les seqüències consens del gen *mdh*. A) Histograma de l'edat de l'ancestre comú més recent (tMRCA). B) Distribució del nombre d'espècies esperat.

4.5.2 A partir de les seqüències de les soques tipus

A partir de les seqüències del gen *mdh* de les soques tipus d'*Aeromonas* s'ha obtingut l'LTT plot (Figura 4.18A) i s'ha calculat la recta de regressió corresponent (Figura 4.18B). A l'observar l'LTT plot s'ha detectat un patró de diversificació temporal constant, amb un bon ajust lineal. El valor de λ estimat a partir del pendent de la recta de regressió ha estat de 0,0160. Mitjançant una prova de màxima versemblança s'han ajustat les dades a un model constant, i s'han calculat els paràmetres $\lambda = 0,0128$ (SE = 0,0018) i $\mu = 0$.

També, s'ha realitzat una simulació a partir de 1000 filogènies per obtenir els corresponents LTT plots (Figura 4.18C). A la figura es pot veure que l'LTT plot generat empíricament (línia contínua negra) es troba dins el 95% dels obtinguts a partir de les filogènies simulades (línies discontinues).

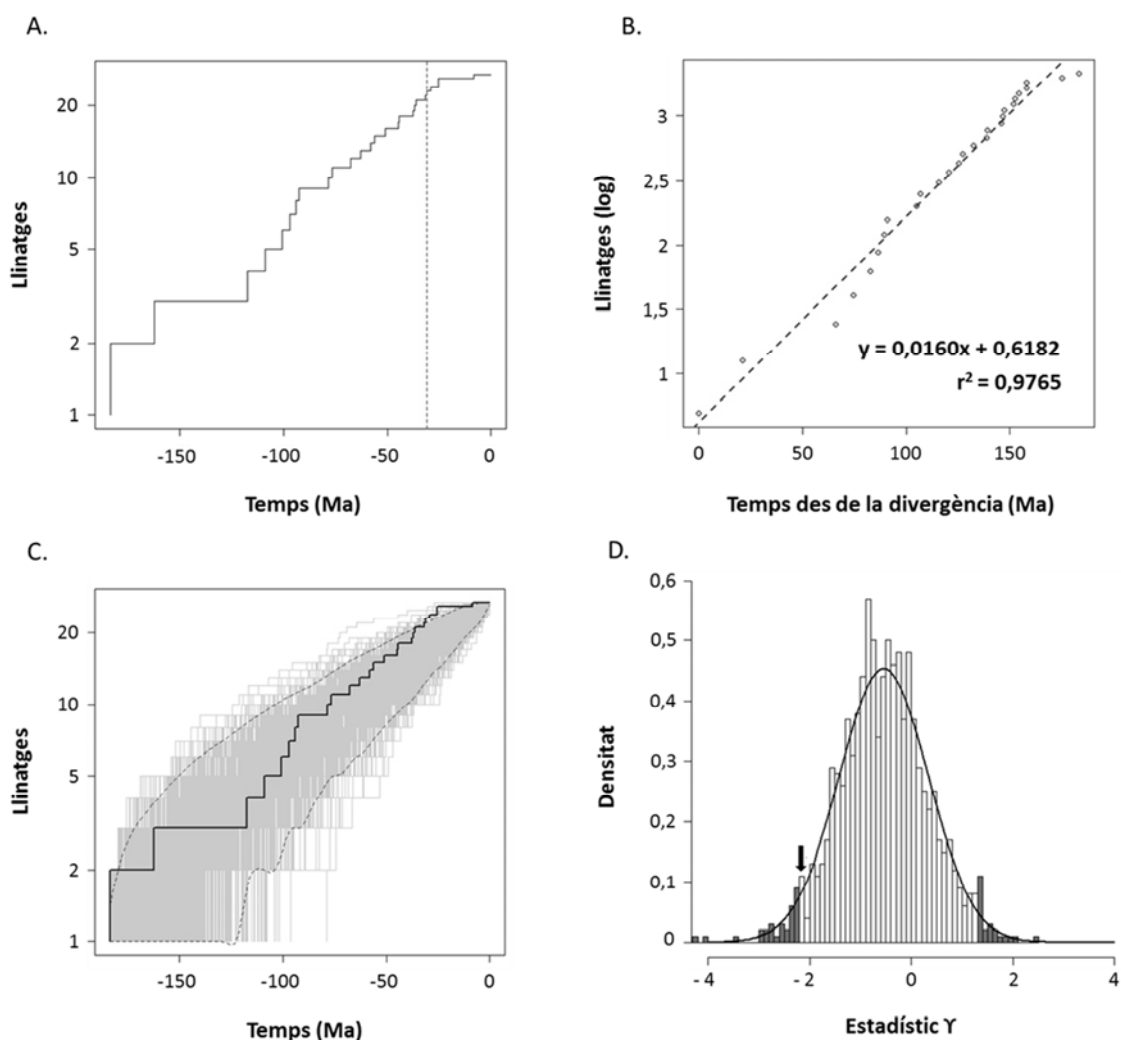


Figura 4.18 Anàlisi del model de diversificació del gènere *Aeromonas* basat en les seqüències del gen *mdh* de les soques tipus. A) LTT plot obtingut a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió de l'LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic Y de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

En relació a l'estadístic Y de Pybus i Harvey, s'ha calculat el valor de Y empíric de -2,1442 i la seva distribució seguint un model de Yule. A la Figura 4.18D s'indica amb una fletxa vertical el valor empíric obtingut, molt proper al límit inferior del 95%, però dins de l'interval de confiança (1,2856 i -2,2325).

Per analitzar el model de diversificació que millor s'ajusta a les seqüències del gen *mdh* de les soques tipus s'ha estimat, amb una prova de màxima versemblança, el millor model de taxa constant i el de taxa variable (Figura 4.19A). A partir dels valors d'AIC obtinguts, el millor model amb taxa de diversificació constant ha estat el model de Yule, i el millor amb taxa variable ha estat el model Yule-2-rate. Després s'ha realitzat el test ΔAIC_{RC} , i els resultats obtinguts estimen que el millor model és el model variable Yule-2-rate, que es caracteritza perquè la taxa de diversificació (λ) presenta dos valors al llarg del temps ($\lambda_1 = 0,0181$ i $\lambda_2 = 0,0029$). S'ha pogut determinar que el punt de canvi (*shift*) en la taxa de diversificació s'hauria produït al final del procés, fa aproximadament 25,50 Ma (línia vertical discontinua en l'LTT plot, Figura 4.18A). Finalment, el resultat del test de Rabosky (Figura 4.19B) ha permès rebutjar el model de diversificació de Yule ($P < 0,05$).

A.

Selecció del model de diversificació					
Models constants			Models variables		
Model	<i>Birth-Death</i>	Yule	DDX	DDL	Yule 2-rate
AIC	149,0760	147,0760	146,6947	142,6957	141,2474

Millor model constant		Millor model variable	
Model	Yule	Model	Yule 2-rate
AIC	147,0760	AIC	141,2474

Test ΔAIC_{RC}	
Model	Yule 2-rate
AIC	141,2474
ΔAIC_{RC}	5,8285

B.

Test de Rabosky		
ΔAIC_{RC}	Valor de P	Decisió
5,8285	0,0128	$\rightarrow P < 0,05$ - Rebutjar hipòtesi nul·la (H_0) de model constant de Yule

Figura 4.19 Determinació del millor model de diversificació del gènere *Aeromonas* basat en les seqüències del gen *mdh* de les soques tipus. A) Selecció del model i resultat del test ΔAIC_{RC} . B) Test de Rabosky.

L'estudi s'ha completat utilitzant el programa BMM per detectar heterogeneïtat en les taxes d'evolució. En la filogènia bayesiana (Figura 4.20A) s'ha pogut observar com les taxes de diversificació (λ) de cadascuna de les branques oscil·len entre 0,017 i 0,012. Aquesta variació ha estat concordant amb el gràfic obtingut de la taxa neta de diversificació (Figura 4.20B), en el que s'ha observat una lleugera disminució en la taxa al llarg del temps. Per veure si aquesta disminució és significativa s'ha determinat quina és la probabilitat de que les dades segueixin un model sense canvis o que presentin un ó mes punts de canvi en la taxa de diversificació, estimant que hi ha una probabilitat de 0,99 que el procés d'especiació es mantingui constant al llarg del temps, i de 0,0070 de que s'hagi produït un *shift*. Aquest resultat està recolzat pel valor del factor de Bayes obtingut de 141,857, que és clarament superior a 100, proporcionant una evidència decisiva a favor de que no s'ha produït cap punt de canvi en la taxa de diversificació.

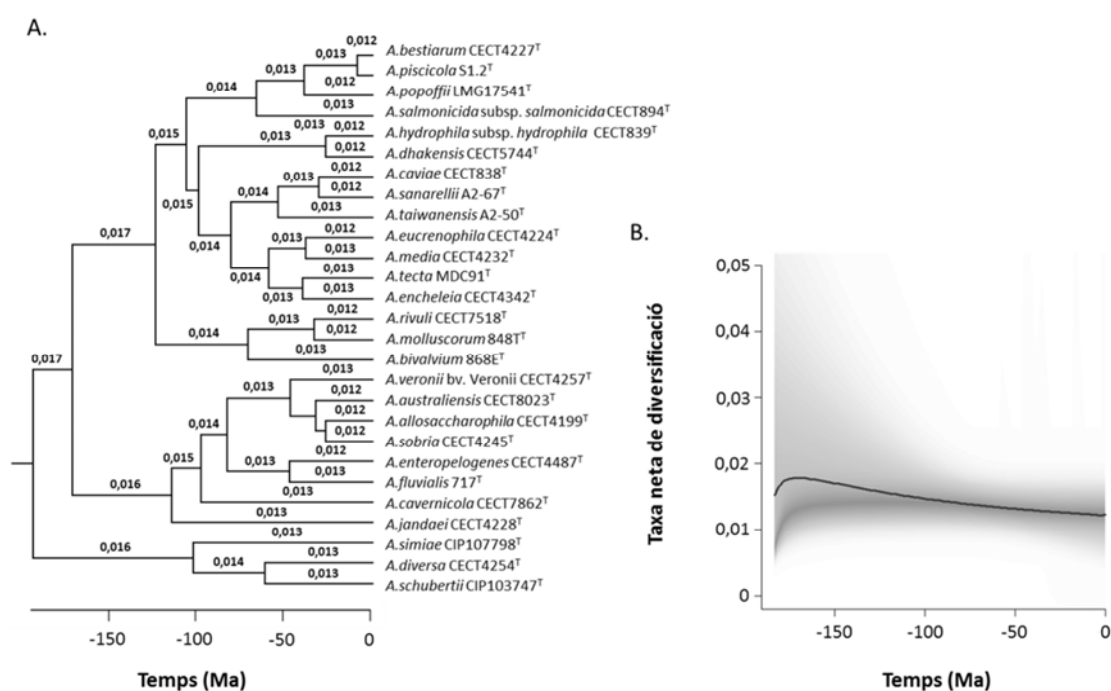


Figura 4.20 Resultats obtinguts amb el programa BMM a partir de les seqüències del gen *mdh* de les soques tipus. A) Arbre bayesià on s'indiquen les taxes de diversificació per cada branca. B) Representació gràfica de la taxa neta de diversificació al llarg del temps.

Finalment s'ha estimat teòricament, seguint un model de Yule (λ empírica = 0,0128 i $\mu = 0$), l'origen i el nombre mig d'espècies actuals esperades d'*Aeromonas* (Figura 4.21). En relació al valor tMRCA obtingut, aquest indica que el gènere hauria aparegut fa aproximadament 224 Ma, valor comparable al determinat a partir de les dades (183 Ma), i que durant el temps que *Aeromonas* ha estat evolucionant s'haurien originat 21 espècies (SD = 14).

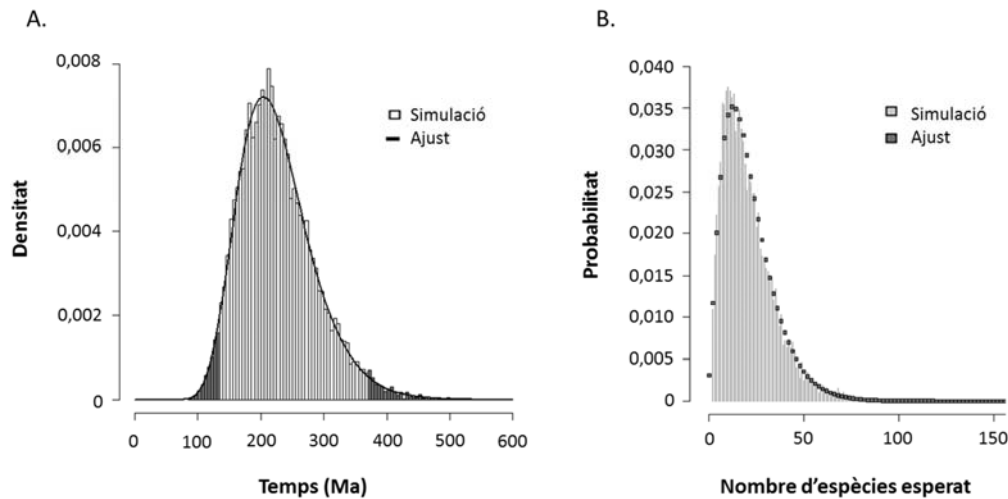


Figura 4.21 Resultats teòrics obtinguts a partir del model de Yule amb les seqüències del gen *mdh* de les soques tipus. A) Histograma de l'edat de l'ancestre comú més recent (tMRCA). B) Distribució del nombre d'espècies esperat.

4.5.3 A partir de l'arbre d'espècies

A partir de l'arbre d'espècies datat del gen *mdh* s'ha generat un LTT plot (Figura 4.22A) que ha mostrat un patró de diversificació de les espècies constant al llarg del temps, que s'ajusta molt bé a la recta de regressió (Figura 4.22B). El valor de λ estimat a partir del pendent de la recta de regressió ha estat de 0,0153. Aquest resultat s'ha confirmat per simulació de 1000 cronogrames obtenint-se els corresponents LTT plots (Figura 4.22C). L'LTT plot generat empíricament es troba dins del 95% dels obtinguts a partir de les filogènies simulades. A l'ajustar les dades a un model de taxa constant, s'ha obtingut un valor de λ de 0,0134 (SE = 0,0019) i un valor de μ de 0, que correspon a un model de Yule.

A la Figura 4.22D es mostren els resultats obtinguts de l'estadístic Y de Pybus i Harvey. En aquesta anàlisi, el valor empíric de Y ha estat de -1,4784 (fletxa vertical), i s'ha localitzat entre el límit superior (1,4374) i inferior (-2,2144) de l'interval de confiança del 95%.

A la Figura 4.23A es mostren els resultats obtinguts de l'anàlisi de selecció del model de diversificació juntament amb el resultat del test ΔAIC_{RC} . El model de Yule ha estat el millor model constant, el de Yule-2-rate el millor model amb taxa de diversificació variable, i després de realitzar el test ΔAIC_{RC} s'ha determinat que el que millor s'ajusta a les dades és el de Yule-2-rate. Els dos valors de λ obtinguts han estat 0,0180 (λ_1) i 0,0034 (λ_2), i el punt de canvi s'hauria produït fa aproximadament 22 Ma (línia vertical discontinua en l'LTT plot, Figura 4.22A). El test de Rabosky (Figura 4.23B) ha permès rebutjar la hipòtesi nul·la de model de Yule, però amb un valor de P (0,0447) molt proper a 0,05.

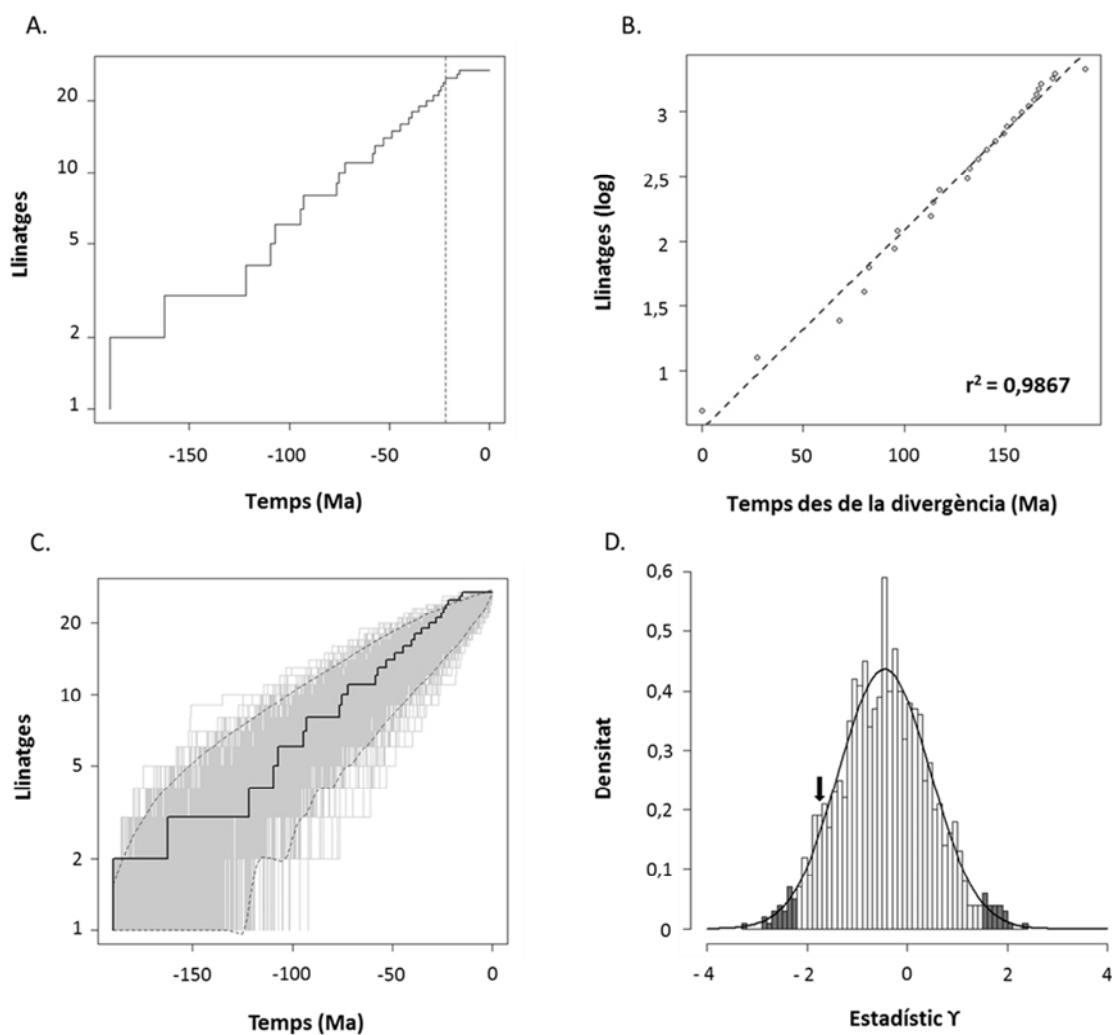


Figura 4.22 Anàlisi del model de diversificació del gènere *Aeromonas* basat en l'arbre d'espècies del gen *mdh*. A) LTT plot obtingut a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió de l'LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic Y de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

A.

Selecció del model de diversificació					
Models constants			Models variables		
Model	<i>Birth-Death</i>	Yule	DDX	DDL	Yule 2-rate
AIC	146,7744	144,7744	145,6834	143,3923	140,7288

Millor model constant		Millor model variable	
Model	Yule	Model	Yule 2-rate
AIC	144,7744	AIC	140,7288

Test ΔAIC_{RC}	
Model	Yule 2-rate
AIC	140,7288
ΔAIC_{RC}	4,0456

B.

Test de Rabosky		
ΔAIC_{RC}	Valor de P	Decisió
4,0456	0,0447	$P < 0,05$ - Rebutjar hipòtesi nul·la (H_0) de model constant de Yule

Figura 4.23 Determinació del millor model de diversificació del gènere *Aeromonas* basat en l'arbre d'espècies del gen *mdh*. A) Selecció del model i resultat del test ΔAIC_{RC} . B) Test de Rabosky.

A la Figura 4.24A es mostra com la taxa de diversificació es manté constant en totes les branques de l'arbre amb un valor de 0,014. Aquesta constància en el valor de λ està en concordança amb el gràfic obtingut de taxa neta de diversificació al llarg del temps (Figura 4.24B). El càlcul de la probabilitat de que no hi hagi cap canvi en el valor de λ ha donat un valor de 0,99, mentre que el d'un punt de canvi és molt baix (0,0089). A més, el factor de Bayes obtingut ha estat de 111,919, valor superior a 100, que suposa una evidència decisiva de taxa de diversificació constant.

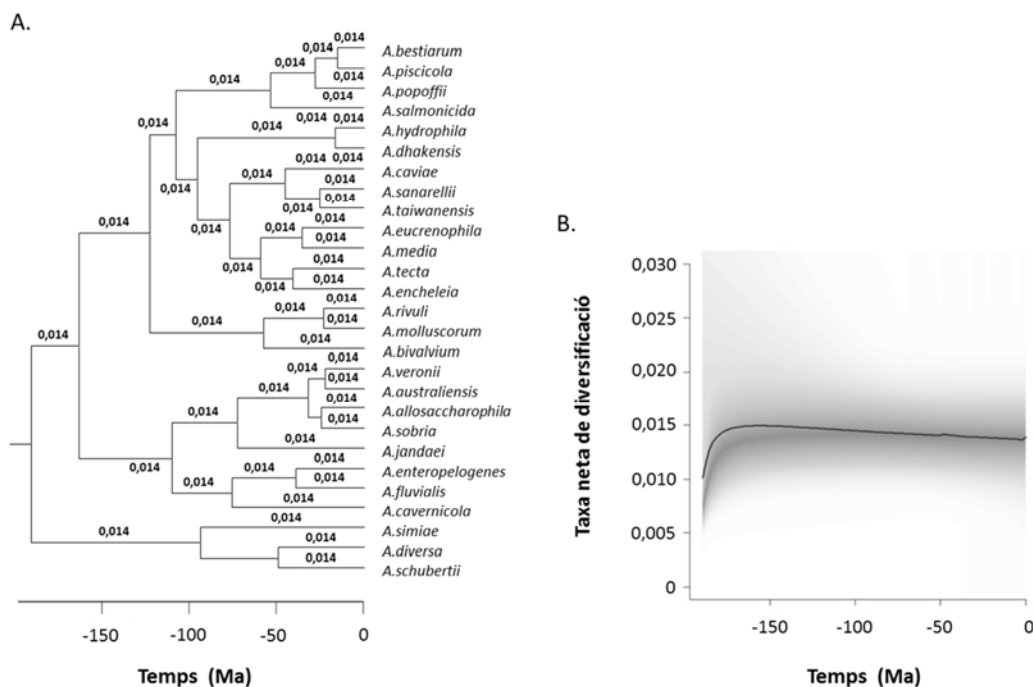


Figura 4.24 Resultats obtinguts amb el programa BAMM a partir de les seqüències del gen *mdh* de l’arbre d’espècies. A) Arbre bayesià on s’indiquen les taxes de diversificació per cada branca. B) Representació gràfica de la taxa neta de diversificació al llarg del temps.

Finalment, s’ha calculat el valor teòric del tMRCA i el valor esperat del nombre mig d’espècies actuals d’*Aeromonas*, considerant un model de Yule (λ empírica = 0,0134 i $\mu = 0$). Els resultats obtinguts (Figura 14.25) indiquen que el gènere *Aeromonas* s’hauria originat fa uns 214 Ma (valor empíric 190 Ma) i estaria format per 25 espècies (SD = 17).

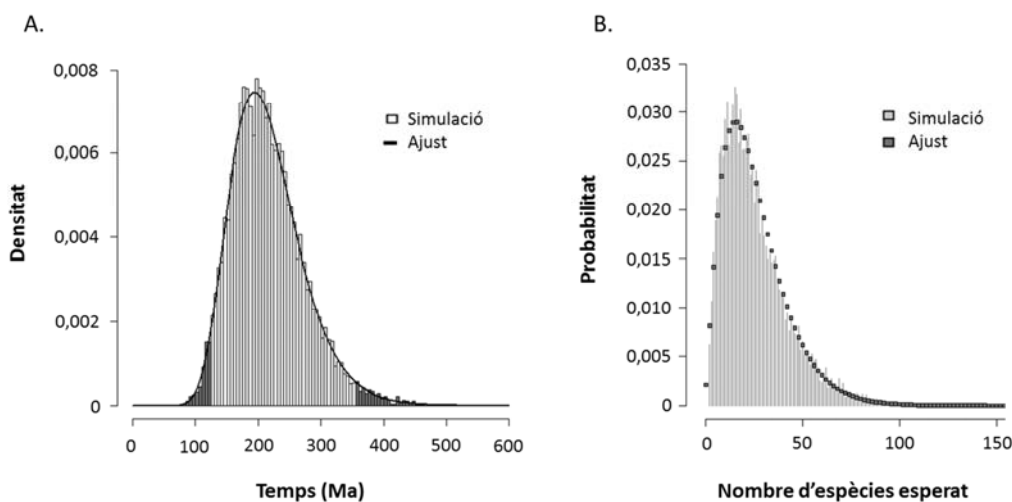


Figura 4.25 Resultats teòrics obtinguts a partir del model de Yule amb l’arbre d’espècies del gen *mdh*. A) Histograma de l’edat de l’ancestre comú més recent (tMRCA). B) Distribució del nombre d’espècies esperat.

5. RESULTATS:
Estudi del gen recombinasa A
d'Aeromonas

5. RESULTATS: ESTUDI DEL GEN RECOMBINASA A d'*Aeromonas*

5.1 OBTENCIÓ DE LES SEQÜÈNCIES

En aquest treball també s'han dissenyat els encebadors i s'han optimitzat les condicions d'amplificació per PCR i de seqüenciació del gen *recA* de totes les espècies d'*Aeromonas* descrites fins al moment d'escriure aquesta memòria, obtenint-se la seqüència nucleotídica completa d'aquest gen per les 150 soques de l'estudi. Els números d'accés del *GenBank* assignats pel conjunt de seqüències s'indiquen a la taula de l'Annex I.

5.1.1. Disseny d'encebadors

Inicialment es van dissenyar encebadors específics a partir de les seqüències conegudes del gen *recA* dels dos genomes d'*Aeromonas* disponibles en aquell moment a la base de dades del *GenBank*: el genoma d'*A.hydrophila* ATCC7966^T (*GenBank accession number*: CP000462, (Seshadri *et al.* 2006)) i el d'*A.salmonicida* subsp. *salmonicida* A449 (*GenBank accession number*: CP000644, (Reith *et al.* 2008)). Pel disseny es va tenir en compte que la seqüència completa del gen era de longitud variable segons l'espècie (1065 pb per *A.hydrophila* i 1062 pb per *A.salmonicida*).

A la Figura 5.1 es mostra de forma esquemàtica l'organització genètica del gen *recA* i els seus gens adjacents, *cinA* (extrem 5') i *recX* (extrem 3'), basant-se en les dades obtingudes d'aquests genomes.

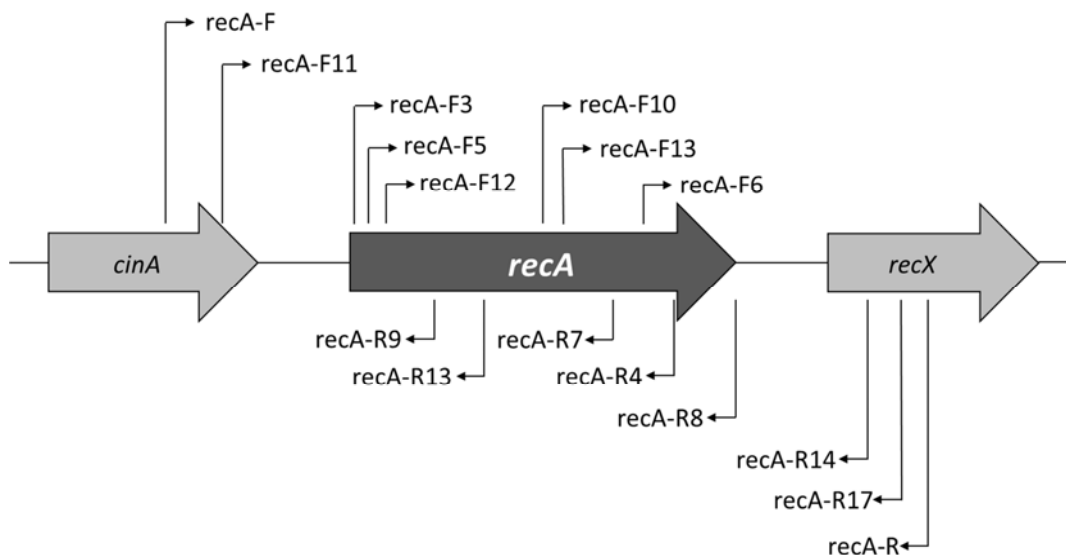


Figura 5.1 Posició dels encebadors utilitzats per a l'amplificació per PCR i la seqüenciació del gen *recA*.

L'amplificació per PCR del gen *recA* de les primeres soques de l'estudi (*A.salmonicida*, *A.hydrophila* i altres espècies properes taxonòmicament) s'ha realitzat amb la parella d'encebadors externs *recA-F* i *recA-R*. A partir dels amplificats obtinguts s'han dissenyat un conjunt d'encebadors interns (*recA-F5*, *recA-F6*, *recA-R7*, *recA-R8*, *recA-R9*) per facilitar la seqüenciació de les dues cadenes de DNA. Posteriorment, degut a la variabilitat que presenta aquest gen sobretot a l'extrem 3', tant a nivell intra- com inter-específic, ha estat necessari dissenyar nous encebadors externs (*recA-F2*, *recA-F3*, *recA-F11*, *recA-R14*, *recA-R15*, *recA-R16*, *recA-R17*) i interns (*recA-F10*, *recA-F12*, *recA-F13*, *recA-2F*, *recA-R4* i *recA-R13*) per a l'amplificació per PCR i la seqüenciació (Figura 5.1, Taula 5.1) alguns dels quals a la pràctica no han funcionat (*recA-F2*, *recA-F15*, *recA-F16*, *recA-2F*). A continuació, a la Taula 5.1 es detallen les principals característiques (seqüència, longitud, sentit i posició) de tots els encebadors dissenyats per a l'amplificació i la seqüenciació del gen *recA*. A la Taula 5.2 s'indiquen les temperatures de fusió (T_m) calculades teòricament (fórmula de Wallace) i les donades pel fabricant, i també les temperatures d'hibridació (T_a) per les reaccions de PCR i de seqüenciació.

Taula 5.1 Característiques dels encebadors per a l'amplificació i la seqüenciació del gen *recA*.

Encebador	Longitud (bp)	Seqüència 5' → 3'	Sentit ^a	Posició ^b
<i>recA-F</i>	18	GTGAGCGGCGAGAGNCTT	F	-370, -353
<i>recA-F3</i>	19	CGGAGATNGGCATGGATC	F	-11, +8
<i>recA-F5</i>	18	GTCTGGGCGACAGCAAGA	F	+80, +97
<i>recA-F6</i>	19	CAAGCACAAGCTGATCGAC	F	+834, +852
<i>recA-F10</i>	20	CTGGAGATCTGCGACATGCT	F	+373, +392
<i>recA-F11</i>	19	GTCGGNACCGTCTGGTTCG	F	-220, -202
<i>recA-F12</i>	18	AGCCATCTCCACCGGTC	F	+111, +128
<i>recA-F13</i>	20	CCTGTGCATCTTCATCAACC	F	+558, +577
<i>recA-F2*</i>	18	AGCAGGCGGTGACAGAGG	F	-129, -112
<i>recA-2F*</i>	19	CCAAGCTGGGNGTCAATGT	F	+311, +329
<i>recA-R</i>	19	GCGTCCCAGCAGATANTTG	R	+1547, +1529
<i>recA-R4</i>	18	CTCACGCAGTCTGGCTTC	R	+972, +955
<i>recA-R7</i>	19	GTCGATCAGCTTGTGCTTG	R	+852, +834
<i>recA-R8</i>	18	CTCTTGTTGCTTTGCGGC	R	+1056, +1039
<i>recA-R9</i>	19	CATGTGCGCAGATTTCCAGC	R	+390, +372
<i>recA-R13</i>	20	GGTTGATGAAGATGCACAGG	R	+577, +558
<i>recA-R14</i>	20	CAGNTCGAANCGGATCTTGA	R	+1385, +1366
<i>recA-R17</i>	17	GCCAGCTCGAACCAGTC	R	+1459, +1443
<i>recA-R15*</i>	20	GCTCTTCAAAGGTNAGTTCG	R	+1155, +1136
<i>recA-R16*</i>	18	CCAGNTGCTCTTCAAAGG	R	+1161, +1144

^a F, forward; R, reverse

^b Posició nucleotídica de cada encebador en relació al genoma d'*A.hydrophila* ATCC7966^T

* Encebadors que no han funcionat

Taula 5.2 Temperatures de fusió (T_m) i d'hibridació (T_a) dels encebadors del gen *recA*.

Encebador	T fusió (°C)		T hibridació (°C)	
	Tèorica	Fabricant	PCR	Seqüenciació
recA-F	59	49	54	53
recA-F3	59	47	54	53
recA-F5	58	47	53	52
recA-F6	58	46	53	52
recA-F10	62	49	57	56
recA-F11	63	51	58	57
recA-F12	58	47	53	52
recA-F13	60	47	55	54
recA-F2*	60	50	55	54
recA-2F*	57	47	52	51
recA-R	59	47	54	53
recA-R4	58	47	53	52
recA-R7	58	46	53	52
recA-R8	56	45	51	50
recA-R9	58	46	53	52
recA-R13	60	47	55	54
recA-R14	56	47	51	50
recA-R17	56	47	51	50
recA-R15*	57	46	52	51
recA-R16*	53	44	48	49

* Encebadors que no han funcionat

5.1.2 Optimització del procés d'amplificació per PCR

Per la majoria de soques de l'estudi (86%) s'ha aconseguit amplificar el gen complet *recA* amb una sola parella d'encebadors externs. No tots els encebadors dissenyats han funcionat en les reaccions de PCR (Taula 5.1). A la Taula 5.3 s'indica cadascuna de les 6 PCRs optimitzades, i per cadascuna la combinació d'encebadors emprada, la temperatura d'hibridació òptima dels encebadors i la longitud de l'amplicó. La PCR1, amb la parella d'encebadors externs recA-F11-recA-R, és la que ha donat millors resultats (72,1%), obtenint-se un amplificat d'aproximadament 1800 pb (Figura 5.2). La PCR2 (9,30%) ha permès amplificar majoritàriament soques de l'espècie *A.caviae*, la PCR3 (6,20%) soques de l'espècie *A.encheleia*, la PCR4 (5,43%) soques de les espècies *A.diversa* i *A.schubertii*, la PCR5 (3,89%), que va ser la primera en estandarditzar-se, algunes soques del complex *A.hydrophila*, i finalment la PCR6 (2,33%) utilitzada per la soca tipus de l'espècie *A.taiwanensis*, entre altres. Per un grup reduït de soques, en el que destaquen *A.simiae*, *A.media*, *A.cavernicola* i *A.dhakensis*, ha estat necessari realitzar dues PCRs amb diferents combinacions d'encebadors externs i interns per obtenir la seqüència completa del gen.

Taula 5.3 Parelles d'encebadors utilitzades per a l'amplificació per PCR del gen *recA*.

PCR	Parelles d'encebadors	Temperatura d'hibridació (°C)	Longitud de l'amplicó (pb) ^a
PCR1	F11-R	56	1767
PCR2	F3-R	54	1558
PCR3	F11-R14	54	1585
PCR4	F3-R17	53 / 58	1470
PCR5	F-R	54	1917
PCR6	F3-R14	54	1376

^a Basat en el genoma d'*A. hydrophila* ATCC7966^T

Per l'estandardització de la barreja de reacció de PCR, les modificacions que s'han portat a terme han afectat a les quantitats de DNA genòmic extret i d'encebadors (Taula 5.4). De la mateixa manera que en el gen *mdh*, també s'ha comprovat que la barreja requereix la presència addicional de MgCl₂ per obtenir millors resultats, que s'aporta amb la solució tamponada de la Taq DNA Polimerasa (Buffer I, 15 mM de MgCl₂) i directament d'una solució de MgCl₂ 25 mM. A més, en quasi tots els casos ha estat necessari addicionar DMSO, que actua com adjuvant per obtenir un bon amplificat, i la quantitat requerida ha variat depenent de la soca. La concentració emprada de la resta de components (dNTPs, solució tamponada Buffer I 10x i Taq DNA polimerasa) no s'ha variat, s'ha mantingut constant en totes les reaccions.

Taula 5.4 Composició de la barreja de reacció de PCR optimitzada per l'estudi del gen *recA*.

Components de la reacció de PCR	Volum (µL)
DNA genòmic	0,5 - 10
Encebador <i>Forward</i> 10 µM	1 - 10
Encebador <i>Reverse</i> 10 µM	1 - 10
Buffer I 10X	5
dNTPs Mix 10 mM	4
MgCl ₂ 25 mM	1
DMSO	0 - 5
Ampli Taq Gold® 5U/µL	0,25
Aigua Milli-Q estèril	qsp. 50

Les condicions optimitzades de les diferents etapes de la reacció de PCR del gen *recA* s'indiquen a la Taula 5.5. De forma general, depenent de la combinació d'encebadors utilitzats, s'ha estandarditzat la seva temperatura d'hibridació. A la Taula 5.3 es mostren les temperatures òptimes de cada PCR. Per obtenir una concentració elevada d'amplificat s'ha augmentat el nombre de cicles requerits per repetir el procés de PCR, variant entre 35 i 38.

Taula 5.5 Condicions optimitzades per l'amplificació per PCR del gen *recA*.

Etapes	Temperatura (°C)	Temps	Cicles
Desnaturalització inicial	95	5 min	1
Desnaturalització	94	45 seg	
Hibridació	53 - 58	1 min	35 - 38
Extensió	72	2 min	
Extensió final	72	10 min	1
Final	4	∞	

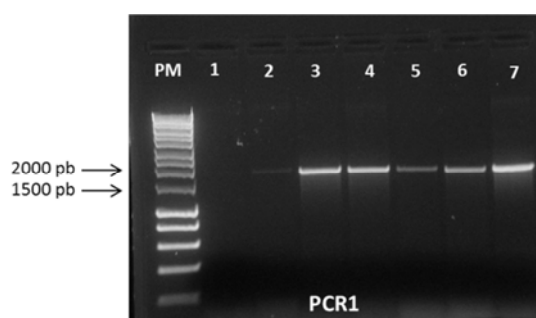


Figura 5.2 Comprovació dels amplificats del gen *recA* obtinguts per PCR (5µL) mitjançant l'electroforesi en gel d'agarosa a l'1%. PM: marcador de pesos moleculars. 7: control positiu.

5.1.3 Optimització del procés de seqüenciació

Per realitzar la reacció de seqüenciació a partir dels amplificats purificats, s'han utilitzat els mateixos encebadors emprats a les PCRs (Taula 5.3), i també altres encebadors interns dissenyats per la seqüenciació (*recA*-F5, *recA*-F6, *recA*-F10, *recA*-F12, *recA*-F13, *recA*-R4, *recA*-R7, *recA*-R8, *recA*-R9, *recA*,R13) degut a la longitud dels amplificats obtinguts (Figura 5.1 i Taula 5.1). En aquest cas, també s'han seqüenciat ambdues cadenes de DNA, per tal de comprovar la fiabilitat de la lectura de les bases. Per obtenir la seqüència completa del gen *recA* ha estat necessari realitzar entre 5 i 7 seqüenciacions, com a mínim dos encebadors diferents per cada cadena. A la Taula 5.2 s'indiquen les temperatures d'hibridació òptimes de tots els encebadors externs i interns utilitzats en les reaccions de seqüenciació.

La composició estandarditzada de la barreja de reacció de seqüenciació s'indica a la Taula 5.6. De la mateixa manera que en el gen *mdh*, prèviament s'ha estimat la concentració de l'amplificat obtingut per PCR al visualitzar la banda d'electroforesi sota llum UV i comparar-la amb un marcador de pesos moleculars (Figura 5.2). De forma general, a cada reacció s'ha afegit 1 µL de l'amplificat purificat i 1,5 µL d'encebador, però en alguns casos que s'han detectat bandes molt intenses o molt febles d'amplificat (concentracions elevades o molt

baixes, respectivament) ha estat necessari variar la proporció DNA i encebador de la barreja de reacció per ajustar les quantitats a una relació 1:2 (bandes intenses) i 1:1 ó 2:1 (bandes molt febles).

Per últim, a la Taula 5.7 s'indiquen les millors condicions de la reacció de seqüenciació del gen *recA*, detallant les temperatures i els temps de cada etapa, així com els cicles realitzats. Les temperatures d'hibridació per la seqüenciació s'han optimitzat restant un grau a les temperatures d'hibridació de PCR de cada encebador, tal com s'indica a la Taula 5.2.

Taula 5.6 Barreja de reacció de seqüenciació optimitzada del gen *recA*.

Components de la reacció de seqüenciació	Volum (µL)
DNA purificat	1 - 2
Encebador 5µM	1 - 2
Buffer Big Dye®	3
Big Dye® Terminator v3.1	1,5
Aigua Milli-Q estèril	qsp. 20

Taula 5.7 Condicions optimitzades per la reacció de seqüenciació del gen *recA*.

Etapes	Temperatura (°C)	Temps	Cicles
Desnaturalització inicial	96	1 min	1
Desnaturalització	96	10 seg	
Hibridació	50 - 57	5 seg	20
Extensió	60	4 min	
Final	4	∞	

5.2 ANÀLISIS DE LES SEQÜÈNCIES

5.2.1 Caracterització del gen *recA*

Després de l'obtenció de la seqüència nucleotídica del gen complet *recA* de totes les soques de l'estudi, s'ha realitzat una anàlisi de les característiques principals del conjunt de seqüències.

La longitud d'aquest gen ha variat segons l'espècie estudiada, oscil·lant entre 1059 i 1068 pb. Cal destacar el cas d'*A. popoffii*, on la majoria de soques (60%) tenen una longitud de 1062 pb, excepte 4 d'elles, 3 de les quals (CECT5250, LMG15742 i LMG17544) han presentat un codó menys (posicions 1015-1017 de l'alineament, 1059 pb), i l'altre aïllat (CECT5210) un de més (posicions 1000-1002 de l'alineament, 1065 pb). En relació a l'espècie *A. tecta*, totes les soques analitzades ha presentat seqüències d'una longitud total de 1062 pb, amb excepció de la soca tipus (MDC91^T) en la que s'ha detectat una mutació al final del gen (posició 1057 de

l'alineament, canvi d'una timina per una guanina) i que ha produït l'aparició d'un codó de terminació (TAG) en el penúltim codó (1059 pb).

A l'estudiar la variabilitat nucleotídica del conjunt de seqüències s'han identificat un total de 651 posicions invariants i 396 posicions polimòrfiques, de les quals 26 han estat llocs variables en una posició (*singleton variable sites*) i 370 llocs parsimòniament informatius (*parsimony informative sites*). Els resultats obtinguts per cada espècie es mostren a la Taula 5.8, amb excepció d'aquelles que només es disposa d'una única soca que no s'ha pogut fer l'anàlisi de variabilitat nucleotídica però sí la determinació del seu contingut en G+C. A l'observar amb detall l'alineament múltiple de seqüències, de forma majoritària els canvis detectats han estat substitucions nucleotídiques, però també sembla que hi hagi algunes petites insercions i/o deletions en la regió final del gen (posicions 1000-1068 de l'alineament). En relació a la proporció de transicions (T_i) i transversions (T_v) en aquest gen, s'ha calculat la taxa de T_i/T_v que ha estat de 3,04, essent més freqüents les transicions que les transversions.

En relació al contingut en G+C, s'ha obtingut un valor mig del 59% pel total de seqüències de l'estudi. A l'analitzar els valors obtinguts de cada espècie (Taula 5.8), *A.sanarellii* ha presentat el valor més alt (61,7%) i *A.cavernicola* el més baix (56,8%).

Per analitzar el grau de saturació del gen *recA*, s'ha estimat la proporció de canvis sinònims per lloc sinònim (d_S) i canvis no sinònims per lloc no sinònim (d_N) respecte la distància genètica (p -distance) per comparació de parelles de tàxons. Els resultats obtinguts s'han representat en forma de gràfic (Figura 5.3), en el qual s'observa que en cap dels dos casos s'ha arribat a la saturació, i per tant, les substitucions que s'estan acumulant encara poden influir en la seva evolució. A l'analitzar amb més detall la figura, s'ha observat tant en les substitucions sinònimes com no sinònimes dues agrupacions, una majoritària que correspon als valors obtinguts a partir de la majoria d'espècies d'*Aeromonas*, i una més petita, dels resultats provinents de la comparació de les espècies *A.simiae*, *A.schubertii* i *A.diversa* amb la resta.

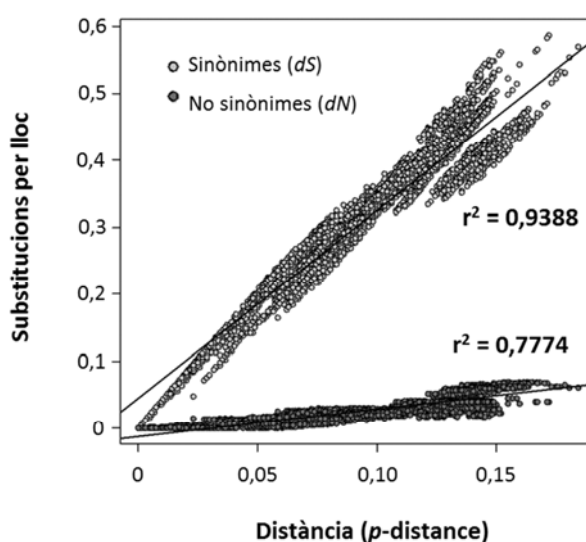


Figura 5.3 Representació gràfica del grau de saturació del gen *recA*.

Taula 5.8 Anàlisi de la variabilitat i contingut en G+C del gen *recA*.

Espècie	Longitud (pb)	Llocs invariants	Llocs variables	Llocs variables en una posició	Llocs parsimòniament informatius	G+C (%mol)
<i>A.allosaccharophila</i>	1062	1021	41	23	18	58,6
<i>A.australiensis</i>	1065	1065	-	-	-	59,0
<i>A.bestiarum</i>	1062	1003	56	10	46	59,1
<i>A.bivalvium</i>	1065	1047	18	18	0	60,8
<i>A.cavernicola</i>	1068	1068	-	-	-	56,8
<i>A.caviae</i>	1065	1005	60	23	37	59,2
<i>A.dhakensis</i>	1065	1007	58	34	24	59,3
<i>A.diversa</i>	1059	1046	13	13	0	60,0
<i>A.encheleia</i>	1062	1020	42	26	16	59,7
<i>A.enteropelogenes</i>	1065	1007	58	42	16	59,2
<i>A.eucrenophila</i>	1062	1037	25	17	8	60,2
<i>A.fluvialis</i>	1065	1065	-	-	-	58,7
<i>A.hydrophila</i>	1065	987	78	48	30	59,4
<i>A.jandaei</i>	1065	1019	46	28	18	58,9
<i>A.media</i>	1062	1012	50	15	35	60,3
<i>A.molluscorum</i>	1065	1053	12	4	8	57,7
<i>A.piscicola</i>	1062	1033	29	18	11	59,0
<i>A.popoffii</i>	1059 1062 1065	1036	23	3	20	57,7
<i>A.rivuli</i>	1065	1053	12	12	0	58,7
<i>A.salmonicida</i>	1062	998	64	36	28	57,7
<i>A.sanarellii</i>	1065	1065	-	-	-	61,7
<i>A.schubertii</i>	1059	1048	11	6	5	60,3
<i>A.simiae</i>	1059	1059	0	0	0	58,7
<i>A.sobria</i>	1059	1002	57	37	20	57,4
<i>A.taiwanensis</i>	1065	1065	-	-	-	61,3
<i>A.tecta</i>	1059 1062	1006	53	43	10	59,9
<i>A.veronii</i>	1065	935	130	53	77	58,5

5.2.2 Tipus de selecció

Per tal de conèixer l'evolució del gen *recA* en el gènere *Aeromonas*, s'ha analitzat el tipus de selecció que opera en aquest gen utilitzant la prova Z (*Z-Test*). Els resultats obtinguts han permès rebutjar la hipòtesi de selecció positiva ($P > 0,05$), però no s'ha pogut discriminar entre selecció neutra o purificadora. Per això, s'ha realitzat el Test de Neutralitat de Tajima, el qual ha permès rebutjar la hipòtesi de neutralitat ($D = 1,129552$, $P > 0,1$), indicant per tant, que segueix un model de selecció negativa o purificadora.

A partir dels canvis detectats en les seqüències s'han estimat les taxes de substitucions sinònimes (dS) i les no sinònimes (dN) de forma global, que han estat de 0,3998 i 0,0247,

respectivament. El valor de dS obtinguts han estat d'un ordre superior als de dN , resultat comparable a l'obtingut pel gen *mdh* (Apartat 4.2.2). També s'ha obtingut un valor d' ω (dN/dS) de 0,0618 ($\omega < 1$), que indicaria que les substitucions que es produeixen en el gen *recA* es troben sota un fort efecte de selecció purificadora. A la Figura 5.4 s'observen gràficament els resultats obtinguts de l'anàlisi de les taxes de substitucions sinònimes (dS) i no sinònimes (dN) codó per codó al llarg del gen. En aquest gràfic es pot veure com el nombre de substitucions sinònimes (dS) augmenta de forma constant al llarg de tota la seqüència, i en canvi el de no sinònimes (dN) l'increment és molt menor, amb regions sense canvis excepte a la regió final del gen (a partir del codó 300) on s'inicia un augment important de canvis no sinònims.

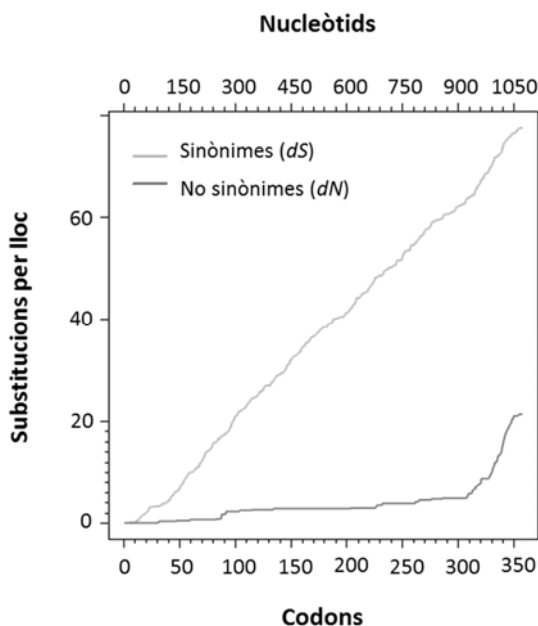


Figura 5.4 Distribució de les substitucions sinònimes i no sinònimes per codons al llarg de les seqüències del gen *recA*.

També s'ha portat a terme una anàlisi de les taxes dS i dN , i el valor ω de les seqüències del gen *recA* a nivell intra- i inter-específic. Els resultats obtinguts es mostren a la Taula 5.9 i la Figura 5.5, respectivament. Per aquelles espècies que només es disposa d'una soca no s'ha pogut realitzar el càlcul a nivell intra-específic. Al considerar els resultats de les seqüències de soques d'una mateixa espècie (Taula 5.9), en tots els casos el nombre de canvis sinònims fixats ha estat més elevat que el de no sinònims. En determinades espècies (*A.allosaccharophila*, *A.bivalvium*, *A.diversa*, *A.eucrenophila*, *A.molluscorum*, *A.piscicola*, *A.rivuli* i *A.schubertii*) només s'han detectat canvis sinònims. Cal destacar el cas d'*A.simiae*, que no s'ha detectat cap tipus de canvi en les seqüències obtingudes de la soca tipus CIP107798^T i de l'aïllat CIP107797, ja que ambdós presenten seqüències idèntiques del gen *recA*. Els resultats obtinguts a nivell inter-específic es mostren a la Figura 5.5, on es representa la correlació entre els canvis no sinònims (dN) i sinònims (dS) i la recta de regressió corresponent. En el cas del gen *recA* s'ha detectat una baixa correlació d'aquestes dues variables, en comparació amb els resultats obtinguts pel gen *mdh* (Apartat 4.2.2).

Taula 5.9 Anàlisi intra-específic dels canvis sinònims (dS) i no sinònims (dN) i el valor ω (dN/dS) del gen *recA*.

Espècie	dS	dN	ω
<i>A.allosaccharophila</i>	0,0866	0	0
<i>A.bestiarum</i>	0,0755	0,0074	0,0977
<i>A.bivalvium</i>	0,0486	0	0
<i>A.caviae</i>	0,0872	0,0002	0,0019
<i>A.dhakensis</i>	0,0831	0,0014	0,0163
<i>A.diversa</i>	0,0529	0	0
<i>A.encheleia</i>	0,0580	0,0003	0,0054
<i>A.enteropelogenes</i>	0,1093	0,0023	0,0208
<i>A.eucrenophila</i>	0,0479	0	0
<i>A.hydrophila</i>	0,0972	0,0003	0,0029
<i>A.jandaei</i>	0,0762	0,0011	0,0139
<i>A.media</i>	0,1139	0,0013	0,0111
<i>A.molluscorum</i>	0,0259	0	0
<i>A.piscicola</i>	0,0399	0	0
<i>A.popoffii</i>	0,0358	0,0007	0,0196
<i>A.rivuli</i>	0,0491	0	0
<i>A.salmonicida</i>	0,0660	0,0016	0,0246
<i>A.schubertii</i>	0,0216	0	0
<i>A.simiae</i>	0	0	-
<i>A.sobria</i>	0,1060	0,0005	0,0048
<i>A.tecta</i>	0,1107	0,0013	0,0115
<i>A.veronii</i>	0,1338	0,0020	0,0151

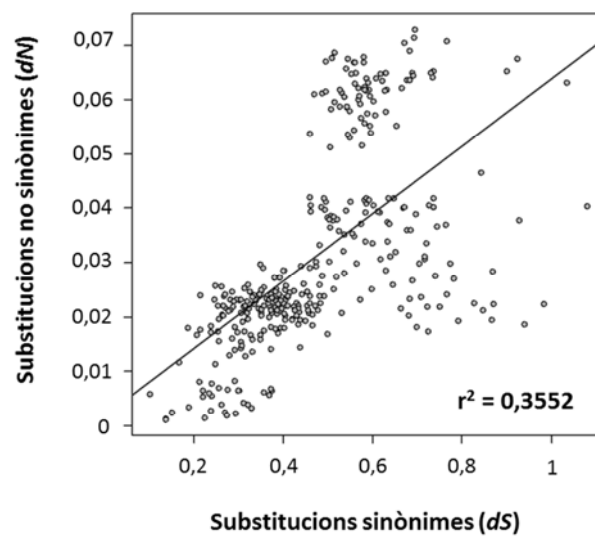


Figura 5.5 Relació entre les substitucions no sinònimes (dN) i sinònimes (dS) a nivell inter-específic del gen *recA*.

5.2.3 Distàncies genètiques intra- i inter-específiques

Per estudiar la variabilitat genètica del gen *recA*, s'han determinat les distàncies genètiques (*p*-distance) a nivell intra- i inter-específic a partir de l'alineament múltiple de seqüències. A la Taula 5.10 es mostren els resultats obtinguts i el nombre de soques de cada espècie. Per aquelles espècies que només es disposa d'una única soca no s'ha pogut estimar la distància intra-específica. *A.veronii* que agrupa un total de 17 soques, és l'espècie que ha presentat el valor de distància intra-específica més elevat (0,0312), en canvi, *A.simiae*, amb dos aïllats, ha donat el valor més baix (0) perquè tal com s'ha comentat anteriorment les dues seqüències obtingudes són idèntiques. A l'analitzar les distàncies inter-específiques comparant per parelles amb la resta d'espècies, *A.simiae* ha obtingut el valor més alt (0,1466) i *A.dhakensis* el més baix (0,0883).

Taula 5.10 Distàncies genètiques (*p*-distance) a nivell intra- i inter-específic del gen *recA*.

Espècie	Distància Intra-específica	Distància Inter-específica	Nombre de soques
<i>A.allosaccharophila</i>	0,0196	0,0954	5
<i>A.australiensis</i>	-	0,0896	1
<i>A.bestiarum</i>	0,0233	0,0926	9
<i>A.bivalvium</i>	0,0115	0,1206	3
<i>A.cavernicola</i>	-	0,1405	1
<i>A.caviae</i>	0,0199	0,1044	10
<i>A.dhakensis</i>	0,0201	0,0883	8
<i>A.diversa</i>	0,0124	0,1411	2
<i>A.encheleia</i>	0,0135	0,0982	8
<i>A.enteropelogenes</i>	0,0261	0,0995	5
<i>A.eucrenophila</i>	0,0111	0,1034	5
<i>A.fluvialis</i>	-	0,0888	1
<i>A.hydrophila</i>	0,0223	0,0885	9
<i>A.jandaei</i>	0,0183	0,0895	6
<i>A.media</i>	0,0260	0,0931	5
<i>A.molluscorum</i>	0,0061	0,1367	5
<i>A.piscicola</i>	0,0093	0,0901	8
<i>A.popoffii</i>	0,0089	0,1004	3
<i>A.rivuli</i>	0,0115	0,1282	2
<i>A.salmonicida</i>	0,0163	0,1014	13
<i>A.sanarellii</i>	-	0,1034	1
<i>A.schubertii</i>	0,0052	0,1374	5
<i>A.simiae</i>	0,0000	0,1466	2
<i>A.sobria</i>	0,0246	0,1020	5
<i>A.taiwanensis</i>	-	0,0968	1
<i>A.tecta</i>	0,0264	0,1089	4
<i>A.veronii</i>	0,0312	0,0910	17

A la Figura 5.6A es representa el gràfic de densitats de la distribució dels valors de distàncies intra- i inter-específics obtinguts a partir de les seqüències del gen *recA*, en el que s'observa una bona separació perquè no hi ha solapament entre les dues distribucions. A més, a la Figura 5.6B es mostra la correlació de distàncies intra- i inter-específiques de cada espècie i la mitjana d'aquestes dues distàncies s'indica amb línies discontinues. En el quadrant superior esquerra apareixen aquelles espècies que presenten una millor separació inter-específica (*A.simiae*, *A.schubertii*, *A.diversa*, *A.molluscorum*, *A.rivuli* i *A.bivalvium*), resultats molt coincidents amb els obtinguts pel gen *mdh*.

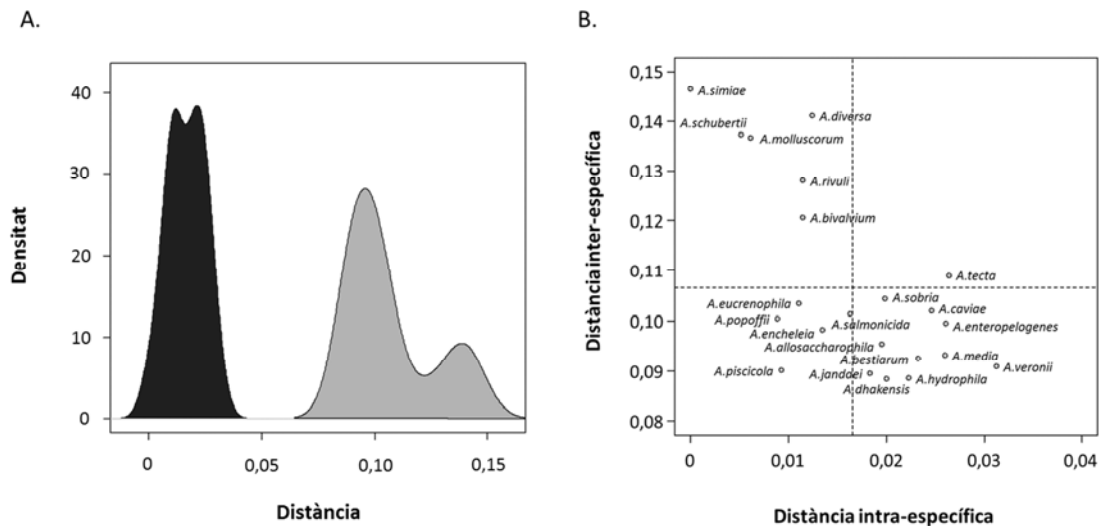


Figura 5.6 Representacions gràfiques dels valors de distància genètica de les seqüències del gen *recA*. A) Gràfic de densitats de les distàncies intra- (negre) i inter-específiques (gris). B) Correlació de les distàncies inter- i intra-específiques de cada espècie, les línies discontinues indiquen els valors mitjans de distàncies.

Finalment, s'ha analitzat si el nombre desigual de soques d'algunes espècies de l'estudi ha afectat a la determinació dels valors de distància. A la Figura 5.7 es mostra la correlació entre les distàncies intra- i inter- específiques i el nombre de soques de cada espècie. Els baixos valors r^2 obtinguts de les rectes de regressió indiquen que no existeix correlació, i per tant, que els resultats no estan influenciats pel nombre de soques.

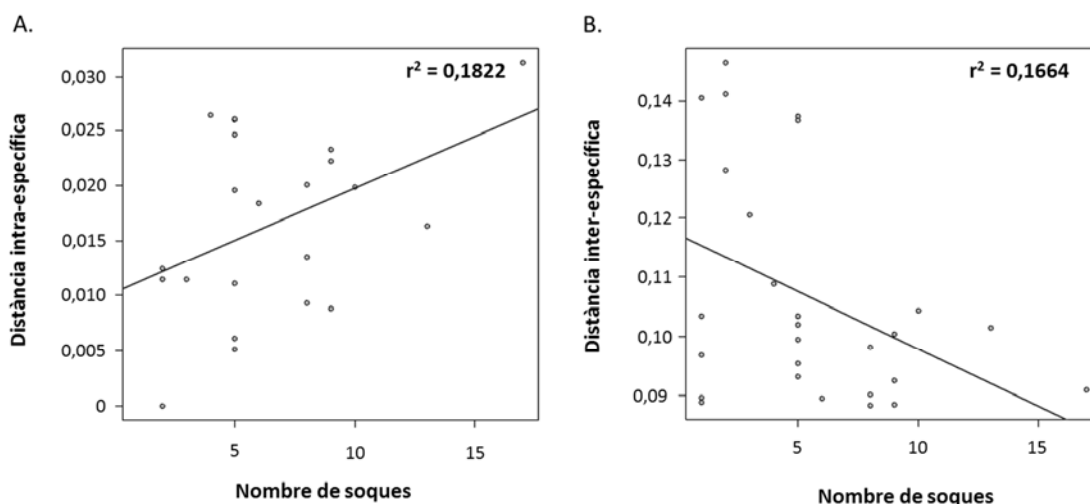


Figura 5.7 Influència del nombre de soques respecte les distàncies genètiques intra- (A) i inter-específiques (B) del gen *recA*.

5.3 FILOGÈNIES

5.3.1 Anàlisi filogenètica del gènere *Aeromonas*

Per establir les relacions evolutives del gènere *Aeromonas* basades en el gen complet *recA* s'han construït dues filogènies, una utilitzant el mètode de màxima versemblança (ML) i l'altre amb el mètode bayesià (IB), pel conjunt de 150 soques de l'estudi (Annex I).

En primer lloc, s'ha determinat si el conjunt de seqüències del gen *recA* evolucionen sota la hipòtesi d'un rellotge molecular. A partir d'una filogènia s'ha realitzat l'anàlisi del test del rellotge molecular amb el programa MEGA6, seguint dues aproximacions, forçant a complir rellotge (Rellotge) i sense forçar (No rellotge). Els resultats obtinguts de l'anàlisi es mostren a la Taula 5.11. El valor de P obtingut ($P \sim 0$) ha permès rebutjar la hipòtesi nul·la de rellotge molecular estricte. Aquest resultat s'ha tingut en compte per a la reconstrucció de les filogènies, ja que quan ha estat possible (IB), s'ha utilitzat un mètode de rellotge molecular relaxat.

Taula 5.11 Anàlisi del test del rellotge molecular pel gen *recA*.

	lnL ^a	Model	+G ^b	+I ^c	Likelihood Ratio Test (LRT)		
					2 (L _{no rellotge} - L _{rellotge})	n-2 ^d	Valor de P
Rellotge	-1194403,28	GTR	0,58	0,62	1181308,68	148	$P \sim 0$
No rellotge	-13094,60		0,98	0,56			

^a lnL, logaritme de la versemblança

^b G, valor estimat del paràmetre α de la distribució gamma

^c I, proporció de posicions invariants

^d n-2, graus de llibertat

En segon lloc, a partir de les seqüències alineades s'ha determinat el model de substitució nucleotídica que millor s'ajusta a les dades, i que ha estat el model més complex, el GTR+G+I. Posteriorment, s'han construït les filogènies de màxima versemblança (ML, Figura 5.8) i d'inferència bayesiana (IB, Figura 5.9). En els nodes dels grups principals es mostren els valors de *bootstrap* (ML) i de probabilitat posterior (IB), que han estat molt elevats (> 70%), indicant la robustesa de les filogènies. De forma general, al comparar les topologies dels arbres obtinguts s'ha vist que no són totalment idèntiques, sinó que presenten diferències en algunes agrupacions d'espècies. Per exemple, el grup d'espècies del complex *A.veronii* forma un clúster independent en l'arbre bayesià però es troba barrejat amb altres espècies no tan relacionades filogenèticament en l'arbre de màxima versemblança.

A l'analitzar amb detall les agrupacions de les dues filogènies, en l'arbre ML s'observa un grup principal format per la majoria d'espècies d'*Aeromonas* i un grup minoritari, que inclou dos clústers diferents, un que agrupa les espècies *A.bivalvium*, *A. molluscorum* i *A.rivuli*, i un segon clúster amb les espècies *A.simiae*, *A.schubertii* i *A.diversa*. En canvi, l'arbre IB també mostra clarament aquests dos grups, un de principal i un de minoritari, però aquest segon grup només conté les 3 espècies *A.simiae*, *A.schubertii* i *A.diversa*, com en el cas del gen *mdh*.

Al comprovar la posició taxonòmica de les soques en les filogènies, totes elles s'han agrupat en el seu clúster corresponent segons l'espècie assignada, amb excepció de les mateixes 6 soques (*A.caviae* CECT5207, *A.hydrophila* CECT398, *A.hydrophila* LMG3767, *A.popoffii* CECT5210, *A.sobria* CECT4250 i *A.veronii* CECT4816) amb incongruències detectades en les filogènies del gen *mdh* (Apartat 4.3.1). Al comparar les posicions taxonòmiques d'aquestes 6 soques en les filogènies dels dos gens, en tots els casos han coincidit en les mateixes agrupacions en els arbres, motiu pel qual s'ha realitzat un estudi fenotípic per confirmar la seva identificació (Apartat 4.3.2).

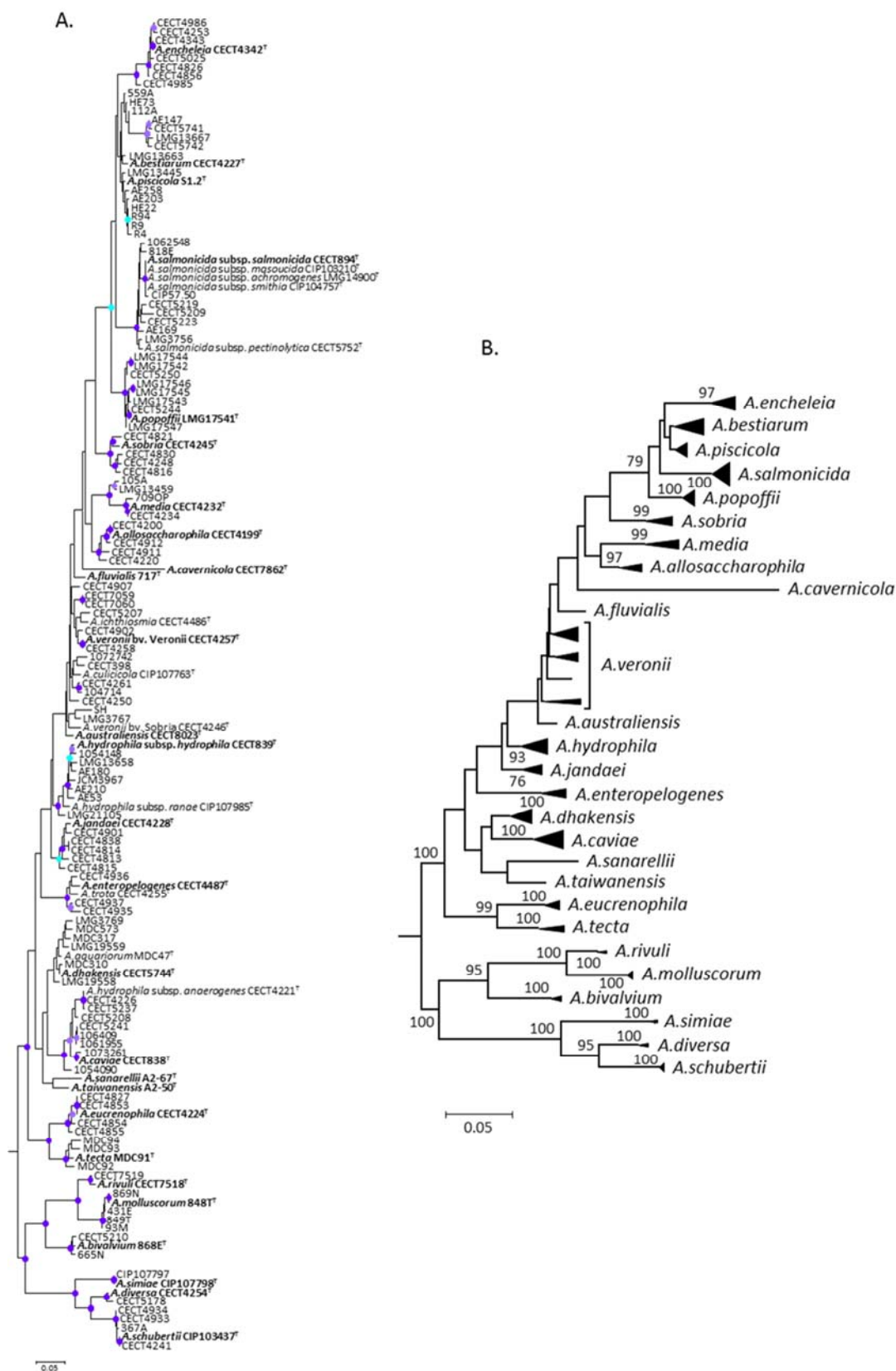


Figura 5.8 Arbre de màxima versemblança (ML) del gen *recA*. A) Arbre amb les 150 soques. B) Arbre col·lapsat per espècies o grans grups. L'escala indica el nombre de substitucions nucleotídiques per lloc. Els nodes mostren els valors de *bootstrap* iguals o superiors al 70% (cian, 70-80%; lila clar, 80-90% i lila fosc, 90-100%).

5.4 RECOMBINACIÓ

A l'analitzar les agrupacions obtingudes en els dos arbres filogenètics del gen *recA* (ML, Figura 5.8 i IB, Figura 5.9), en l'arbre de ML s'ha detectat un grup de 4 aïllats en una branca més profunda dins del clade corresponent a l'espècie *A.bestiarum* (Figura 5.10A). En el cas de l'arbre d'IB (Figura 5.10B), aquestes soques també formen part d'un grup independent de la resta. Aquestes 4 soques d'*A.bestiarum* (AE147, CECT5741, CECT5742 i LMG13667) han estat obtingudes de col·leccions, procedeixen d'ambients aquàtics, i han estat aïllades de zones geogràfiques diferents (Annex I).

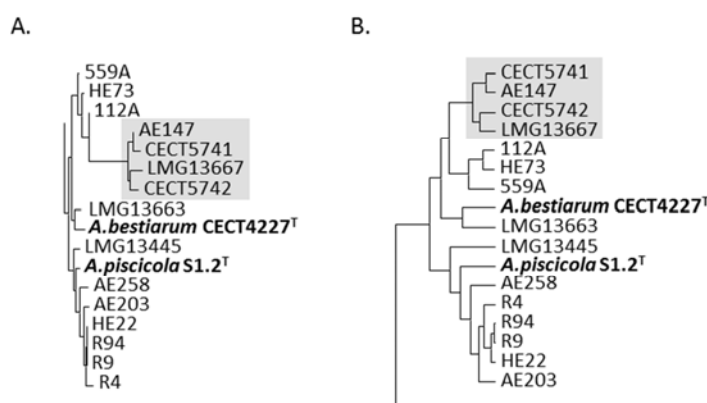


Figura 5.10 Ampliació del clade de l'espècie *A.bestiarum* de les filogenies obtingudes per màxima versemblança (A) i inferència bayesiana (B) del gen *recA* complet. En gris s'assenyalen les 4 soques problema.

5.4.1 Anàlisi de recombinació del gen *recA*

Per determinar si aquesta incongruència filogenètica detectada pot ser conseqüència d'un esdeveniment de recombinació, primer s'ha portat a terme una anàlisi exploratòria de les seqüències, i després, a partir dels resultats obtinguts, s'ha decidit realitzar un estudi més ampli utilitzant mètodes de detecció de recombinació amb programes bioinformàtics i altres metodologies clàssiques, com la determinació del contingut en G+C i de l'ús de codons, entre d'altres.

5.4.1.1 Anàlisi exploratòria de les seqüències

Visualitzant l'alineament múltiple de seqüències del gen *recA*, a la regió terminal d'aquest gen s'han detectat algunes diferències entre els 4 aïllats problema (verd) i la resta de soques d'*A.bestiarum* (Figura 5.11A) tot i que al comparar-les entre elles han estat idèntiques, excepte la soca LMG13667 que ha presentat un canvi en la posició 1014 (timina per citosina). Seguidament, s'ha generat un gràfic de diagrama de punts que mostra visualment només les posicions polimòrfiques (segregants) al llarg de la seqüència completa del gen pel conjunt de soques de l'estudi (Figura 5.11B). El perfil obtingut de les seqüències del clade *A.bestiarum*

també ha posat de manifest aquestes diferències nucleotídiques a la regió terminal del gen per les 4 soques a estudi.

A.

<i>A.bestiarum</i> CECT4227T	AGACTGCGTGAGCTGCTGCTCTCCGGTGATGCGCCGGC---CGAGAAGCCGGTGGCGGCT
<i>A.bestiarum</i> 112A	AGACTGCGTGAGCTGCTGCTCTCCGGTGATGCGCCGGC---CGAGAAGCCGGTGGCGGCT
<i>A.bestiarum</i> 559A	AGACTGCGTGAGCTGCTGCTCTCCGGTGATGCGCCAGC---CGAGAAGCCGGTGGCGGCT
<i>A.bestiarum</i> HE73	AGACTGCGTGAGCTGCTGCTCTCCGGTGATGCGCCGGC---CGAGAAGCCGGTGGCGGCT
<i>A.bestiarum</i> LMG13663	AGACTGCGTGAGCTGCTGCTCTCCGGTGATGCGCCGGC---CGAGAAGCCGGTGGCGGCT
<i>A.bestiarum</i> AE147	AGACTGCGTGAGCTGCTGCTCTCGGGGCGCCGCCCGGTGGACGACAAGCCTGCCGCTATC
<i>A.bestiarum</i> CECT5741	AGACTGCGTGAGCTGCTGCTCTCGGGGCGCCGCCCGGTGGACGACAAGCCTGCCGCTATC
<i>A.bestiarum</i> CECT5742	AGACTGCGTGAGCTGCTGCTCTCGGGGCGCCGCCCGGTGGACGACAAGCCTGCCGCTATC
<i>A.bestiarum</i> LMG13667	AGACTGCGTGAGCTGCTGCTCTCGGGGCGCCGCCCGGTGGACGACAAGCCTGCCGCTATC
	***** ** *
<i>A.bestiarum</i> CECT4227T	GACGCTGACGAGCTCGAAGCCGAAAGCGAACAAGAGTTCGAATAA
<i>A.bestiarum</i> 112A	GACGCTGACGAGCTCGAAGCCGAAAGCGAACAAGAGTTCGAATAA
<i>A.bestiarum</i> 559A	GACGCTGACGAGCTCGAAGCCGAAAGCGAACAAGAGTTCGAATAA
<i>A.bestiarum</i> HE73	GACGCTGACGAGCTCGAAGCCGAAAGCGAACAAGAGTTCGAATAA
<i>A.bestiarum</i> LMG13663	GACGCTGACGAGCTCGAAGCCGAAAGCGAACAAGAGTTCGAATAA
<i>A.bestiarum</i> AE147	GAGGCCGACGAGTTGAGCCTGAAGGCGA---AGAACTCGAGTAA
<i>A.bestiarum</i> CECT5741	GAGGCCGACGAGTTGAGCCTGAAGGCGA---AGAACTCGAGTAA
<i>A.bestiarum</i> CECT5742	GAGGCCGACGAGTTGAGCCTGAAGGCGA---AGAACTCGAGTAA
<i>A.bestiarum</i> LMG13667	GAGGCCGACGAGTTGAGCCTGAAGGCGA---AGAACTCGAGTAA
	** *

B.

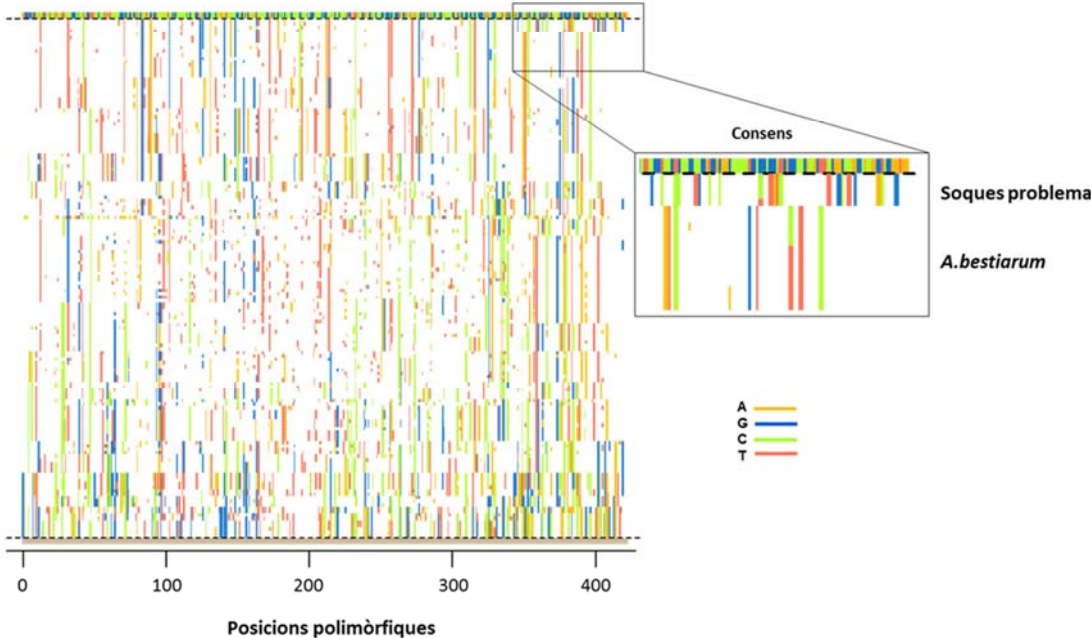


Figura 5.11 Visualització de les seqüències del gen *recA* de les soques d'*A.bestiarum*. A) Alineament múltiple de seqüències corresponent a la regió terminal del gen (posició 961-final). El primer aïllat correspon a la soca tipus i els 4 últims (verd) a les soques problema. B) Diagrama de punts de les posicions polimòrfiques al llarg de la seqüència completa del gen per a cadascuna de les soques de l'estudi. A la part superior es mostra la seqüència consens obtinguda a partir de totes les seqüències. Els llocs polimòrfics de cada soca es marquen amb un color diferent segons la base que varia (A, groc; G, blau; C, verd; T, vermell) respecte la seqüència consens. A la dreta de la imatge, es mostra una ampliació de la regió d'interès.

5.4.1.2 Mètodes de detecció amb programes bioinformàtics

5.4.1.2.1 Programa RAT

Per saber si aquesta regió terminal correspon a un fragment adquirit per recombinació s'ha realitzat una primera aproximació d'aquest procés en el clade *A.bestiarum* i l'abast de la regió afectada utilitzant el programa *Recombination Analysis Tool* (RAT). Els resultats proporcionats per aquesta anàlisi es mostren en forma de gràfic de distàncies genètiques (Figura 5.12), a partir de la comparació de la seqüència de la soca problema AE147 respecte la resta (5 pertanyents al clade *A.bestiarum* i 3 als aïllats problema). En aquest gràfic s'aprecia clarament l'existència de dos grups diferents de soques en relació a la part terminal del gen *recA* (a partir de la posició 955 de l'alineament), on els valors de distàncies disminueixen entre la soca problema AE147 i la resta de soques del clade *A.bestiarum*, amb excepció de les altres 3 soques problema. En aquest gràfic l'entrecreuament de línies s'interpreta com un esdeveniment de recombinació, i en aquest cas només es detecta un possible inici.

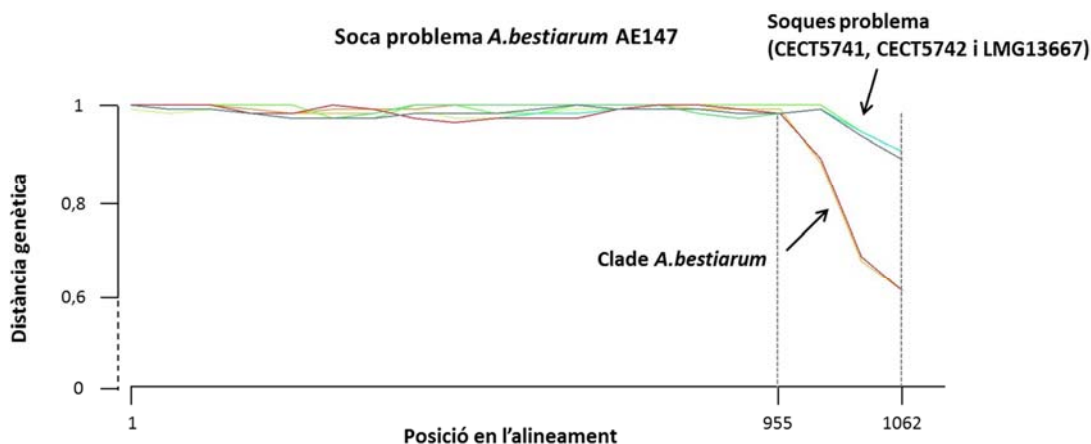


Figura 5.12 Resultat de sortida del programa RAT del clade *A.bestiarum* del gen *recA*. El gràfic representa la distància genètica (eix Y) respecte les posicions de l'alineament (eix X). Les línies amb colors indiquen cadascuna de les seqüències analitzades d'*A.bestiarum* respecte la seqüència de la soca problema *A.bestiarum* AE147.

5.4.1.2.2 Programa RDP4

Els resultats obtinguts del programa RAT no han permès descartar un possible origen exogen d'aquest fragment terminal, per això s'ha analitzat el conjunt de seqüències utilitzant el programa RDP4 que consta de varis mètodes de detecció de recombinació. Aquesta anàlisi permet estudiar quines seqüències analitzades podrien ser recombinants, proporcionar uns punts d'inici i final de recombinació (*breakpoints*) aproximats, i els possibles parentals, el principal (*major parent*) i el recombinant (*minor parent*).

Taula 5.12 Esdeveniments de recombinació detectats amb el programa RDP4 pel conjunt de seqüències del gen *recA*.

Mètode	Esdeveniments	Parental principal (%)	Parental recombinant (%)	Valor de <i>P</i>
RDP	4	<i>A.bestiarum</i> (98,5 - 99,0)	Desconegut	$9,6 \times 10^{-7}$
GENECONV	4	<i>A.bestiarum</i> (98,1 - 99,9)	Desconegut	$1,9 \times 10^{-5}$
BootScan	4	<i>A.bestiarum</i> (98,1 - 99,9)	Desconegut	$2,6 \times 10^{-7}$
MaxChi	4	<i>A.bestiarum</i> (98,1 - 99,9)	<i>A.eucrenophila</i> (92,9 - 94,7)	$9,8 \times 10^{-4}$
Chimaera	4	<i>A.bestiarum</i> (98,1 - 99,9)	Desconegut	$2,9 \times 10^{-5}$
Siscan	4	<i>A.bestiarum</i> (97,6 - 98,0)	Desconegut	$1,8 \times 10^{-8}$
3Seq	4	<i>A.bestiarum</i> (98,1 - 99,9)	<i>A.eucrenophila</i> (94,7 - 95,7)	$4,5 \times 10^{-7}$

Els resultats obtinguts utilitzant 7 mètodes de detecció diferents (RDP, GENECONV, BootScan, MaxChi, Chimaera, SiScan i 3Seq) s'exposen a la Taula 5.12. Es considera una possible soca recombinant, aquella que presenta un esdeveniment en la seqüència que sigui significatiu (amb un valor de $P < 0,05$) en al menys dos mètodes de detecció aplicats. En tots els casos s'ha identificat un fenomen de recombinació significatiu en la part terminal de les seqüències corresponents a les 4 soques problemes d'*A.bestiarum* (AE147, CECT5741, CECT5742 i LMG 13667), com es pot observar a la Figura 5.13. El programa ha estat capaç d'identificar *A.bestiarum* com a parental principal, amb una probabilitat molt elevada (> 97%). En canvi, la identificació del parental recombinant ha estat dubtosa ja que el resultat ha variat entre desconegut o *A.eucrenophila* depenent del mètode utilitzat (Taula 5.12). Només dos dels 7 mètodes assajats (MaxChi i 3Seq) han suggerit que *A.eucrenophila* podria ser el possible parental donador de la regió recombinant, amb valors de probabilitat elevats (92,9-95,7%). També s'ha identificat de forma aproximada el punt d'inici de recombinació (entre les posicions 794-933 de l'alineament), que després de l'observació amb detall de l'alineament s'ha assignat a la posició 979. En cap cas s'ha detectat el punt final del fragment recombinant.

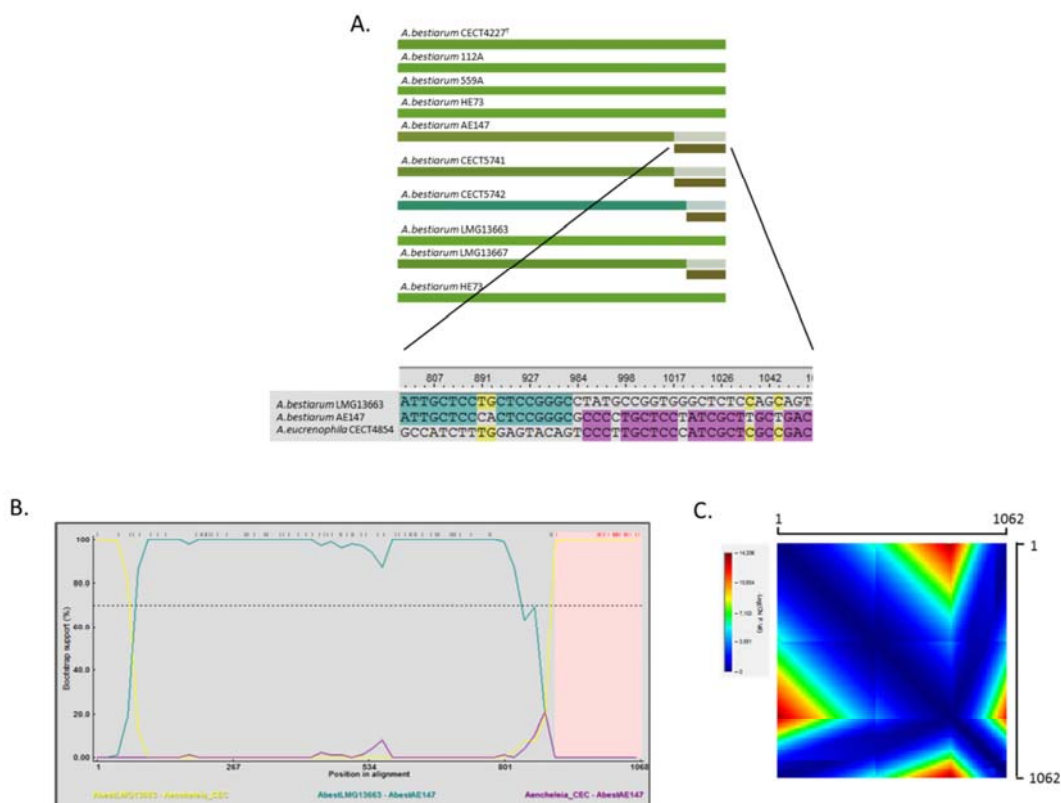


Figura 5.13 Anàlisi de detecció de recombinació del gen *recA* utilitzant el programa RDP4. A) Resultat de sortida del mètode RDP que mostra la identificació dels possibles esdeveniments de recombinació en la part terminal de les seqüències de les 4 soques problema d'*A.bestiarum*. A la part inferior es detallen les posicions nucleotídiques del fragment recombinant de la soca AE147 respecte els seus possibles parents. B) Gràfic obtingut pel mètode BootScan que detecta el punt d'inici de recombinació al final de la seqüència de la soca problema AE147 i els seus possibles parents. C) Matriu MaxChi que determina la localització òptima dels *breakpoints*, en aquest cas només mostra la identificació del punt d'inici de recombinació en la seqüència de la soca AE147.

5.4.1.3 Mètodes de detecció tradicionals

5.4.1.3.1 Incongruències filogenètiques

L'evidència de recombinació també s'ha confirmat al generar filogènies considerant de manera separada la seqüència que correspon a la regió no recombinant (posicions 1-978, Figura 5.14A) i la del fragment recombinant (posicions 979-final, Figura 5.14B) pels 4 aïllats recombinants i les soques tipus de totes les espècies d'*Aeromonas*. S'han construït arbres de màxima versemblança (ML) amb el programa PhyML, utilitzant el millor model de substitució que s'ajusta a les seqüències, GTR+G+I (Figura 5.14A) i TN93+G+I (Figura 5.14B).

Com era d'esperar, en l'arbre que correspon a la regió no recombinant s'ha observat que les 4 soques problema s'agrupen en el mateix clúster que la soca tipus d'*A.bestiarum*. En canvi, en l'arbre generat a partir de la regió recombinant, aquestes soques no es situen amb

A.bestiarum, sinó en un altre grup independent, juntament amb les espècies *A.eucrenophila* i *A.tecta*, indicant una relació més propera.

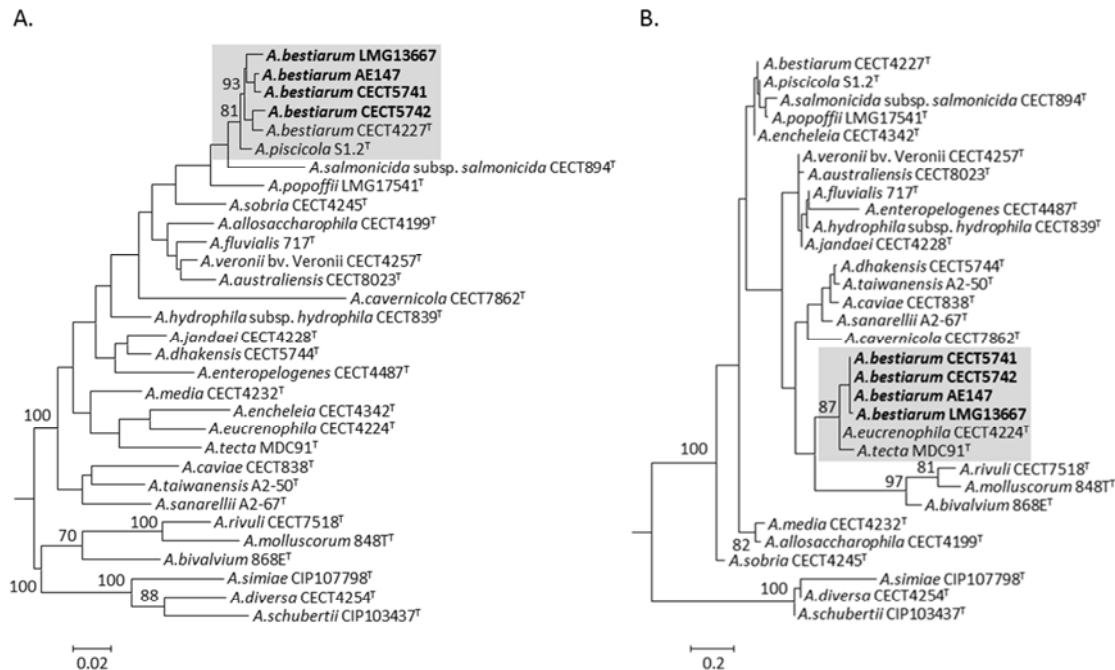


Figura 5.14 Arbres de màxima versemblança construïts a partir de seqüències intragèniques del gen *recA* dels aïllats recombinants i les soques tipus d'*Aeromonas*. A) Regió no recombinant (posicions 1-978). B) Fragment recombinant (posicions 979-final). La posició taxonòmica de les soques recombinants s'assenyala en negreta i fons gris. Els nodes mostren els valors de *bootstrap* superiors al 70%. L'escala indica el nombre de substitucions nucleotídiques per posició.

5.4.1.3.2 Contingut en guanina i citosina (G+C)

Per confirmar que la regió terminal del gen *recA* de les 4 soques recombinants d'*A.bestiarum* correspon a un fragment adquirit per recombinació s'ha estudiat el percentatge en G+C (%mol) d'aquesta regió (posicions 979-final) i s'ha comparat amb el contingut total del gen complet i la regió no recombinant (posicions 1-978). Els resultats obtinguts per les 4 soques recombinants (gris), així com per les espècies *A.bestiarum* (5 soques), *A.eucrenophila* (5 soques) i *A.tecta* (4 soques) es mostren a la Taula 5.13. A l'analitzar els percentatges, s'ha observat que al considerar el gen complet i la regió no recombinant, els 4 aïllats presenten uns valors quasi idèntics als del clade *A.bestiarum*. En canvi, a l'examinar el resultat de la regió terminal del gen, les soques recombinants presenten un valor clarament superior al del grup *A.bestiarum*, que recolzaria l'evidència del possible origen exogen d'aquest fragment.

Taula 5.13 Anàlisi del contingut en G+C (%mol) del gen complet *recA* i de regions intragèniques de les 4 soques recombinants i espècies relacionades.

Espècie	<i>recA</i>		
	Gen complet (1-1062)	Regió no recombinant (1-978)	Regió recombinant (979-1062)
<i>A.bestiarum</i> (n=5)	59,1	58,9	61,7
Recombinants (n=4)	59,2	58,8	64,0
<i>A.eucrenophila</i> (n=5)	60,2	59,7	66,4
<i>A.tecta</i> (n=4)	59,9	59,3	67,0

5.4.1.3.3 Composició d'aminoàcids i ús de codons

De la mateixa manera, s'ha dut a terme un estudi de la composició d'aminoàcids i de l'ús de codons de les seqüències del gen complet *recA* i de les dues regions intragèniques (no recombinant i recombinant) pel mateix grup de soques. Els resultats obtinguts es mostren a les taules de l'Annex V i VI.

En relació al patró de la composició d'aminoàcids, les 4 soques problema han presentat diferències significatives amb les soques del clade *A.bestiarum* al considerar el gen complet i la regió no recombinant, però si que se n'han detectat en la regió terminal (Annex V). Les soques recombinants tenen una proporció més baixa dels aminoàcids alanina (Ala) i serina (Ser), i més elevada de glicina (Gly) i prolina (Pro), en comparació a les obtingudes pel grup *A.bestiarum*. A més, aquests aïllats presenten l'aminoàcid isoleucina (Ile), que està absent en *A.bestiarum* i present en *A.eucrenophila*, i no contenen glutamina (Gln), com *A.eucrenophila* i *A.tecta*, però en canvi si que forma part de la composició d'*A.bestiarum*.

També s'ha calculat la freqüència d'ús de codons, per determinar si hi ha diferències en el codó majoritari utilitzat per codificar cada aminoàcid (Annex VI). Els resultats han estat coincidents amb els obtinguts en la composició d'aminoàcids. No s'han observat diferències significatives al comparar les freqüències d'ús de codons dels aïllats recombinants amb el clade d'*A.bestiarum* al considerar el gen complet o la regió no recombinant, però en canvi, s'han detectat diferències a l'analitzar la regió final del gen. Els resultats mostren que en aquesta regió, les soques del clade *A.bestiarum* utilitzen en igual proporció tres triplets diferents GCT/GCC/GCG per codificar l'alanina, mentre que les soques recombinants, *A.eucrenophila* i *A.tecta* només n'utilitzen un, i és sempre el mateix GCC. També s'ha observat l'ús de triplets diferents pels següents aminoàcids: àcid glutàmic (Glu), glicina (Gly), fenilalanina (Phe), prolina (Pro) i serina (Ser), i que en la majoria de casos coincideixen amb els emprats per *A.eucrenophila* i *A.tecta*.

Amb tots els resultats obtinguts de l'estudi de recombinació de les seqüències del gen *recA*, s'ha confirmat la presència d'una regió recombinant de 81 pb en la part terminal del gen en 4 soques d'*A.bestiarum*. Només s'ha identificat el punt d'inici de recombinació (posició 979) però

no el final. A més, s'ha identificat el parental principal com *A.bestiarum*, però en canvi, el parental recombinant no ha quedat del tot clar, podria estar relacionat amb les espècies *A.eucrenophila* i *A.tecta*.

5.4.2 Anàlisi de recombinació dels gens *recA+recX*

Per intentar trobar el punt final del fragment adquirit i el possible parental recombinant, s'ha ampliat la seqüenciació del gen *recA* en el seu extrem 3' terminal fins arribar al gen adjacent *recX* (Figura 5.1). Aquest procés s'ha realitzat per les 4 soques recombinants (AE147, CECT5741, CECT5742, LMG13667) i un grup representatiu de les espècies *A.bestiarum* (CECT4227^T, 112A, 559A, HE73, LMG13663), *A.eucrenophila* (CECT4224^T, CECT4827, CECT4853, CECT4854, CECT4855) i *A.tecta* (MDC91^T, MDC92, MDC93, MDC94). Per cada soca es va obtenir la seqüència de la regió no codificant entre els dos gens, i aproximadament, més dos terços del gen *recX*. Per les posteriors anàlisi de detecció de recombinació s'ha utilitzat doncs la seqüència nucleotídica que inclou el gen complet *recA*, la regió intergènica, i el gen parcial *recX*, i que s'ha anomenat com seqüència *recA+recX*.

5.4.2.1 Anàlisi exploratòria de les seqüències

S'ha realitzat un alineament múltiple de les 18 seqüències *recA+recX* obtingudes (posicions 1-1522) i el diagrama de punts corresponent (Figura 5.15), observant en ambdós casos que la part central presenta un patró comú per les 4 soques recombinants i *A.eucrenophila*, considerablement diferent del grup d'*A.bestiarum*. Aquesta regió central correspondria a la part final del gen *recA* detectada com a recombinant, la regió no codificant intergènica, i l'inici del gen *recX*.

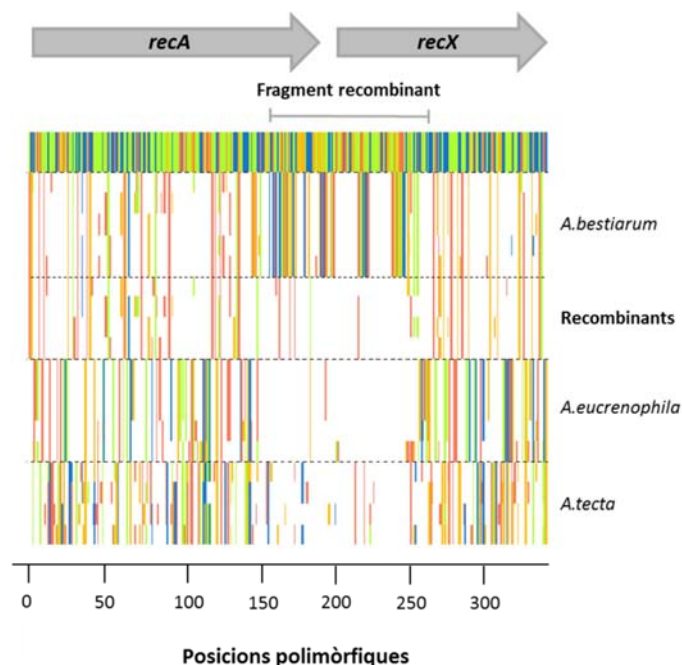


Figura 5.15 Diagrama de punts de les posicions polimòrfiques de les seqüències *recA+recX*. A la part superior es mostra la seqüència consens obtinguda a partir de totes les seqüències. Els llocs polimòrfics de cada soca es marquen amb un color diferent segons la base que varia (A, groc; G, blau; C, verd; T, vermell) respecte la seqüència consens.

5.4.2.2 Mètodes de detecció amb programes bioinformàtics

5.4.2.2.1 Programa RAT

La primera aproximació per acotar el final de la regió recombinant s'ha realitzat amb el programa RAT. Els resultats de l'anàlisi de la seqüència *recA+recX* de la soca recombinant AE147 al comparar-la amb la resta de seqüències obtingudes (3 soques recombinants, 5 *A.bestiarum*, 5 *A.eucrenophila* i 4 *A.tecta*) es mostren en un gràfic de distàncies (Figura 5.16). De forma general, s'aprecia clarament un entrecruament de línies, que indica un esdeveniment de recombinació entre el final del gen *recA* i el principi del gen *recX* per les 4 soques recombinants, donant unes posicions aproximades dels *breakpoints* (posicions 989-1292 de l'alineament). A més, el possible parental recombinant podria ser *A.eucrenophila* o *A.tecta*, perquè en aquesta regió la distància obtinguda al comparar la soca recombinant AE147 amb les seqüències d'aquestes espècies ha estat bastant similar, diferint considerablement al comparar-la amb les del clade *A.bestiarum*.

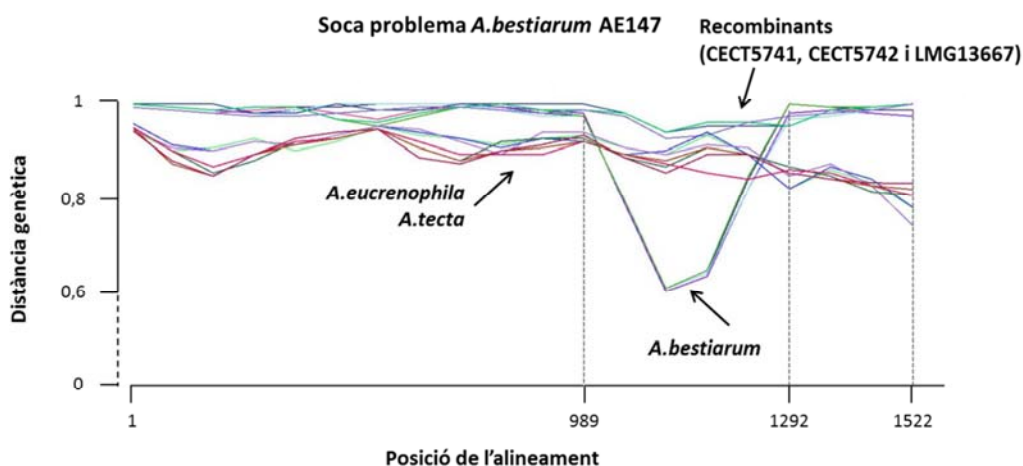


Figura 5.16 Resultat de sortida del programa RAT del *A.bestiarum* per les seqüències *recA+recX*. Els eixos en el gràfic representen les distàncies genètiques (eix Y) respecte les posicions de l'alineament (eix X). Les línies amb colors indiquen cadascuna de les seqüències analitzades respecte la seqüència de la soca problema *A.bestiarum* AE147.

5.4.2.2.2 Programa RDP4

A l'analitzar el conjunt de seqüències *recA+recX* amb el programa RDP4, els resultats obtinguts han corroborat l'existència d'aquesta regió recombinant pels 4 aïllats problema (Figura 5.17). Dels mètodes assajats, en 5 d'ells (GENECONV, BootScan, MaxChi, Chimaera, 3Seq) han detectat de forma significativa ($P < 0,05$) els 4 fragments recombinants (Taula 5.14). El programa ha identificat en tots els casos els dos parentals amb probabilitats molt elevades, *A.bestiarum* com parental principal (> 98%) i *A.eucrenophila* com parental recombinant (> 94%). En relació als *breakpoints*, el programa també ha identificat de forma aproximada el punt d'inici (entre les posicions 885-986 de l'alineament), ja detectada anteriorment en

l'anàlisi de les seqüències del gen *recA*, i el punt final de recombinació (entre les posicions 1188-1235). Després de visualitzar amb detall aquests regions en l'alineament de les seqüències *recA+recX*, s'han assignat les posicions 979 (final del gen *recA*) i 1235 (inici del gen *recX*) com els punts d'inici i final del fragment recombinant.

Taula 5.14 Esdeveniments de recombinació detectats amb el programa RDP4 pel conjunt de seqüències *recA+recX*.

Mètode	Esdeveniments	Parental principal (%)	Parental recombinant (%)	Valor de P
GENECONV	4	<i>A.bestiarum</i> (98,7 - 99,3)	<i>A.eucrenophila</i> (96,7 - 97,3)	$1,3 \times 10^{-17}$
BootScan	4	<i>A.bestiarum</i> (98,8 - 99,1)	<i>A.eucrenophila</i> (94,9 - 97,5)	$8,8 \times 10^{-14}$
MaxChi	4	<i>A.bestiarum</i> (98,2 - 99,6)	<i>A.eucrenophila</i> (96,2 - 97,3)	$5,4 \times 10^{-13}$
Chimaera	4	<i>A.bestiarum</i> (98,2 - 99,6)	<i>A.eucrenophila</i> (96,7 - 97,3)	$3,1 \times 10^{-12}$
3Seq	4	<i>A.bestiarum</i> (98,7 - 99,3)	<i>A.eucrenophila</i> (96,7 - 97,3)	$3,4 \times 10^{-25}$

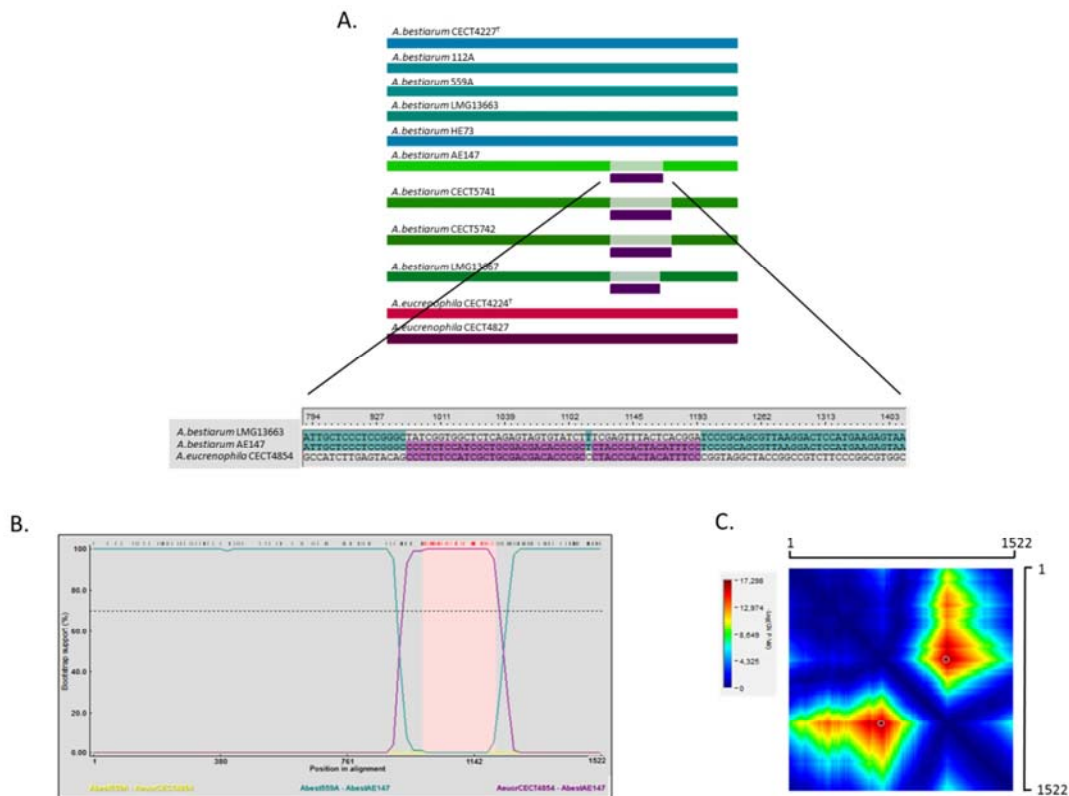


Figura 5.17 Anàlisi de detecció de recombinació de la seqüència *recA+recX* utilitzant el programa RDP4. A) Resultat de sortida del mètode RDP que identifica els esdeveniments de recombinació en les seqüències de les 4 soques recombinants d'*A.bestiarum*. En la part inferior es detalla les posicions nucleotídiques del fragment recombinant de la soca AE147 respecte els seus possibles parentals. B) Gràfic obtingut pel mètode BootScan que detecta els punts d'inici i final de recombinació de la seqüència de la soca problema AE147 i els seus possibles parentals. C) Matriu MaxChi que determina la localització òptima dels dos *breakpoints* en la seqüència de la soca AE147.

5.4.2.2.3 Programa SplitsTree4

Les relacions filogenètiques entre diferents organismes es representen freqüentment en forma d'arbres filogenètics, admetent que la seva evolució és un procés de bifurcació o ramificació. Però en alguns casos, les dades experimentals no s'ajusten a una representació en forma d'arbre, com per exemple en el cas de l'existència de transferència lateral d'informació genètica entre diferents tàxons. Per això, s'ha aplicat el mètode *split decomposition* al conjunt de seqüències *recA+recX*, el qual representa el procés evolutiu com un diagrama no jeràrquic en forma de xarxa (*split*), a més s'ha estimat la significació dels resultats amb el test Phi. Amb aquest mètode es pot detectar la presència de recombinació quan apareix una estructura reticulada en forma de xarxa en el gràfic.

A la Figura 5.18 es mostra la xarxa filogenètica obtinguda amb el programa SplitsTree4, la qual presenta en la part central una estructura reticulada que agrupa a les 4 soques d'*A.bestiarum* detectades com a recombinants (negreta), i que estan molt relacionades amb els clades *A.bestiarum* (esquerra) i *A.eucrenophila* (dreta) com possibles parentals. El test Phi, amb un valor de P de $1,7 \times 10^{-16}$, també ha demostrat que hi ha una evidència estadísticament significativa de recombinació per aquest grup de soques.

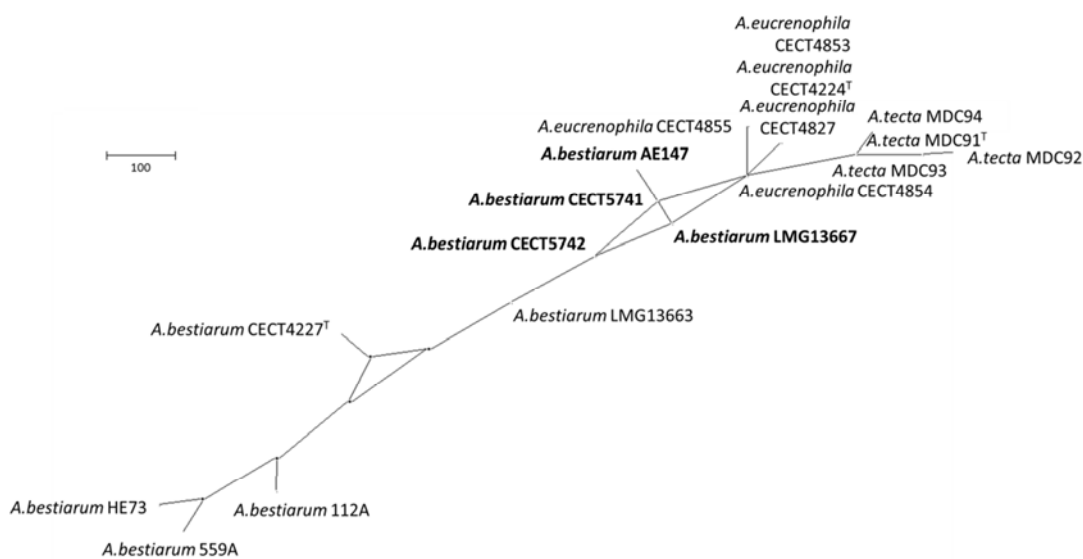


Figura 5.18 Anàlisi *split decomposition* de les seqüències *recA+recX* de les soques recombinants i espècies relacionades.

5.4.2.3 Mètodes de detecció tradicionals

5.4.2.3.1 Contingut en guanina i citosina (G+C)

Per examinar l'evidència de recombinació en el gen *recX* s'ha determinat el percentatge en G+C (%mol) considerant la seqüència obtinguda del gen (> 80%, posicions 1-416) i les regions recombinant (posicions 1-129) i no recombinant (posicions 130-416). A la Taula 5.15 s'indiquen els resultats obtinguts per les 4 soques (gris), així com per les espècies *A.bestiarum* (5 soques),

A.eucrenophila (5 soques) i *A.tecta* (4 soques). S'han obtingut resultats molt coincidents als del gen *recA*. A l'analitzar el percentatge mig de G+C de la seqüència del gen i de la regió no recombinant dels aïllats, aquests presenten valors bastant similars als obtinguts pel clade *A.bestiarum*, però són molt diferents als d'*A.eucrenophila*. En quan als resultats de la regió recombinant del gen *recX*, aquestes soques presenten un valor mig de G+C més semblant al del clade *A.eucrenophila*.

Taula 5.15 Anàlisi del contingut en G+C (%mol) del gen *recX* i de regions intragèniques de les 4 soques recombinants i espècies relacionades.

Espècie	<i>recX</i>		
	Gen (1-416)	Regió recombinant (1-129)	Regió no recombinant (130-416)
<i>A.bestiarum</i> (n=5)	62,2	63,7	62,2
Recombinants (n=4)	62,6	63,0	61,8
<i>A.eucrenophila</i> (n=5)	64,2	62,9	64,7
<i>A.tecta</i> (n=4)	62,0	61,4	62,3

5.4.2.3.2 Composició d'aminoàcids i ús de codons

També s'ha analitzat la composició d'aminoàcids i l'ús de codons tenint en compte la seqüència del gen *recX* i les dues regions intragèniques (recombinant i no recombinant) pel mateix grup de soques. Els resultats obtinguts es mostren a les taules de l'Annex VII i VIII. En les dues anàlisis les 4 soques problema només han presentat diferències significatives respecte al clade *A.bestiarum* a la regió recombinant del gen *recX*. Analitzant amb detall la composició d'aminoàcids d'aquesta regió, s'ha observat que aquestes soques presenten una proporció més baixa dels aminoàcids alanina (Ala), àcid aspàrtic (Asp), àcid glutàmic (Glu), leucina (Leu), arginina (Arg) i treonina (Thr), i més alta de fenilalanina (Phe), metionina (Met) i valina (Val) respecte el grup d'*A.bestiarum*, un patró molt similar al d'*A.eucrenophila*. Aquests aïllats també tenen l'aminoàcid prolina (Pro), que està absent en *A.bestiarum* i present en *A.eucrenophila* i *A.tecta*, i igual que en *A.eucrenophila* i *A.tecta* no contenen glutamina (Gln), a diferència d'*A.bestiarum*. Al comparar els resultats obtinguts de freqüències d'ús de codons a la regió recombinant, el patró de les soques recombinants s'assembla més a les d'*A.eucrenophila*, utilitzant els mateixos triplets pels aminoàcids àcid aspàrtic (Asp), àcid glutàmic (Glu), prolina (Pro) i serina (Ser), que difereix del clade *A.bestiarum*.

5.4.2.3.3 Incongruències filogenètiques

S'han construït filogènies a partir de les seqüències obtingudes del gen *recX* (posicions 1-416, Figura 5.19A), i dels fragments recombinant (posicions 1-129, Figura 5.19B) i no recombinant (posicions 130-416, Figura 5.19C) de les 4 soques problema i d'un grup representatiu de les espècies *A.bestiarum*, *A.eucrenophila* i *A.tecta*. S'han generat arbres de màxima versemblança (ML) amb el programa PhyML, utilitzant el millor model de substitució que s'ajusta a les dades, HKY+G+I (Figures 5.19A i C) i K2+G+I (Figura 5.19B). Les filogènies obtingudes han posat de manifest que l'únic cas en que aquestes soques no s'agrupen amb el clade *A.bestiarum* és

quan es considera la seqüència de la regió recombinant (Figura 5.19B), que forma un clúster junt amb *A. eucrenophila*. Aquests resultats han estat coincidents als obtinguts pel gen *recA*.

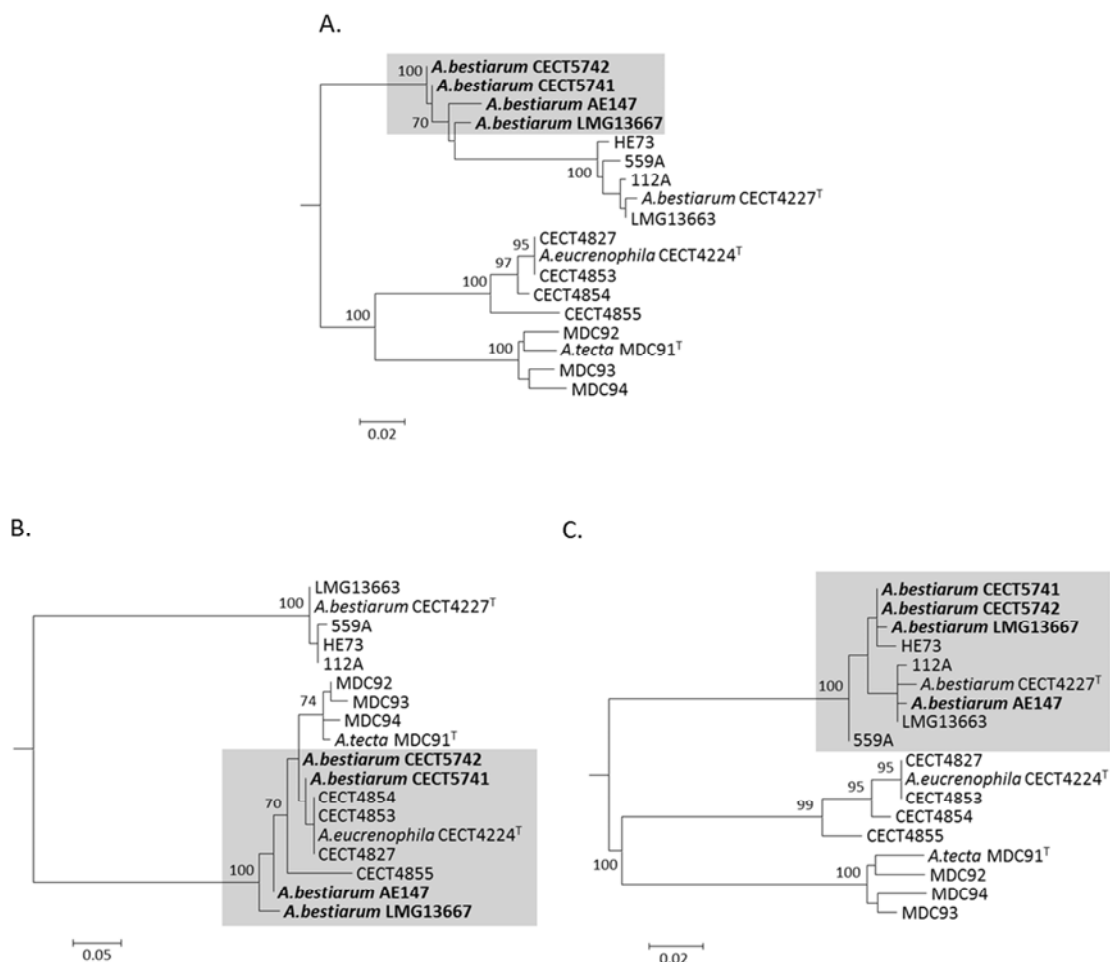
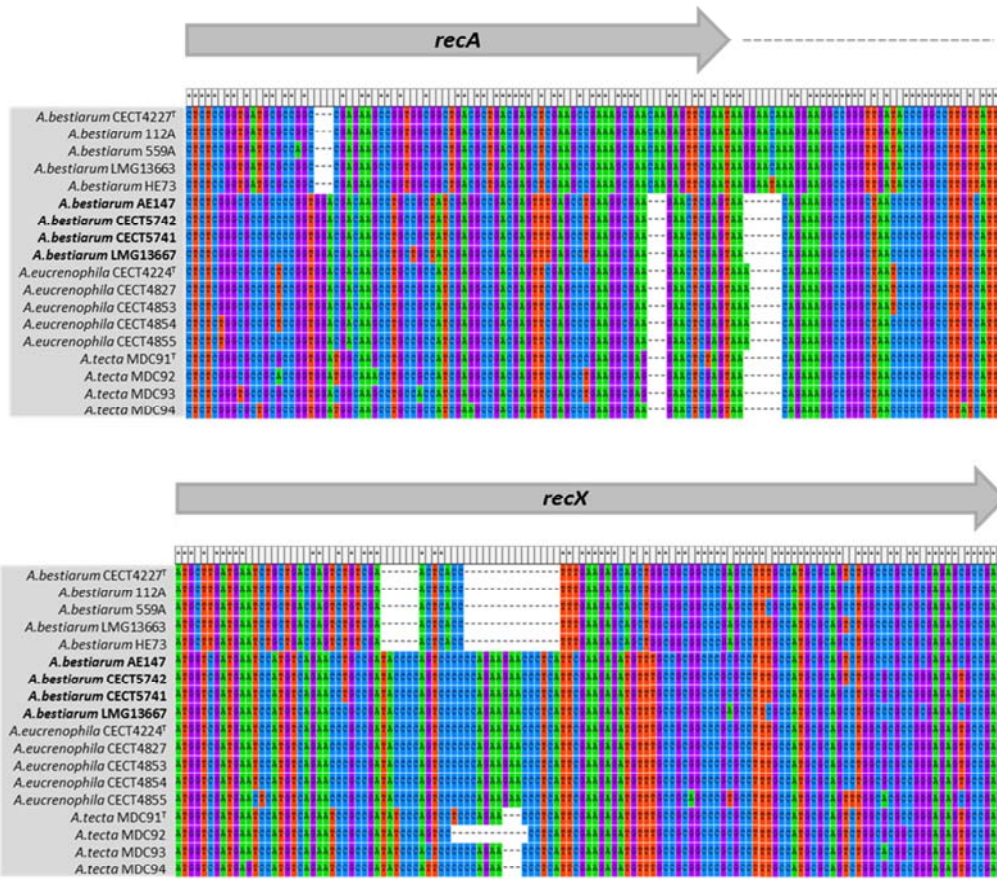


Figura 5.19 Arbres de màxima versemblança construïts a partir de les seqüències del gen *recX* i de regions intragèniques de les soques problema i espècies relacionades. A) Gen *recX* (posicions 1-416). B) Regió recombinant (posicions 1-129). C) Regió no recombinant (posicions 130-416). La posició taxonòmica de les soques recombinants s'assenyala en negreta i fons gris. Els nodes mostren els valors de *bootstrap* superiors al 70%. L'escala indica el nombre de substitucions nucleotídiques per posició.

5.4.3 Caracterització del fragment recombinant

En resum, a partir dels resultats obtinguts de totes les anàlisis, s'ha identificat la presència d'un fragment de possible origen exogen en 4 soques ambientals d'*A. bestiarum* (AE147, CECT5741, CECT5742 i LMG13667) d'origen geogràfic diferent (Finlàndia, Alemanya, Suïssa i Estats Units, respectivament). Com es pot observar en l'alineament (Figura 5.20A), la seqüència nucleotídica d'aquesta regió té una longitud de 248 pb que correspon a la part terminal del gen *recA* i a l'inici del gen *recX*.

A.



B.

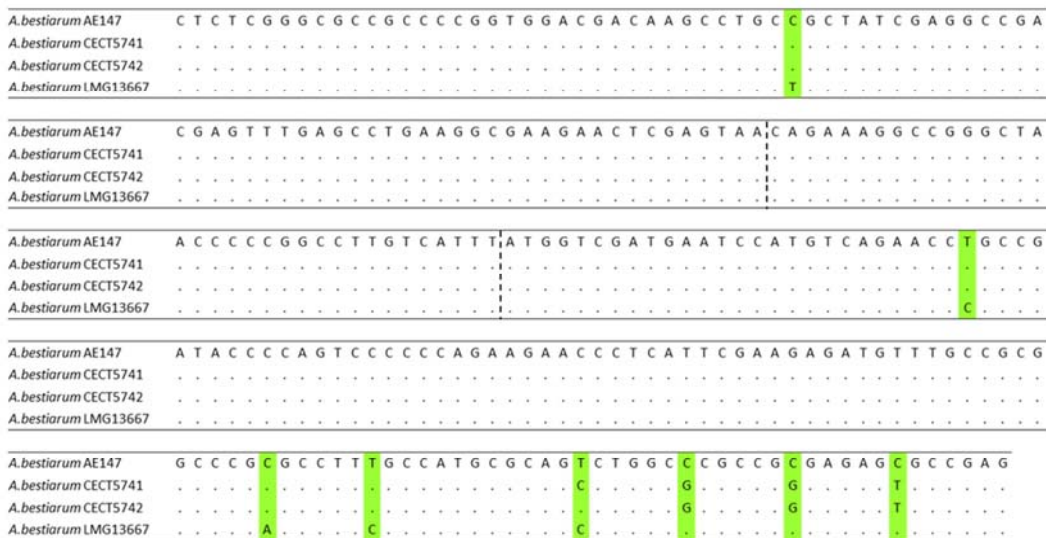


Figura 5.20 Visualització de les seqüències nucleotídiques del fragment recombinant. A) Alineament múltiple de seqüències amb el programa ClustalW (MEGA6) dels 4 aïllats recombinants (negreta) i de soques de les espècies *A.bestiarum*, *A.eucrenophila* i *A.tecta*. B) Diferències nucleotídiques (verd) de les 4 seqüències corresponents a les soques recombinants. Les línies discontinues verticals assenyalen el final del gen *recA* i l'inici del gen *recX*.

Tots els fragments han estat pràcticament idèntics, diferint només entre 1-8 pb (Figura 5.20B). La seqüència de la regió recombinant que correspon al gen *recA* ha estat pràcticament igual pels 4 aïllats, amb excepció d'un canvi en la posició 1014 de la soca LMG13667. Però, quan s'ha considerat la regió recombinant del gen *recX*, la variació nucleotídica detectada ha estat major. També, s'han pogut identificar per diferents mètodes (RAT, RDP4, SplitsTree4, contingut G+C, composició d'aminoàcids, ús de codons, filogènies) els parents principal (*A.bestiarum*) i recombinant (*A.eucrenophila*) (Figura 5.21), destacant que l'origen de la regió recombinant és una altre espècie del gènere *Aeromonas*. Amb la determinació del percentatge del contingut en G+C de la totalitat del fragment recombinant s'han confirmat aquests resultats perquè els 4 aïllats recombinants presenten un valor de 62,9 %mol que és molt similar a l'obtingut per *A.eucrenophila* (63,2 %mol), però en canvi és molt diferent al del clade *A.bestiarum* (60,7 %mol).

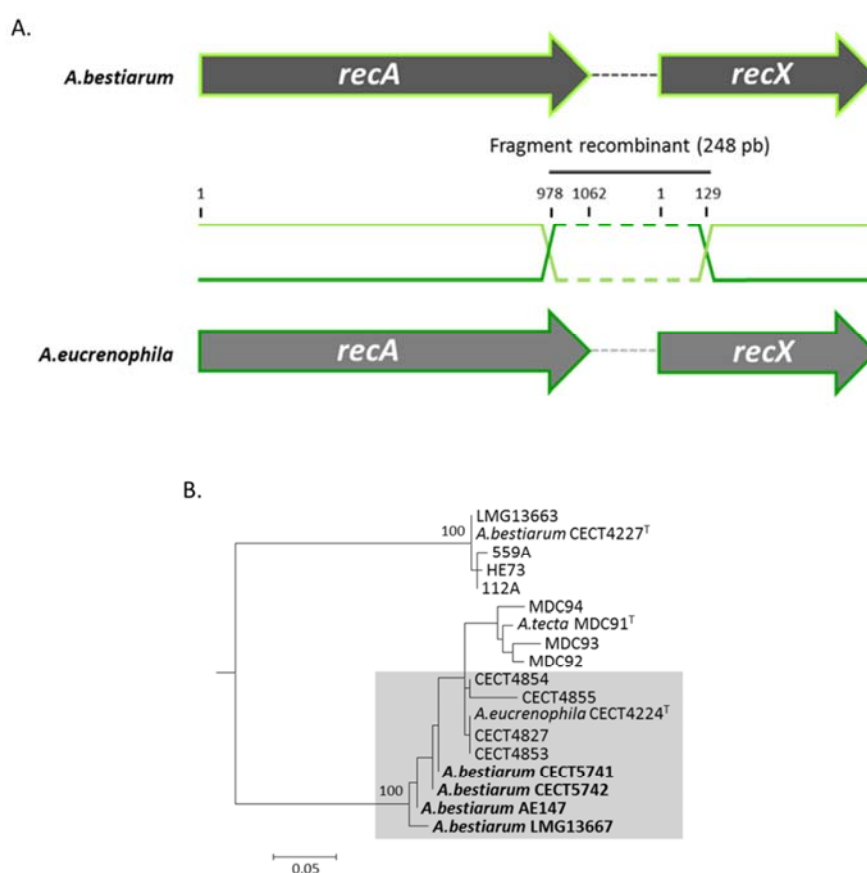


Figura 5.21 Identificació del parental recombinant. A) Representació esquemàtica de l'esdeveniment de recombinació i les soques parentals. B) Arbre de màxima versemblança construït a partir de les posicions 979-1235 de l'alineament de les seqüències *recA* + *recX* que correspon al fragment recombinant. La posició taxonòmica de les soques recombinants s'assenyala en negreta i fons gris. Els nodes mostren els valors de *bootstrap* superiors al 70%. L'escala indica el nombre de substitucions nucleotídiques per posició.

5.5 PREDICCIÓ DE LA PROTEÏNA RecA D'*Aeromonas*

La proteïna RecA té un paper biològic important, com a principal responsable dels processos de recombinació homòloga del DNA i com activadora del sistema de reparació d'emergència o resposta SOS dels procariotes. Per tractar de veure si l'adquisició d'un fragment recombinant pot afectar a l'estructura i la funció de la proteïna RecA, s'ha realitzat la predicció de l'estructura tridimensional d'una soca recombinant (CECT5741) i de la soca tipus de l'espècie *A.bestiarum* (CECT4227^T).

5.5.1 Modelatge molecular per homologia

Per determinar l'estructura de la proteïna RecA de les dues soques d'*A.bestiarum* s'han emprat programes informàtics utilitzant la tècnica de modelatge per homologia que consisteix en seleccionar proteïnes homòlogues que comparteixin una estructura similar per alineament estructural a partir d'un banc de dades. A continuació, s'expliquen detalladament els resultats obtinguts en cada etapa de la predicció de la proteïna RecA d'*Aeromonas*.

5.5.1.1 Reconeixement del motlle

La cerca de proteïnes homòlogues de RecA d'origen bacterià a la base de dades *Protein Data Bank* (PDB, <http://www.rcsb.org/pdb/home/home.do>) ha generat com a resultat un total de 60 estructures que corresponen a les següents espècies: *Mycobacterium smegmatis* (30), *Mycobacterium tuberculosis* (14), *Escherichia coli* (14), *Thermotoga marítima* (1), *Deinococcus radiodurans* (1), i una estructura parcial de l'extrem C-terminal d'*Escherichia coli* K12. Totes elles resoltes experimentalment a través de diferents tècniques.

Aquesta cerca a la base PDB també s'ha realitzat a través dels programes Phyre2 i SWISS-MODEL. El programa Phyre2 ha seleccionat com a resultat de la cerca els 20 millors motlles (*E-values* superiors), destacant els 11 primers corresponents a estructures d'*E.coli* amb percentatges de similitud superiors al 50% (59-83%) amb les seqüències problema (soques recombinant i tipus). Els resultats de la cerca mitjançant el programa SWISS-MODEL es mostren a la Taula 5.16, identificant també les estructures d'*E.coli* com les més idònies per ser utilitzades en el modelatge de RecA d'*A.bestiarum*, amb els valors més alts obtinguts de percentatges d'identitat, similitud i cobertura (*coverage*).

A la Taula 5.17 es detallen les característiques generals de totes les estructures de RecA d'*E.coli* que estaven disponibles a la base de dades PDB per realitzar l'estudi de predicció. La identificació de les millors estructures d'*E.coli* per a ser utilitzades com a motlle apareixen marcades en gris, 2REB, 3CMW i 3CMT. Totes elles són estructures obtingudes per difracció de raigs X, la primera cristal·litzada sense DNA (2REB), la segona i la tercera (3CMW i 3CMT) amb una cadena simple i una doble de DNA, respectivament.

Taula 5.16 Resultats de la cerca d'estructures motlles amb el programa SWISS-MODEL per la soca recombinant *A.bestiarum* CECT5741.

Motlles	Identitat	Similitud	Covertura
<i>Escherichia coli</i>	77,5 - 80,3	52 - 53	92 - 96
<i>Mycobacterium smegmatis</i>	66,6 - 66,9	49 - 50	92
<i>Mycobacterium tuberculosis</i>	65,1	49	93
<i>Thermotoga maritima</i>	60,3	47	92
<i>Deinococcus radiodurans</i>	58,7	46	96

Taula 5.17 Principals característiques de les estructures tridimensionals de la proteïna RecA d'*E.coli* disponibles a la base de dades PDB.

PDB	DNA ^a	Cofactors i lligands ^b	Tècnica de cristal·lització	Referències
3CMT	dsDNA	MG, ALF, ADP	Difracció raigs X (3,15Å)	(Chen <i>et al.</i> 2008)
3CMU	ssDNA	MG, ALF, ADP	Difracció raigs X (4,2Å)	(Chen <i>et al.</i> 2008)
3CMW	ssDNA	MG, ALF, ADP	Difracció raigs X (2,8Å)	(Chen <i>et al.</i> 2008)
3CMX	dsDNA	MG, ALF, ADP	Difracció raigs X (3,4Å)	(Chen <i>et al.</i> 2008)
3CMV	-	MG, ANP	Difracció raigs X (4,3Å)	(Chen <i>et al.</i> 2008)
1N03	-	ADP	Microscopia electrònica	(Vanloock <i>et al.</i> 2003)
2REC	-	-	Microscopia electrònica	(Yu & Egelman 1997)
1XMS	-	MN, ANP	Difracció raigs X (2,1Å)	(Xing & Bell 2004b)
1XMV	-	MG, ADP	Difracció raigs X (1,9Å)	(Xing & Bell 2004b)
1U94	-	CA	Difracció raigs X (1,9Å)	(Xing & Bell 2004b)
1U98	-	Glicerol + SO ₄ ²⁻	Difracció raigs X (2,0Å)	(Xing & Bell 2004b)
1U99	-	PO ₄ ³⁻	Difracció raigs X (2,6Å)	(Xing & Bell 2004b)
1REA	-	-	Difracció raigs X (2,7Å)	(Story & Steitz 1992)
2REB	-	-	Difracció raigs X (2,3Å)	(Story & Steitz 1992)
1AA3	-	-	Resonància magnètica nuclear	(Aihara <i>et al.</i> 1997)

^a ds, DNA de cadena doble; ss, DNA de cadena simple

^b MG, magnesi; ALF, tetrafluoraluminat; ADP, adenosina 5-difosfat; ANP, ester adenilat de l'àcid fosfoaminofosfònic; MN, magnaès; CA, ió calci; SO₄²⁻, ió sulfat; PO₄³⁻, ió fosfat

5.5.1.2 Alineament

L'alineament múltiple de seqüències aminoacídiques és una etapa important en el procés de modelatge, perquè és bastant probable que les regions conservades presentin estructures similars. S'ha realitzat un primer alineament (Annex IX) amb les seqüències de les soques recombinant (CECT5741) i tipus d'*A.bestiarum* (CECT4227^T) i les seqüències de 5 estructures resoltes de gèneres bacterians diferents obtingudes del PDB (2REB, *E.coli*; 1UBG, *Mycobacterium smegmatis*; 1MO4, *Mycobacterium tuberculosis*; 3HR8, *Thermotoga maritima* i 1XP8 *Deinococcus radiodurans*). El resultat d'aquest alineament ha permès confirmar que la

seqüència d'*E.coli* és la que presenta més posicions conservades al comparar-la amb les d'*A.bestiarum*.

En segon lloc, també s'han generat alineaments de seqüències per parelles, considerant les seqüències d'*A.bestiarum* respecte els 3 models d'*E.coli* seleccionats (2REB, 3CMW i 3CMT) per comprovar que presenten percentatges d'identitat elevats i que són bons candidats com a proteïnes motlle pel modelatge. A la Figura 5.22 es mostra el resultat de l'alineament corresponent a les seqüències de la soca recombinant (CECT5741) i de l'estructura 3CMW d'*E.coli*, en el que es pot observar que el percentatge d'identitat és elevat (79%). Entremig de les dues seqüències s'indica la similitud (gris) detectada entre ambdues, de manera que si l'aminoàcid és el mateix es conserva el símbol, si l'aminoàcid és diferent però manté la càrrega positiva s'indica amb el símbol +, i finalment es deixa en blanc si és totalment diferent.

```

Score = 520 bits (1340), Expect = e-147
Identities = 262/328 (79%), Positives = 292/328 (89%)

CECT5741 MDQNKQKALAAALGQIEKQFGKGSIMRLGDSKTMDEIAISTGSLSLDVALGIGGLPCGRI
      +D+NKQKALAAALGQIEKQFGKGSIMRLG+ ++MD+E ISTGSLSLD+ALG GGLP GRI
3CMW IDENKQKALAAALGQIEKQFGKGSIMRLGEDRSMDEVETIISTGSLSLDIALGAGGLPMGRI

CECT5741 VEIYGPRESSGKTTTLTLQVIAEAQKKGVCAFI DAEHALDPI YAAKLG VNVDDLLISQPDT
      VEIYGPRESSGKTTTLTLQVIA AQ++GK CAFIDAEHALDPIYA KLGV++D+LL SQPDT
3CMW VEIYGPRESSGKTTTLTLQVIAAAQREGKTCAFI DAEHALDPI YARKLGVDIDNLLCSQPDT

CECT5741 GEQALEICDMLVRSNAVDVIVVDSVAALTPKAEIEGEMGDSHVGLQARLMSQALRKL TAN
      GEQALEICD L RS AVDVI+VDSVAALTPKAEIEGE+GDSH+GL AR+MSQA+RKL N
3CMW GEQALEICDALARSGAVDVIVVDSVAALTPKAEIEGEIGDSHMG LAARMMSQAMRKLGN

CECT5741 IKNANCLCIFINQIRMKIGVMFGSPETTTGGNALKFYASVRLDIRRIGAIKEGDEVVGN E
      +K +N L IFINQIRMKIGVMFG+PETTTGGNALKFYASVRLDIRRIGA+KEG+ VVG+E
3CMW LKQSNLTLIFINQIRMKIGVMFGNPETTTGGNALKFYASVRLDIRRIGAVKEGENNVGSE

CECT5741 TRVKVVKNVAPPFKQAEFQIFYGVGISKEGELVDLGVKHKLIDKAGAWYSYNGEKIGQG
      TRVKVVKNK+A PFKQAEFQI YG GI+ GELVDLGVK KLI+KAGAWYSY GEKIGQG
3CMW TRVKVVKNKIAAPFKQAEFQILYGEGINFYGELVDLGVKEKLI EKAGAWYSYKGEKIGQG

CECT5741 KANVMKLFSENKAMAAEVEARLRELLS
      KAN +N A E+E ++RELLS
3CMW KANATAWLKDNPETAKEIEKKVRELLS

```

Figura 5.22 Alineament de la seqüència aminoacídica de RecA d'*A.bestiarum* CECT5741 i de l'estructura 3CMW d'*E.coli* amb el programa JPred.

5.5.1.3 Construcció del model

En la tercera etapa es comença a construir un model tridimensional a partir de l'alineament entre la seqüència de la proteïna problema i l'estructura de la proteïna motlle. En aquest treball, la construcció del model 3D ha inclòs les prediccions de les estructures secundària i terciària de la seqüència aminoacídica de RecA de la soca tipus d'*A.bestiarum* (CECT4227^T) i un aïllat recombinant (CECT5741). Aquestes prediccions s'han realitzat utilitzant tres models diferents de RecA d'*E.coli* com a motlle: 2REB, que es troba lliure en forma de monòmer i

sense DNA, i 3CMW i 3CMT, que estan polimeritzant en forma de filaments, adoptant un estat inactiu (3CMW, DNA de cadena simple) o actiu (3CMT, DNA de cadena doble).

5.5.1.3.1 Predicció de l'estructura secundària

L'estructura secundària es refereix a l'estructura que adopta espacialment la seqüència aminoacídica. Un polipèptid té molts grups capaços de formar ponts d'hidrogen entre els àtoms que formen l'enllaç peptídic, i per tant, es poden formar de manera espontània elements d'estructura secundària, com làmines β i hèlixs α , que queden connectats entre si per llaços (*loops*) i per regions intrínsecament desorganitzades. Per l'estudi s'ha obtingut la predicció de l'estructura secundària de les dues seqüències de RecA d'*A.bestiarum* a partir de dues aproximacions diferents: A) directament des de la pròpia seqüència i B) utilitzant les estructures d'*E.coli* 2REB, 3CMW i 3CMT com a motlles.

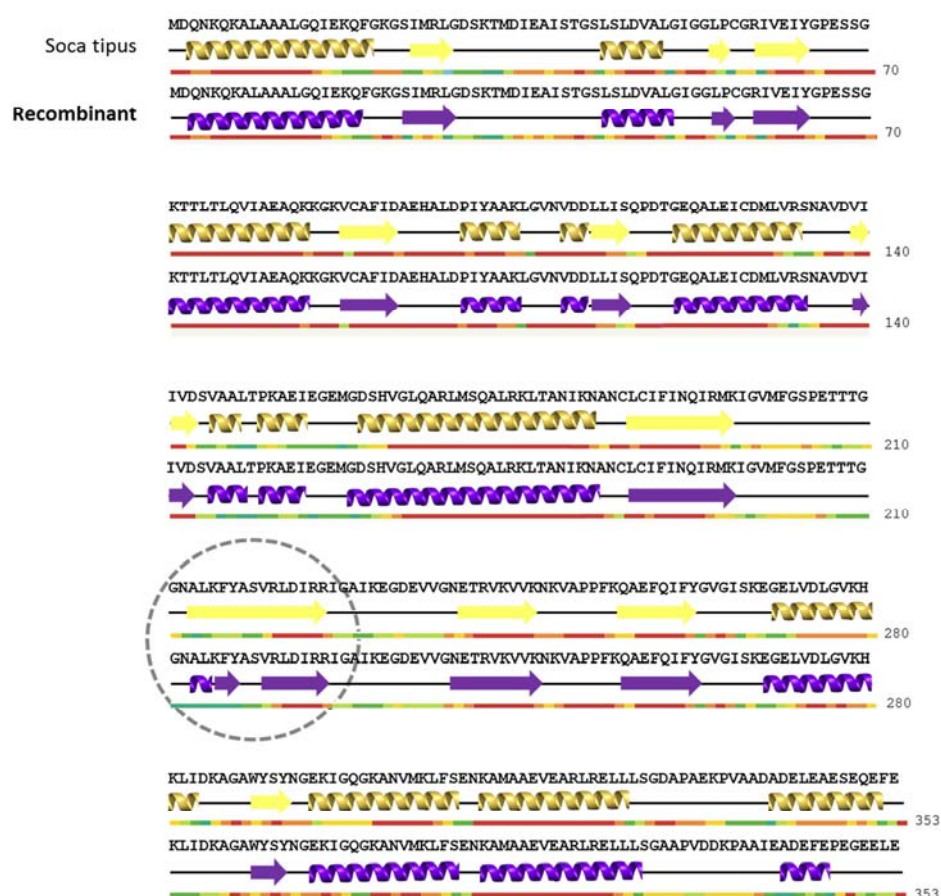


Figura 5.23 Predicció de l'estructura secundària de la proteïna RecA de la soca tipus d'*A.bestiarum* CECT4227^T i la soca recombinant CECT5741 a partir de la seqüència aminoacídica. La línia situada immediatament per sota de la predicció correspon al grau de confiança de l'estructura secundària obtinguda. Els colors càlids indiquen un elevat grau de confiança i els colors freds menor.

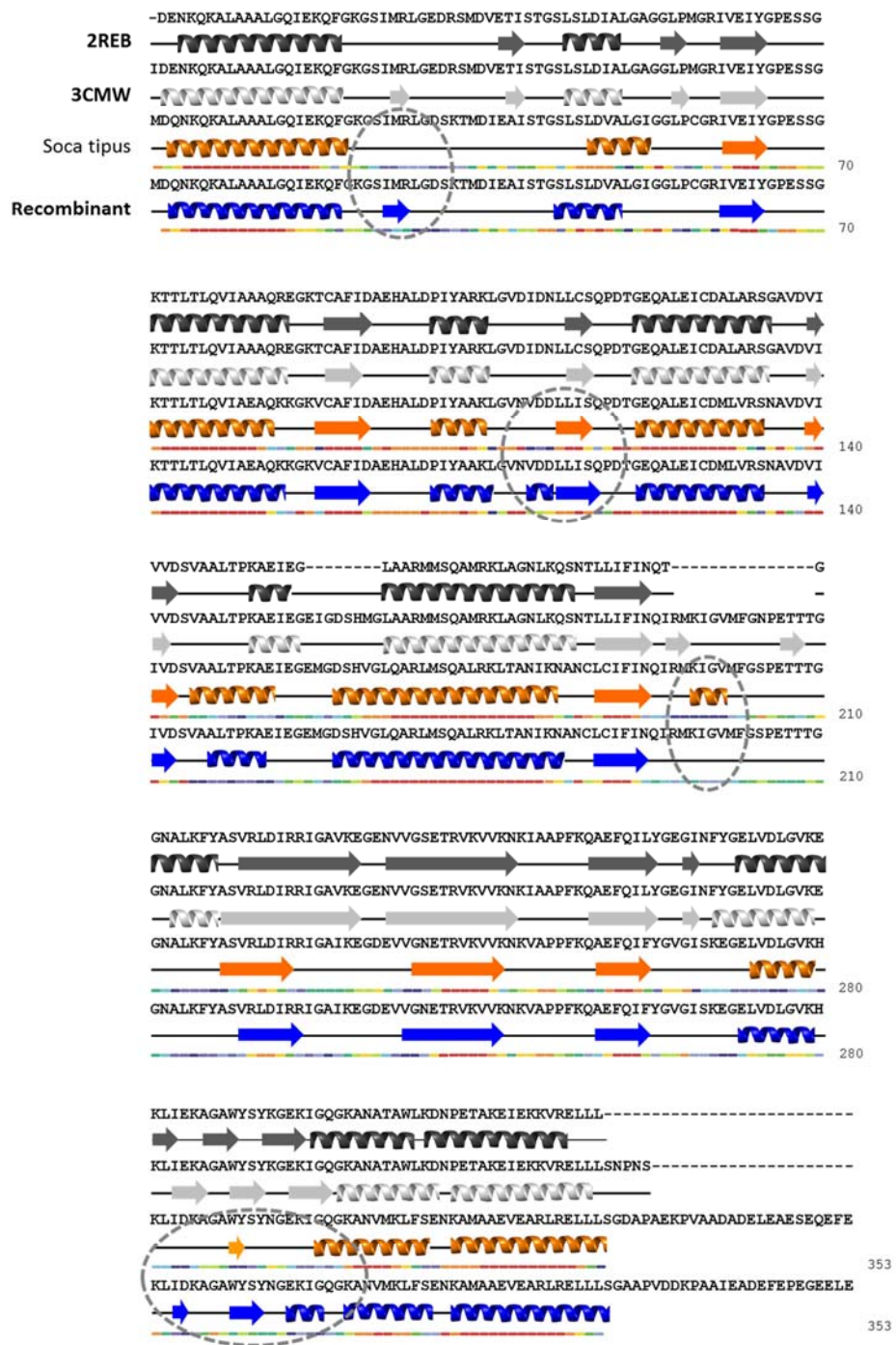


Figura 5.24 Predicció de l'estructura secundària de la proteïna RecA de la soca tipus d'*A.bestiarum* CECT4227^T i la soca recombinant CECT5741 a partir dels motlles 2REB i 3CMW. La línia situada immediatament per sota de la predicció correspon al grau de confiança de l'estructura secundària obtinguda. Els colors càlids indiquen un elevat grau de confiança i els colors freds menor.

A la Figura 5.23 es representa la comparació de l'estructura secundària de RecA obtinguda per la soca tipus d'*A.bestiarum* CECT4227^T (groc) i la soca recombinant CECT5741 (lila). Segons la predicció, i tal com era d'esperar, els elements d'estructura secundària obtinguts per ambdues proteïnes d'*A.bestiarum* han estat pràcticament iguals (ja que presenten seqüències gairebé idèntiques) amb excepció d'una regió al voltant de les posicions 210-230 (cercle) amb un baix grau de confiança. També s'ha determinat el percentatge d'estructures (hèlixs α i làmines β) en cadascuna de les prediccions, i que en els dos casos ha estat d'un 41% d'hèlixs α , un 20% de làmines β i un 23% de desordre.

A més, s'han generat les prediccions de l'estructura secundària d'ambdues soques d'*A.bestiarum* a partir dels 3 motlles de RecA d'*E.coli*, obtenint resultats idèntics per cada soca utilitzant els diferents models. A la Figura 5.24 es mostra la predicció per les dues soques problema (soca tipus i recombinant) juntament amb les estructures dels motlles utilitzats (2REB i 3CMW) a la part superior. Les 4 diferències detectades al comparar les dues prediccions s'assenyalen amb cercles de línies discontinues, i presenten en alguns casos un grau de confiança baix.

5.5.1.3.2 Predicció de l'estructura terciària

Per analitzar el possible efecte del fragment recombinant a l'extrem C-terminal de la proteïna RecA, s'han obtingut les estructures tridimensionals de RecA de les dues soques d'*A.bestiarum* a partir dels 3 motlles d'*E.coli* seleccionats, donant lloc a 6 prediccions que s'han anomenat com 4227_2REB, 4227_3CMW, 4227_3CMT, 5741_2REB, 5741_3CMW i 5741_3CMT, en funció de la soca modelada (soca tipus 4227 i recombinant 5741) i el motlle utilitzat (2REB, 3CMW i 3CMT). S'han utilitzat els programes Phyre2 i SWISS-MODEL per construir les estructures, donant resultats comparables, per això, només es presenten els resultats obtinguts amb el Phyre2.

5.5.1.3.2.1 Obtenció dels models tridimensionals

La predicció estructural corresponent a les dues soques problema d'*A.bestiarum* utilitzant com a motlle la proteïna RecA d'*E.coli* sense DNA (2REB) es visualitzen a la Figura 5.25. Els resultats mostren que al comparar ambdues estructures (Figura 5.25A) no han presentat diferències significatives, essent pràcticament idèntiques, i presentant 13 hèlixs α i 10 làmines β (Figura 5.25B). En canvi, si que se n'han detectat en la superposició de la predicció de la soca recombinant (5741_2REB) i el motlle d'*E.coli* 2REB (Figura 5.25C). En aquesta cas, la principal diferència es troba en la làmina β_0 situada en l'extrem N-terminal i la làmina β_5 situada al domini central de 2REB que no han estat modelades en la predicció de la soca recombinant. També s'han observat diferències en relació als llocs d'unió al DNA, ja que tant el *Loop* L1 com el *Loop* L2, absents en el motlle 2REB, si que han estat modelats com dues làmines β i un gir en la predicció de la soca recombinant (5741_2REB).

Les prediccions obtingudes de la proteïna RecA de les dues soques d'*A.bestiarum* a partir del motlle amb una cadena de DNA (3CMW) s'observen a la Figura 5.26. En aquest cas tampoc s'han observat diferències al superposar les estructures de les dues soques d'*A.bestiarum* (Figura 5.26A). Però a diferència dels resultats de l'anterior motlle, en la superposició entre la proteïna de la soca recombinant (5741_3CMW) i el propi motlle d'*E.coli* 3CMW no s'han detectat diferències, han estat pràcticament idèntiques (Figura 5.26B). També s'han generat les prediccions de les soques d'*A.bestiarum* amb el motlle de doble cadena de DNA (3CMT), estructures que no es mostren perquè han estat altament coincidents a les obtingudes a partir del motlle 3CMW.

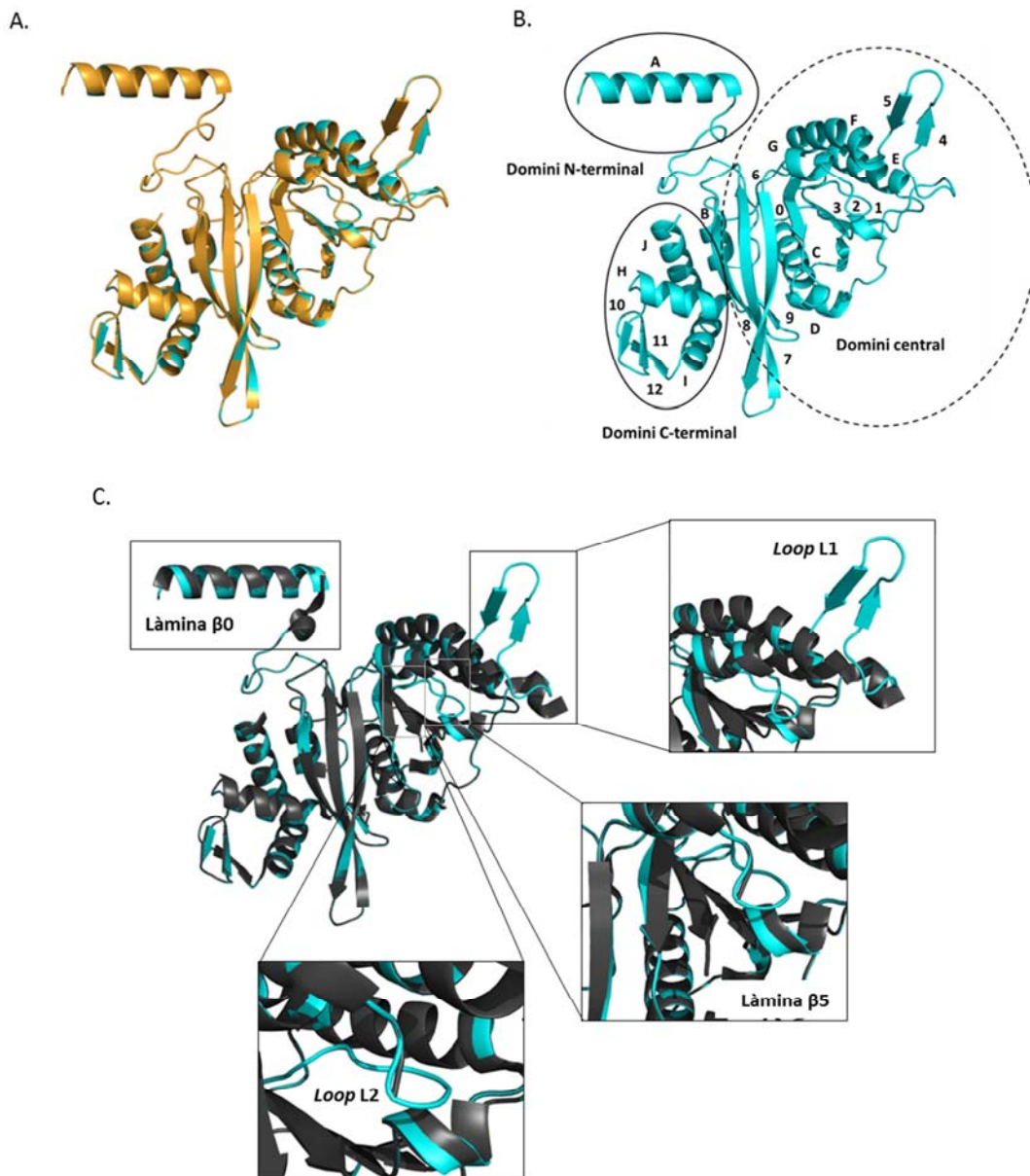


Figura 5.25 Predicció de les estructures 3D de la proteïna RecA de la soca tipus d'*A.bestiarum* CECT4227^T i de l'aïllat recombinant CECT5741 a partir del motlle 2REB. A) Superposició entre 5741_2REB (cian) i 4227_2REB (taronja). B) Localització dels dominis (cercles) i identificació de les hèlixs α i làmines β amb números i lletres, respectivament. C) Superposició entre 5741_2REB (cian) i el motlle d'*E.coli* 2REB (gris) amb ampliacions de les diferències observades.

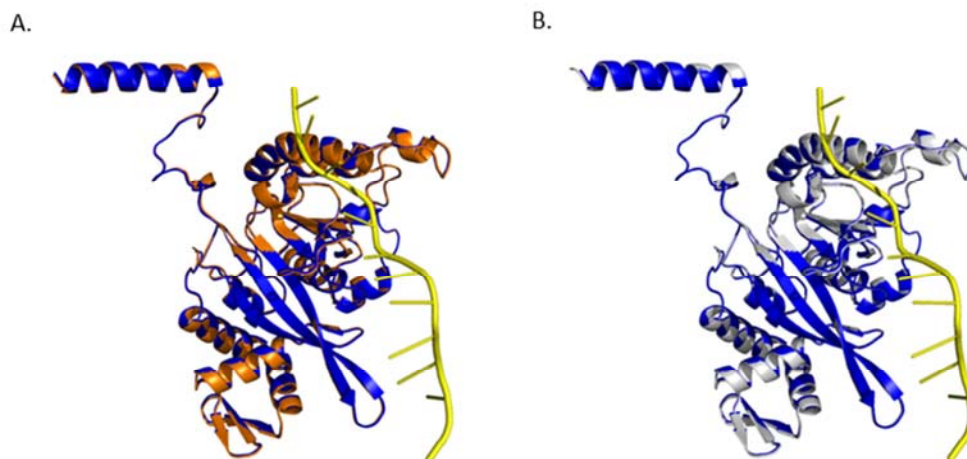


Figura 5.26 Predicció de les estructures 3D de la proteïna RecA de la soca tipus d'*A.bestiarum* CECT4227^T i de l'aïllat recombinant CECT5741 a partir del motlle 3CMW. A) Superposició de 5741_3CMW (blau) i 4227_3CMW (taronja). B) Superposició de 5741_3CMW (blau) i el motlle d'*E.coli* 3CMW (gris). En groc es mostra la cadena de DNA.

Finalment, també s'ha realitzat una comparació dels models tridimensionals de RecA de la soca recombinant obtinguts a partir dels motlles d'*E.coli* sense DNA (2REB) i amb DNA (3CMW). A la Figura 5.27 es pot observar la superposició d'aquestes dues estructures (Figura 5.27C) i les diferències observades que afecten a l'extrem N-terminal de la proteïna, al domini central i als dominis d'unió al DNA. Pel que fa a l'extrem N-terminal s'ha observat com l'orientació en l'espai d'aquest domini és diferent segons l'estructura considerada, ja que en la predicció 5741_3CMW (blau) semblaria estar més obert, i a més s'observa el que podria ser una petita làmina β . També s'han detectat diferències als *Loops* L1 i L2 relacionats com llocs d'unió al DNA. En relació a la quantitat de làmines β en el domini central de la proteïna, s'ha trobat una diferència significativa, l'estructura generada en absència de DNA (cian) s'ha predit 9 làmines β , i en presència de DNA (blau) un total de 10.

No ha estat possible realitzar una predicció de l'estructura tridimensional de l'extrem C-terminal de la proteïna RecA corresponent a la regió recombinant en cap predicció realitzada ja que les estructures d'*E.coli* que s'han utilitzat de motlle tenen aquesta regió final desordenada (2REB) o van ser cristal·litzades sense aquesta regió final (3CMW i 3CMT), i per tant, no presenten cap estructura. No obstant, s'ha considerat oportú utilitzar els models 3CMW i 3CMT com a motlle, per determinar si la presència del DNA afecta a l'estructura.

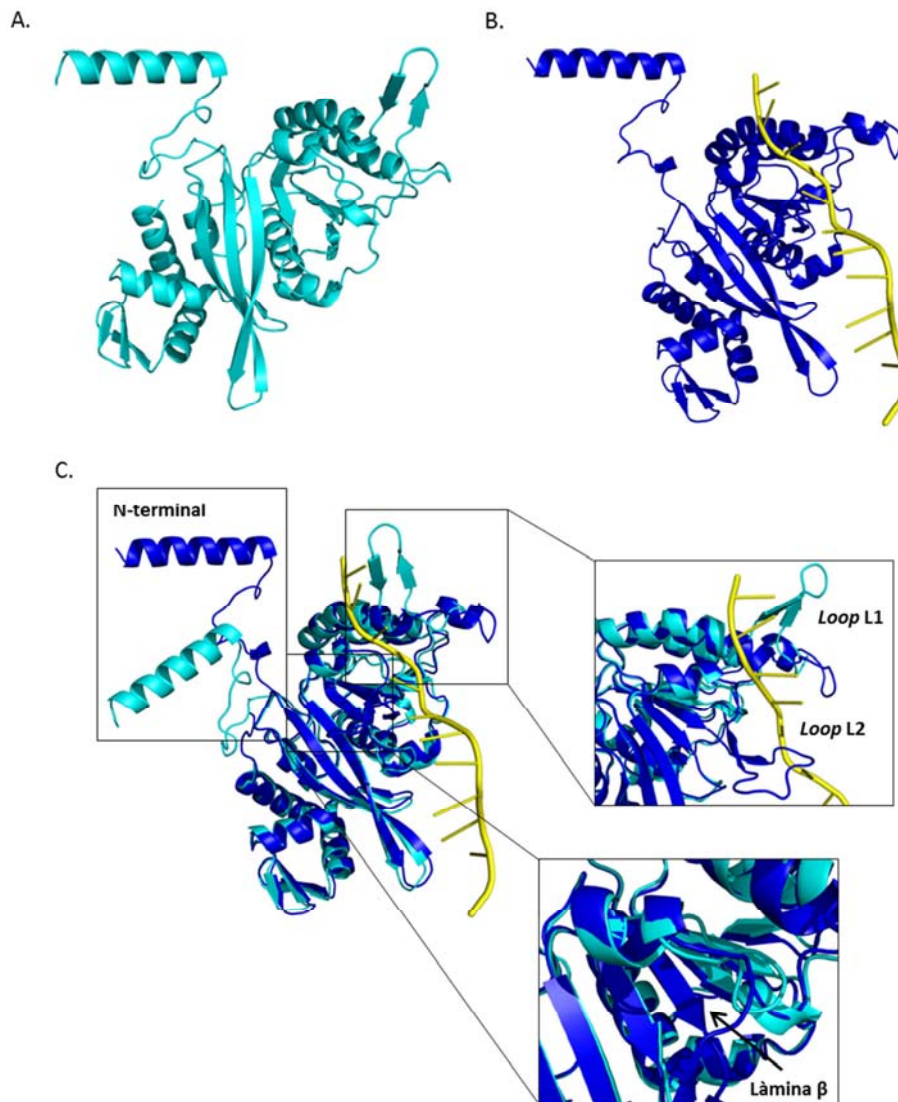


Figura 5.27 Predicció de les estructures 3D de la proteïna RecA de la soca recombinant CECT5741 a partir dels motlles 2REB i 3CMW. A) Predicció de 5741_2REB (cian). B) Predicció de 5741_3CMW (blau). C) Superposició de 5741_2REB (cian) i 5741_3CMW (blau) amb ampliacions de les diferències observades. En groc es mostra la cadena de DNA.

5.2.1.3.2.2 Model tridimensional de RecA de la soca recombinant *A.bestiarum* CECT5741

Per completar l'estudi, s'ha realitzat una caracterització dels dominis estructurals i funcionals més rellevants de la proteïna RecA de la soca recombinant a partir del motlle d'*E.coli* amb DNA (3CMW), per veure si afectaria la presència del fragment recombinant.

A la Figura 5.28A s'assenyalen amb cercles els 3 dominis estructurals (N-terminal, central, C-terminal) que presenta l'estructura de RecA de la soca recombinant i la identificació de les hèlixs α i les làmines β amb números i lletres, respectivament. En relació al domini N-terminal, que juga un paper molt important en la interacció proteïna-proteïna, en *E.coli* es situa entre les posicions 1-33 i consta d'una làmina β i una hèlix α , igual que la predicció obtinguda. El domini central és una regió molt conservada entre les diferents espècies bacterianes, i és on es troben

els llocs d'unió a l'ATP i al DNA. En *E.coli*, aquest motiu comença a la posició 34 i finalitza a la posició 269 i està formada per 8 làmines β antiparal·leles flanquejades per 4 hèlixs α , mentre que en el model de la soca recombinant (5741_3CMW) aquest domini central inclou 7 hèlixs α i 10 làmines β . Finalment, el domini C-terminal, que en *E.coli* es troba entre les posicions 270-352 amb 3 hèlixs α i 2 làmines β , la predicció de l'estructura obtinguda mostra que estaria formada per 3 hèlixs α i 3 làmines β .

Pel que fa als dominis funcionals de RecA, destaquen principalment dos: el domini d'unió a ATP i el d'unió a DNA. Així mateix, cadascun d'aquests dominis es divideix en diferents subdominis que tenen una funció molt concreta. També s'han identificat altres residus relacionats amb algunes de les funcions més importants.

A la Figura 5.28B es mostren els dominis funcionals d'unió a ATP (Walker A, Walker B i MAW) de la predicció de la proteïna RecA de la soca recombinant, i també la comparació de la seqüència dels tres motius de la soca recombinant i d'*E.coli*. En relació al motiu Walker A (vermell), en *E.coli* es localitza a la regió compresa entre els residus 47-74, està descrit que es caracteritza per tenir la seqüència consens GPESGKT entre les posicions 66-73, i els aminoàcids Gly₆₆, Gly₇₁, Lys₇₂ i Thr₇₃ semblen ser imprescindibles per al manteniment de la funció. En la seqüència de la soca recombinant, el motiu Walker A presenta la mateixa seqüència consens que *E.coli*, però està situat en posicions lleugerament diferents (64-71). El segon domini d'unió a ATP, el motiu Walker B (verd), es tracta d'una regió més petita que es troba entre les posicions 140-150 de la seqüència d'*E.coli*, en la que la regió funcional correspon concretament a 6 residus localitzats entre les posicions 139-144. En el cas d'*A.bestiarum*, aquest domini presentaria la mateixa seqüència que *E.coli* però es localitzaria a la posició 137-142, adoptant una conformació de làmina β . L'últim dels 3 motius d'unió a ATP és el MAW ó *Make ATP Work* (groc), que en la seqüència d'*E.coli* correspon a una regió de 24 residus (posicions 42-65), 8 dels quals (Ala₅₀, Gly₅₂, Gly₅₄, Gly₅₅, Pro₅₇, Gly₅₉, Arg₆₀ i Glu₆₃) estan altament conservats entre les diferents seqüències de RecA bacterianes. En la predicció de la proteïna RecA de la soca recombinant, aquest domini es situaria entre les posicions 40-63, presentant una seqüència lleugerament diferent a *E.coli*, a excepció dels 8 residus invariants, i adoptaria una estructura d'hèlix α i làmina β .

A la Figura 5.28C s'assenyalen els dos dominis d'unió al DNA en la predicció de RecA de la soca recombinant, anomenats com *Loop* L1 o lloc d'unió secundari i *Loop* L2 o lloc d'unió primari (taronja). La posició del *Loop* L1 (rosa) és lleugerament diferent entre *E.coli* i *A.bestiarum*, situant-se concretament entre les posicions 157-164 i 155-162, respectivament. Al comparar les seqüències del *Loop* L1 entre l'*E.coli* i la soca recombinant, s'observen dos canvis aminoacídics, tot i que es tracta d'una regió altament conservada. Pel que fa al *Loop* L2 o lloc d'unió al DNA primari (taronja), en *E.coli* es situa entre les posicions 195-209, mentre que per la soca recombinant aquest domini està en les posicions 193-207, i al comparar les seqüències presenta una única diferència (canvi d'una asparagina per serina). La predicció obtinguda també mostra com aquesta regió adoptaria una conformació de dues làmines β .

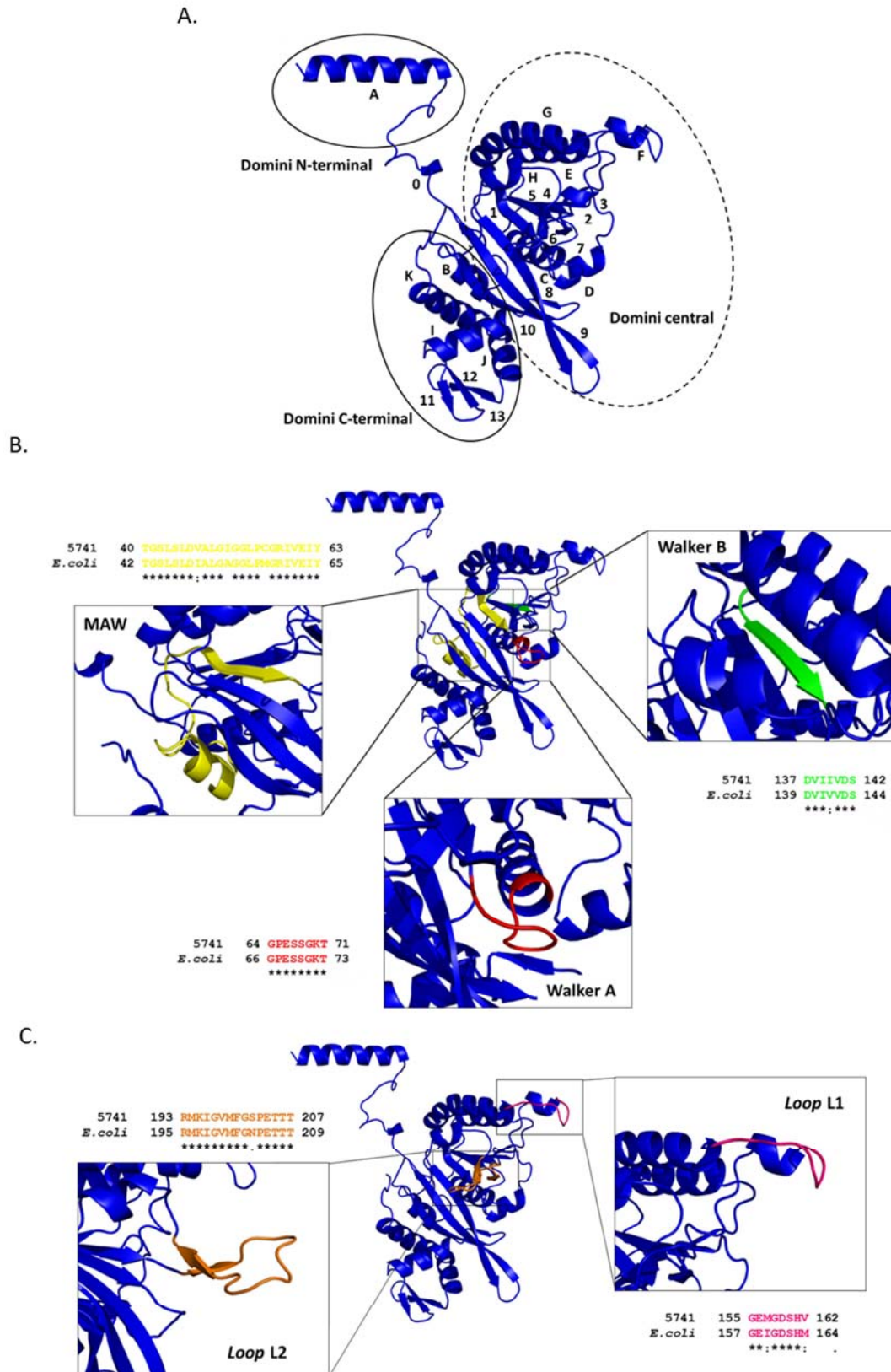


Figura 5.28 Caracterització dels dominis estructurals i funcionals de la proteïna RecA de la soca recombinant CECT5741 a partir del motlle 3CMW. Predicció de 5741_3CMW (blau). A) Dominis estructurals. B) Dominis funcionals d'unió a ATP. C) Dominis funcionals d'unió al DNA.

També, s'ha analitzat la localització d'altres residus importants relacionats amb la interacció amb molècules d'ATP, de DNA i entre subunitats en la predicció de RecA de la soca recombinant (Figura 5.29). En el cas d'*A.bestiarum* (Figura 5.29A), presenta la mateixa seqüència consens que *E.coli*, està situat pràcticament igual (posicions 93-98) i conté els mateixos residus Ala₉₃ (violeta), Glu₉₄ (verd) i Asp₉₈ (groc) que jugarien un paper important en la interacció amb molècules d'ATP. En segon lloc, per la proteïna RecA d'*E.coli* també cal destacar dos residus (Tyr₁₀₃ i Lys₁₈₃) i una seqüència consens (posicions 233-243) implicats en la interacció de la proteïna amb el DNA, que es localitzen fora dels dos dominis propis d'unió a DNA. A la Figura 5.29B es pot observar la predicció de la soca recombinant, que presenta aquests dos residus importants Tyr₁₀₁ (gris) i Lys₁₈₁ (vermell) situats en una posició lleugerament diferent, i la seqüència consens que té tres canvis aminoacídics respecte la d'*E.coli*.

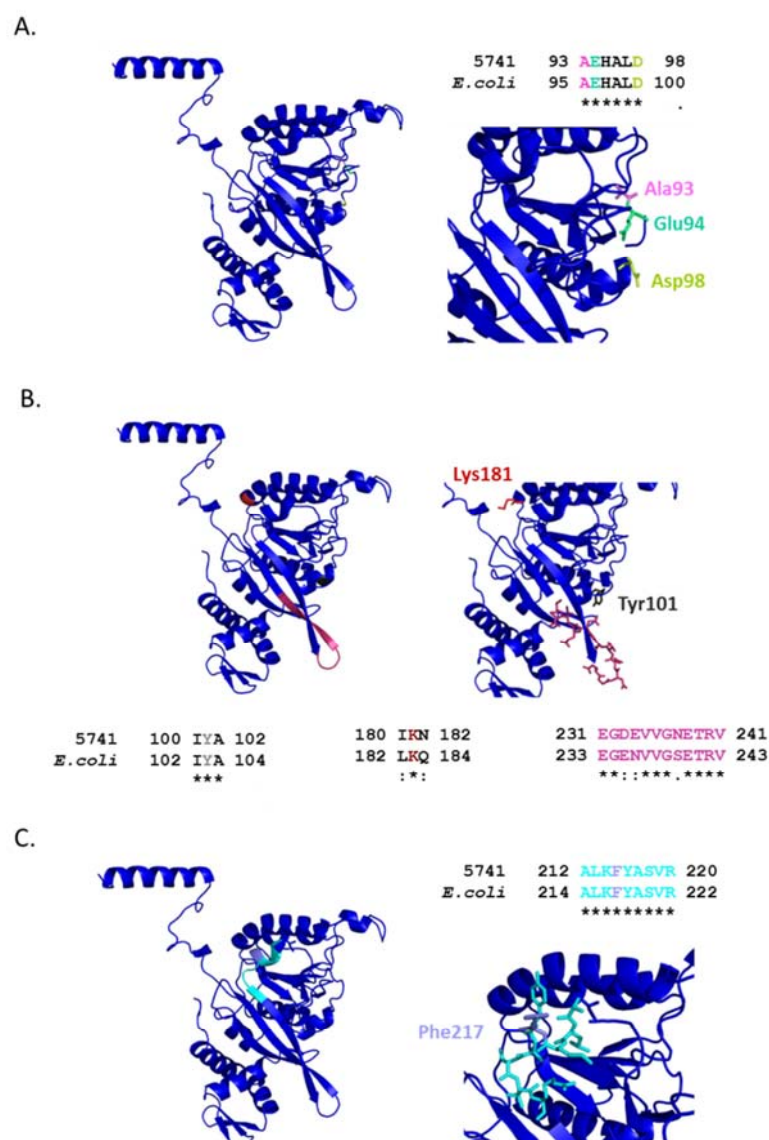


Figura 5.29 Localització d'altres residus i dominis amb importància funcional de la proteïna RecA de la soca recombinant CECT5741 a partir del motlle 3CMW. A) Implicats en la unió a ATP. B) Implicats en la unió al DNA. C) Implicats en el contacte entre subunitats.

Una altra de les característiques rellevants de la proteïna RecA que també s'ha de tenir en compte és el seu caràcter oligomèric, ja que s'ha descrit en *E.coli* un domini altament conservat (214-222) que estaria implicat en aquesta funció, i que presenta la seqüència consens $_{214}\text{Ala-Leu-Lys-Phe-[Phe/Tyr]-[Ala-Ser]-[Asp/Ser/Thr]-[Ile/Leu/Met/Gln/Val]-Arg}_{222}$. D'aquests residus, el més important és la Phe₂₁₇ i és el que menys tolerància mostra als canvis, ja que només pot ser substituït per una tirosina si es vol mantenir la funció. A la Figura 5.29C es mostra que aquesta seqüència consens en la predicció de la soca recombinant (212-220) que és idèntica a la d'*E.coli*.

Finalment, sobre l'extrem C-terminal de RecA on es troba el punt d'inici del fragment recombinant, actualment no es disposa de cap estructura de la proteïna RecA d'*E.coli* que presenti aquesta regió final resolta. Aquesta regió sempre apareix desordenada degut a la gran mobilitat que impedeix la seva cristal·lització. No obstant, es coneix la seva seqüència.

L'anàlisi detallada de l'extrem C-terminal a partir de l'alineament dels 25 últims aminoàcids de les seqüències de la soca tipus d'*A.bestiarum* (CECT4227^T), de l'aïllat recombinant (CECT5741) i d'*E.coli* ha mostrat que, malgrat tots tres tenen seqüències diferents, presenten una gran quantitat d'aminoàcids amb càrrega negativa (Figura 5.30A). Concretament, en la seqüència d'*E.coli* hi ha 7 aminoàcids carregats negativament (3 àcids glutàmics (D) i 4 àcids aspàrtics (E)) i les dues soques d'*A.bestiarum* un total de 10 (3 àcids aspàrtics (E) i 7 àcids glutàmics (D)) però en posicions diferents.

A.

```

A.bestiarum CECT4227T  GDAPAEKPVAAADAELEAESEQEFE
A.bestiarum CECT5741  GAAPVDDKPAAI EADEFEP EGEELE
E.coli              LSNPNSTP D FSVDDSEGV AETNE D F

```

B.

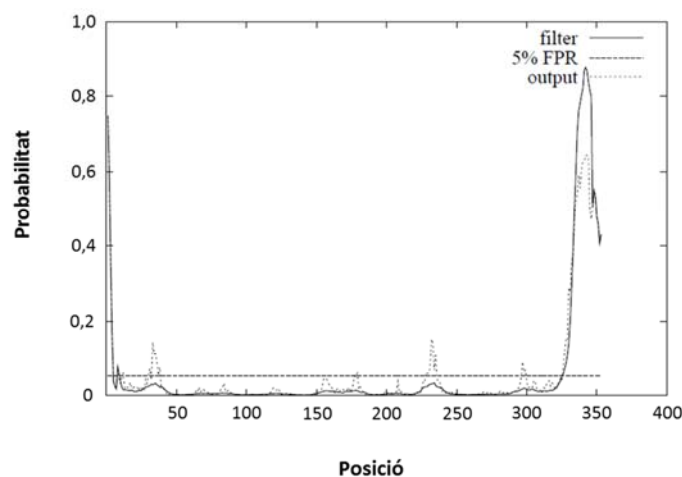


Figura 5.30 Anàlisi de la seqüència de la regió C-terminal de la proteïna RecA de la soca recombinant *A.bestiarum* CECT5741. A) Alineament corresponent als últims 25 aminoàcids. B) Representació gràfica del grau de desordre al llarg de tota la seqüència.

Com que el fragment recombinant localitzat en l'extrem C-terminal no s'ha pogut visualitzar en les prediccions d'estructures d'aquest estudi perquè el motlle utilitzat presenta una regió desordenada o perquè no està cristal·litzat, s'ha fet una anàlisi del grau de desordre de la seqüència per comprovar-ho. A la Figura 5.30B es mostra el patró de desordre obtingut al llarg de tota la seqüència de la proteïna RecA de la soca recombinant amb el programa DISOPRED on s'ha observat com aquesta regió final presenta una elevada probabilitat d'estar desordenada. S'han obtingut resultats idèntics a l'analitzar la seqüència de la soca tipus d'*A.bestiarum*.

5.5.1.4 Validació del model

La validació dels models tridimensionals obtinguts per predicció s'ha realitzat amb dos programes diferents: el ProQ que determina la qualitat a través dels paràmetres LGScore i el MaxSub, i el programa SWISS-MODEL, que per fer-ho estima dos paràmetres, QMEAN i GMQE (Apartat 3.18.1).

Els resultats obtinguts per totes les estructures de la proteïna RecA de les soques tipus i recombinant d'*A.bestiarum* amb els 3 motlles d'*E.coli* han estat molt coincidents, presentant valors de LGScore entre 4,57 i 4,60 i valors de MaxSub entre 0,40 i 0,41 que indiquen que els models obtinguts són bons (Apartat 3.18.1). Els resultats de la validació de models amb el programa SWISS-MODEL també han estat bons, amb valors de QMEAN4 entre 0,73 i 0,78 i valors de GMQE de 0,80 i 0,81 (Apartat 3.18.1).

5.5.1.5 Llocs d'unió dels lligands i perfils d'hidrofobicitat

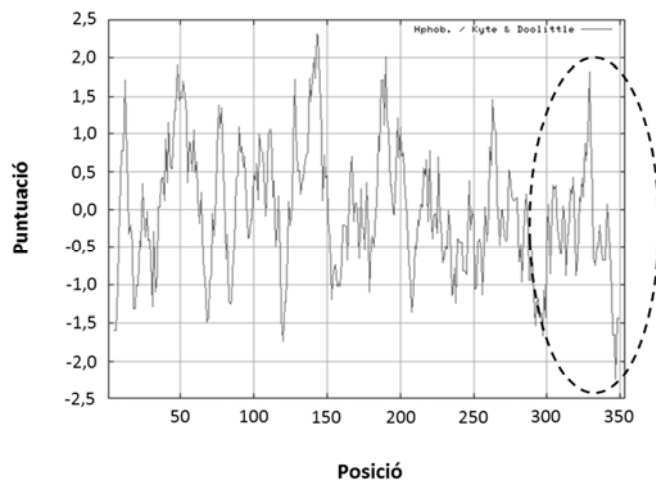
Per completar aquest estudi, també s'ha realitzat una anàlisi amb el programa 3DligandSite per caracteritzar quins dels aminoàcids que conformen la proteïna podrien estar en contacte amb diferents lligands. Aquest programa ha identificat 22 aminoàcids (Taula 5.18) que estarien en contacte amb cofactors com l'ADP i ions com el zinc, el magnesi i el calci, a una distància entre 0-0,67 Àngstroms. En cap cas afecta a aminoàcids de l'extrem C-terminal de la proteïna RecA, i per tant, l'adquisició del fragment recombinant no tindria cap influència en els llocs d'unió actius de lligands.

Taula 5.18 Identificació dels possibles aminoàcids implicats en la unió a diferents lligands.

Lligand	Aminoàcid
ADP / Mg	Ser ₆₇ , Ser ₆₈ , Gly ₆₉ , Lys ₇₀ , Thr ₇₁ , Thr ₇₂ , Asp ₉₈ , Tyr ₁₀₁ , Asn ₂₃₈ , Ile ₂₆₀ , Tyr ₂₆₁ , Gly ₂₆₃
ADP / Zn	Lys ₂₄₆ , Asn ₂₄₇ , Lys ₂₄₈ , Ala ₂₅₀ , Pro ₂₅₁ , Pro ₂₅₂
Zn	Ile ₂₂₆ , Glu ₂₃₉
Mg	Glu ₉₄
Ca	Asn ₁₃₄

També, la caracterització del caràcter hidrofòbic de les seqüències proteiques permet tenir indicis sobre la seva topologia, i té utilitat en la predicció de regions que estan probablement exposades a la superfície. Una de les escales d'hidrofobicitat més emprades és la de Kyte-Doolittle, en la que les regions amb valors inferiors a 0 es consideren de caràcter hidrofòbiques. La Figura 5.31 mostra el perfil d'hidrofobicitat per la soca recombinant, on s'observa que la major part dels aminoàcids de l'extrem C-terminal i, sobretot la part final, presenten valors inferiors a 0 en aquesta escala, i per tant, indicant que una part important d'ells són hidrofòbics.

Figura 5.31 Representació gràfica del perfil d'hidrofobicitat de la seqüència de la soca recombinant *A.bestiarum* CECT5741 utilitzant el programa ProtScale.



5.6 DATACIÓ

Per conèixer l'origen de l'ancestre comú de les espècies actuals del gènere *Aeromonas* s'ha fet un estudi de datació basat en les seqüències del gen *recA* complet. Per tal de realitzar-lo, s'han utilitzat les mateixes aproximacions (a partir de les 150 seqüències, de les 27 seqüències consens, de les 27 seqüències de les soques tipus, i de l'arbre d'espècies), metodologies (cronogrames construïts amb el programa BEAST amb un model de rellotge molecular relaxat), i punt de calibratge (140 Ma per *E.coli* i *S.typhimurium*) que en el cas del gen *mdh* (Apartat 4.4).

5.6.1 A partir de les 150 seqüències

A partir del conjunt de seqüències del gen *recA* de l'estudi s'ha generat el corresponent cronograma, utilitzant el model GTR+G+I com el millor model de substitució que s'ajusta a les dades. A la Figura 5.32 es mostra el cronograma obtingut, el qual presenta uns valors molt alts de probabilitat posterior en els nodes dels grups principals, que recolzen la robustesa de la filogènia. Els temps de divergència de l'ancestre del gènere *Aeromonas* i de les principals agrupacions s'indiquen a la Taula 5.19.

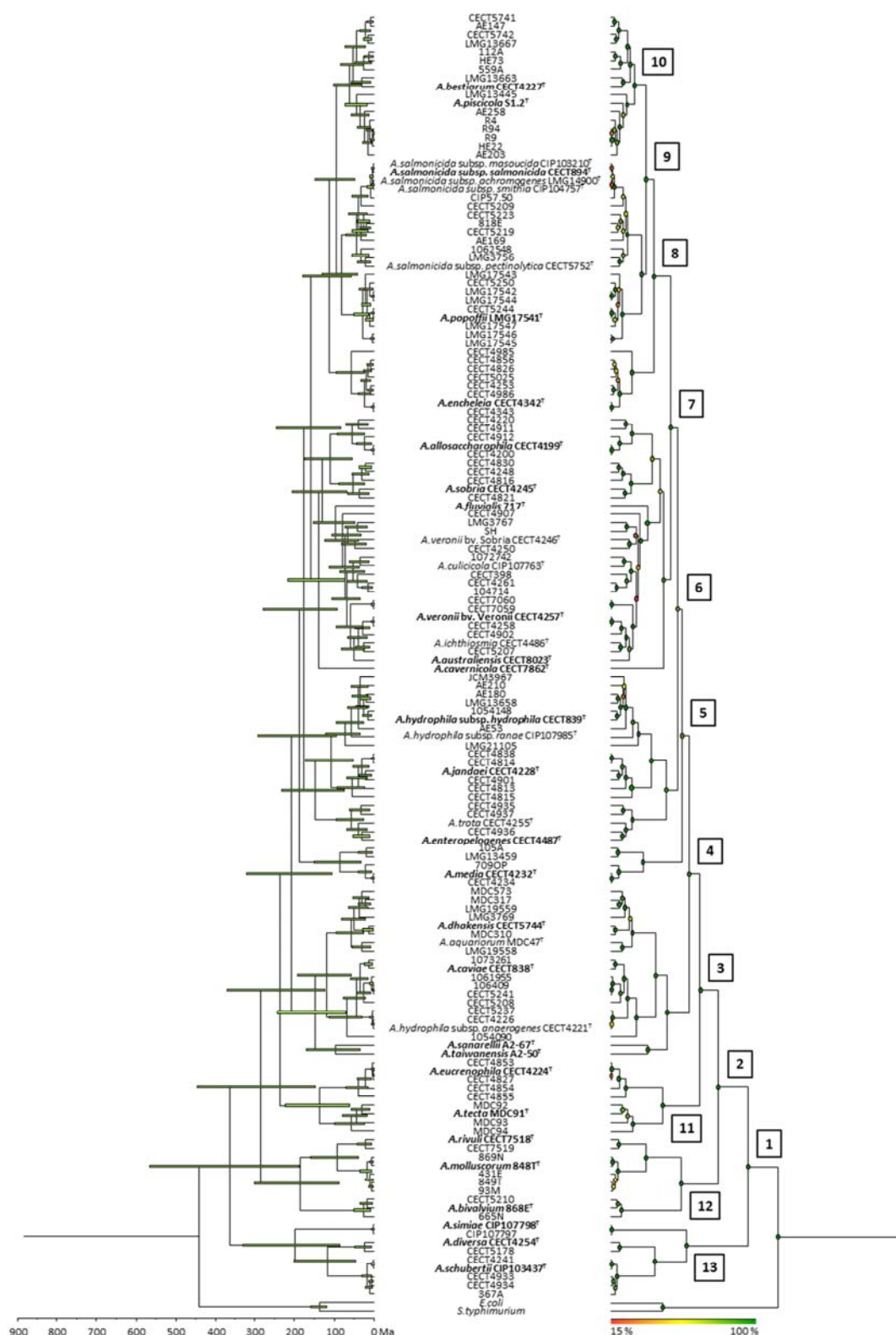


Figura 5.32 Cronograma del gènere *Aeromonas* basat en les 150 seqüències del gen *recA*. A l'esquerra es mostra el cronograma a escala temporal (Ma), i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 5.19 Edat de divergència estimada dels principals grups del cronograma obtingut a partir de les 150 seqüències del gen *recA*.

Clúster	Edat	IC inferior	IC superior
1	362,7	188,7	566,3
2	284,6	148,0	445,3
3	236,1	124,1	369,5
4	207,9	105,5	320,7
13	199,3	86,7	329,9
5	188,8	95,6	293,0
12	186,5	87,9	302,1
6	177,5	93,2	278,3
7	157,8	82,4	246,7
11	136,8	61,7	224,4
8	113,1	56,5	178,5
9	93,6	47,5	148,6
10	61,8	29,3	100,7

A partir de l'anàlisi realitzada amb les 150 seqüències del gen *recA*, s'ha pogut determinar que l'ancestre comú del gènere *Aeromonas* hauria aparegut fa aproximadament 363 Ma. Com en el cas del gen *mdh*, aquesta primera separació hauria originat un llinatge que conté la majoria d'espècies actuals, i un segon que ha donat lloc a 3 espècies *A.diversa*, *A.schubertii* i *A.simiae*. També s'han estimat les diversificacions de les espècies més recents: *A.bestiarum*-*A.piscicola* (61 Ma), *A.salmonicida*-*A.popoffii* (81 Ma), *A.molluscorum*-*A.bivalvium* (92 Ma), *A.fluvialis*-*A.veronii* (95 Ma) i *A.sanarellii*-*A.taiwanensis* (97 Ma). Per últim, s'ha determinat la taxa evolutiva del gen *recA* en $3,1 \times 10^{-4}$ substitucions/lloc/milió d'any amb una desviació estàndard de $2,5 \times 10^{-6}$.

5.6.2 A partir de les seqüències consens

Per realitzar l'anàlisi de datació s'ha construït un arbre bayesià a partir de les 27 seqüències consens de totes les espècies d'*Aeromonas*, seguint el millor model de substitució, el GTR+G+I. A la Figura 5.33 es mostra el cronograma obtingut, on s'observa que la majoria dels nodes presenten valors de probabilitat posterior superiors al 50%. L'edat estimada dels principals nodes i els seus intervals de confiança (IC) s'indiquen a la Taula 5.20. Els resultats obtinguts a partir de les seqüències consens suggereixen que l'ancestre comú del gènere *Aeromonas* s'hauria originat fa aproximadament 356 Ma, i les diversificacions més recents es van produir quan es van separar les espècies *A.bestiarum*-*A.piscicola* (23 Ma), *A.australiensis*-*A.veronii* (27 Ma) i *A.sanarellii*-*A.taiwanensis* (42 Ma).

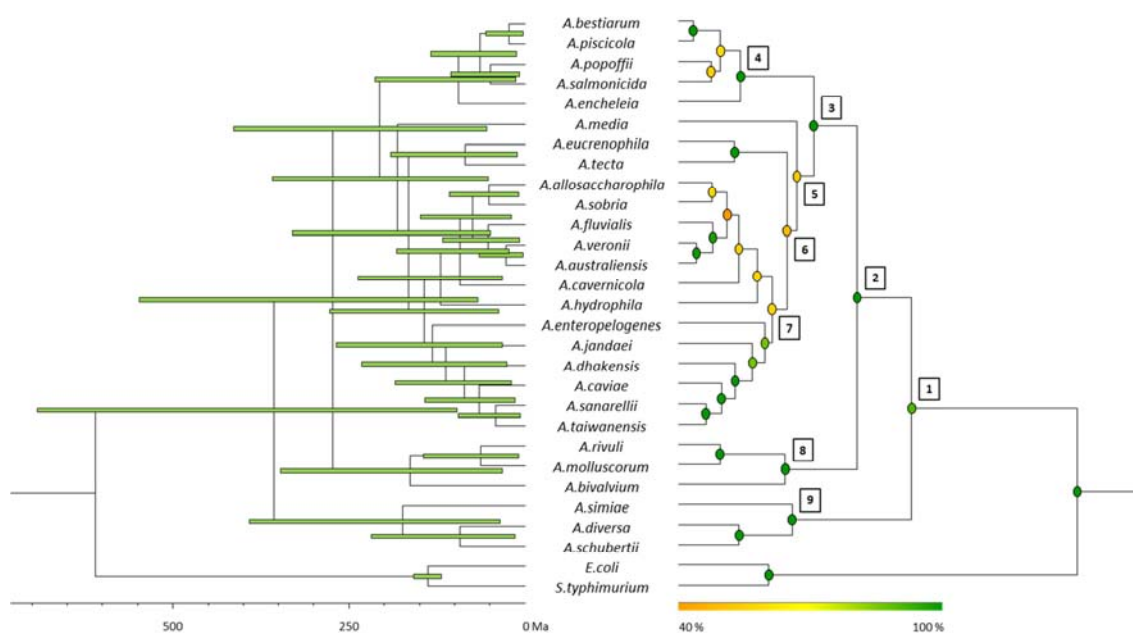


Figura 5.33 Cronograma del gènere *Aeromonas* basat en les seqüències consens del gen *recA* de les diferents espècies. A l'esquerra es mostra el cronograma a escala temporal (Ma) i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 5.20 Edat de divergència estimada dels principals grups del cronograma obtingut a partir de les seqüències consens del gen *recA*.

Clúster	Edat	IC inferior	IC superior
1	356,3	96,2	692,6
2	273,0	67,6	548,0
3	206,6	54,5	413,6
5	180,9	52,8	358,5
9	173,8	35,2	391,7
6	165,8	49,9	329,8
8	163,1	32,0	346,8
7	142,8	37,9	277,0
4	94,8	13,5	212,4

5.6.3 A partir de les seqüències de les soques tipus

A partir de les seqüències del gen *recA* de les soques tipus de totes les espècies d'*Aeromonas*, s'ha construït un cronograma utilitzant el model GTR+G+I com a millor model de substitució, i el mateix punt de calibratge que en els casos anteriors (140 Ma). L'arbre bayesià datat (Figura 5.34) presenta en la majoria dels nodes valors de probabilitat posterior superiors al 50%. A la

Taula 5.21 es mostra un llistat d'edats estimades dels grups principals i els seus intervals de confiança (IC). En base a aquests resultats, la diversificació del gènere *Aeromonas* s'hauria produït fa aproximadament 373 Ma. També s'han estimat les diversificacions més recents que corresponen a les separacions de les espècies *A.bestiarum*-*A.piscicola* (24 Ma), *A.australiensis*-*A.veronii* (27 Ma) i *A.sanarellii*-*A.taiwanensis* (39 Ma).

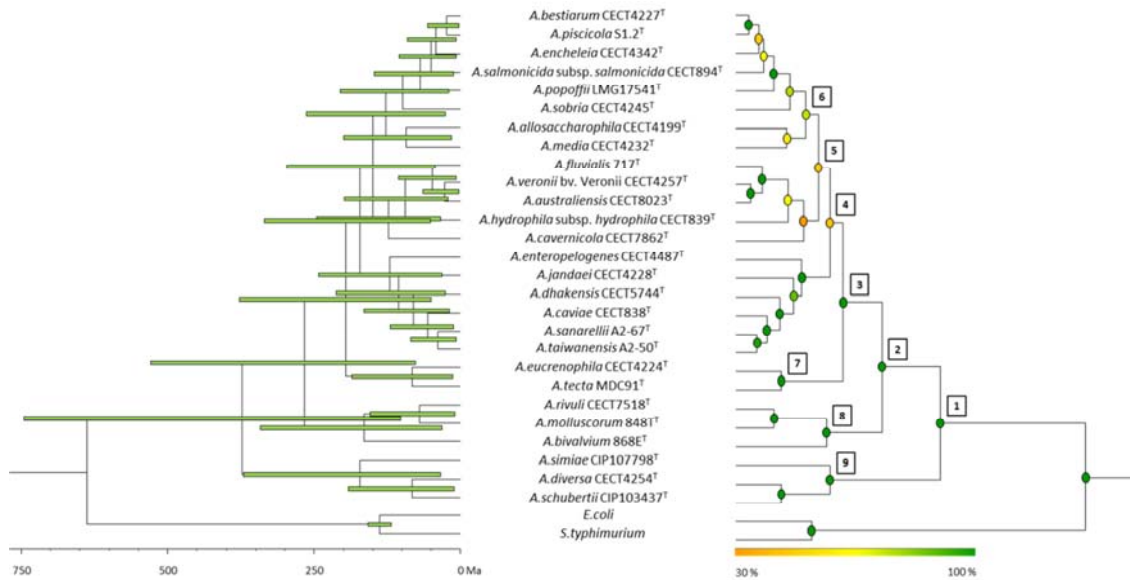


Figura 5.34 Cronograma del gènere *Aeromonas* basat en les seqüències del gen *recA* de les soques tipus. A l'esquerra es mostra el cronograma a escala temporal (Ma) i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 5.21 Edat de divergència estimada dels grups principals del cronograma obtingut a partir de les seqüències de les soques tipus del gen *recA*.

Clúster	Edat	IC inferior	IC superior
1	372,6	102,0	745,8
2	266,6	76,5	529,0
3	195,8	49,6	377,7
4	171,3	50,8	335,3
9	171,2	34,4	370,8
8	165,1	31,6	341,4
5	150,3	43,7	296,8
6	127,4	25,7	263,4
7	83,0	13,4	185,5

5.6.4 A partir de l'arbre d'espècies

Per últim, s'ha construït un arbre d'espècies datat a partir de les seqüències del gen *recA*, considerant el model GTR+G+I com a millor model d'evolució. El cronograma obtingut es mostra a la Figura 5.35, i una gran majoria de nodes principals presenten valors de probabilitat posterior superiors al 50%. Els temps de divergència estimats dels grups principals i els seus intervals de confiança (IC) s'indiquen a la Taula 5.22. Segons els resultats de l'arbre d'espècies del gen *recA*, l'origen del gènere *Aeromonas* es trobaria al voltant dels 348 Ma. Les diversificacions d'espècies més recents es van produir quan es van separar *A.australiensis*-*A.veronii* (27 Ma), *A.bestiarum*-*A.piscicola* (34 Ma), *A.salmonicida*-*A.popoffii* (40 Ma) i *A.sanarellii*-*A.taiwanensis* (41 Ma).

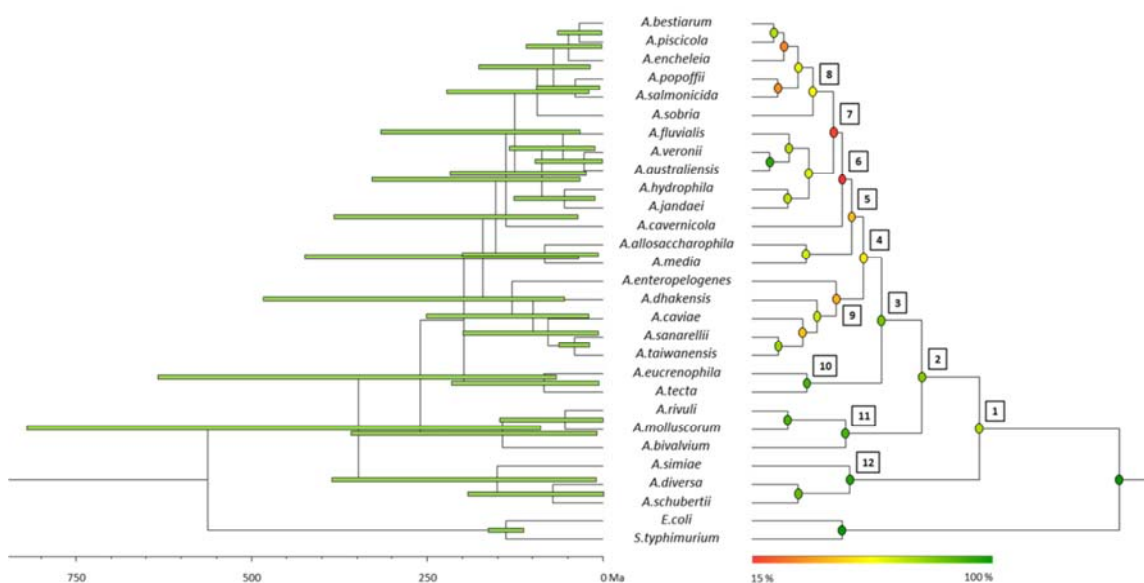


Figura 5.35 Cronograma de l'arbre d'espècies del gènere *Aeromonas* obtingut a partir de les seqüències del gen *recA*. A l'esquerra es mostra el cronograma a escala temporal (Ma) i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 5.22 Edat de divergència estimada dels grups principals de l'arbre d'espècies d'*Aeromonas* obtingut a partir de l'arbre d'espècies del gen *recA*.

Clúster	Edat	IC inferior	IC superior
1	348,0	96,1	822,4
2	260,3	69,1	630,1
3	198,3	56,1	485,6
4	170,8	45,5	427,1
12	150,2	15,5	381,1
5	152,8	42,0	381,5
11	143,2	11,7	359,4
6	138,4	37,0	329,8
9	129,3	24,2	318,3
7	125,1	34,5	315,2
8	93,1	22,9	233,4
10	84,2	7,7	215,7

5.7 ANÀLISI DE LA DIVERSIFICACIÓ

Com ja s'ha comentat en els resultats presentats pel gen *mdh* (Apartat 4.5), per realitzar l'anàlisi del model temporal de diversificació d'*Aeromonas* és necessari utilitzar un arbre ultramètric datat generat amb una única seqüència per cada espècie. Pel gen *recA* s'ha utilitzat la mateixa metodologia emprada pel gen *mdh*, construint cronogrames per inferència bayesiana a partir de les mateixes aproximacions (a partir de les 27 seqüències consens, de les 27 seqüències de les soques tipus, i de l'arbre d'espècies). En tots els casos, s'ha determinat el model i la taxa de diversificació del gènere *Aeromonas*, per després poder fer un estudi comparatiu dels resultats obtinguts per ambdós gens (Apartat 6.7).

5.7.1 A partir de les seqüències consens

Per fer una primera aproximació del model de diversificació s'ha elaborat un gràfic LTT plot (Figura 5.36A), que ha evidenciat un patró constant al llarg temps, amb un bon ajust lineal de la recta de regressió (Figura 5.36B). A partir del pendent de la recta de regressió s'ha obtingut un valor estimat de λ de 0,0088. També, a partir del cronograma i ajustant les dades a un model constant mitjançant una prova de màxima versemblança, s'ha obtingut un valor de taxa de diversificació (λ) que en aquest cas ha estat de 0,0074 (SE = 0,0011) i una taxa d'extinció $\mu = 0$, que indicaria un model de Yule.

A més, s'ha realitzat una simulació de 1000 filogènies, utilitzant el model de Yule, el valor empíric de λ i el nombre d'espècies analitzades, per avaluar si els resultats obtinguts són significatius. A partir dels 1000 cronogrames generats s'han obtingut els corresponents LTT plots (Figura 5.36C). En la simulació s'observa que l'LTT plot obtingut a partir de les dades

empíriques (línia contínua negra) es situa dins del 95% dels LTT plots generats per simulació (línies discontinues).

Per verificar que el model de diversificació s'ajusta a un model constant, també s'ha calculat l'estadístic γ de Pybus i Harvey, i s'ha obtingut un valor empíric de γ de -1,7817. En la Figura 5.36D es representa la distribució dels valors de γ obtinguts per simulació de 1000 filogènies seguint un model constant de Yule, on el valor empíric de γ (fletxa vertical) es troba dins els límits dels intervals de confiança del 95% (1,3431 i -2,3055).

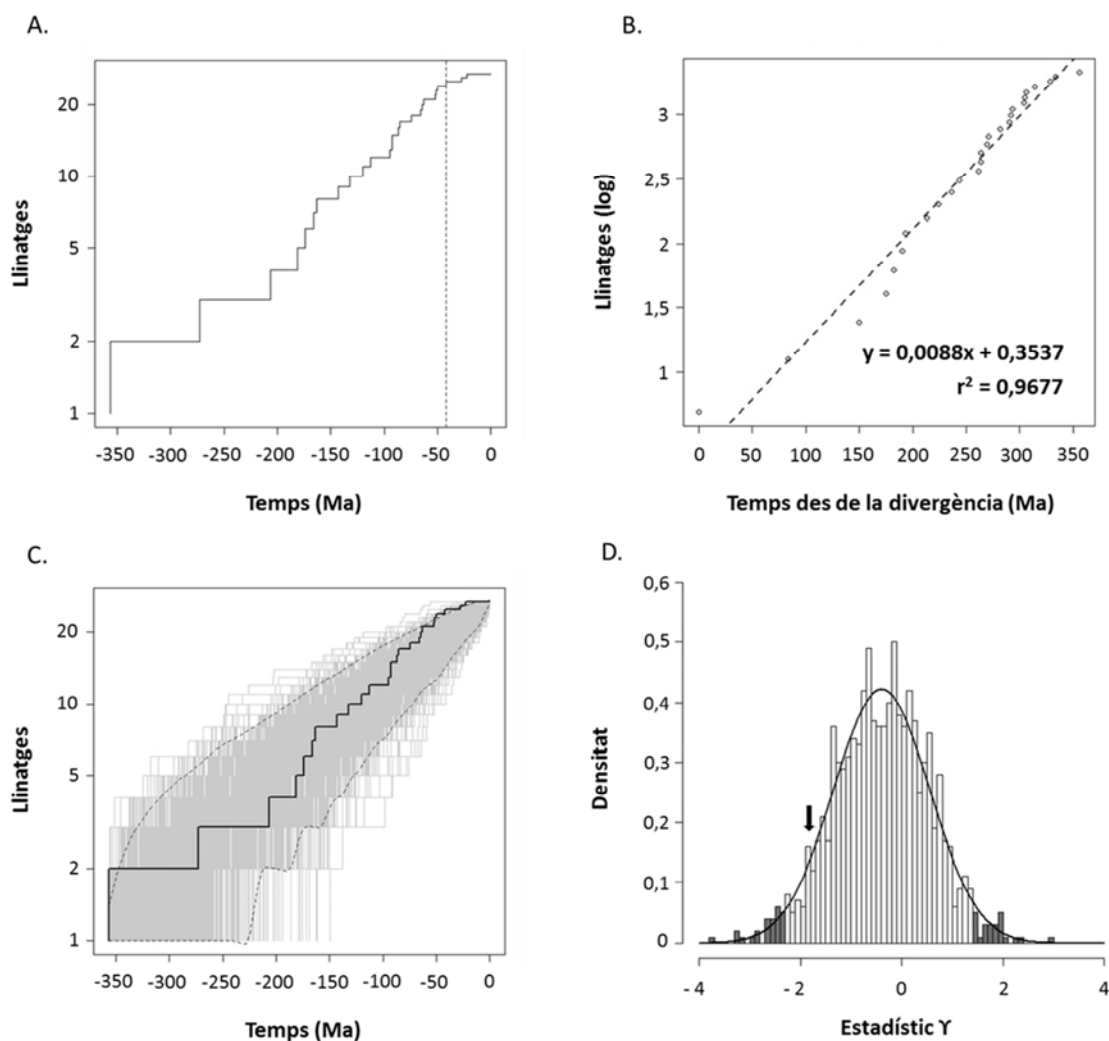


Figura 5.36 Anàlisi del model de diversificació del gènere *Aeromonas* basat en les seqüències consens del gen *recA*. A) LTT plot a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió del LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic γ de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

Segons l'anàlisi de selecció del model de diversificació (Figura 5.37A), el millor model amb taxa constant és el model de Yule mentre que el millor amb taxa variable és el Yule 2-rate. Mitjançant el test ΔAIC_{RC} s'ha determinat que el millor model de diversificació seria el Yule-2-

rate, el qual presenta dues taxes de diversificació 0,0102 (λ_1) i 0,0018 (λ_2), i el punt de canvi (*shift*) s'hauria produït fa aproximadament 36 Ma (línia discontinua de l'LTT plot, Figura 5.36A). S'ha estimat la fiabilitat del valor ΔAIC_{RC} obtingut (4,7810) per simulació amb el test de Rabosky (Figura 5.37B), obtenint un valor de P de 0,0313, amb el que cal rebutjar la hipòtesi nul·la de model de Yule ($P < 0,05$).

Figura 5.37 Determinació del millor model de diversificació del gènere *Aeromonas* basat en les seqüències consens del gen *recA*. A) Selecció del model i resultat del test ΔAIC_{RC} . B) Test de Rabosky.

A.

Selecció del model de diversificació					
Models constants			Models variables		
Model	<i>Birth-Death</i>	Yule	DDX	DDL	Yule 2-rate
AIC	176,3422	174,3422	175,2445	171,7184	169,5611

Millor model constant		Millor model variable	
Model	Yule	Model	Yule 2-rate
AIC	174,3422	AIC	169,5611

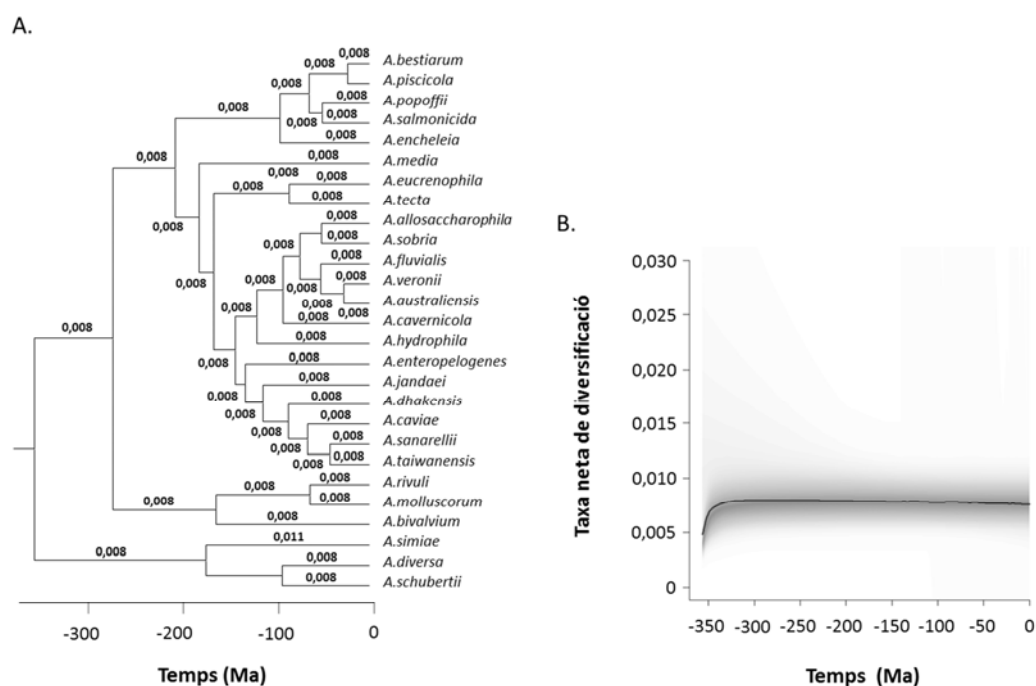
Test ΔAIC_{RC}	
Model	Yule 2-rate
AIC	169,5611
ΔAIC_{RC}	4,7810

B.

Test de Rabosky		
ΔAIC_{RC}	Valor de P	Decisió
4,7810	0,0313	$P < 0,05$ - Rebutjar hipòtesi nul·la (H_0) de model constant de Yule

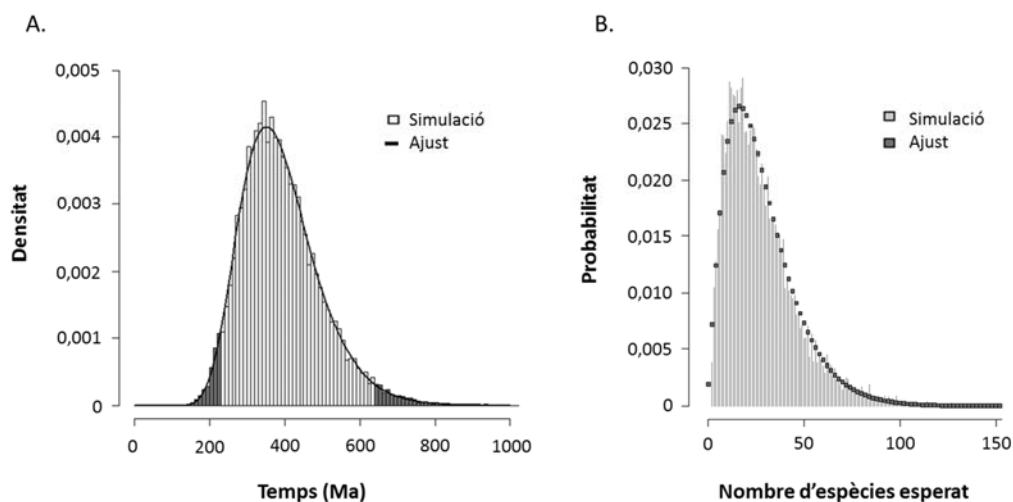
L'estudi s'ha completat amb una anàlisi bayesiana amb el programa BMM per detectar canvis de taxes de diversificació. A la Figura 5.38A es mostren els valors de les taxes de diversificació obtingudes per cada branca de l'arbre bayesià, que presenten un valor de λ constant de 0,008, excepte en un cas (0,011). Aquest resultat ha estat concordant amb el perfil constant obtingut en la Figura 5.38B, en el que es representa la taxa neta de diversificació al llarg del temps. També s'ha determinat quina és la probabilitat de que les dades segueixin un model sense *shift* en la taxa de diversificació o que presenti un ó mes *shifts*, estimant que hi ha una probabilitat de 0,99 que el procés es mantingui constant al llarg del temps, i de 0,0064 de que s'hagi produït un punt de canvi en el valor de λ . Finalment, s'ha calculat el factor de Bayes, com a mesura de suport estadístic de les taxes estimades, i s'ha obtingut un valor de 154,53 (> 100), que dóna una evidència decisiva de que no s'ha produït cap punt de canvi en la taxa de diversificació.

Figura 5.38 Resultats obtinguts amb el programa BAMM a partir de les seqüències consens del gen *recA*. A) Arbre bayesià on s'indiquen les taxes de diversificació per cada branca. B) Representació gràfica de la taxa neta de diversificació al llarg del temps.



Per últim, s'ha realitzat el càlcul teòric de l'origen i el nombre d'espècies actuals esperades del gènere *Aeromonas* si seguissin un model de diversificació de Yule (λ empírica = 0,0074 i $\mu = 0$). El valor tMRCA obtingut situaria l'origen d'aquest gènere fa 387 Ma (Figura 5.39A), valor pràcticament idèntic a l'empíric (356 Ma). En relació al nombre d'espècies esperades actualment des de l'origen d'aquest gènere (Figura 5.39B), s'ha obtingut un valor mig de 28 espècies (SD = 19).

Figura 5.39 Resultats teòrics obtinguts a partir del model de Yule amb les seqüències consens del gen *recA*. A) Histograma de l'edat de l'ancestre comú més recent (tMRCA). B) Distribució del nombre d'espècies esperat.



5.7.2 A partir de les seqüències de les soques tipus

En primer lloc, a partir del cronograma corresponent s'ha realitzat el gràfic LTT plot (Figura 5.40A), que evidencia un patró de diversificació temporal constant al llarg del temps. La recta de regressió obtinguda indica un bon ajust lineal (Figura 5.40B). El valor de λ estimat a partir del pendent de la recta de regressió ha estat de 0,0086. Fent servir una prova de màxima versemblança s'han ajustat les dades a un model constant, i s'han calculat els paràmetres $\lambda = 0,0075$ (SE = 0,0011) i $\mu = 0$. Els resultats obtinguts indiquen que seguiria un model de Yule.

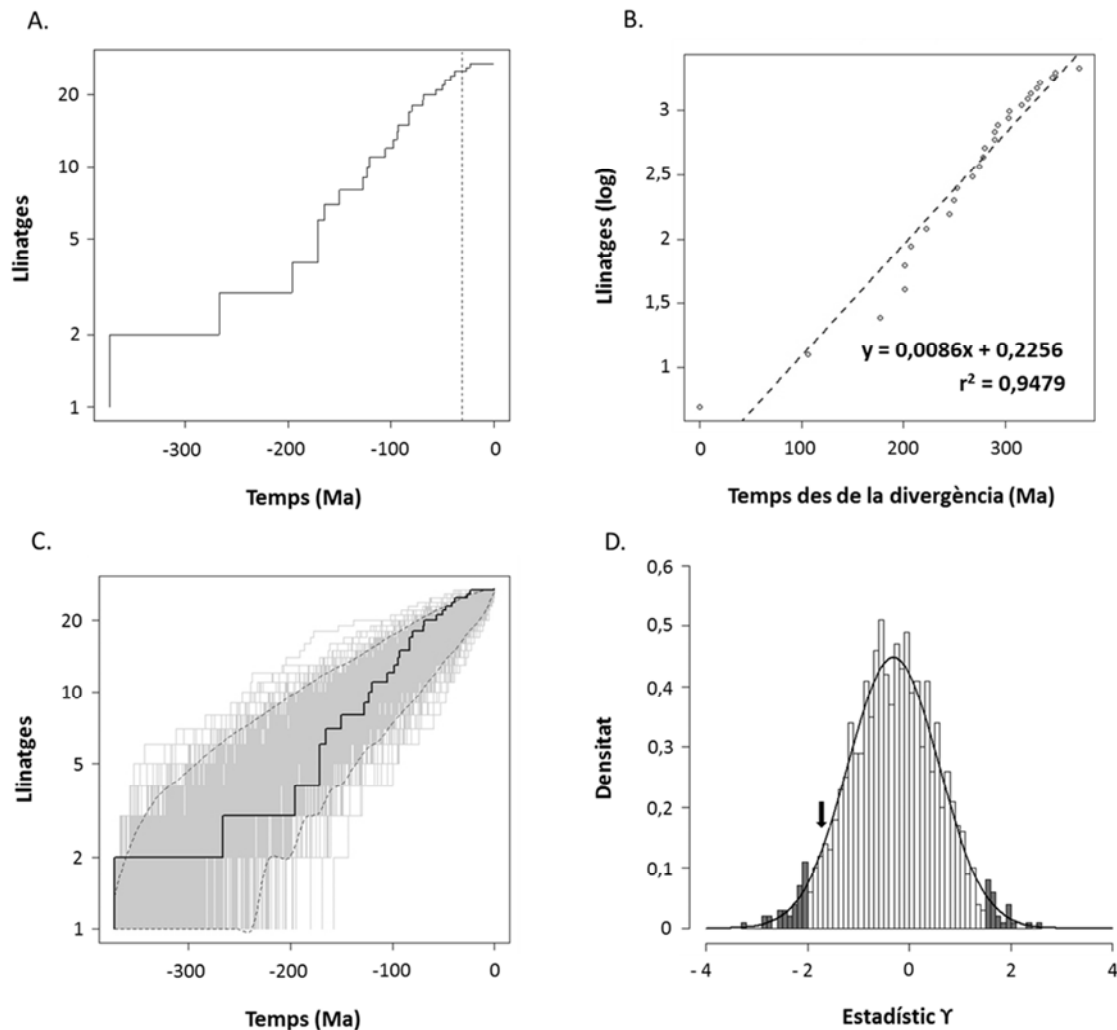


Figura 5.40 Anàlisi del model de diversificació del gènere *Aeromonas* basat en les seqüències del gen *recA* de les soques tipus. A) LTT plot a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió del LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic Y de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

Per comprovar la fiabilitat d'aquests resultats, s'ha realitzat una simulació de 1000 cronogrames, utilitzant el valor λ empíric i el model de Yule, per obtenir els corresponents LTT plots. A la Figura 5.40C es pot veure que l'LTT plot generat amb les dades empíriques (línia continua negra) es situa dins del 95% dels obtinguts a partir de les filogènies simulades (línies discontinues).

També s'ha calculat l'estadístic Υ de Pybus i Harvey per les dades analitzades, que ha sigut de -1,6593. En la distribució de valors de Υ (Figura 5.40D) s'assenyala el valor empíric amb una fletxa vertical, el qual es troba dins de l'interval de confiança del 95% (1,5412 i -2,1504).

Posteriorment s'ha analitzat quin és el millor model de diversificació (Figura 5.41A) que s'ajusta a les dades empíriques. Primer, mitjançant una prova de màxima versemblança, s'ha estimat el millor model amb taxa de diversificació constant, que ha estat el model de Yule, i el millor amb taxa variable, en aquest cas, el model Yule-2-rate. Després s'ha realitzat el test ΔAIC_{RC} que ha determinat que el millor model dels dos seleccionats és el de Yule-2-rate. Els dos valors de λ obtinguts han estat 0,0101 (λ_1) i 0,0019 (λ_2), i el punt de canvi s'hauria produït fa aproximadament 39 Ma (línia vertical discontinua en l'LTT plot, Figura 5.40A). Per avaluar si el resultat obtingut és significatiu, s'ha realitzat el test de Rabosky (Figura 5.41B), obtenint un valor de P (0,0561) que permet acceptar la hipòtesi nul·la de model de Yule.

A.

Selecció del model de diversificació					
Models constants			Models variables		
Model	<i>Birth-Death</i>	Yule	DDX	DDL	Yule 2-rate
AIC	175,6124	173,6124	174,9036	171,514	169,7764

Millor model constant		Millor model variable	
Model	Yule	Model	Yule 2-rate
AIC	173,6124	AIC	169,7764

Test ΔAIC_{RC}	
Model	Yule 2-rate
AIC	169,7764
ΔAIC_{RC}	3,8359

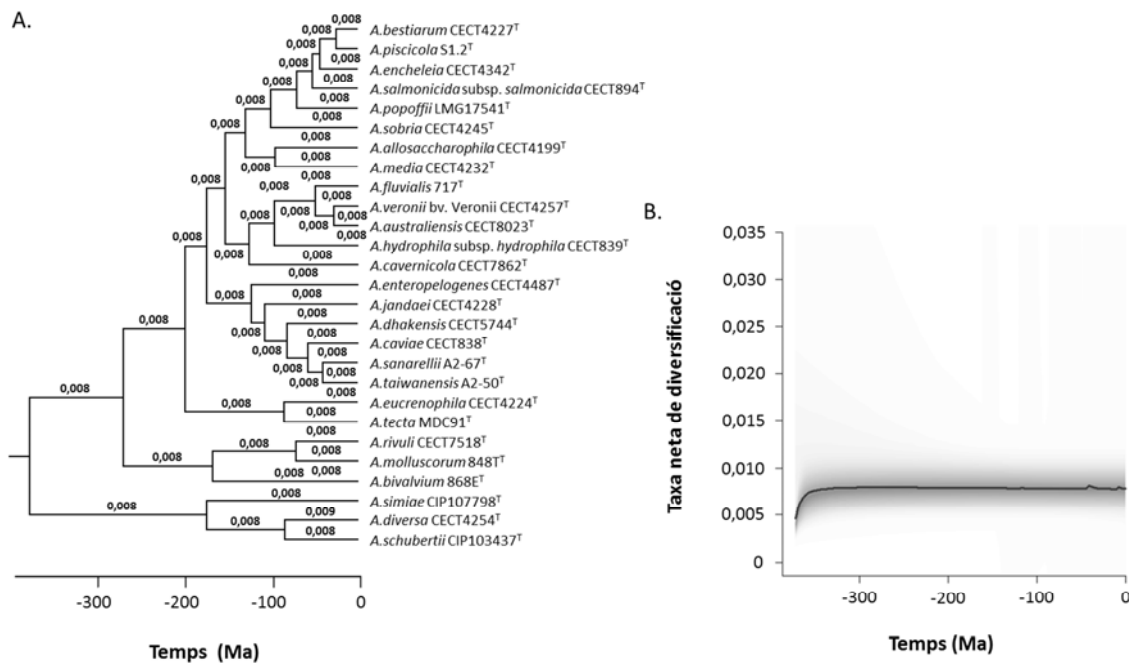
B.

Test de Rabosky		
ΔAIC_{RC}	Valor de P	Decisió
3,8359	0,0561	$\rightarrow P > 0,05$ - Acceptar hipòtesi nul·la (H_0) de model constant de Yule

Figura 5.41 Determinació del millor model de diversificació del gènere *Aeromonas* basat en les seqüències del gen *recA* de les soques tipus. A) Selecció del model i resultat del test ΔAIC_{RC} . B) Test de Rabosky.

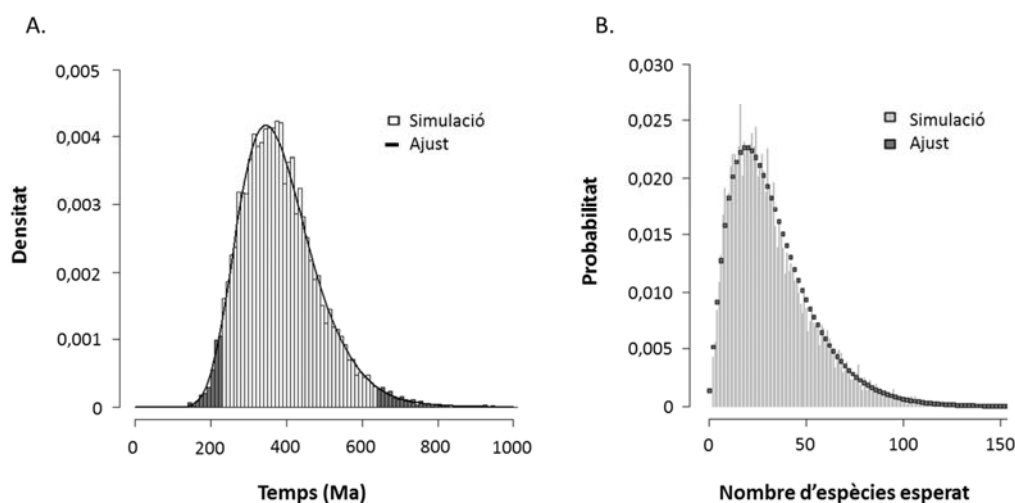
La majoria dels resultats indiquen que seguiria un model de diversificació constant de Yule, però per acabar de confirmar aquest resultat s'ha completat l'estudi determinant l'heterogeneïtat de la taxa de diversificació. A la Figura 5.42A es mostren els valors de les taxes de diversificació obtingudes per cada branca de l'arbre bayesià, que presenten un valor de λ constant de 0,008, excepte en un cas (0,009). Aquest resultat coincideix amb el perfil constant que s'observa en la Figura 5.42B, en el que es representa la taxa neta de diversificació al llarg del temps. Per avaluar si aquests resultats són significatius s'ha calculat la probabilitat de que les dades segueixen un model de diversificació constant (sense *shift*) o que presenti un ó mes *shifts*, estimant que hi ha una probabilitat d'1 que el procés es mantingui constant al llarg del temps, i de 0,0050 de que s'hagi produït un punt de canvi. També, s'ha calculat el factor de Bayes i s'ha obtingut un valor de 194,122 (superior a 100), que proporciona una evidència decisiva de que no s'ha produït cap punt de canvi en la taxa de diversificació i seguiria un model constant.

Figura 5.42 Resultats obtinguts amb el programa BAMM a partir de les seqüències del gen *recA* de les soques tipus. A) Arbre bayesià on s'indiquen les taxes de diversificació per cada branca. B) Representació gràfica de la taxa neta de diversificació al llarg del temps.



Finalment s'ha calculat teòricament, seguint un model de Yule (λ empírica = 0,0075 i $\mu = 0$), l'edat de l'ancestre comú més recent (tMRCA) i el nombre d'espècies actuals esperades pel gènere *Aeromonas* (Figura 5.43). Els resultats obtinguts estimen que aquest gènere hauria aparegut fa 382 Ma, valor comparable al determinat a partir de les dades (373 Ma), i que el valor mig del nombre d'espècies esperat estaria constituït per 33 espècies (SD = 22).

Figura 5.43 Resultats teòrics obtinguts a partir del model de Yule amb les seqüències del gen *recA* de les soques tipus. A) Histograma de l'edat de l'ancestre comú més recent (tMRCA). B) Distribució del nombre d'espècies esperat.



5.7.3 A partir de l'arbre d'espècies

A partir del cronograma de l'arbre d'espècies s'ha obtingut l'LTT plot (Figura 5.44A), el qual indica un patró temporal de diversificació constant, amb un bon ajust lineal a la recta de regressió (Figura 5.44B). El valor de λ estimat a partir del pendent de la recta de regressió ha estat de 0,0091. El resultat obtingut s'ha confirmat per simulació de 1000 cronogrames obtenint-se els corresponents LTT plots (Figura 5.44C). L'LTT plot generat a partir de les dades (línia contínua) s'ha situat dins del 95% dels LTT plots obtinguts per simulació (línies discontinues). A l'ajustar les dades a un model de taxa constant, mitjançant una prova de màxima versemblança, s'ha estimat el valor de λ que ha estat de 0,0078 (SE = 0,0011) i un valor de μ de 0, que correspondria a un model de Yule.

S'ha determinat l'estadístic Υ de Pybus i Harvey, obtenint un valor empíric de Υ de -1,7983. En la distribució de l'estadístic Υ obtingut per simulació de 1000 filogènies seguint un model constant de Yule (Figura 5.44D), s'indica el valor empíric de Υ amb una fletxa vertical, i es situa entre el límit superior (1,4057) i inferior (-2,1220) de l'interval de confiança del 95%.

Els resultats obtinguts de l'anàlisi de selecció del model de diversificació juntament amb el resultat del test ΔAIC_{RC} es mostren a la Figura 5.45A. El model de Yule ha estat el millor model constant, mentre que el model de Yule-2-rate el millor model amb taxa de diversificació variable. Després s'ha dut a terme el test ΔAIC_{RC} , que ha estimat que el que millor s'ajusta a les dades és el de Yule-2-rate. En aquest cas, els dos valors de λ obtinguts han estat de 0,0107 (λ_1) i 0,0018 (λ_2), i el punt de canvi en les taxes s'hauria produït fa 39,80 Ma (línia discontinua en l'LTT plot, Figura 5.44A). El resultat obtingut del test de Rabosky (Figura 5.45B) permet rebutjar la hipòtesi nul·la de model de Yule, amb un valor de P (0,0289) inferior a 0,05.

S'ha completat l'estudi amb una anàlisi bayesià per determinar possibles heterogeneïtats en les taxes de diversificació davant els resultats desiguals que s'han obtingut. A la Figura 5.46A es mostra com els valors de taxa de diversificació en cadascuna de les branques de l'arbre bayesià obtingut es manté constant en tots els casos, amb un valor de λ de 0,008. Aquest patró de taxa constant també s'aprecia en el gràfic obtingut de taxa neta de diversificació al llarg del temps (Figura 5.46B). La probabilitat de que no hi hagi cap canvi (*shift*) en el valor de λ ha estat d'1, mentre que la de produir-se un canvi (*shift*) ha estat molt baixa (0,0054). Aquest resultat ha estat recolzat pel factor de Bayes obtingut de 183,1974, valor superior a 100, que suposa una evidència decisiva de taxa de diversificació constant.

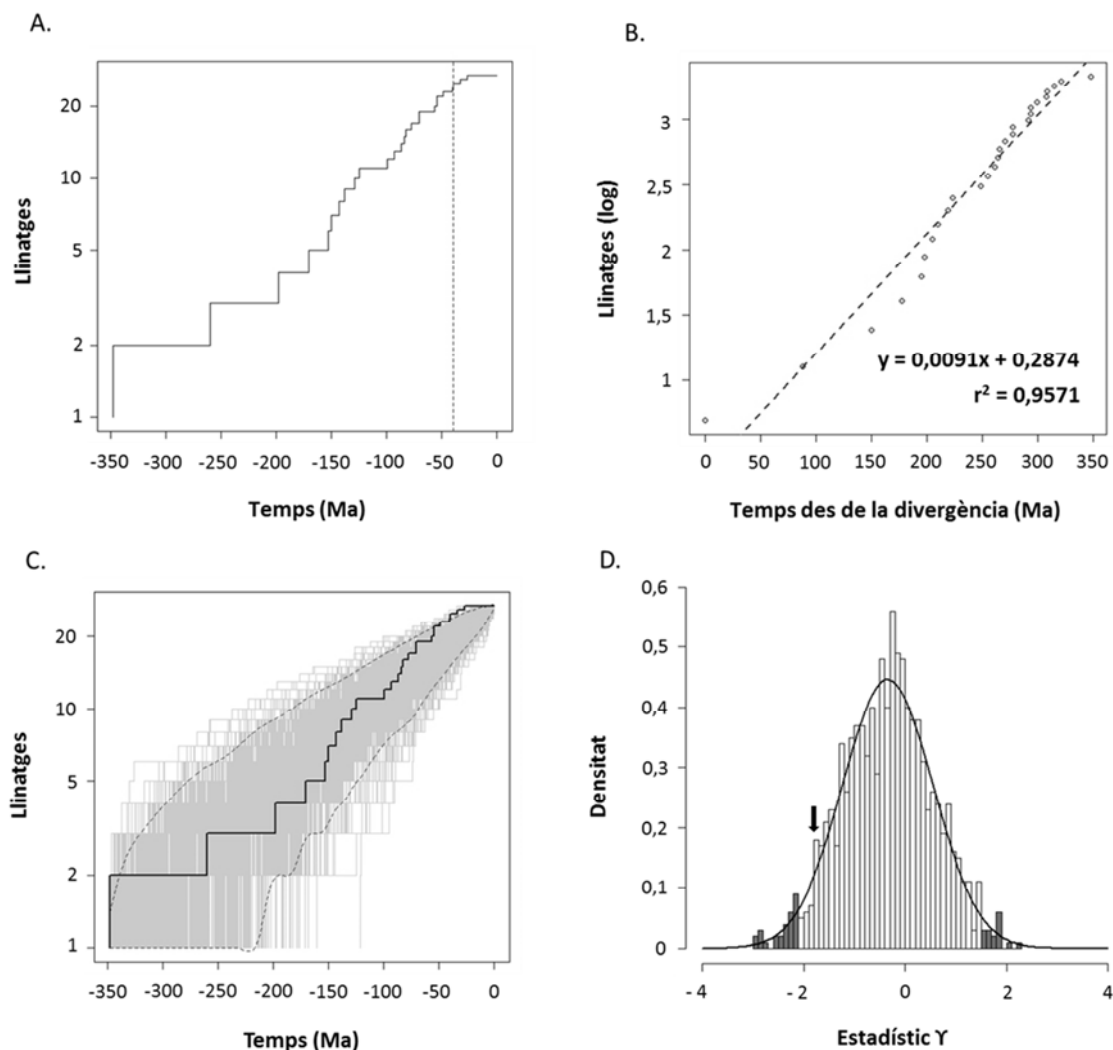


Figura 5.44 Anàlisi del model de diversificació del gènere *Aeromonas* basat en l'arbre d'espècies del gen *recA*. A) LTT plot a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió de l'LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic Y de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

Figura 5.45 Determinació del millor model de diversificació del gènere *Aeromonas* basat en l'arbre d'espècies del gen *recA*. A) Selecció del model i resultat del test ΔAIC_{RC} . B) Test de Rabosky.

A.

Selecció del model de diversificació					
Models constants			Models variables		
Model	<i>Birth-Death</i>	Yule	DDX	DDL	Yule 2-rate
AIC	173,8551	171,8551	172,9174	169,1575	166,8557

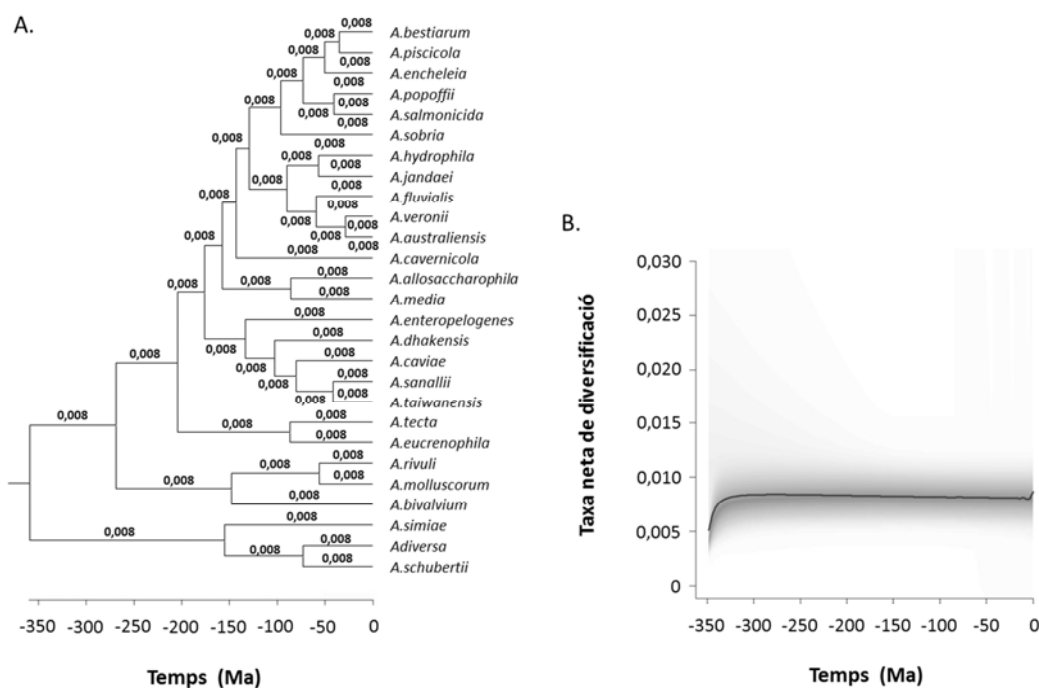
Millor model constant		Millor model variable	
Model	Yule	Model	Yule 2-rate
AIC	171,8551	AIC	166,8557

Test ΔAIC_{RC}	
Model	Yule 2-rate
AIC	166,8557
ΔAIC_{RC}	4,9994

B.

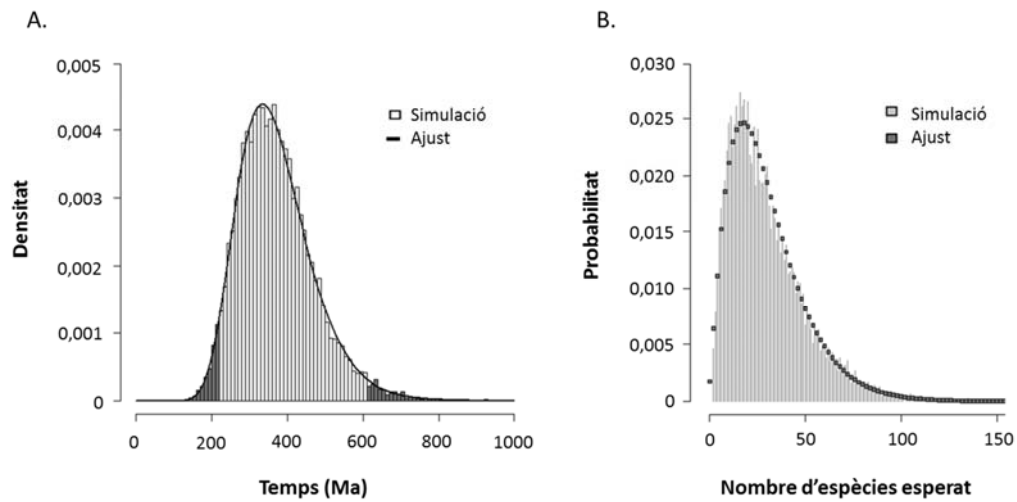
Test de Rabosky		
ΔAIC_{RC}	Valor de <i>P</i>	Decisió
4,9994	0,0289	→ <i>P</i> < 0,05 - Rebutjar hipòtesi nul·la (<i>H</i> ₀) de model constant de Yule

Figura 5.46 Resultats obtinguts amb el programa BAMM a partir de les seqüències del gen *recA* de l'arbre d'espècies. A) Arbre bayesià on s'indiquen les taxes de diversificació per cada branca. B) Representació gràfica de la taxa neta de diversificació al llarg del temps.



Per últim, s'ha determinat teòricament l'edat de l'ancestre comú més recent (tMRCA) i el valor esperat del nombre d'espècies d'*Aeromonas*, considerant un model de Yule (λ empíric = 0,0078 i $\mu = 0$). Els resultats obtinguts (Figura 5.47) indiquen que el gènere *Aeromonas* hauria aparegut fa 368 Ma (valor empíric 348 Ma) i estaria format per 30 espècies (SD= 21).

Figura 5.47 Resultats teòrics obtinguts a partir del model de Yule amb l'arbre d'espècies del gen *recA*. A) Histograma de l'edat de l'ancestre comú més recent (tMRCA). B) Distribució del nombre d'espècies esperat.



6. RESULTATS:

Estudi conjunt dels dos gens

6. RESULTATS: ESTUDI CONJUNT DELS DOS GENS

Actualment, les seqüències concatenades de dos o més gens s'utilitzen per fer reconstruccions filogenètiques. En aquest capítol, s'han analitzat les filogènies d'*Aeromonas* construïdes a partir dels concatenats de seqüències dels gens *mdh* i *recA*. També, a partir dels cronogrames obtinguts de les diferents filogènies analitzant conjuntament els dos gens s'ha fet l'estudi de datació i del model d'evolució que seguiria aquest gènere bacterià.

6.1 FILOGÈNIES

6.1.1 Anàlisi filogenètica del gènere *Aeromonas*

A partir de l'alineament de les 150 seqüències concatenades dels dos gens s'ha determinat quin és el millor model de substitució nucleotídica que s'ajusta a les dades, que ha estat el model GTR+G+I. S'han construït filogènies utilitzant els mètodes de màxima versemblança (Figura 6.1) i d'inferència bayesiana (Figura 6.2), obtenint-se valors de bootstrap (ML) i de probabilitat posterior (IB) alts en els nodes dels grups majoritaris. Al comparar les topologies dels dos arbres filogenètics, en general són pràcticament idèntiques, excepte l'espècie *A. encheleia* que en l'arbre ML es situa dins del complex d'espècies *A. hydrophila* (*A. salmonicida*, *A. bestiarum*, *A. piscicola*, *A. popoffii*) i en l'arbre IB en el clúster amb *A. eucrenophila*, *A. tecta* i *A. media*.

Al comparar amb més detall les filogènies obtingudes s'han detectat clarament dos grups. Un grup principal que conté la majoria d'espècies d'*Aeromonas*, i que les agrupa en dos clústers diferenciats. Un primer grup que està format per les espècies del complex *A. hydrophila* (*A. salmonicida*, *A. bestiarum*, *A. piscicola* i *A. popoffii*), juntament amb *A. eucrenophila*, *A. tecta*, *A. media*, *A. caviae*, *A. sanarellii*, *A. taiwanensis*, *A. hydrophila* i *A. dhakensis*, i un grup d'espècies ben cohesionat format per *A. bivalvium*, *A. molluscorum* i *A. rivuli*. Un segon clúster amb les espècies del complex *A. veronii* amb les espècies *A. cavernicola*, *A. jandaei* i *A. enteropelogenes*. A més del grup majoritari, en tots dos casos, s'ha observat un segon grup més petit, que porta més temps divergint, amb les espècies *A. simiae*, *A. schubertii* i *A. diversa*.

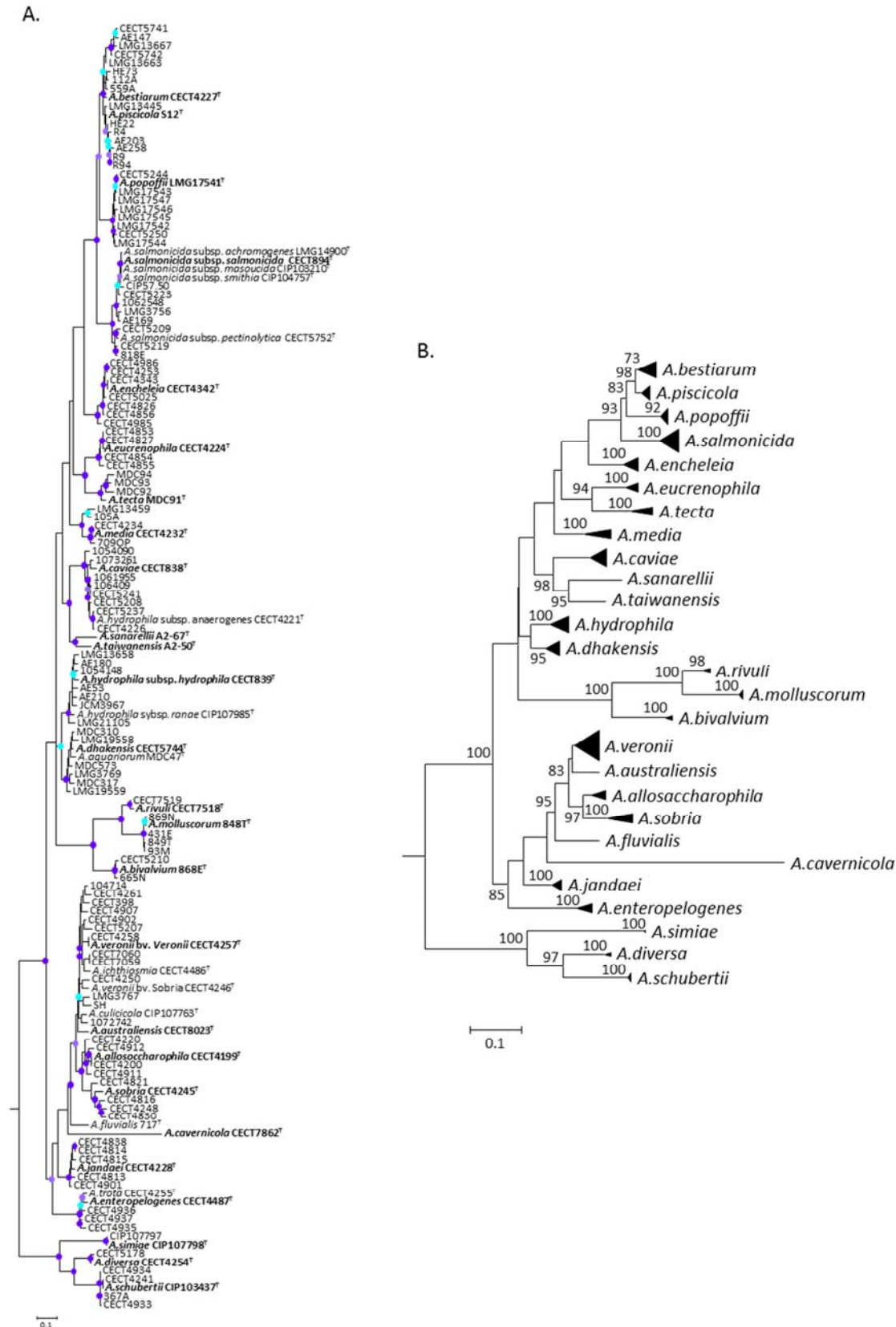


Figura 6.1 Arbre de màxima versemblança (ML) del concatenat de seqüències dels gens *mdh* i *recA*. A) Arbre amb les 150 soques. B) Arbre col·lapsat per espècies o grans grups. L'escala indica el nombre de substitucions nucleotídiques per lloc. Els nodes mostren els valors de *bootstrap* iguals o superiors al 70% (cian, 70-80%; lila clar, 80-90% i lila fosc, 90-100%).

6.2 DATACIÓ

S'ha realitzat l'estudi de datació del gènere *Aeromonas* a partir del concatenat dels dos gens *mdh* i *recA* de les 150 soques de l'estudi, del concatenat de les 27 seqüències consens, del concatenat de les 27 seqüències de les soques tipus de totes les espècies, i de l'arbre d'espècies generat amb les seqüències dels dos gens. S'han construït cronogrames per inferència bayesiana amb el programa BEAST, utilitzant el millor model d'evolució per les dades, un model de rellotge molecular relaxat, i com a punt de calibratge el temps de diversificació d'*E.coli* i *Salmonella* emprat en les datacions anteriors (140 Ma).

6.2.1 A partir del concatenat de les 150 seqüències

S'ha generat un cronograma del gènere *Aeromonas*, utilitzant el model de substitució GTR+G+I com el que millor s'ajusta per al concatenat de les 150 seqüències (Figura 6.3). Els valors de probabilitat posterior dels nodes principals han estat elevats, indicant que l'arbre construït és robust. Els temps de divergència dels principals grups i els seus intervals de confiança (IC) es mostren a la Taula 6.1.

Taula 6.1 Edat de divergència estimada dels principals grups del cronograma obtingut a partir del concatenat de les 150 seqüències dels gens *mdh* i *recA*.

Clúster	Edat	IC inferior	IC superior
1	353,2	134,8	637,6
2	267,4	111,1	477,2
3	225,6	88,9	396,2
9	216,0	86,4	387,3
10	174,0	59,3	322,3
8	125,6	44,9	225,7
4	115,9	42,3	216,7
7	109,2	35,3	203,4
5	80,9	31,1	152,8
6	39,7	17,8	89,7

Segons els resultats obtinguts, s'ha estimat que l'ancestre comú del gènere *Aeromonas* hauria aparegut fa aproximadament 353 Ma, i que la primera separació hauria donat lloc a un clade format per 3 espècies (*A.diversa*, *A.schubertii* i *A.simiae*) independent de la resta d'espècies actuals d'*Aeromonas*. Les diversificacions més recents es van produir quan es van separar les espècies *A.bestiarum*-*A.piscicola* (49 Ma), *A.rivuli*-*A.molluscorum* (64 Ma) i *A.sanarellii*-*A.taiwanensis* (75 Ma) i *A.australiensis*-*A.veronii* (79 Ma).

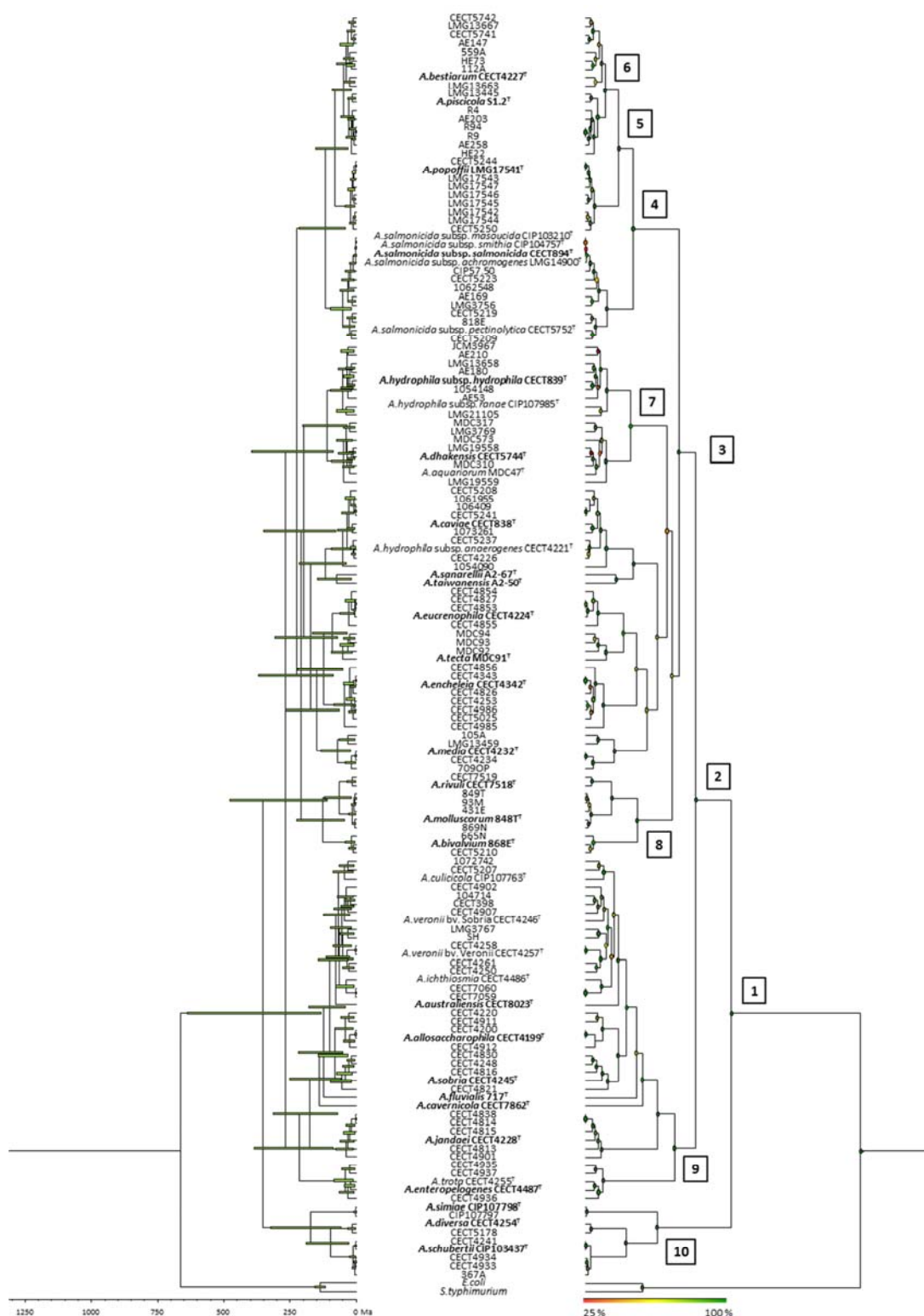


Figura 6.3 Cronograma del gènere *Aeromonas* basat en el concatenat de les 150 seqüències dels gens *mdh* i *recA*. A l'esquerra es mostra el cronograma a escala temporal (Ma), i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

6.2.2 A partir del concatenat de les seqüències consens

Per estimar l'edat de diversificació d'*Aeromonas* a partir del concatenat de les 27 seqüències consens dels dos gens s'ha construït un arbre bayesià amb el millor model de substitució per les dades (GTR+G+I). A la Figura 6.4 es mostra el cronograma obtingut, i la gran majoria dels nodes presenten valors de probabilitat posterior al voltant del 100%. L'edat de divergència dels nodes principals s'indica a la Taula 6.2, i estima que aquest gènere hauria començat a divergir fa aproximadament 247 Ma. Les diversificacions més recents es van produir amb la separació de les següents espècies: *A.bestiarum*-*A.piscicola* (14 Ma), *A.allosaccharophila*-*A.sobria* (25 Ma) i *A.australiensis*-*A.veronii* (31 Ma), i *A.rivuli*-*A.molluscorum* (34 Ma).

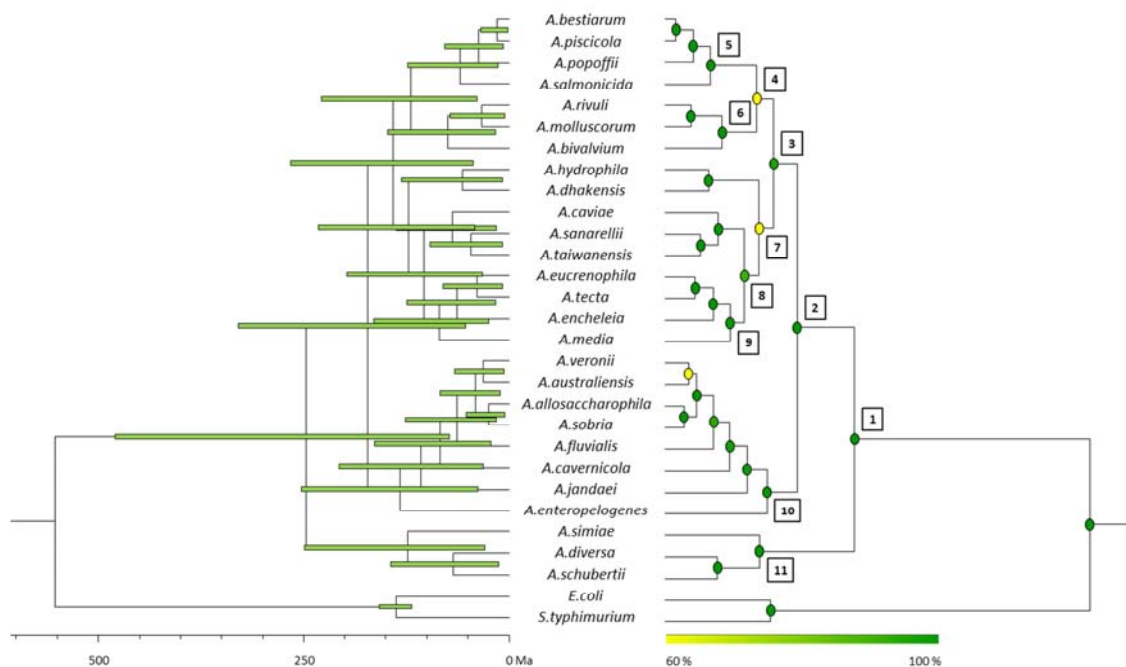


Figura 6.4 Cronograma del gènere *Aeromonas* basat en el concatenat de les seqüències consens dels gens *mdh* i *recA*. A l'esquerra es mostra el cronograma a escala temporal (Ma), i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 6.2 Edat de divergència estimada dels principals grups del cronograma obtingut a partir del concatenat de les seqüències consens dels gens *mdh* i *recA*.

Clúster	Edat	IC inferior	IC superior
1	246,8	72,7	479,8
2	172,4	53,4	330,0
3	141,5	44,1	266,0
10	132,9	38,1	252,9
11	123,2	29,5	249,0
7	122,3	41,9	232,5
4	119,7	39,4	227,9
8	103,4	32,7	197,8
9	84,8	24,5	164,7
6	74,7	16,2	147,4
5	59,5	13,7	123,0

6.2.3 A partir del concatenat de les seqüències de les soques tipus

S'ha determinat l'origen del gènere *Aeromonas* a partir del concatenat de les 27 seqüències corresponents a les soques tipus de totes les espècies. S'ha generat un cronograma considerant el model GTR+G+I com el model de substitució que millor s'ajusta a les seqüències (Figura 6.5), obtenint-se valors de probabilitat posteriors molt elevats. A la Taula 6.3 s'indica l'edat dels principals nodes i els seus intervals de confiança (IC). El temps de divergència estimat per *Aeromonas* ha sigut de 256 Ma, i les diversificacions d'espècies més recents es van produir per *A.bestiarum*-*A.piscicola* (14 Ma), *A.allosaccharophila*-*A.sobria* (29 Ma) i *A.rivuli*-*A.molluscorum* (37 Ma).

Taula 6.3 Edat de divergència estimada dels principals grups del cronograma obtingut a partir del concatenat de les seqüències de les soques tipus dels gens *mdh* i *recA*.

Clúster	Edat	IC inferior	IC superior
1	255,5	77,0	492,8
2	180,8	56,5	339,4
3	149,6	45,2	279,5
9	140,2	40,7	269,0
10	128,3	28,2	263,5
4	125,0	43,1	241,1
8	121,4	33,2	231,1
5	114,0	34,5	217,3
7	76,2	19,2	150,5
6	61,9	15,0	123,0

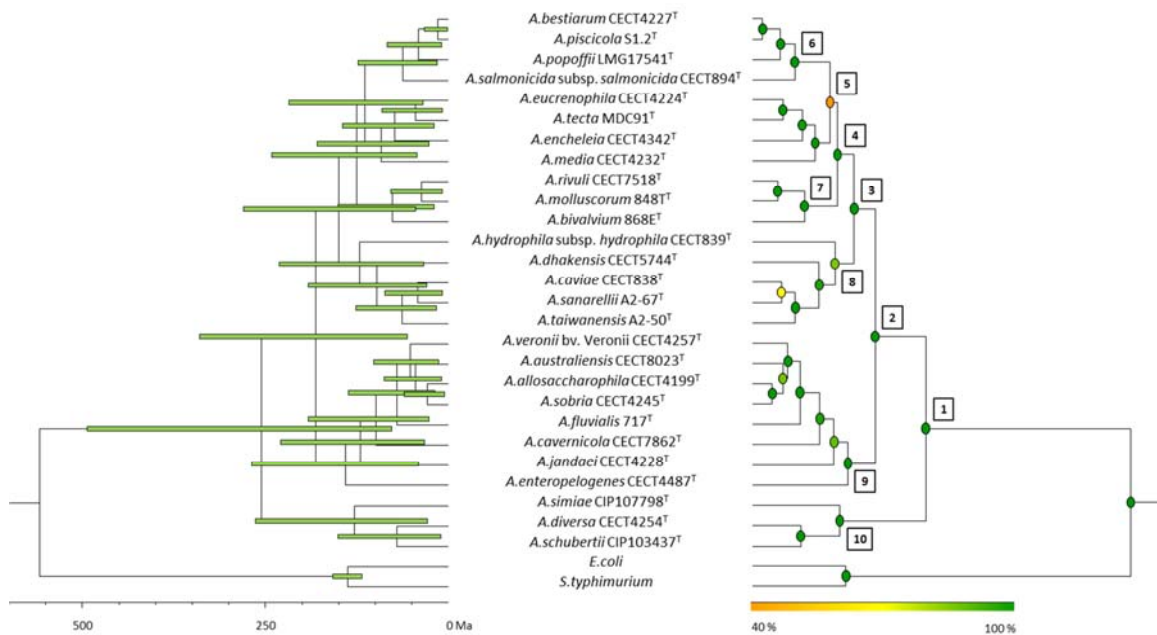


Figura 6.5 Cronograma del gènere *Aeromonas* basat en el concatenat de les seqüències de les soques tipus dels gens *mdh* i *recA*. A l'esquerra es mostra el cronograma a escala temporal (Ma), i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

6.2.4 A partir de l'arbre d'espècies dels dos gens

Finalment, s'ha construït l'arbre d'espècies del gènere *Aeromonas* a partir de les seqüències dels dos gens de l'estudi, utilitzant el model TN93+G+I per *mdh* i el GTR+G+I per *recA* com a millors models d'evolució. El cronograma obtingut i els valors de probabilitat posteriors en els nodes principals es mostren a la Figura 6.6. Els valors estimats de datació dels grups principals s'indiquen a la Taula 6.4, que suggereix que l'edat de divergència del gènere *Aeromonas* es situaria fa aproximadament 237 Ma. També s'ha estimat el temps de diversificació de les espècies més recents: *A. bestiarum*-*A. piscicola* (20 Ma), *A. hydrophila*-*A. dhakensis* (23 Ma), *A. australiensis*-*A. veronii* (24 Ma), *A. sanarellii*-*A. taiwanensis* (31 Ma), *A. rivuli*-*A. molluscorum* (32 Ma) i *A. allosaccharophila*-*A. sobria* (32 Ma).

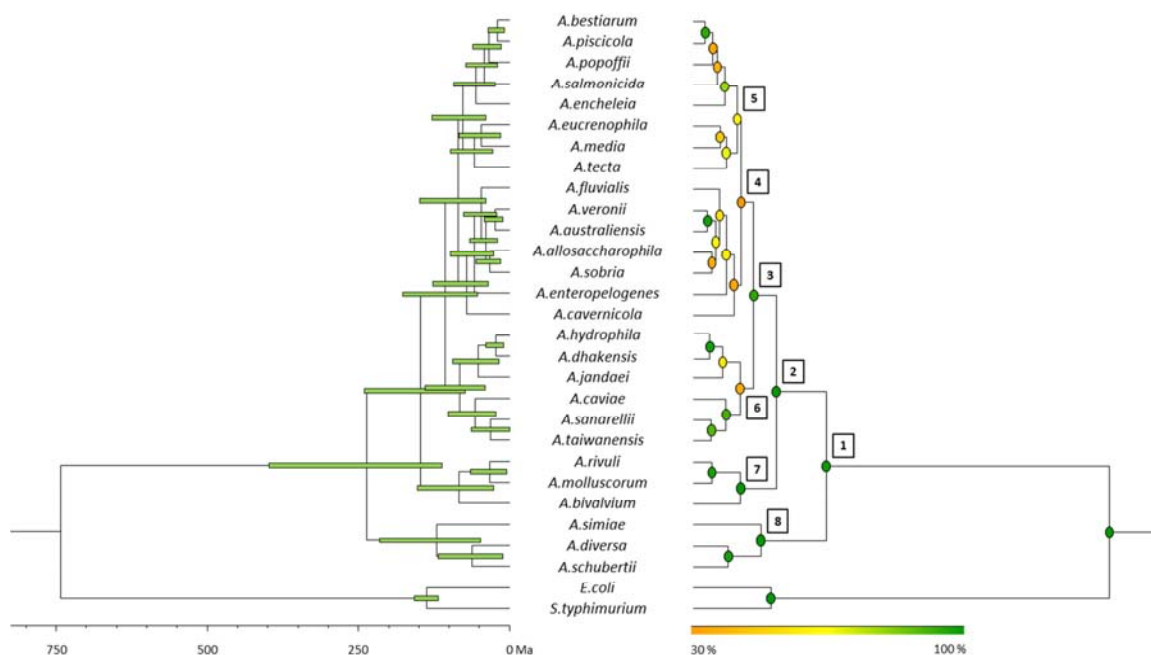


Figura 6.6 Cronograma del gènere *Aeromonas* basat en l'arbre d'espècies dels gens *mdh* i *recA*. A l'esquerra es mostra el cronograma a escala temporal (Ma), i les barres dels nodes indiquen la incertesa de la probabilitat posterior (*Highest Posterior Density*, HPD) expressada amb un nivell de confiança del 95%. A la dreta s'observa el mateix cronograma indicant en els nodes els valors de probabilitat posterior (%) amb una escala de colors, i es numeren els nodes dels grups principals.

Taula 6.4 Edat de divergència estimada dels principals grups del cronograma obtingut a partir de l'arbre d'espècies dels gens *mdh* i *recA*.

Clúster	Edat	IC inferior	IC superior
1	236,6	111,5	397,2
2	147,3	73,3	240,1
8	120,0	47,4	214,1
3	106,9	53,1	176,2
4	84,7	39,1	148,3
7	83,4	26,8	151,7
6	82,9	40,0	139,6
5	77,6	38,7	128,5

6.3 ANÀLISI DE LA DIVERSIFICACIÓ

Per portar a terme aquest estudi amb les seqüències dels dos gens, com ja s'ha explicat anteriorment (Apartat 4.6 i 5.6) s'han construït cronogrames per inferència bayesiana, considerant una única seqüència per espècie. Amb les mateixes metodologies i aproximacions (a partir dels concatenats de les seqüències consens i de les seqüències de les soques tipus, i de l'arbre d'espècies dels dos gens), per tots els casos s'ha estimat el model i la taxa de diversificació del gènere *Aeromonas*.

6.3.1 A partir del concatenat de les seqüències consens

Per visualitzar el patró de diversificació d'*Aeromonas*, s'ha realitzat un LTT plot (Figura 6.7A), en el que s'observa com l'aparició dels llinatges s'ha produït de manera constant al llarg del temps. La recta de regressió obtinguda (Figura 6.7B) presenta un bon ajust lineal, amb un valor de λ estimat del pendent de la recta de 0,0129.

També, s'ha realitzat una simulació a partir de 1000 filogènies per obtenir els corresponents LTT plots (Figura 6.7C), amb els que s'ha comprovat que l'LTT plot generat empíricament (línia continua negra) es troba dins del 95% (línies discontinues) dels obtinguts a partir de les filogènies simulades. A partir del cronograma, mitjançant una prova de màxima versemblança, s'han ajustat les dades a un model constant, i s'ha obtingut un valor de taxa de diversificació (λ) de 0,0103 (SE = 0,0014) i de taxa d'extinció (μ) de 0.

Per confirmar que s'ajusta a un model constant, també s'ha calculat l'estadístic Υ de Pybus i Harvey, obtenint un valor empíric de -2,1015, que es troba dins de l'interval de confiança del 95% (1,2546 i -2,1779) de la distribució de valors Υ estimats a partir de 1000 filogènies seguint el model de Yule (Figura 6.7D).

Per tal de determinar quin és el model de diversificació que millor s'ajusta a la dades, s'han analitzat diferents models amb taxes constants i variables al llarg del temps amb un prova de màxima versemblança (Figura 6.8A). Segons els valors d'AIC obtinguts, el millor model constant és el de Yule, i el millor model variable el de Yule-2-rate. El resultat del test ΔAIC_{RC} mostra que el millor model és el Yule-2-rate amb dues taxes ($\lambda_1 = 0,0144$ i $\lambda_2 = 0,0024$) i el punt de canvi s'hauria produït fa aproximadament 39 Ma (línia vertical discontinua en l'LTT plot, Figura 6.7A). Per determinar la fiabilitat d'aquest resultat s'ha realitzat el test de Rabosky (Figura 6.8B), obtenint un valor de P de 0,0311, que indica que s'ha de rebutjar la hipòtesi nul·la de model constant ($P < 0,05$) i acceptar el model Yule-2-rate.

S'ha fet l'anàlisi per detectar i quantificar l'heterogeneïtat de les taxes d'evolució amb el programa BAMM. A la Figura 6.9A es mostren els resultats de la taxa de diversificació estimada per cada branca de la filogènia bayesiana, amb valors que oscil·len entre 0,010 i 0,012, resultat coincident amb el perfil que s'ha observat en el gràfic de la taxa neta de diversificació (Figura 6.9B). Per veure la significació d'aquests resultats, s'ha calculat la probabilitat que no es produeixi cap canvi (0,99) o que se'n produeixi un (0,0086) en la taxa de diversificació, i el

factor de Bayes (114,942), que al donar un valor superior a 100 proporciona una evidència decisiva de que no s'ha produït cap canvi.

Per últim, considerant un model de diversificació constant de Yule (amb λ empírica = 0,0103 i $\mu = 0$), s'ha calculat el valor teòric de l'ancestre comú més recent d'*Aeromonas* i el nombre mig d'espècies esperat a partir de les dades del concatenat de seqüències consens (Figura 6.10). Els resultats teòrics obtinguts estimen que l'origen d'aquest gènere s'hauria produït fa 279 Ma (valor empíric 247 Ma) i hauria donat lloc a 25 espècies (SD = 17).

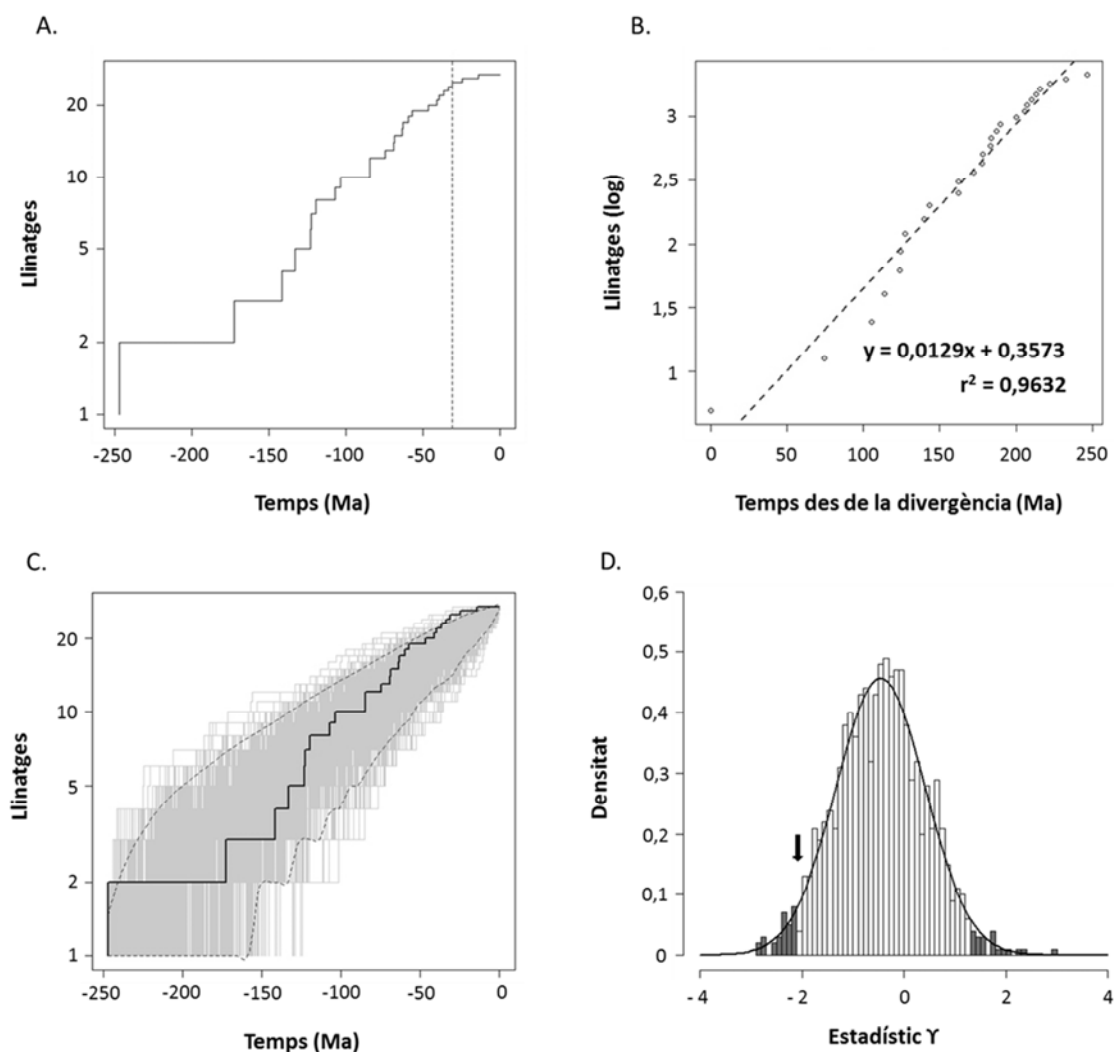


Figura 6.7 Anàlisi del model de diversificació del gènere *Aeromonas* basat en el concatenat de les seqüències consens dels dos gens. A) LTT plot obtingut a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió de l'LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic Y de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

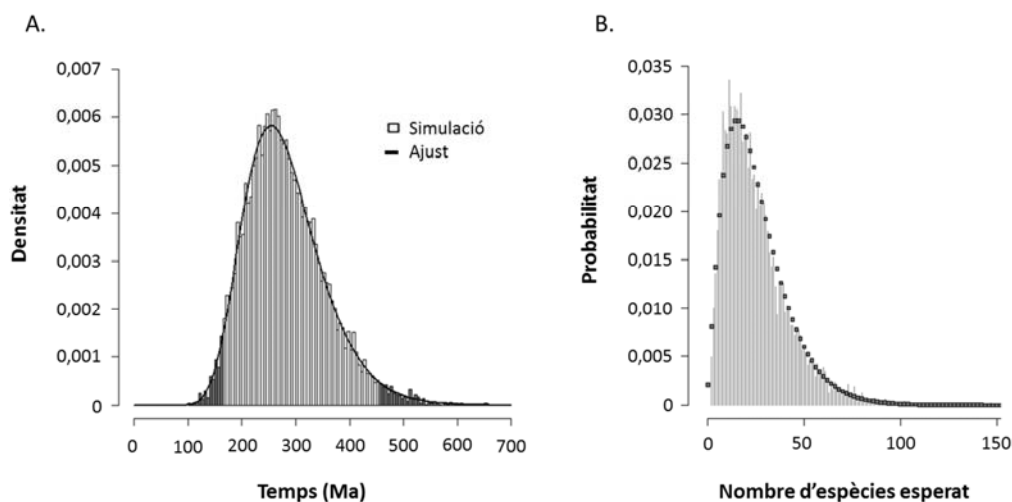


Figura 6.10 Resultats teòrics obtinguts a partir del model de Yule amb el concatenat de les seqüències consens dels dos gens. A) Histograma de l'edat de l'ancestre comú més recent (tMRCA). B) Distribució del nombre d'espècies esperat.

6.3.2 A partir del concatenat de les seqüències de les soques tipus

Per estudiar el patró de diversificació temporal s'ha generat el corresponent LTT plot a partir de les seqüències concatenades dels dos gens de les soques tipus de les espècies d'*Aeromonas*, que com s'observa en el gràfic es manté constant al llarg del temps (Figura 6.11A). S'ha obtingut un bon ajust lineal de la recta de regressió (Figura 6.11B), amb un valor de λ estimat, a partir del pendent de la recta, de 0,0125. Mitjançant una prova de màxima versemblança s'han ajustat les dades a un model constant, i s'han calculat els valors de λ i μ que han estat de 0,0096 (SE = 0,0013) i de 0, respectivament. A la Figura 6.11C es pot veure el resultat de la simulació de 1000 filogènies amb els corresponents LTT plots generats, on l'LTT plot empíric (línia continua negra) es situa dins del 95% (línies discontinues) dels obtinguts a partir de les filogènies simulades.

Els resultats de l'estadístic Υ de Pybus i Harvey es mostren a la Figura 6.11D, on s'indica el valor de Υ empíric obtingut de -2,4016 (fletxa negra) i la distribució de valors de Υ seguint un model de Yule. En aquest cas, el valor empíric està situat molt proper al límit inferior però fora de l'interval de confiança del 95% (1,1664 i -2,3514), i per tant, podria no seguir un model constant.

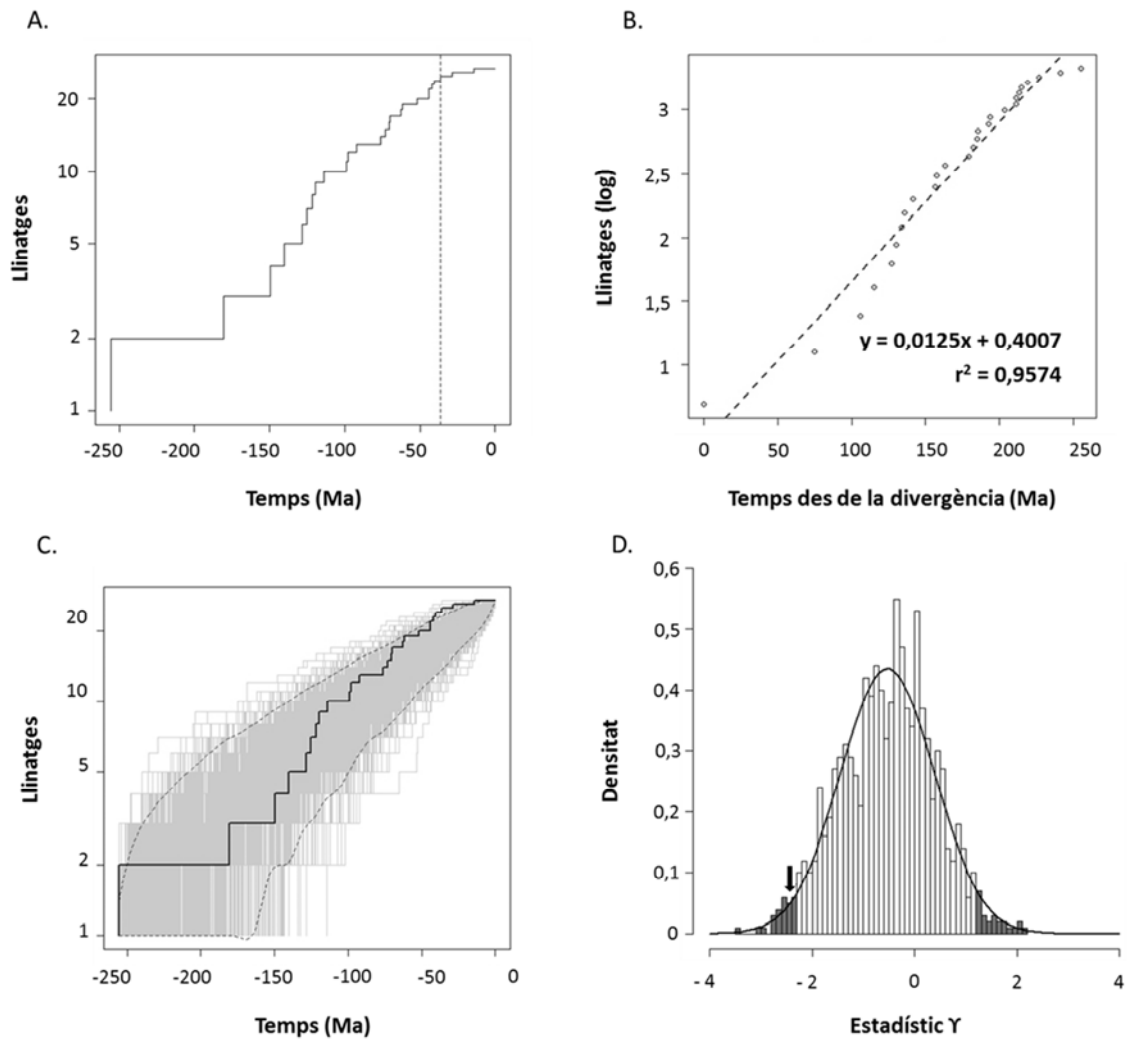


Figura 6.11 Anàlisi del model de diversificació del gènere *Aeromonas* basat en el concatenat de les seqüències de les soques tipus dels dos gens. A) LTT plot obtingut a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió de l'LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic Y de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

Mitjançant una prova de màxima versemblança i els corresponents valors d'AIC s'ha determinat el millor model amb taxa constant i variable (Figura 6.12A) que han estat el model de Yule i el Yule-2-rate, respectivament. Posteriorment, s'ha realitzat el test ΔAIC_{RC} per determinar quin d'aquest dos models és el que millor s'ajusta a les dades, essent el model de Yule-2-rate amb dos valors de λ al llarg del temps ($\lambda_1 = 0,0140$ i $\lambda_2 = 0,0020$) i el canvi s'hauria produït fa 36,62 Ma aproximadament (línia discontinua, Figura 6.11A). El resultat del test de Rabosky (Figura 6.12B), amb un valor de $P = 0,0079$ ($P < 0,05$), ha permès rebutjar la hipòtesi nul·la de model constant, i s'hauria d'acceptar l'alternativa.

A.

Selecció del model de diversificació					
Models constants			Models variables		
Model	<i>Birth-Death</i>	Yule	DDX	DDL	Yule 2-rate
AIC	163,6251	161,6251	161,2376	155,8100	154,3886

Millor model constant		Millor model variable	
Model	Yule	Model	Yule 2-rate
AIC	161,6251	AIC	154,3886

Test ΔAIC_{RC}	
Model	Yule 2-rate
AIC	154,3886
ΔAIC_{RC}	7,2364

B.

Test de Rabosky		
ΔAIC_{RC}	Valor de P	Decisió
7,2364	0,0079	$\rightarrow P < 0,05$ - Rebutjar hipòtesi nul·la (H_0) de model constant de Yule

Figura 6.12 Determinació del millor model de diversificació del gènere *Aeromonas* basat en el concatenat de les seqüències de les soques tipus dels dos gens. A) Selecció del model i resultat del test ΔAIC_{RC} . B) Test de Rabosky.

Degut als resultats no coincidents, s'ha realitzat l'anàlisi per detectar heterogeneïtats en les taxes, per tal de confirmar si el model seguit és un model constant o no. Per això, s'han determinat els valors de λ de cada branca de la filogènia bayesiana (Figura 6.13A), que han variat lleugerament entre 0,013-0,009. A la Figura 6.13B s'observa la taxa neta de diversificació que inicialment baixa lleugerament i després s'ha mantingut constant al llarg del temps. També, s'ha calculat la probabilitat de que es produeixi o no un canvi en la taxa de diversificació i el factor de Bayes com a suport estadístic dels resultats obtinguts. La probabilitat que es produeixi un canvi ha sigut molt baixa, de 0,0066, i de que es mantingui constant sense cap canvi de 0,99. El factor de Bayes obtingut de 154,53 recolza amb una evidència decisiva que la taxa neta de diversificació s'ha mantingut constant al llarg del temps.

Per últim, s'ha calculat quina és l'edat teòrica d'aparició de l'ancestre comú més recent d'*Aeromonas* i el nombre mig d'espècies esperat, seguint un model de diversificació de Yule (λ empírica = 0,0096 i $\mu = 0$). Els resultats obtinguts es mostren a les Figures 6.14A i B, indicant que aquest gènere hauria aparegut fa aproximadament 300 Ma (255 Ma obtingut empíricament) i estaria constituït per un nombre mig de 23 espècies (SD = 16).

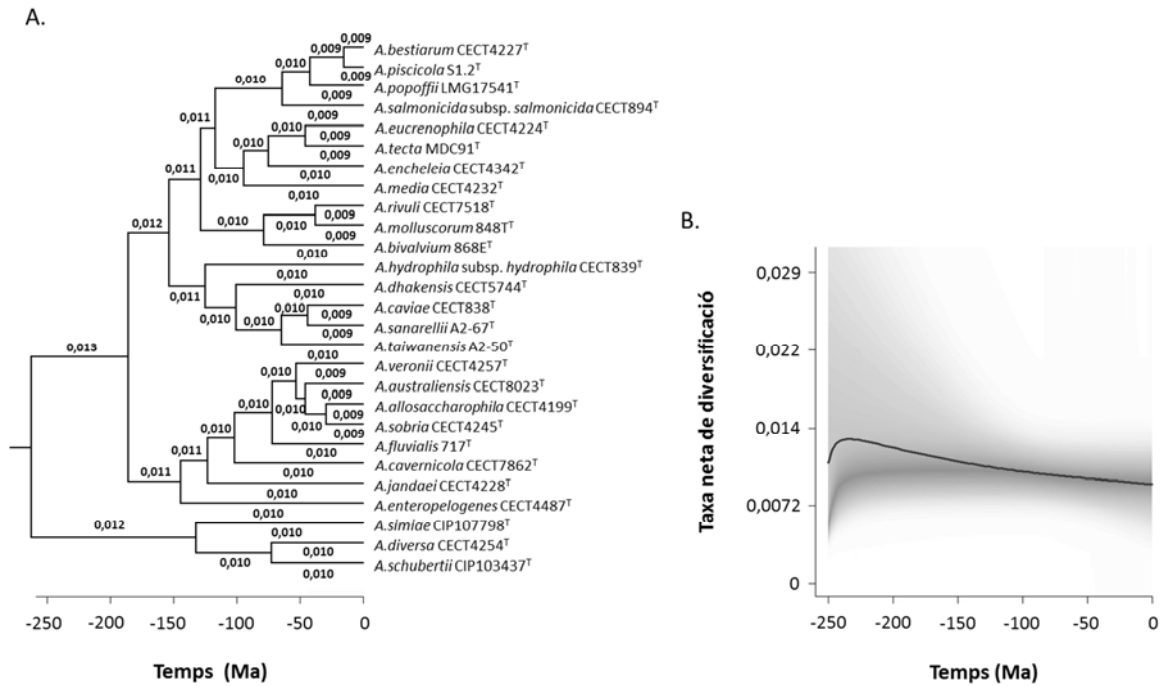


Figura 6.13 Resultats obtinguts amb el programa BAMM a partir del concatenat de les seqüències de les soques tipus dels dos gens. A) Arbre bayesià on s'indiquen les taxes de diversificació per cada branca. B) Representació gràfica de la taxa neta de diversificació al llarg del temps.

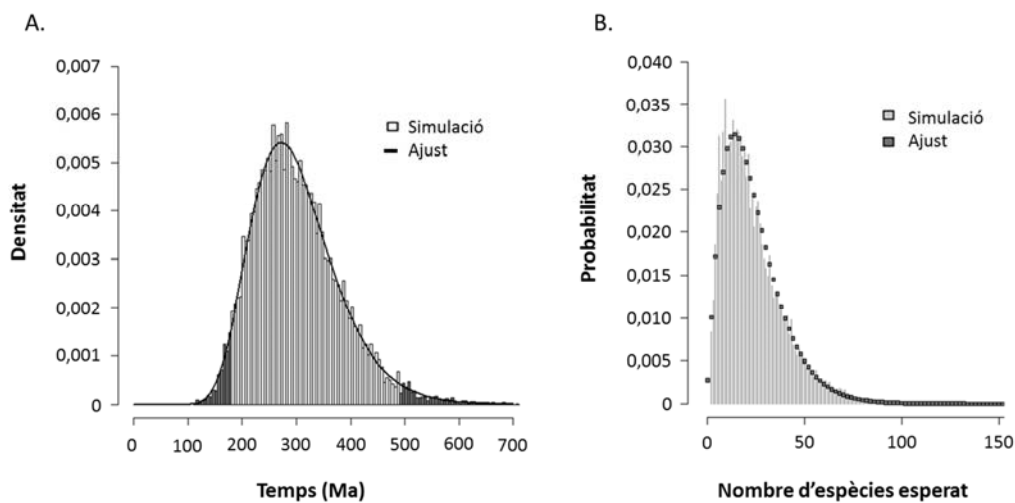


Figura 6.14 Resultats teòrics obtinguts a partir del model de Yule amb el concatenat de les seqüències de les soques tipus dels dos gens. A) Histograma de l'edat de l'ancestre comú més recent (tMRCA). B) Distribució del nombre d'espècies esperat.

6.3.3 A partir de l'arbre d'espècies dels dos gens

A partir del cronograma obtingut de l'arbre d'espècies dels gens *mdh* i *recA* s'ha construït el gràfic LTT plot (Figura 6.15A), on es visualitza un patró de diversificació temporal constant al llarg del temps. Els resultats s'han ajustat a una recta de regressió lineal (Figura 6.15B) amb un

valor d' r^2 inferior (0,9051) als obtinguts a partir de filogènies amb seqüències concatenades. El valor de λ estimat a partir del pendent de la recta de regressió ha estat de 0,0142. Per verificar aquests resultats, a partir de 1000 filogènies simulades s'han generat els corresponents LTT plots (Figura 6.15C), i l'LTT plot obtingut empíricament (línia contínua) s'ha situat dins del 95% (línies discontinues) del total d'LTT plots simulats. A l'ajustar les dades del cronograma a un model de taxa constant, mitjançant una prova de màxima versemblança, s'ha obtingut un valor de λ de 0,0127 (SE = 0,0018) i de μ de 0, que correspon a un model de Yule.

El valor empíric de l'estadístic Υ de Pybus i Harvey ha estat de -1,8420 (Figura 6.15D), que s'ha situat entre els límits superior (1,4931) i inferior (-1,9386) de l'interval de confiança del 95% de la distribució, corrobora també que seguiria un model de taxa constant.

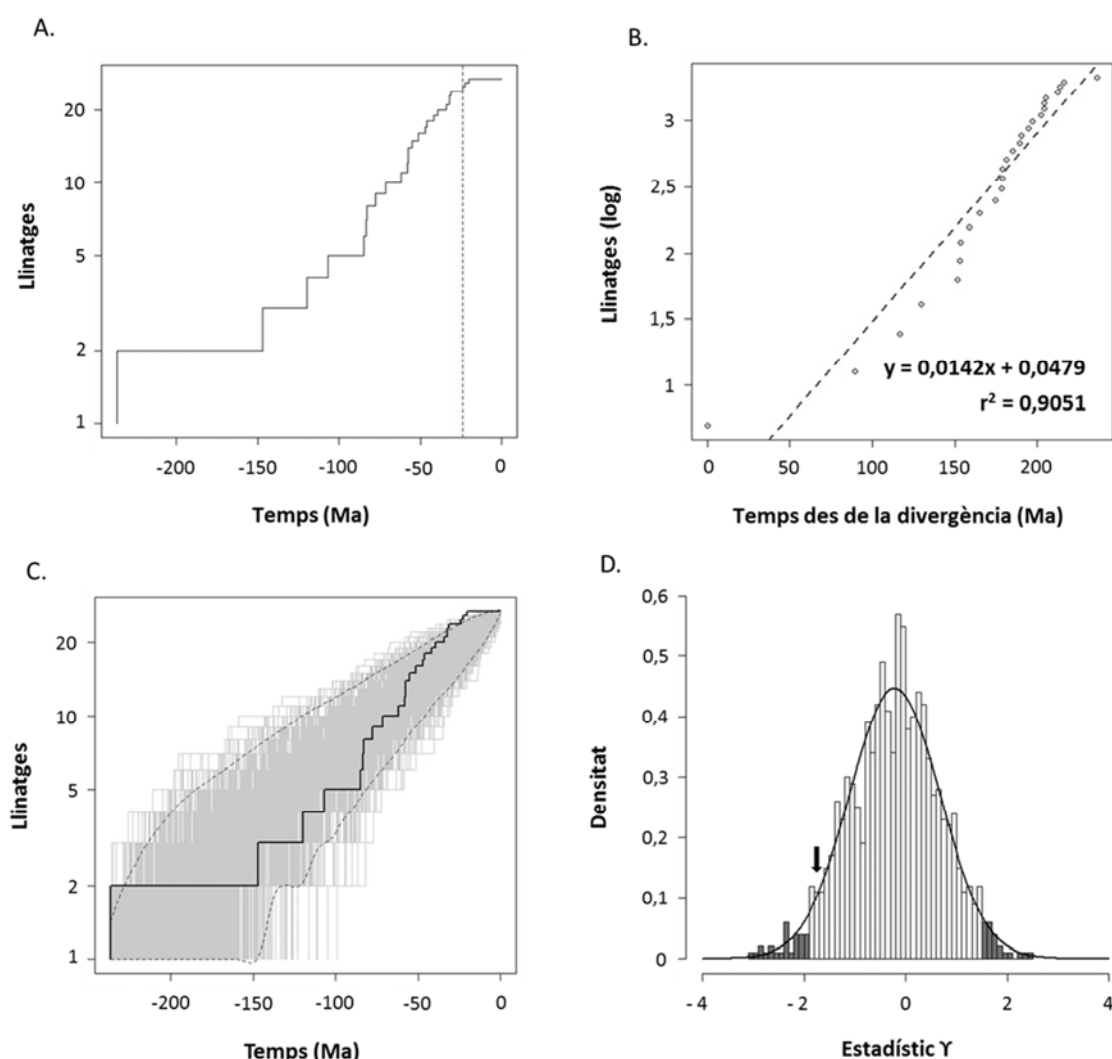


Figura 6.15 Anàlisi del model de diversificació del gènere *Aeromonas* basat en l'arbre d'espècies dels dos gens. A) LTT plot obtingut a partir de les dades empíriques. B) Ajust lineal mitjançant la recta de regressió de l'LTT plot obtingut. C) LTT plots obtinguts de la simulació de 1000 cronogrames. D) Distribució de l'estadístic Υ de Pybus i Harvey. La línia contínua negra indica l'ajust a la distribució normal reduïda.

Els resultats de l'anàlisi de selecció del millor model de diversificació de taxa constant i variable, i del test ΔAIC_{RC} es mostren a la Figura 6.16A. Dels dos millors models obtinguts (Yule com a millor model constant i Yule-2-rate com a millor model variable), el que millor s'ajusta a la dades és el Yule-2-rate, el qual hauria presentat dues taxes de diversificació de 0,0175 (λ_1) i 0,0030 (λ_2), i el canvi s'hauria produït fa 24,27 Ma aproximadament (línia discontinua de l'LTT plot, Figura 6.15A). Després, s'ha realitzat el test de Rabosky (Figura 6.16B) per determinar que el resultat obtingut té significació estadística, i ha permès rebutjar la hipòtesi nul·la de model constant de Yule amb un valor de $P < 0,05$.

A.

Selecció del model de diversificació					
Models constants			Models variables		
Model	<i>Birth-Death</i>	Yule	DDX	DDL	Yule 2-rate
AIC	149,5955	147,5955	149,1043	144,7005	140,9955

Millor model constant		Millor model variable	
Model	Yule	Model	Yule 2-rate
AIC	147,5955	AIC	140,9955

Test ΔAIC_{RC}	
Model	Yule 2-rate
AIC	140,9955
ΔAIC_{RC}	6,5999

B.

Test de Rabosky		
ΔAIC_{RC}	Valor de P	Decisió
6,5999	0,0149	$\rightarrow P < 0,05$ - Rebutjar hipòtesi nul·la (H_0) de model constant de Yule

Figura 6.16 Determinació del millor model de diversificació del gènere *Aeromonas* basat en l'arbre d'espècies dels dos gens. A) Selecció del model i resultat del test ΔAIC_{RC} . B) Test de Rabosky.

Al fer l'anàlisi per detectar heterogeneïtats de taxes, s'ha calculat la taxa de diversificació per a cadascuna de les branques de l'arbre bayesià (Figura 6.17A), i s'ha pogut observar com les taxes s'han mantingut constants al llarg del temps en totes les branques, amb un únic valor de λ de 0,013. Aquest resultat ha estat coincident amb el perfil obtingut de la taxa neta de diversificació temporal (Figura 6.17B). També s'ha calculat la probabilitat que no es produeix cap canvi (0,99) i que se'n produeixi un (0,0084) confirmant els resultats anteriors. A més, el factor de Bayes obtingut ha estat de 117,661, valor superior a 100, que suposa una evidència decisiva de que no s'hauria produït cap canvi en la taxa de diversificació al llarg del procés d'evolució.

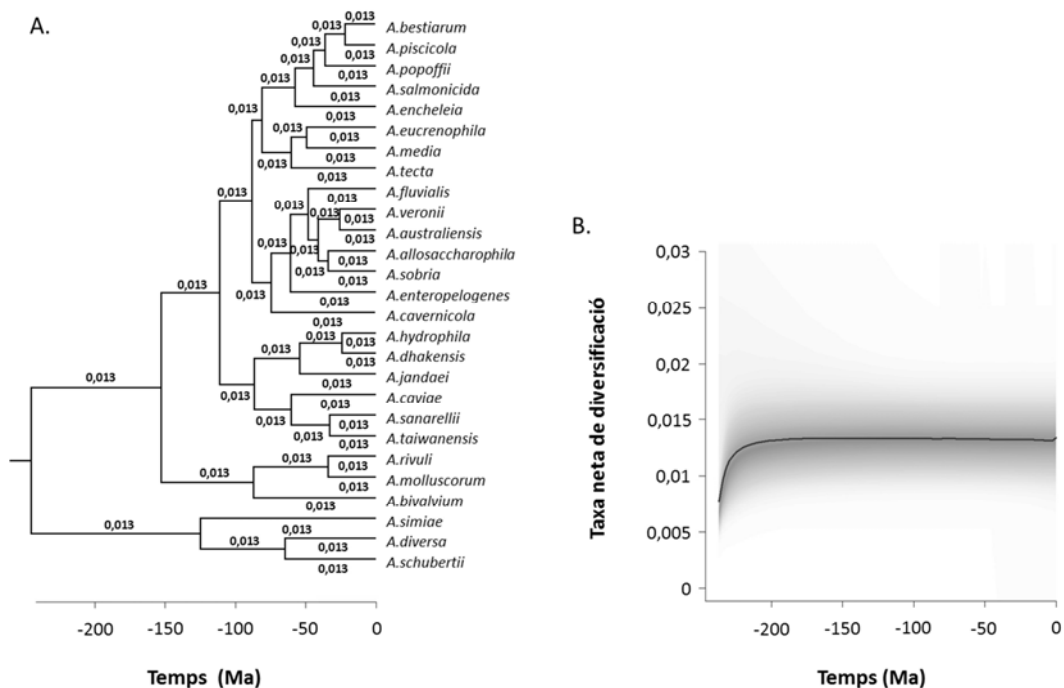


Figura 6.17 Resultats obtinguts amb el programa BAMM a partir de l'arbre d'espècies dels dos gens. A) Arbre bayesià on s'indiquen les taxes de diversificació per cada branca. B) Representació gràfica de la taxa neta de diversificació al llarg del temps.

Per últim, s'ha calculat el valor teòric del tMRCA d'*Aeromonas* i el valor mig d'espècies esperat, considerant que segueix un model de Yule (λ empírica = 0,0127 i $\mu = 0$). Els resultats obtinguts (Figura 6.18) estimen que aquest gènere s'hauria originat fa aproximadament 227 Ma, comparable al valor empíric obtingut (237 Ma), i que estaria format per 40 espècies (SD = 28).

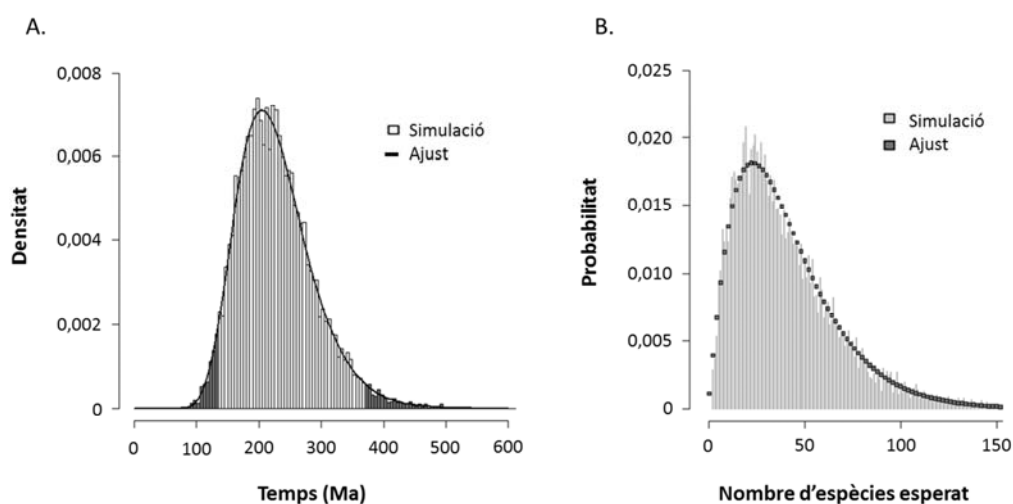


Figura 6.18 Resultats teòrics obtinguts a partir del model de Yule amb l'arbre d'espècies dels dos gens. A) Histograma de l'edat de l'ancestre comú més recent (tMRCA). B) Distribució del nombre d'espècies esperat.

7. DISCUSSIÓ

7. DISCUSSIÓ

Els resultats obtinguts en aquest estudi han donat resposta als principals objectius de la tesi doctoral. Aquests objectius han pretès contribuir a clarificar la història evolutiva del gènere *Aeromonas*. Primerament, s'ha tractat d'avaluar el potencial dels gens *mdh* i *recA* per resoldre les relacions filogenètiques de les diferents espècies del gènere a partir de les seqüències de 150 soques. En segon lloc, s'han construït els corresponents cronogrames de les filogènies obtingudes mitjançant una anàlisi d'inferència bayesiana, utilitzant com a punt de calibratge la datació acceptada comunament de la divergència d'*E.coli* i *Salmonella*. Tercer, el potser més ambiciós objectiu, ha estat la descripció de la radiació adaptativa del gènere *Aeromonas*, estudiant els temps de divergència, la taxa de diversificació i l'ajust del procés seguit als diferents models de diversificació temporal.

7.1 ANÀLISI DELS MARCADORS

Un aspecte que cal destacar del treball és el fet de que s'han seqüenciat gens sencers. Això és especialment rellevant degut a que no totes les regions dels gens codificadors de proteïnes estan sotmeses a un mateix tipus de pressió selectiva, i en conseqüència, hi ha regions més lliures d'acumular canvis mentre que d'altres són més conservades. Per aquest motiu, si s'analiza únicament un fragment, pot ser que no sigui representatiu de la diversitat del gen. Els gens escollits per aquest treball, han estat utilitzats prèviament en altres anàlisis filogenètiques analitzant seqüències parcials d'entre 400-600 pb, tant en el cas d'*Aeromonas* (Sepe *et al.* 2008) com d'altres espècies bacterianes (Nalin *et al.* 2000; Stine *et al.* 2000; Torriani *et al.* 2001; Payne *et al.* 2005; Rossi *et al.* 2006; Weng *et al.* 2009; Pascual *et al.* 2010; Gruber *et al.* 2011; Zbinden *et al.* 2011). L'altre fet diferencial respecte de treballs previs, és que aquest estudi s'ha fet a partir d'un nombre molt elevat de soques (150) representatives de totes les espècies i subespècies del gènere *Aeromonas*.

Els dos gens escollits, *mdh* i *recA*, són gens conservats que codifiquen per funcions essencials per la cèl·lula, que es troben en una còpia única i estan presents en totes les espècies del gènere. El gen *recA* ha estat utilitzat en treballs previs d'*Aeromonas*, però mai s'havia obtingut la seqüència completa del gen, només un fragment de 348 pb (Sepe *et al.* 2008). Pel que fa al gen *mdh*, aquest ha estat utilitzat únicament pel nostre grup, en un treball amb les soques tipus d'*Aeromonas* (Farfán *et al.* 2010) i un altre de genètica de poblacions amb les espècies del complex *A.hydrophila* (Fusté *et al.* 2012), però mai s'havia utilitzat un nombre tant elevat de soques de totes les espècies. A la Taula 7.1 es resumeixen les principals característiques dels dos gens estudiats, destacant que en el cas de la *mdh*, totes les soques han presentat una mateixa longitud de seqüència (936 pb) i no s'han observat ni insercions ni delecions, mentre que en el gen *recA* les seqüències han presentat una longitud variable (1059-1068 pb), fins i tot entre soques d'una mateixa espècie degut a la presència d'insercions i/o delecions.

Taula 7.1 Principals característiques dels gens *mdh* i *recA*

Caracterització del gen		
	<i>mdh</i>	<i>recA</i>
Longitud (pb)	936	1059-1068
Posicions invariants	566	651
Posicions polimòrfiques	370	396
Delecions/Insercions	No	Si
<i>Ti/Tv</i>	2,88	3,04
Contingut en G+C (%mol)	60,4	59,0
Saturació	No	No
Tipus de selecció	Purificadora	Purificadora
Substitucions sinònimes (<i>dS</i>)	0,4064	0,3998
Substitucions no sinònimes (<i>dN</i>)	0,0223	0,0247
ω (<i>dN/dS</i>)	0,0549	0,0618
Filogènia		
Relotge molecular estricta	No	No
Millor model d'evolució	TN93+G+I	GTR+G+I

Tant en el cas del gen *mdh* com en el *recA* s'ha vist que es tracta de gens sotmesos a selecció purificadora, en els que el nombre de substitucions sinònimes i no sinònimes i els valors d' ω obtinguts han estat similars en els dos casos. A l'analitzar el grau de saturació de les substitucions sinònimes i no sinònimes s'ha comprovat que en cap cas estan saturades, demostrant, per tant, que es tracta de gens adequats per construir filogènies. La distribució d'ambdós tipus de substitucions estimades codó per codó al llarg de la seqüència, varia lleugerament en els dos gens analitzats, observant-se que en el cas del *recA* presenta un augment important especialment de substitucions no sinònimes en la regió terminal del gen (Figura 5.4). Aquesta regió C-terminal, tal com es comenta en altres apartats, sembla que permet acumular canvis més fàcilment. Els valors de distància intra- i inter-específica determinats a partir de les seqüències dels gens han mostrat que en els dos casos tenen un bon poder de discriminació de les espècies d'*Aeromonas*, tal com es pot observar clarament en els gràfics de densitat de distàncies (Figures 4.6A i 5.6A). Totes aquestes característiques permeten afirmar que tant *mdh* com *recA* han demostrat ser bons marcadors moleculars per avaluar la variabilitat genètica i inferir les relacions filogenètiques de les espècies del gènere *Aeromonas*.

7.2 FILOGÈNIES d'*Aeromonas*

Els mètodes de reconstrucció triats: màxima versemblança (ML) i inferència bayesiana (IB), són els que han demostrat ser més rigorosos per a la construcció de filogènies. Es tracta de mètodes que basen la seva anàlisi en estimes de valors de màxima versemblança per fer les agrupacions dels diferents llinatges i construir l'arbre, tot i que ho fan a partir de dues suposicions totalment diferents. En la filogènia per ML, donat un conjunt de seqüències i un model de substitució el que es pretén és trobar la topologia d'un únic arbre en el que la longitud de cadascuna de les branques maximitzi la probabilitat de les dades. En la filogènia per IB, es genera una gran quantitat d'arbres i s'utilitza la probabilitat de distribució posterior

per veure quantes vegades es repeteix un node en cadascun dels arbres. Això ho fa per a cada node, i d'aquesta manera es construeix l'arbre MCC (*maximum clade credibility*) que dona com a resultat final. Les dues aproximacions són totalment diferents, i per això, la coincidència dels clades majoritaris en els arbres construïts amb els dos mètodes indica la robustesa de les filogènies obtingudes quan s'utilitzen els marcadors *mdh* i *recA*.

El mètode de ML és computacionalment més complex, i per tant, cal tenir-ho en compte quan s'utilitza un nombre important de dades en l'anàlisi, mentre que el d'IB gràcies a la utilització que fa de les cadenes de Markov a través del mètode de Monte Carlo (MCMC) aconseguix simplificar els càlculs. A més, aquest últim té l'avantatge que permet escollir, entre altres paràmetres, diferents models de rellotge molecular per l'anàlisi, com és el que s'ha utilitzat en l'estudi amb els gens *mdh* i *recA* (model de rellotge molecular relaxat), ja que s'ha demostrat que el nombre de substitucions que presenten al llarg del temps no s'ajusta en cap dels dos casos a un model de rellotge molecular estricte.

Assumint que la cohesió dels grups filogenètics majoritaris dins els procariotes és degut fonamentalment a la transmissió vertical i a l'ascendència comuna, és possible construir filogènies robustes, en les que la majoria d'espècies bacterianes es defineixen com llinatges evolutius discrets, que reflecteixen la història evolutiva dels bacteris utilitzant seqüències de gens conservats. Això, no exclou l'existència de transmissió horitzontal, que de fet es produeix, i té conseqüències evolutives importants, però no està clar que tingui un gran impacte en les filogènies construïdes a partir de gens conservats, degut a l'aïllament ecològic i als obstacles a l'adquisició de DNA extern pels mecanismes de restricció de les cèl·lules, provocant l'aïllament sexual. Un altre punt a considerar, és que tal com s'ha demostrat en diferents grups bacterians, la possibilitat de recombinació homòloga disminueix de manera exponencial amb la divergència de les seqüències, de manera que una divergència en les seqüències d'un 10% redueix la taxa de recombinació entre elles per un factor de 100 (Majewski *et al.* 2000; Fraser *et al.* 2007; Achtman & Wagner 2008; Caro-Quintero & Konstantinidis 2012; Shapiro *et al.* 2012).

L'anàlisi filogenètica del gènere *Aeromonas*, ha corroborat l'origen monofilètic d'aquest grup de bacteris. A més, en tots els casos, s'ha obtingut un bon suport estadístic amb valors alts de *bootstrap* (ML) o de probabilitat posterior (IB) en tots els nodes dels grups majoritaris, que corroboren que els arbres obtinguts són robusts.

En el cas dels arbres generats amb les seqüències del gen *mdh*, les dues aproximacions realitzades, ML i IB, han donat lloc a filogènies totalment coincidents. Cal destacar que la distribució dels clades principals coincideix també de forma majoritària amb altres filogènies construïdes prèviament per *Aeromonas* amb altres gens (Soler *et al.* 2004; Nhung *et al.* 2007; Farfán *et al.* 2010), tot i que s'han observat algunes diferències que poden ser explicades per haver utilitzat mètodes de reconstrucció diferents (basats en distàncies). Al comparar els resultats obtinguts amb altres estudis d'*Aeromonas* que han emprat mètodes de màxima versemblança (Colston *et al.* 2014; Lorén *et al.* 2014), els resultats han estat pràcticament idèntics, fins i tot quan s'ha comparat amb el treball de Colston i col·laboradors (2014) que utilitzen seqüències concatenades de 16 gens *housekeeping* obtinguts a partir de la

seqüenciació de genomes sencers de 56 soques d'*Aeromonas* en el que s'inclouen totes les soques tipus, excepte *A.cavernicola*. En relació al gen *recA*, s'han observat algunes diferències en els dos mètodes de reconstrucció (Apartat 5.3.1), l'arbre d'IB és el que ha donat una agrupació de les espècies més coincident amb l'obtingut pel gen *mdh*.

Finalment, en considerar les filogènies construïdes a partir de les seqüències concatenades dels dos gens, aquestes han mostrat també algunes petites discrepàncies en funció del mètode de reconstrucció. En l'arbre de ML del concatenat, de manera similar a l'obtingut pel gen *recA*, l'espècie *A.enceleia* s'ha agrupat juntament amb el clade que inclou *A.bestiarum*, *A.piscicola*, *A.popoffii* i *A.salmonicida*. En canvi, el construït a partir d'IB, els clades obtinguts han estat més coincidents amb l'obtingut com a resultat de la filogènia del gen *mdh*.

En general, en tots els casos analitzats el mètode d'inferència bayesiana és el que ha donat filogènies més fiables. Aquest fet es podria explicar perquè les seqüències estudiades no compleixen la condició de rellotge molecular estricte. En les filogènies construïdes pel mètode de ML no es poden utilitzar models de rellotge alternatius, per això, no s'haurien ajustat tant fidelment a la realitat com les generades per inferència bayesiana, que si que s'han emprat un model de rellotge relaxat.

En aquest estudi, s'ha confirmat la bona agrupació de les espècies considerades sinònimes, com és el cas d'*A.culicicola* i *A.ichthiosmia* dins el complex *A.veronii*, o la justificació de la descripció de noves espècies com *A.dhakensis* (Beaz-Hidalgo *et al.* 2013). En general les diferents soques de cada espècie s'han agrupat en clades coherents, excepte 6 soques (*A.caviae* CECT5207, *A.hydrophila* CECT398, *A.hydrophila* LMG3767, *A.popoffii* CECT5210, *A.sobria* CECT4250, *A.veronii* CECT4816) que s'ha vist que estaven mal identificades (Apartat 4.3.2). Cinc d'aquestes soques s'han reclassificat mitjançant diferents proves, però hi ha hagut una soca (*A.veronii* CECT4816), que fenotípicament no ha pogut ser assignada a cap espècie en concret i s'ha inclòs en l'anomenat "complex *A. veronii*", un grup en el que la separació entre espècies és, en general, difícil (Taula 4.13).

7.3 ANÀLISI DE LA RECOMBINACIÓ EN EL GEN *recA*

A l'anàlisi realitzada a partir de les seqüències del gen *recA*, s'ha detectat dins del clade d'*A.bestiarum*, 4 soques (AE147, CECT5741, CECT5742 i LMG13667) que en l'arbre de ML s'han agrupat en una branca més profunda, encara que conjuntament amb la resta de soques de l'espècie. A l'examinar amb detall les seves seqüències, s'ha vist que en la regió terminal del gen, presentaven un fragment de 81 pb molt diferent de la resta de soques d'*A.bestiarum*. Aquest fragment, malgrat el diferent origen de les soques (tres d'elles aigües de Finlàndia, Suïssa i Estats Units i la quarta d'origen ambiental d'Alemanya), ha estat idèntic en totes elles excepte per una diferència en un nucleòtid en la posició 1014 de la soca *A.bestiarum* LMG13667.

S'ha pogut determinar el possible origen d'aquesta regió, que s'hauria adquirit per recombinació a partir d'una soca d'*A.eucrenophila*. El fragment adquirit (248 pb) s'estendria més enllà del final del gen *recA* abastant part del gen adjacent (*recX*). Malgrat la presència d'aquest fragment recombinant la filogènia s'ha mantingut, confirmant l'expressat anteriorment sobre l'impacte de la recombinació en les filogènies bacterianes construïdes a partir de gens conservats.

En la bibliografia s'han descrit alguns fets de recombinació en *Aeromonas* lligats al gen *recA* i altres gens. Silver i col·laboradors (2011) van ser els primers en descriure un fenomen de recombinació per aquest gènere. En el seu estudi van utilitzar únicament aïllats d'origen clínic pertanyents al complex *A.veronii*. L'evidència d'un possible fenomen de recombinació va sorgir com a conseqüència de l'observació d'incongruències filogenètiques observades en els arbres obtinguts per a cadascun dels gens considerats (*gyrB*, *recA*, *dnaJ*, *chiA*, *ascFG*, *ascV*, *aexT* i *aexU*) i l'arbre filogenètic corresponent a les seqüències concatenades. Aquest possible fenomen de recombinació va ser comprovat posteriorment utilitzant dos mètodes diferents, el *Phi Test* (Bruen *et al.* 2006) i el GARD (Kosakovsky Pond *et al.* 2006) detectant recombinació en els gens *ascFG*, *aexU*, *ascV*, *gyrB* i *dnaJ*.

El mateix any, Martino i col·laboradors (2011) van descriure un segon esdeveniment de recombinació en *Aeromonas*. En aquest cas van considerar totes les soques tipus de les espècie pertanyents al gènere descrites en aquell moment juntament amb un total de 77 soques, totes elles provinents de peixos, mol·luscs i crustacis. Es va realitzar un estudi MLST amb sis gens (*gyrB*, *groL*, *gltA*, *metG*, *ppsA* i *recA*). L'anàlisi de recombinació es va realitzar utilitzant xarxes filogenètiques amb el programa *SplitsTree*, el *Phi Test* i 5 mètodes per a la detecció de recombinació implementats en el programa RDP3. El resultat de l'anàlisi de les xarxes filogenètiques va suggerir que cap dels gens inclosos en l'estudi excepte *recA* estaven afectats significativament per un fenomen de recombinació. A més, la topologia dels arbres filogenètics obtinguts per a cadascun dels gens va ser molt similar. També es va calcular la taxa de recombinació/mutació amb el programa eBURST, utilitzat per identificar complexes clonals, la qual va suggerir una taxa de recombinació baixa. En base als resultats obtinguts, aquests autors van concloure que l'impacte de la recombinació en el gènere *Aeromonas* és baix.

Més recentment han aparegut dos treballs més sobre recombinació en *Aeromonas* (Roger *et al.* 2002, Farfán *et al.* 2013). En tots ells es fa referència a fenòmens de recombinació detectats a partir d'incongruències en les filogènies en algunes de les soques analitzades. Els autors conclouen que la recombinació està present en el gènere *Aeromonas*, almenys en algunes soques, però amb un freqüència baixa.

7.4 MODELATGE DE LA PROTEÏNA RecA

Per tal de veure si la regió adquirida per aquestes soques podria tenir algun impacte en l'estructura i/o la funcionalitat de la proteïna s'ha realitzat un modelatge molecular per homologia de la proteïna RecA de la soca tipus d'*A.bestiarum* CECT4227^T i una soca recombinant, *A.bestiarum* CECT5741. S'ha escollit una estructura de RecA d'*E.coli* obtinguda de les bases de dades, que ha estat identificada com el motlle que presentava una major

identitat (77,5-80%) i similitud (52-53%) de tots els disponibles, ja que en les bases de dades no es disposa d'un gran nombre d'estructures de procariotes. L'alineament de la seqüència aminoacídica (Figura 5.20), ha mostrat diferències entre les dues soques d' *A.bestiarum* i *E.coli* especialment en la regió terminal que sembla ser una regió que permet acumular més canvis, tan sinònims com no sinònims (Figura 5.4).

L'alineament s'ha realitzat amb diferents estructures cristal·litzades amb DNA de cadena senzilla i de cadena doble per tal d'observar possibles diferències en la predicció. Els resultats obtinguts per les dues seqüències d'*A.bestiarum* (soca tipus i recombinant), han estat pràcticament idèntics tant en la predicció de l'estructura secundària com terciària. Pel que fa als dominis estructurals de la proteïna RecA se'n poden diferenciar 3: el domini central, el domini N-terminal i el domini C-terminal, aquest dos últims considerablement més petits que el domini central. Aquests dominis es mantenen gairebé iguals en les estructures corresponents a les diferents espècies bacterianes. No obstant, són estructuralment molt diferent als seus homòlegs UvsX (bacteriòfag T4), RadA (arquees) i Dmc1 i Rad51 (llevats i eucariotes) (Figura 7.1)(Lusetti & Cox 2002; Cox 2007a; Wang *et al.* 2008).

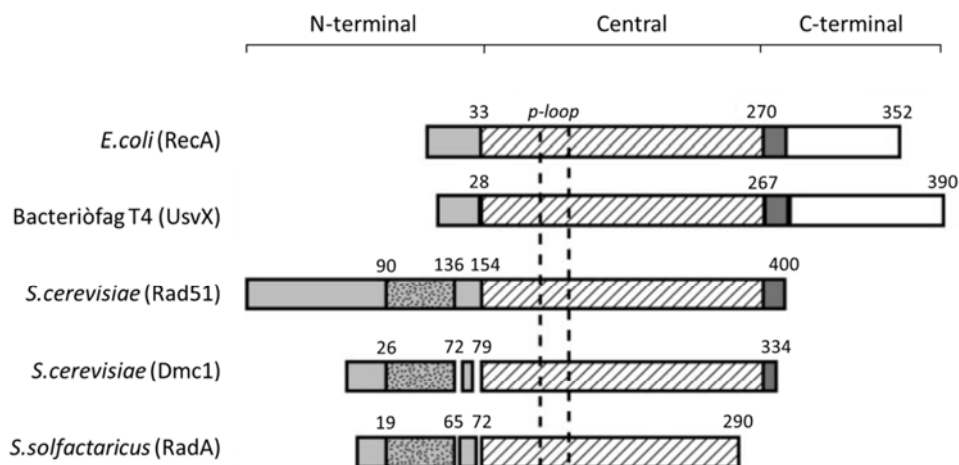


Figura 7.1 Dominis estructurals de la proteïna RecA i homòlegs de les diferents espècies bacterianes (Lusetti & Cox 2002).

Tot i que estructuralment totes aquestes proteïnes es caracteritzen perquè formen un filament molt similar sobre una cadena de DNA la principal diferència resideix en que RecA inclou un domini C-terminal absent en les altres i pel que fa al domini N-terminal s'observa tot el contrari, és a dir, RadA, Rad51 i Dmc1 presenten una regió que no comparteixen amb RecA.

Al comparar la predicció de l'estructura secundària de les soques tipus i recombinant d'*A.bestiarum*, s'han observat algunes discrepàncies (Figura 5.23) marcades amb un cercle en la figura, però corresponen a zones amb un grau de confiança en la predicció baix. De manera semblant, al comparar les dues soques d'*A.bestiarum* amb els models 2REB i 3CMW d'*E.coli* també s'han observat discrepàncies (Figura 5.24) que en la majoria dels casos corresponen també a regions a on la predicció indica un grau de confiança baix.

Malgrat que s'han determinat algunes diferències entre la proteïna RecA d'*A.bestiarum* i *E.coli*, no s'han observat diferències pel que fa als principals dominis funcionals de la proteïna, d'unió a ATP i els dos d'unió a DNA (de cadena senzilla i cadena doble) excepte en un dels llocs d'unió a DNA, el *Loop* L1, que en *A.bestiarum* presenta dues làmines β que no apareixen a *E.coli*. Aquests resultat però cal considerar-los amb precaució, perquè podria tractar-se d'un artefacte originat pel mètode de modelatge triat.

Una de les principals característiques de RecA és que els monòmers d'aquesta proteïna es poden trobar lliures o bé polimeritzant en direcció 5'-3' donant lloc a un filament helicoïdal la qual cosa fa que adoptin dues conformacions. Una vegada formen el filament, també es diferencien dos estats, l'inactiu i l'actiu (Cox 2007b). El filament inactiu correspon a la forma comprimida i està format per únicament monòmers amb o sense DNA i ADP. El filament inactiu es caracteritza perquè cada volta correspon a 6,2 monòmers i la distància entre cadascuna de les voltes del filament helicoïdal (també anomenat *pitch*) és d'aproximadament de 68Å. En canvi, el filament actiu correspon a la forma extensa, és a dir, a les molècules de RecA unides a DNA de cadena senzilla o doble juntament amb ATP o una molècula anàloga. En aquest estat, cada volta del filament correspon també a 6,2 monòmers però la distància entre cadascuna de les voltes és de 95Å (Bell 2005). Es per aquest motiu que s'ha realitzat el modelatge de les dues proteïnes RecA d'*A.bestiarum* amb una molècula sense DNA, una molècula amb DNA de cadena simple i una de cadena doble (2REB, 3CMW i 3CMT).

Alguns estudis han proposat que l'extrem C-terminal jugaria un paper important en la unió al DNA, ja que s'ha postulat que els residus entre les posicions 270-352 d'aquest domini formarien part d'una porta d'entrada a través de la qual el DNA accedeix al centre del filament (Kurumizaka *et al.* 1996). Altres autors com Cox i col·laboradors (McGrew & Knight 2003; Egger *et al.* 2003; Lusetti *et al.* 2003a) concreten encara més aquesta regió situant-la entre els aminoàcids 329-352. La seva proposta consisteix en que en absència o presència de determinats cofactors, aquesta regió adopta una conformació tancada o oberta que impedeix o afavoreix l'entrada del DNA (Lusetti *et al.* 2003a; b). La gran quantitat d'aminoàcids carregats negativament que hi ha també s'ha proposat com a possible zona d'interacció amb el DNA modulant al·lostèricament la proteïna per alteració de la conformació del domini central (Egger *et al.* 2003; Drees *et al.* 2004; Bell 2005).

Els resultats obtinguts mostren també que la regió recombinant presentaria un elevat grau de desordre i estaria constituït majoritàriament per aminoàcids carregats negativament. A més, tot i la gran variabilitat d'aquesta regió, aquesta característica es trobaria també a la regió final de totes les seqüències del gen *recA* de les altres espècies d'*Aeromonas* la qual cosa recolzaria la hipòtesi que podria actuar com una possible zona de modulació de l'acció de la proteïna. Sembla per tant que la regió recombinant no afectaria directament la unió de DNA, necessària per a la funció de la proteïna.

7.5 DATACIÓ DE L'ORIGEN d'*Aeromonas*

La idea de que els canvis mutacionals que es produeixen en la seqüència d'un gen són proporcionals al temps que fa que porten evolucionant (hipòtesi del rellotge molecular) va obrir les portes a poder realitzar datacions, a partir de seqüències gèniques, en aquelles espècies en les que no es disposa d'un registre fòssil, com és el cas dels procarïotes, o és escàs i la informació que aporta és esbiaixada. A més, els primers representants dels diferents llinatges s'espera que siguin pràcticament indistingibles i com a conseqüència d'això el registre fòssil tendeix a subestimar, i sovint considerablement, l'edat de diversificació de dos llinatges. En absència de registre fòssil, els estudis a datació a partir de dades moleculars s'han basat en punts de calibratge corresponents a esdeveniments biogeoquímics, altres dades moleculars, etc. Les edats determinades a partir d'estudis considerant diferents aproximacions no sempre han coincidit, la qual cosa ha estat objecte de molts debats sobre els possibles biaixos inherents al tipus de dades i els mètodes analítics utilitzats (Battistuzzi *et al.* 2010) de manera que per tal de determinar la utilitat d'aquestes estimacions, molts investigadors han comparat els temps obtinguts considerant diferents calibratges i programes informàtics en diferents tàxons i gens (Hedges & Kumar 2004; Ho *et al.* 2005; Linder *et al.* 2005; Rutschmann 2006; Pérez-Losada *et al.* 2006; Hug & Roger 2007; Lepage *et al.* 2007)

Com s'ha esmentat anteriorment, per determinar l'edat de diversificació és necessari utilitzar com a mínim un punt de calibratge tot i que seria desitjable poder-ne utilitzar més d'un, sempre i quan els escollits aportin dades fiables. En aquest estudi se n'ha utilitzat únicament un (no es disposa de més), i tot i que a priori podria fer pensar en un cert biaix no semblaria ser el cas, ja que els valors de datació obtinguts en les diferents aproximacions han estat perfectament equiparables. La utilització d'un únic punt de calibratge pot suposar no obstant una incertesa què és molt difícil de minimitzar en absència de punts de calibratge més fiables i la impossibilitat d'acceptar la hipòtesi d'una evolució seguint un rellotge molecular per a les seqüències. De tota manera si la filogènia obtinguda és robusta, com en aquest treball, és possible inclús amb un únic punt de calibratge fer afirmacions útils sobre el temps de divergència (Kuo & Ochman 2009).

El punt de calibratge utilitzat ha estat el temps de divergència entre *E.coli* i *S.typhimurium* proposat per Ochman i Wilson de 120-160 Ma. Aquesta dada coincideix aproximadament amb l'aparició del principal nínxol d'*E.coli*, l'intestí dels mamífers. Una dada similar per la divergència d'*E.coli* i *Salmonella* (100-130 Ma) s'ha obtingut considerant que les proteïnes distribuïdes universalment evolucionen a ritme similars en els bacteris entèrics i en mamífers (Doolittle *et al.* 1996) i un marge més ample es va obtenir en altres estudis en que el temps de divergència va ser determinat a partir d'evidències biogeoquímiques relacionades amb la diversificació dels cianobacteris (57-176 Ma) (Battistuzzi *et al.* 2004).

L'objectiu del treball no ha estat realitzar una datació precisa de l'origen del gènere *Aeromonas* i els diferents clades que conformen la filogènia d'aquest gènere. La datació dels diferents nodes és necessària per realitzar l'estudi quantitatiu d'un dels aspectes menys estudiat i més interessant dels marcadors filogenètics: el model de diversificació que expliqui

l'evolució de cadascun dels gens estudiats, en aquest cas els gens *mdh* i *recA* en el gènere *Aeromonas*.

Tots els cronogrames obtinguts de l'anàlisi de datació d'aquest treball suggereixen que l'ancestre comú del gènere *Aeromonas* hauria aparegut entre l'era del Paleozoic, i concretament entre els períodes Carbonífer (358 Ma) i Permià (299 Ma), i l'era del Mesozoic, en el període Triàsic (252 Ma). Les estimacions fetes sobre l'edat de l'ancestre comú varien en funció del conjunt de seqüències considerat (seqüències consens, seqüències de les soques tipus o l'arbre d'espècies) o del gen utilitzat (*mdh*, *recA* o ambdós) en la determinació (Taula 7.2). En tots els casos s'ha pogut observar que a partir de l'ancestre comú, hi va haver una primera separació d'un grup de tres espècies (*A.simiae*, *A.diversa* i *A.schubertii*) del grup majoritari. Aquestes tres espècies mantenen característiques diferents a la resta molt probablement degut a que s'han mantingut com un grup segregat després de la seva primària separació. Finalment, en referència a les espècies de diversificació més recent, cal comentar que en tots els casos, excepte en el cas del cronograma corresponent a l'arbre d'espècies del gen *recA*, aquestes han estat *A.bestiarum*-*A.piscicola*, *A.hydrophila*-*A.dhakensis* i algunes espècies del grup *A.veronii*.

Taula 7.2 Edats de divergència empíriques i teòriques estimades a partir de les seqüències.

Datació	Subgrup	tMRCA (Ma)	
		Empíric	Teòric
<i>mdh</i>	Seqüències consens	184	217
	Seqüències de les soques tipus	183	224
	Arbre d'espècies	190	214
<i>recA</i>	Seqüències consens	356	387
	Seqüències de les soques tipus	373	382
	Arbre d'espècies	348	368
<i>mdh+recA</i>	Concatenat de les seqüències consens	247	279
	Concatenat de les seqüències de les soques tipus	256	300
	Arbre d'espècies	237	227

Els resultats obtinguts, concorden perfectament amb els publicats per Lorén i col·laboradors (2014). Aquests autors, van determinar el temps de divergència del gènere *Aeromonas* a partir de 5 gens codificadors de proteïnes (*cpn60*, *dnaJ*, *gyrB*, *mdh* i *rpoD*) corresponents a la majoria de les espècies tipus i subespècies descrites en aquell moment i considerant el mateix punt de calibratge que l'utilitzat en aquest estudi. Segons les dades que van obtenir, l'origen d'*Aeromonas* se situaria entre 248-266 Ma en funció del mètode de reconstrucció del cronograma utilitzat (*Maximum Likelihood Approach* o MLA i Bayesià, respectivament). En aquest treball, els valors de datació obtinguts varien també lleugerament en funció de les seqüències analitzades i el gen escollit (Taula 7.2). Cal destacar que els valors de temps de divergència obtinguts a partir dels cronogrames de les seqüències dels dos gens (*mdh+recA*), entre 237-256 Ma, són pràcticament idèntics als determinats per Lorén i col·laboradors.

7.6 MODEL I TAXA DE DIVERSIFICACIÓ EN *Aeromonas*

Malgrat la gran quantitat de dades moleculars obtingudes en els darrers temps, aquesta informació generalment no ha estat utilitzada en estudis per determinar el model i els mecanismes d'especiació dels procariotes ni per clarificar les diferents hipòtesis que existeixen actualment relacionat amb el concepte d'espècies procariota (Rossello-Mora & Amann 2001; Cohan 2002; Fraser *et al.* 2009). Per això, les controvèrsies relacionades amb el món procariota causen més discussions i són més polèmiques que les que fan referència als eucariotes ja que els patrons evolutius han estat àmpliament documentats en organismes superiors com animals i plantes però són molt menys coneguts en microorganismes tot i que constitueixen la major part de la diversitat existent. Els models macroevolutius establerts pels eucariotes escassament s'han aplicat als procariotes (Drummond *et al.* 2006) i l'origen i la manera com es diversifica un llinatge bacterià segueix sent encara ignorat. A la literatura hi ha únicament algunes referències sobre la diversificació en els bacteris (Martin *et al.* 2004; Vinuesa *et al.* 2009; Morlon *et al.* 2013, Loren *et al.*, 2014) i en la majoria dels casos, les anàlisis no han estat tant completes com en el cas dels organismes superiors. Degut a aquesta manca de dades, el coneixement obtingut de l'estudi de la taxa i el patró de diversificació d'un gènere bacterià a partir de diferents aproximacions com les realitzades en aquest treball podria ser útil per entendre millor l'evolució dels procariotes ja que tenen papers rellevants en la salut i l'economia, així com en els processos que afecten als ecosistemes naturals (Morlon *et al.* 2013). Una part d'aquest treball ha consistit en determinar el model i la taxa de diversificació d'*Aeromonas* aplicant mètodes que, encara que poc emprats en procariotes, han mostrat la seva utilitat en eucariotes. Per això s'ha proposat tractar aquests aspectes de la diversificació del gènere *Aeromonas*, estudiant l'evolució dels gens *mdh* i *recA* en les espècies actualment descrites d'aquest gènere.

Per abordar l'estudi de diversificació és necessari utilitzar tal com s'ha comentat únicament una seqüència per cada espècie. Aquest fet podria produir un cert grau de distorsió en els resultats degut al biaix que es pot produir a l'escollir les seqüències. En aquest treball, s'han analitzat tres conjunts de dades diferents per tal d'avaluar la importància d'aquest biaix; les seqüències consens de totes les soques pertanyents a una mateixa espècie, les seqüències de les soques tipus i els arbres d'espècies de cada gen per separat i considerant els dos gens alhora.

Els diferents mètodes descrits per l'estudi del model i la taxa de diversificació temporal poden incloure també imprecisions en les determinacions encara que d'un origen diferent al de l'elecció de les soques. Efectivament, per estudiar i quantificar la variació temporal de les taxes de diversificació, s'han emprat un conjunt d'eines estadístiques, cada vegada més complexes (Rabosky 2006b). Típicament s'utilitzen les longituds de les branques d'una filogènia o un cronograma per estimar el temps dels esdeveniments d'especiació que s'han produït (nodes) i els temps transcorreguts entre els diferents fets d'especiació, per tal de comparar-los amb les distribucions esperades sota un procés de ramificació estocàstica (per exemple, el model de Yule) en el que les taxes han estat constants al llarg del temps (Wollenberg *et al.* 1996). Tots aquests mètodes presenten incerteses, fonamentalment en la determinació de la longitud de les branques, que tot i ser quantificables, sovint fan difícil decidir quin model és el que millor

descriu les dades. Per tal de minimitzar aquestes incerteses, per realitzar aquest estudi de la diversificació temporal, s'han utilitzat diferents mètodes amb plantejaments i fonaments estadístics diferents: una anàlisi qualitativa i quantitativa dels gràfics semilogarítmics del nombre de llinatges al llarg del temps (LTT plot), la determinació de l'estadístic Υ de Pybus i Harvey i un anàlisi del model de cladogènesi que millor s'ajusta a les dades mitjançant una aproximació de màxima versemblança descrita per Rabosky (Nee 2001; Rabosky 2006a; Morlon *et al.* 2010, 2011; Stadler 2010a). A més s'ha utilitzat el programa BAMM que permet realitzar proves específiques de la hipòtesi de cladogènesi mitjançant inferència bayesiana com una altre aproximació per tal de completar l'estudi.

La determinació del nombre esperat de llinatges al llarg del temps és un mètode àmpliament utilitzat per caracteritzar la diversificació d'un clade com una funció del temps (Nee *et al.* 1994; Ricklefs 2007). Els LTT plots derivats dels cronogrames bayesians han evidenciat que, en tots els casos, l'aparició de llinatges al llarg del temps en *Aeromonas* s'ajusta raonablement bé a una línia recta que suggereix una taxa constant de diversificació. A més, els gràfics no mostren en cap cas canvis abruptes que podrien indicar l'existència d'una major diversificació en el passat (*push of the past*) o un augment en la taxa de diversificació en el present (*pull of the present*). Aquests resultats estarien d'acord amb els obtinguts per Martin i col·laboradors al determinar el patró de diversificació d'una àmplia varietat de procariotes a partir de les seqüències del 16S rRNA, que en tots els casos van demostrar que era constant al llarg del temps (Martin *et al.* 2004). Més recentment Morlon i col·laboradors van utilitzar diferents seqüències per determinar el model de diversificació de *Borrelia burgdorferi sensu lato* observant, en contrast amb l'estudi anterior, que el procés no era constant, i presentava radiacions explosives seguides de ràpides baixades en les taxes de diversificació (Morlon *et al.* 2013). Aquest diferent comportament podria explicar-se si es considera que *B.burgdorferi* està associat a un hoste que podria limitar el flux gènic entre poblacions aïllades i donar un patró de diversificació similar al dels eucariotes.

Les taxes de diversificació obtingudes han variat entre 0,0074-0,0160 espècies/Ma segons el gen i el tipus de seqüències considerades en l'anàlisi, remarcant que en el cas de *recA* els valors de λ han estat inferiors als de *mdh* i als obtinguts amb els dos gens conjuntament (Taula 7.3). Els valors de λ amb major dispersió corresponen als obtinguts a partir de la recta de regressió, la qual cosa és lògica ja que tot i que l'ajust a la recta ha estat bo en tots els casos (valors d' r^2 entre 0,9051-0,9867) aquest no ha estat total.

Els valors de λ determinats en el treball, són semblants als obtinguts prèviament per Lorén i col·laboradors (2014), però tal com els autors comenten, serien valors més baixos que els determinats per altres organismes com peixos (0,075), ocells (0,186) mamífers (0,075) o plantes (0,083). No s'han pogut comparar les dades amb les d'altres procariotes perquè no hi ha cap treball sobre aquest tema més que l'esmentat anteriorment. Aquesta baixa taxa d'especiació semblaria no estar relacionada amb la taxa de substitució calculada ($3,1 \times 10^{-4}$ /lloc/Ma per *mdh* i $3,8 \times 10^{-4}$ /lloc/Ma per *recA*) que és similar a l'obtinguda per diferents espècies tant de procariotes com d'eucariotes (Doolittle *et al.* 1996; Degnan *et al.* 2004).

Taula 7.3 Valors estimats de λ obtinguts a partir de tres aproximacions diferents.

Taxa de diversificació (λ)				
	Subgrup	ML	Regressió lineal	IB*
<i>mdh</i>	Seqüències consens	0,0133	0,0156	0,013 - 0,016
	Seqüències de les soques tipus	0,0128	0,0160	0,012 - 0,017
	Arbre d'espècies	0,0134	0,0153	0,014
<i>recA</i>	Seqüències consens	0,0074	0,0088	0,008
	Seqüències de les soques tipus	0,0075	0,0086	0,008
	Arbre d'espècies	0,0078	0,0091	0,008
<i>mdh+recA</i>	Concatenat de les seqüències consens	0,0103	0,0129	0,010 - 0,012
	Concatenat de les seqüències de les soques tipus	0,0096	0,0125	0,009 - 0,013
	Arbre d'espècies	0,0127	0,0142	0,013

* Només es mostra la variació superior al 2%

Per confirmar si la taxa de diversificació és realment constant o ha variat al llarg del temps s'ha utilitzat el mètode de màxima versemblança per ajustar els temps de divergència (*branching times*) derivats dels cronogrames a diferents models de diversificació constants i variables (Rabosky 2006a; b) i s'ha determinat el model que millor s'ajusta a les dades a partir dels valors d'AIC obtinguts per cadascun d'ells, mitjançant el ΔAIC_{RC} . Tal com suggereix Rabosky, s'ha calculat la significació del valor d' ΔAIC_{RC} , simulant 5000 filogènies amb el mateix nombre d'espècies i la taxa de diversificació que les obtingudes amb les dades, i sota el model de Yule. El valor de P calculat de les distribucions resultants, mostra que en la majoria de casos, excepte les seqüències consens d'*mdh* i les de les soques tipus de *recA*, s'ha d'acceptar la hipòtesi alternativa de model de Yule 2-rate. Segons aquest model, al llarg del procés de cladogènesi s'hauria produït un canvi en la taxa de diversificació, és a dir, la diversificació s'hauria produït amb dues λ diferents. En aquest cas sembla ser que seria cap al final del procés, en els últims 20-40 Ma, quan s'hauria produït una desacceleració en el procés d'especiació que fins aleshores s'hauria mantingut constant.

Per corroborar aquests resultats s'ha comparat l'LTT plot empíric amb els obtinguts a partir de 1000 arbres simulats sota un model de Yule amb el mateix nombre d'espècies i taxa de diversificació obtingudes experimentalment. En tots els casos l'LTT plot obtingut a partir de les dades empíriques, entra dins els límits del 95% dels LTTplots de les filogènies simulades. Si s'analitzen amb detall en els gràfics, sembla que efectivament s'observa una petita inflexió en l'LTT plot que correspondria als darrers anys del procés de diversificació i que indicaria una possible desacceleració en aquest procés.

Un altre aproximació que s'ha realitzat, és la determinació de l'estadístic Υ de Pybus i Harvey per analitzar la constància de les taxes de diversificació. Tot i que els valors de Υ obtinguts han estat negatius, suggerint una possible desacceleració en les taxes de diversificació al llarg del temps, tots ells, excepte en el cas de les seqüències de les soques tipus d'*mdh+recA*, s'han situat dins de l'interval de confiança del 95% de la distribució de valors d'aquest estadístic

obtinguts per simulació de 1000 filogènies sota un model constant de Yule, i per tant, s'hauria d'acceptar que segueixen un model de Yule.

Finalment, per poder estimar estadísticament quina és la probabilitat de que les dades segueixin un model sense canvis (*shifts*) en la taxa de diversificació o amb un o més *shifts* s'ha utilitzat el programa BMM. Aquest programa permet a més, fer altres determinacions relatives al model de diversificació seguit. Cal remarcar que en tots els casos (*mdh*, *recA* i *mdh+recA*) els cronogrames corresponents a l'arbre d'espècies han presentat un valor de λ idèntic en totes les branques. En el cas dels cronogrames relatius a seqüències de soques tipus o seqüències consens, el valor de la taxa de diversificació ha presentat variacions en alguna branca (en ocasions en una única) encara que petites (Taula 7.4). Aquestes variacions es poden observar també en el gràfic corresponent a la taxa neta de diversificació al llarg del temps. Pel que fa al factor de Bayes, en tots els casos s'han obtingut valors superiors a 100, proporcionant una evidència decisiva, segons la interpretació del mètode, en contra de l'existència de canvis en els valors de λ .

Taula 7.4 Resultats de l'anàlisi de diversificació.

Anàlisi de diversificació							
	Subgrup	LTT plot	Regressió lineal (r^2)	Estadístic Υ	Test de Rabosky	Anàlisi BMM	
						0 shift	1 shift
<i>mdh</i>	SC*	Constant	0,9764	Yule	Yule	0,99	0,0084
	SST*	Constant	0,9765	Yule	Yule-2-rate	0,99	0,0070
	AE*	Constant	0,9867	Yule	Yule-2-rate	0,99	0,0089
<i>recA</i>	SC*	Constant	0,9677	Yule	Yule-2-rate	0,99	0,0064
	SST*	Constant	0,9479	Yule	Yule	1	0,0050
	AE*	Constant	0,9571	Yule	Yule-2-rate	1	0,0054
<i>mdh+ recA</i>	CSC*	Constant	0,9632	Yule	Yule-2-rate	0,99	0,0086
	CSST*	Constant	0,9574	No Yule	Yule-2-rate	0,99	0,0066
	AE*	Constant	0,9051	Yule	Yule-2-rate	0,99	0,0084

* SC, seqüències consens; SST, seqüències de les soques tipus; AE, arbre d'espècies; CSC, concatenat de les seqüències consens; CSST, concatenat de les seqüències de les soques tipus.

Resumint, amb els resultats obtinguts en aquesta anàlisi, es pot dir que *Aeromonas* hauria seguit un procés d'especiació constant d'acord a un model de Yule tot i que en els últims anys, podria ser possible, però poc probable (els valors de probabilitat d'un punt de canvi en la taxa de diversificació han estat en tots els casos entre 0,0050-0,0089), que s'hagi produït una desacceleració en el procés d'especiació.

Al calcular d'acord amb un model d'especiació de Yule els valors teòrics de tMRCA s'ha observat que els valors obtinguts han estat, en tots els casos, molt similars als determinats empíricament a partir dels cronogrames construïts (Taula 7.2). De manera semblant s'ha determinat també el nombre mig d'espècies que seria d'esperar que s'haguessin produït al llarg dels anys en que *Aeromonas* ha estat evolucionant i una vegada més els resultats teòrics

obtinguts han estat perfectament comparables als obtinguts experimentalment (Taula 7.5). Aquests resultats reafirmarien les conclusions de l'anàlisi realitzat.

A la Figura 7.2 es pot observar l'evolució del nombre de llinatges del gènere *Aeromonas* en els darrers 250 Ma, determinada a partir de l'arbre d'espècies dels dos gens. Segons suggereixen els resultats, i tal com s'ha comentat prèviament, l'evolució temporal de la diversificació d'*Aeromonas* sembla haver seguit un model exponencial. En aquesta figura s'ha superposat la gràfica a la clàssica corba de diversificació dels gèneres animals fòssils recopilats per Sepkoski (Sepkoski 2002). A partir d'aquestes dades, l'autor va establir un sèrie d'aspectes interessants de l'evolució dels animals com són entre altres, les extincions periòdiques o el concepte de "faunes evolutives".

Taula 7.5 Nombre d'espècies teòric obtingut a partir de les diferents aproximacions realitzades amb els gens *mdh* i *recA*.

Nombre d'espècies			
	Subgrup	Esperat	Desviació estàndard
<i>mdh</i>	Seqüències consens	23	16
	Seqüències de les soques tipus	21	14
	Arbre d'espècies	25	17
<i>recA</i>	Seqüències consens	28	19
	Seqüències de les soques tipus	33	22
	Arbre d'espècies	30	21
<i>mdh+ recA</i>	Concatenat de les seqüències consens	25	17
	Concatenat de les seqüències de les soques tipus	23	16
	Arbre d'espècies	40	28

Tal com es pot veure en la Figura 7.2 des del Permià-Triàsic la diversificació del gènere *Aeromonas* hauria transcorregut paral·lelament al augment del nombre dels gèneres animals. Aquesta diversificació hauria començat després de la gran extinció de fa aproximadament 250 Ma quan es va produir l'extinció de més del 90% de la vida terrestre i marina (Sahney & Benton 2008). Tot i que cal considerar que aquesta correlació no demostra cap relació causa efecte, és evident que l'augment de les espècies d'*Aeromonas* és paral·lel a la diversificació de la fauna Meso-Cenozoica. L'explosiva proliferació d'organismes multicel·lulars i la seva decisiva influència en l'estructura i funció dels ecosistemes va proporcionar un nou univers de nínxols ecològics potencials pels llinatges bacterians presents. La bona correlació observada entre el nombre de gèneres animals i la diversificació d'*Aeromonas* en els darrers 250 Ma podria explicar aquesta acurada correlació. Seria molt interessant ampliar aquest tipus d'anàlisi a altres gèneres bacterians per corroborar si es tracta d'un singularitat del gènere *Aeromonas*, cosa que sembla poc plausible, o si els altres taxa presentarien un comportament semblant.

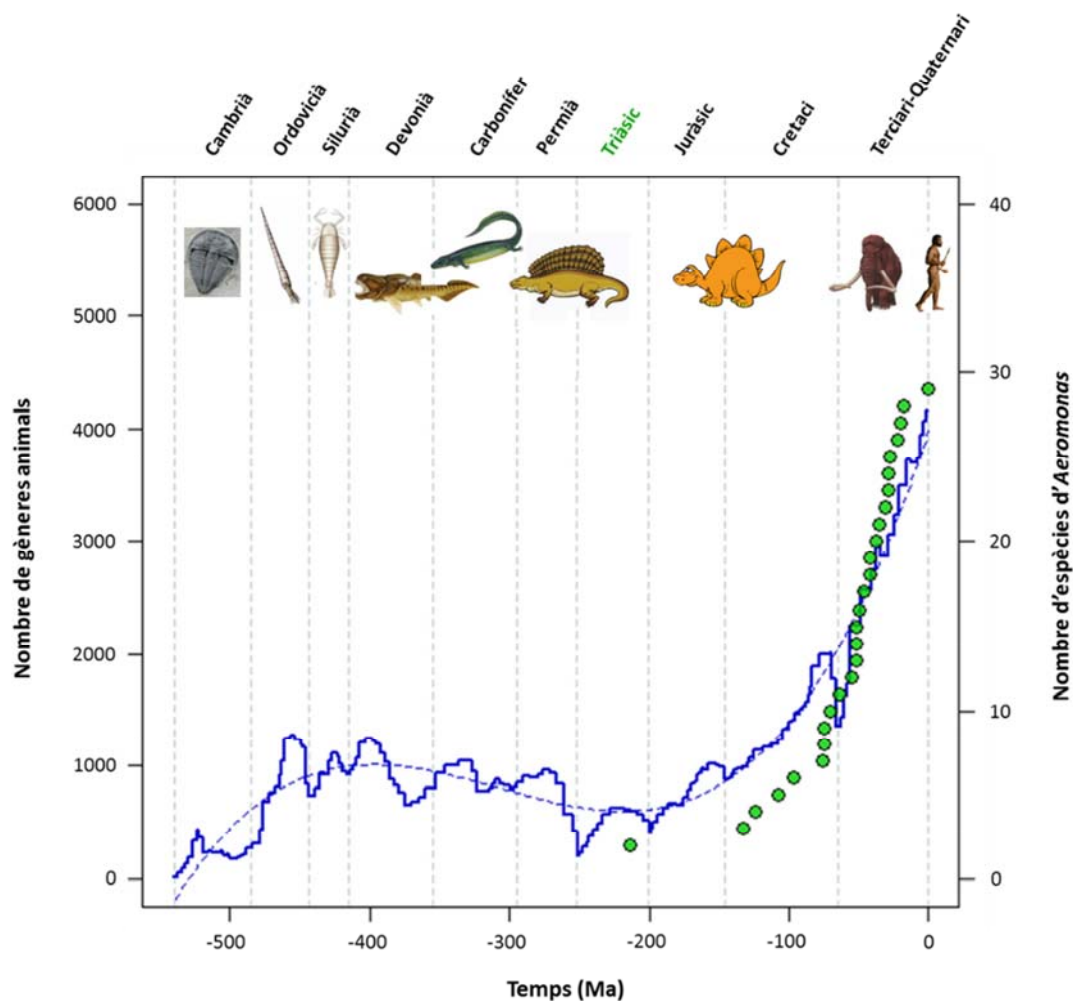


Figura 7.2 Diversificació al llarg del temps del gènere *Aeromonas* (verd) comparada amb la dels gèneres animals (blau) recopilats per Sepkoski (Sepkoski 2002; Rohde & Muller 2005).

Concloent, el treball descrit en aquesta memòria doctoral demostra que els gens codificadors de proteïna *mdh* i *recA* són uns bons marcadors filogenètics per establir filogènies robustes del gènere *Aeromonas*. L'anàlisi d'aquestes filogènies suggereix que l'especiació d'aquest gènere ha seguit models de diversificació constant, tot i que amb taxes significativament menors que les determinades pels organismes eucariotes. Aquest fet, tot i que tindrà que ser verificat per altres gèneres bacterians, planteja dos aspectes molt importants en l'estudi de la macroevolució dels organismes procarïotes: la similitud, també suggerida per altres autors (Shapiro *et al.* 2012) dels mecanismes d'especiació d'eucariotes i procarïotes, i el manteniment en aquests últims, tot i l'indubtable existència d'un cert grau de transferència genètica lateral, de les unitats fenotípiques i genotípiques que coneixem com espècies.

8. CONCLUSIONS

8. CONCLUSIONS

- 1.- Els gens estudiats, *mdh* i *recA*, han demostrat ser bons marcadors moleculars per les filogènies del gènere *Aeromonas*.
- 2.- L'evolució dels gens *mdh* i *recA* en el gènere *Aeromonas* no segueix un model de rellotge molecular estricta.
- 3.- Aquesta anàlisi filogenètica del gènere *Aeromonas* ha confirmat l'origen monofilètic d'aquest grup de bacteris. En tots els casos, s'ha obtingut un bon suport estadístic en els nodes dels grups majoritaris que corroboren que els arbres filogenètics generats són robusts.
- 4.- En el cas d'*Aeromonas* la inferència bayesiana ha proporcionat filogènies més coherents que les obtingudes per màxima versemblança.
- 5.- El resultat d'aquesta anàlisi filogenètica ha conduït a la reclassificació de 5 soques: *A.caviae* CECT5207, *A.hydrophila* CECT398, *A.hydrophila* LMG3767 i *A.sobria* CECT4250 en *A.culicicola*, i una soca d'*A.popoffii* CECT5210 en *A.bivalvium*.
- 6.- S'han detectat 4 soques d'*A.bestiarum* (AE147, CECT5741, CECT5742 i LMG13667) en les que s'ha identificat un fragment recombinant de 248 pb que correspon a la part terminal del gen *recA* i l'inici del gen adjacent *recX*. A més, s'han identificat els punts d'inici i final de la recombinació i els parentals principal i recombinant que han resultat ser *A.bestiarum* i *A.eucrenophila*, respectivament. El fragment recombinant no afecta a l'estructura ni a les funcions principals de la proteïna RecA, tot i que podria influir en la polimerització de les diferents subunitats.
- 7.- L'origen de l'ancestre comú del gènere *Aeromonas*, determinat a partir de diferents aproximacions, es situaria entre 370 i 180 Ma. També s'ha datat l'origen dels diferents clades que formen el gènere, que mostra que les darreres diversificacions s'haurien produït, aproximadament, en els últims 40 Ma.
- 8.- Les taxes de substitució nucleotídiques obtingudes per cada gen són similars ($3,1 \times 10^{-4}$ i $3,8 \times 10^{-4}$ substitucions/lloc/Ma), i del mateix ordre que les descrites per a diferents espècies tant d'eucariotes com de procariotes.
- 9.- Amb els resultats obtinguts en l'anàlisi de diversificació suggereixen que *Aeromonas* hauria seguit un procés d'especiació constant seguint el model de Yule, tot i que potser, que en els últims anys es possible però poc probable s'hagi produït una desacceleració en el procés d'especiació.
- 10.- Els diferents mètodes emprats per determinar la diversificació d'*Aeromonas* han donat resultats similars i comparables.

11.- Els valors de la taxa de diversificació han estat similars en tots els casos, i aproximadament igual a 0,01 espècies/Ma. Aquest valor és clarament inferior als descrits per diferents grups d'eucariotes.

12.- Els resultats obtinguts suggereixen que la diversificació del gènere *Aeromonas* s'origina en els últims 250 Ma, després de la gran extinció del Permià-Triàsic, i hauria transcorregut paral·lela a l'augment del nombre de gèneres animals.

9. ANNEXES

9. ANNEXES

Annex I. Soques utilitzades en l'estudi

Principals característiques de les 150 soques d'*Aeromonas* analitzades i el número d'accés del *GenBank* de les seqüències gèniques obtingudes.

Espècie / Complex ^a / Soca	Altres designacions	Font / Origen geogràfic	Nº d'accés del <i>GenBank</i>	
			<i>mdh</i>	<i>recA</i>
CECT4199 ^T	<i>A. veronii</i>	Angües malaltes de piscifactoria / Espanya	HM163292	KM260565
CECT4200		Angües malaltes de piscifactoria / Espanya	KM507368	KM260566
CECT4220		Femta / EUA	KM507369	KM260567
CECT4911		Femta / Suïssa	KM507370	KM260568
CECT4912		Femta / Suïssa	KM507371	KM260569
CECT8023 ^T		Aigua de reg / Austràlia	KM507372	KM260570
CECT4227 ^T		Peix malalt	HM163294	KM260547
112A		Aigua no potable / Espanya	JN660159	KM260548
559A		Aigua potable / Espanya	JN660162	KM260549
AE147		Aigua de llac / Finlàndia	JN660167	KM260550
AE203		Aigua de llac / Finlàndia	JN660169	JN660285
AE258		Aigua de llac / Finlàndia	JN660171	JN660286
CECT5741		Ambiental / Alemanya	JN660175	KM260551
CECT5742		Aigua / Suïssa	JN660176	KM260552
HE22		Aigua del llac Salajärvi / Finlàndia	JN660178	JN660294
HE73		Aigua del llac Iso-Kukka / Finlàndia	JN660179	KM260553
LMG13445		Humà / Alemanya	JN660180	JN660296
LMG13663		Intestí de salmó juvenil platejat	JN660183	KM260554
LMG13667		Aigua / EUA	JN660187	KM260555
868E ^T		Escopinya / Espanya	HM163295	KM260571
665N		Navalla / Espanya	KM507373	KM260572
CECT7862 ^T	<i>Aeromonas</i> sp.	Aigua de rierol d'una caverna / República Txeca	KM507374	KM260573

CECT838 ^T		Larves epizootiques en un conill d'indies	HM163296	KM260574
CECT4221 ^T	<i>A. hydrophila</i> subsp. <i>anaerogenes</i>	Emulsions d'oli utilitzades	HM163304	KM260575
106409		Clínic, humà / Espanya	KM507375	KM260576
1054090		Clínic, humà / Espanya	KM507376	KM260577
1061955		Clínic, humà / Espanya	KM507377	KM260578
1073261		Clínic, humà / Espanya	KM507378	KM260579
CECT4226		Emulsió d'oli	KM507379	KM260580
CECT5207		Sang humana / Espanya	KM507380	KM260581
CECT5208		Abscés humà / Espanya	KM507381	KM260582
CECT5237		Femta humana / Espanya	KM507382	KM260583
CECT5241		Femta humana / Espanya	KM507383	KM260584
LMG13459		Peix infectat	KM507384	KM260585
CECT5744 ^T	<i>A. hydrophila</i> subsp. <i>dhakensis</i>	Femta de nen amb diarrea / Bangladesh	HM163305	JN660346
LMG3769	<i>A. hydrophila</i> subsp. <i>hydrophila</i>	Aigua de riu / Índia	JN660226	JN660347
LMG19558		Femta de nen amb diarrea / Bangladesh	JN660228	JN660349
LMG19559		Femta de nen amb diarrea / Bangladesh	JN660229	JN660350
MDC47 ^T	<i>A. aquariorum</i>	Auari de peixos ornamental / Portugal	HM163293	JN660357
MDC310	<i>A. aquariorum</i>	Peix tropical / Portugal	JN660236	JN660358
MDC317	<i>A. aquariorum</i>	Peix tropical / Portugal	JN660237	JN660359
MDC573	<i>A. aquariorum</i>	Hospital Vega Baja / Espanya	JN660238	JN660360
CECT4254 ^T	<i>Aeromonas</i> sp. HG13	Ferida de cama humana / EUA	HM163298	KM260586
CECT5178	<i>Aeromonas</i> sp.	Ferida de cama humana / EUA	KM507385	KM260587
CECT4342 ^T		Anguilles europees sanes / Espanya	HM163299	KM260588
CECT4253	<i>Aeromonas</i> sp. HG11	Sutura de turmell humà / Nova Zelanda	HM163300	KM260589
CECT4343		Anguilles sanes / Espanya	KM507386	KM260590
CECT4826		Aigua d'una font natural / Regne Unit	KM507387	KM260591
CECT4856		Aigua del riu Mohawk / EUA	KM507388	KM260592
CECT4985		Ambient hospitalari / Alemanya	KM507389	KM260593

CECT4986	Aigua potable de pou / Finlàndia	KM507390	KM260594
CECT5025	Aigua potable de pou / Finlàndia	KM507391	KM260595
CECT4487 ^T	Femta humana / Índia	HM163301	KM260596
CECT4255 ^T	Femta humana / Índia	HM163325	KM260597
CECT4935	Apèndix humà / EUA	KM507392	KM260598
CECT4936	Femta humana / Tailàndia	KM507393	KM260599
CECT4937	Femta humana / Indonèsia	KM507394	KM260600
CECT4224 ^T	Peix d'aigua dolça	HM163302	KM260556
CECT4827	Ascites de carpa	KM507395	KM260557
CECT4853	Ascites de carpa	KM507396	KM260558
CECT4854	Aigua de font urbana / Alemanya	KM507397	KM260559
CECT4855	Aigua de font rural / Alemanya	KM507398	KM260560
717 ^T	Aigua del riu Muga / Espanya	HM163303	KM260601
CECT839 ^T	<i>A. hydrophila</i> subsp. <i>hydrophila</i>	HM163306	JN660318
CIP107985 ^T	Fetge d'una granota amb septicèmia / Tailàndia	HM163307	JN660345
1054148	<i>A. hydrophila</i> subsp. <i>ranae</i>	JN660200	JN660319
AE53	Clínic, humà / Espanya	JN660203	JN660322
AE180	Aigua de llac / Finlàndia	JN660206	JN660325
AE210	Aigua de llac / Finlàndia	JN660207	JN660326
CECT398	Aigua de llac / Finlàndia	KM507399	KM260602
JCM3967	Femta de nen amb diarrea / EUA	JN660217	JN660336
LMG3767	Desconegut	KM507400	KM260603
LMG13658	Humà / Índia	JN660221	JN660340
LMG21105	Femta / Suïssa	JN660225	JN660344
CECT4228 ^T	Llacuna d'aigües residuals / Marroc	HM163309	KM260604
CECT4813	Femta de pacient amb diarrea / EUA	KM507401	KM260605
CECT4814	Femta	KM507402	KM260606
CECT4815	Desconegut	KM507403	KM260607
	Femta		
<i>A. jandaei</i> (n=6)			

CECT4838	Sang / EUA	KM507404	KM260608
CECT4901	Ferida de cama / EUA	KM507405	KM260609
CECT4232 ^T	Efluent de piscifactoria	HM163310	KM260610
105A	Aigua no potable / Espanya	KM507406	KM260611
709OP	Ostra / Espanya	KM507407	KM260612
CECT4234	Estany de piscifactoria	KM507408	KM260613
848 ^T	Tellina / Espanya	HM163311	KM260614
93M	Musclo / Espanya	KM507409	KM260615
431E	Escopinya / Espanya	KM507410	KM260616
849T	Tellina / Espanya	KM507411	KM260617
869N	Navalla / Espanya	KM507412	KM260618
S1.2 ^T	Salmó atlàntic salvatge infectat / Espanya	HM163312	JN660307
R4	Peix malalt / Espanya	JN660191	JN660308
R9	Peix malalt / Espanya	JN660192	JN660309
R94	Peix malalt / Espanya	JN660193	JN660310
LMG17541 ^T	Planta de producció d'aigua potable / Bèlgica	HM163313	JN660311
CECT5210	Aigua de mar / Espanya	KM507413	KM260619
CECT5244	Aigua del riu Noguera Pallaresa / Espanya	KM507414	KM260620
CECT5250	Aigua de la riera de Merlès / Espanya	KM507415	KM260621
LMG17542	Planta de producció d'aigua potable / Bèlgica	JN660194	JN660312
LMG17543	Planta de producció d'aigua potable / Bèlgica	JN660195	JN660313
LMG17544	Planta de producció d'aigua potable / Bèlgica	JN660196	JN660314
LMG17545	Planta de producció d'aigua potable / Bèlgica	JN660197	JN660315
LMG17546	Aigua potable / Regne Unit	JN660198	JN660316
LMG17547	Planta de tractament d'aigua potable / Regne Unit	JN660199	JN660317
CECT7518 ^T	Aigua de regió càrstica / Alemanya	JN215542	KM260622
CECT7519	Aigua de regió càrstica / Alemanya	KM507416	KM260623

CECT894 ^T	<i>A. salmonicida</i> subsp. <i>salmonicida</i>	Salmó atlàntic del riu Cletter / Regne Unit	HM163317	JN660361
CECT5752 ^T	<i>A. salmonicida</i> subsp. <i>pectinolytica</i>	Aigua d'una cisterna / Espanya	HM163316	JN660399
CIP103210 ^T	<i>A. salmonicida</i> subsp. <i>masoucida</i>	Sang del cor d'un Sakuramasou	HM163315	JN660398
CIP104757 ^T	<i>A. salmonicida</i> subsp. <i>smithia</i>	Llaga de panerola / Regne Unit	HM163318	JN660400
LMG14900 ^T	<i>A. salmonicida</i> subsp. <i>achromogenes</i>	Truita comuna del riu Dee / Regne Unit	HM163314	JN660397
818E		Escopinya / Espanya	JN660249	JN660372
1062548	<i>A. salmonicida</i> / <i>A. hydrophila</i> (n=13)	Clínic, humà / Espanya	JN660252	JN660375
AE169		Aigua del mar Bàltic / Finlàndia	JN660253	JN660376
CECT5209	<i>A. salmonicida</i> subsp. <i>salmonicida</i>	Aigua de mar / Espanya	JN660258	JN660381
CECT5219		Pastís / Espanya	JN660262	JN660385
CECT5223		Musclo / Espanya	JN660265	JN660388
CIP57.50		Soca de control de qualitat	JN660271	JN660394
LMG3756	<i>A. hydrophila</i> subsp. <i>hydrophila</i>	Clínic, humà	JN660272	JN660395
A2-67 ^T	<i>A. sanarellii</i> (n=1)	Clínic, ferida / Taiwan	HM163319	KM260624
CIP103437 ^T		Abscés d'un front humà / EUA	HM163320	KM260625
367A		Aqüicultura / Sud Àfrica	KM507417	KM260626
CECT4241	<i>A. schubertii</i> (n=5)	Abscés d'un front	KM507418	KM260627
CECT4933		Ferida de la cama d'un home de 47 anys	KM507419	KM260628
CECT4934		Ferida de la cama d'una dona de 36 anys	KM507420	KM260629
CIP107798 ^T	<i>A. simiae</i> (n=2)	Femta de mono / França	HM163321	KM260630
CIP107797		Femta de mono / França	KM507421	KM260631
CECT4245 ^T		Carpa / França	HM163322	KM260632
CECT4248		Peix	KM507422	KM260633
CECT4250	<i>A. sobria</i> (n=5)	Peix	KM507423	KM260634
CECT4821		Luci	KM507424	KM260635
CECT4830		Peix	KM507425	KM260636
A2-50 ^T	<i>A. taiwanensis</i> (n=1)	Clínic, ferida / Taiwan	HM163323	KM260637

MDC91 ^T	Femta d'un nen de 5 anys / Suïssa	HM163324	KM260561
MDC92	Hisop de superfície d'una truita	KM507426	KM260562
MDC93	Aigua de l'aixeta	KM507427	KM260563
MDC94	Femta d'un adult asimptomàtic	KM507428	KM260564
CECT4257 ^T	<i>A.veronii</i> bv. <i>Veronii</i>	HM163327	KM260638
CECT4246 ^T	<i>A.veronii</i> bv. <i>Sobria</i>	HM163326	KM260639
CIP107763 ^T	<i>A.culicicola</i>	HM163297	KM260640
CECT4486 ^T	<i>A.ichthiosmia</i>	HM163308	KM260641
104714	Aigües superficials	KM507429	KM260642
1072742	Clínic, humà / Espanya	KM507430	KM260643
<i>A.veronii</i> / <i>A.veronii</i>	Clínic, humà / Espanya	KM507431	KM260644
(n=14)	Femta diarreica / EUA	KM507432	KM260645
CECT4261	Sinus maxil·lar / EUA	KM507433	KM260646
CECT4816	Ronyó de turbot	KM507434	KM260647
CECT4902	Aeromonas sp.	KM507435	KM260648
CECT4907	Ambiental / Alemanya	KM507436	KM260649
CECT7059	Femta / Suïssa	KM507437	KM260650
CECT7060	Subministrament d'aigua potable / Espanya	KM507438	KM260651
SH	Subministrament d'aigua potable / Espanya		
	Intestí mig de <i>Culex quinquefasciatus</i> / Índia		
	Intestí mig de <i>Aeromonas</i> al qual pertany.		

^a Per cada soca s'indica el nom d'espècie i el complex d'espècies d'*Aeromonas* al qual pertany.

Annex II. Medis de cultiu

▪ Agar de Citrat de Christensen:

Composició (en g/L)	
Citrat sòdic (Panreac 141653)	3,0
Glucosa (Sigma G7021)	0,2
Extracte de llevat (Oxoid LP0021)	0,5
Cisteïna hidrocloreada (Merck 9013290)	0,1
Citrat fèrric amònic (ICN 158040)	0,4
Potassi dihidrogen fosfat (Merck 1.04873)	1,0
Clorur de sodi (Sigma 31434)	5,0
Tiosulfat de sodi (Merck 1.06509)	0,08
Roig de fenol (Sharlau 120130)	0,012
Agar (Oxoid LP0011)	15,0
pH final	6,9 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.
Distribuir en tubs i esterilitzar en autoclau a 121°C durant 15 minuts.
Refredar i solidificar en posició inclinada (pendent en pic de flauta).

▪ Agar d'Elastina:

Composició (en g/L)	
Peptona (Difco 0118-01)	10,0
Elastina (Sigma E1625)	10,0
Clorur de calci (Panreac 141219)	11,6
Tris (Fluka 93349)	7,2
Agar (Oxoid LP0011)	20,0
pH final	7,8 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.
Esterilitzar en autoclau a 121°C durant 20 minuts.
Distribuir en plaques de Petri i deixar refredar i solidificar.

▪ Agar d'Esculina:

Composició (en g/L)	
Peptona (Difco 0118-01)	10,0
Esculina (Fluka 02350)	1,0
Citrat fèrric amònic (ICN 158040)	1,0
Agar (Oxoid LP0011)	20,0
pH final	7,4 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.
Esterilitzar en autoclau a 121°C durant 20 minuts.
Distribuir en plaques de Petri i deixar refredar i solidificar.

▪ **Agar de Midó (Adsa 1-283):**

Composició (en g/L)	
Peptona	5,0
Extracte de carn	3,0
Midó soluble	2,0
Agar	15,0
pH final	7,2 ± 0,1

Barrejar els components i portar a ebullició fins a dissolució total.
Esterilitzar en autoclau a 121°C durant 20 minuts.
Distribuir en plaques de Petri i deixar refredar i solidificar.

▪ **Agar de Mueller-Hinton (Oxoid CM0337):**

Composició (en g/L)	
Infusió de carn	300,0
Peptona de caseïna	17,5
Midó	1,5
Agar	17,0
pH final	7,3 ± 0,1

Barrejar els components i portar a ebullició fins a dissolució total.
Esterilitzar en autoclau a 121°C durant 20 minuts.
Distribuir en plaques de Petri i deixar refredar i solidificar.

▪ **Agar de Sulfídric de L-Cisteïna:**

MEDI A	
Composició (en g/L)	
Peptona (Difco 0118-01)	10,0
Clorur de potassi (Fluka 60129)	4,0
Agar (Oxoid LP0011)	15,0
pH final	7,2 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.
Distribuir en tubs i esterilitzar en autoclau a 121°C durant 15 minuts.
Mantenir a una temperatura d'entre 45-50°C.

MEDI B**Composició (en g/L)**

L-cisteïna (Merck 9013290)	1,0
Citrat fèrric amònic (ICN 158040)	1,0

Barrejar els components i esterilitzar per filtració (xeringa amb filtre de 0,2 µm).

Afegir medi B als tubs amb medi A (1B:10A), barrejar suaument i deixar refredar i solidificar.

- **Agar de Triptona de Soja, TSA (Pronadisa 1068):**

Composició (en g/L)

Peptona de caseïna	15,0
Peptona de soja	5,0
Clorur de sodi	5,0
Agar	15,0
+ Clorur de sodi (Sigma 71379)	10,0
pH final	7,3 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.

Esterilitzar en autoclau a 121°C durant 15 minuts.

Distribuir en plaques de Petri o en tubs inclinats i deixar refredar i solidificar.

- **Brou Base de Descarboxilacions Moeller:**

Composició (en g/L)

Peptona (Difco 0118-01)	5,0
Extracte de carn (Sharlau 07-075)	5,0
Dextrosa (Difco 0155-17)	0,5
Púrpura de bromocresol (Panreac 171546)	0,01
Vermell de cresol (Sigma 114472)	0,005
Piridoxal (Merck 7527)	0,005
+ L-lisina monohidroclorat (Merck 1147954)	10,0
+ L-ornitina monohidroclorat (Fluka 75470)	10,0
pH final	6,0 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.

Distribuir en tubs i esterilitzar en autoclau a 121°C durant 10 minuts.

Afegir oli de vaselina (Sigma 18512) estèril a tots els tubs.

▪ Brou Base de Fermentacions:

Composició (en g/L)	
Peptona (Difco 0118-01)	10,0
Extracte de carn (Sharlau 07-075)	3,0
Púrpura de bromocresol (Panreac 171546)	0,04
+ L-Arabinosa (Sigma A3256)	10,0
pH final	6,8 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.
Distribuir en tubs i esterilitzar en autoclau a 121°C durant 20 minuts.

▪ Brou de Mueller-Hinton (Oxoid CM0405):

Composició (en g/L)	
Infusió de carn	300,0
Peptona de caseïna	17,5
Midó	1,5
pH final	7,3 ± 0,1

Barrejar els components i portar a ebullició fins a dissolució total.
Distribuir en tubs i esterilitzar en autoclau a 121°C durant 10 minuts.

▪ Brou de Triptona de Soja (Oxoid CM0087):

Composició (en g/L)	
Triptona	10,0
Clorur de sodi	5,0
pH final	7,5 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.
Distribuir en tubs i esterilitzar en autoclau a 121°C durant 15 minuts.

▪ Brou de Triptona de Soja, TSB (CM0129):

Composició (en g/L)	
Peptona de caseïna	17,0
Peptona de soja	3,0
Clorur de sodi	5,0
Glucosa	2,5
Fosfat dipotàssic monobàsic	2,5
+ Clorur de sodi (Sigma 71379)	10,0
pH final	7,3 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.
Distribuir en tubs i esterilitzar en autoclau a 121°C durant 15 minuts.

- **Brou de Voges-Proskauer - Roig de metil (VPRM)** (Cultimed 413789.1210):

Composició (en g/L)	
Peptona	7,0
Glucosa	5,0
Fosfat monopotàssic	5,0
pH final	6,9 ± 0,2

Barrejar els components i portar a ebullició fins a dissolució total.
Distribuir en tubs i esterilitzar en autoclau a 121°C durant 15 minuts.

En tots els casos, els medis de cultiu emprats s'han controlat durant 24h a 30°C.

Annex III. Solucions, reactius i colorants▪ **Cristall Violeta:**

SOLUCIÓ A	
Composició (en g/L)	
Cristall Violeta (Fluka 61135)	100,0

La solució A es prepara amb etanol absolut (Panreac 141086).

SOLUCIÓ B	
Composició (en g/L)	
Oxalat amònic (Panreac 131136)	10,0

Barrejar els les solucions A i B preparades prèviament.

És recomanable filtrar el colorant abans d'utilitzar-lo.

Solució utilitzada en la tinció de Gram.

▪ **Lugol:**

Composició (en g/L)	
Iodur de potassi (Panreac 131542)	16,6
Iode resublimat (Panreac 141771)	3,3

Solució utilitzada en la tinció de Gram i la hidròlisi de midó.

▪ **Safranina:**

Composició (en mL)	
Solució alcohòlica de safranina (Panreac 251622) al 2,5%	10,0
Aigua destil·lada	100,0

Solució utilitzada en la tinció de Gram.

▪ **Reactiu de Barrit:**

SOLUCIÓ A	
Composició (en g/L)	
α -naftol (Fluka 70442)	60,0

La solució A es prepara amb etanol absolut (Panreac 141086).

SOLUCIÓ B**Composició (en g/L)**

Hidròxid de potassi (Panreac 121515)	40,0
--------------------------------------	------

Barrejar els les solucions A i B preparades prèviament.

Reactiu utilitzat a la prova del Voges-Proskauer.

- **Reactiu de Kovacs:**

Composició (en g/100 mL)

Para-dimetilaminobenzaldehyd (Panreac 161293)	5,0
Alcohol amílic (Panreac 141079)	75,0
Àcid clorhídric (Panreac 131020)	25,0

Barrejar el para-dimetilaminobenzaldehyd i l'alcohol amílic.

Escalfar a 60°C al bany maria.

Afegir l'àcid clorhídric pur gota a gota en campana de gasos

Reactiu utilitzat per a la prova de l'indol.

- **Reactiu d'Oxidasa:**

Composició (en g/L)

Tetrametil-p-fenilendiamoni (Fluka 87890)	0,1
---	-----

Preparar en el moment de realitzar la prova.

Reactiu utilitzat per a la prova de l'oxidasa.

- **Reactiu de Roig de metil:**

Composició (en g/L)

Roig de metil (Panreac 171617)	40,0
--------------------------------	------

El reactiu es prepara amb etanol absolut (Panreac 141086).

Reactiu utilitzar per a la prova del roig de metil.

Annex IV. Electroforesi en gel d'agarosa

- **Gel d'agarosa a l'1%:**

Composició	
Agarosa	0,5 g
TBE (1X)	50 mL
GelRed™	3 µL

- **Àcid etilendiamintetraacètic, EDTA 0,5M:**

Composició (en g/L)	
Àcid etilendiamintetraacètic (Fluka 03677)	186,1
pH final	8,0

Ajustar el pH a 8 addicionant hidròxid de sodi (Fluka 71690).

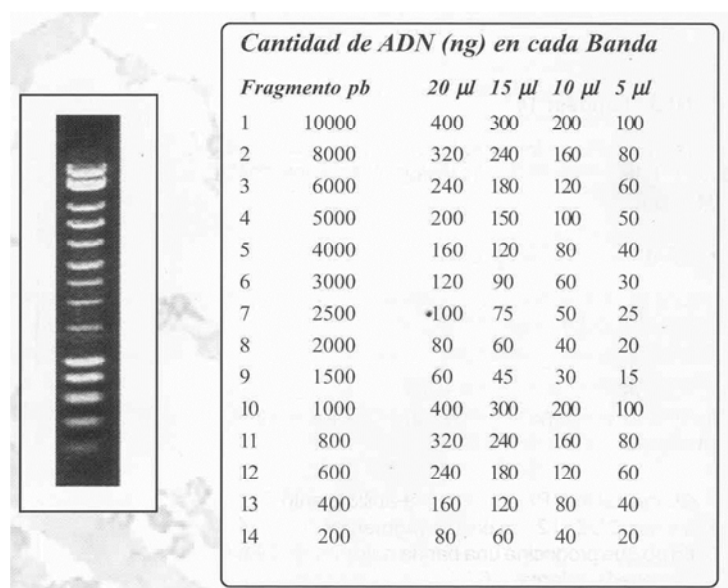
Solució utilitzada en el tampó de càrrega.

- **Tampó de càrrega:**

Composició	
EDTA 0,5M pH=8	1,0 mL
Glicerol (Merck 644 K3268091)	2,5 mL
Blau de bromofenol (Merck 180382)	0,025 g
Aigua Milli-Q estèril	qsp. 5 mL

Solució utilitzada per córrer les mostres en el gel d'electroforesi.

- **Patró de pesos moleculars utilitzat (Real® Escala N°1 RBMM1, Durviz)**



Annex V. Composició d'aminoàcids del gen *recA*

Freqüències d'aminoàcids expressades en percentatges per les 4 soques recombinants d'*A.bestiarum* (en gris) i un grup representatiu de soques de les espècies *A.bestiarum*, *A.eucrenophila* i *A.tecta*, considerant les seqüències del gen *recA* complet i de regions intragèniques (no recombinant i recombinant).

recA - Gen complet																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	11,3	1,4	5,7	8,5	2,8	9,3	0,8	7,7	8,2	9,1	2,8	3,7	2,8	4,2	3,4	5,4	3,4	7,4	0,3	1,7
Recombinants	11,0	1,4	5,7	8,5	2,8	9,6	0,8	7,9	8,2	9,1	2,8	3,7	3,1	4,0	3,4	5,1	3,4	7,4	0,3	1,7
<i>A.eucrenophila</i>	11,3	1,4	5,7	8,5	2,8	9,6	0,8	7,6	8,2	9,1	2,8	3,7	3,1	4,0	3,4	5,1	3,7	7,1	0,3	1,7
<i>A.tecta</i>	11,3	1,4	5,4	8,4	2,8	9,9	0,9	7,7	8,2	9,1	2,8	3,7	3,1	4,0	3,4	4,8	4,0	7,1	0,3	1,7

recA- Regió no recombinant (1-978)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	10,4	1,5	5,2	7,1	2,8	9,8	0,9	8,3	8,6	9,2	3,0	4,0	2,5	4,3	3,7	5,2	3,7	7,7	0,3	1,8
Recombinants	10,4	1,5	5,2	7,1	2,8	9,8	0,9	8,3	8,6	9,2	3,1	4,0	2,5	4,3	3,7	5,2	3,7	7,7	0,3	1,8
<i>A.eucrenophila</i>	10,7	1,5	5,2	7,1	2,8	9,8	0,9	8,0	8,6	9,2	3,1	4,0	2,5	4,3	3,7	5,2	4,0	7,4	0,3	1,8
<i>A.tecta</i>	10,7	1,5	5,2	7,1	2,8	9,8	0,9	8,0	8,6	9,2	3,1	4,0	2,5	4,3	3,7	4,9	4,3	7,4	0,3	1,8

recA - Regió recombinant (979-final)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	22,2	0,0	11,1	25,9	3,7	3,7	0,0	0,0	3,7	7,4	0,0	0,0	7,4	3,7	0,0	7,4	0,0	3,7	0,0	0,0
Recombinants	18,5	0,0	11,1	25,9	3,7	7,4	0,0	3,7	3,7	7,4	0,0	0,0	11,1	0,0	0,0	3,7	0,0	3,7	0,0	0,0
<i>A.eucrenophila</i>	18,5	0,0	11,1	25,9	3,7	7,4	0,0	3,7	3,7	7,4	0,0	0,0	11,1	0,0	0,0	3,7	0,0	3,7	0,0	0,0
<i>A.tecta</i>	17,8	0,0	7,5	25,2	3,7	11,2	0,0	3,7	3,7	7,5	0,0	0,0	11,2	0,0	0,0	3,7	0,9	3,7	0,0	0,0

Abreviatures: Ala, alanina; Cys, cisteïna; Asp, àcid aspàrtic; Glu, àcid glutàmic; Phe, fenilalanina; Gly, glicina; His, histidina; Ile, isoleucina; Lys, lisina; Leu, leucina; Met, metionina; Asn, asparragina; Pro, prolina; Gln, glutamina; Arg, arginina; Ser, serina; Thr, treonina; Val, valina; Trp, triptòfan i Tyr, tirosina.

Annex VI. Ús de codons del gen *recA*

Codons utilitzats per les 4 soques recombinants d'*A.bestiarum* (en gris) i un grup representatiu de soques de les espècies *A.bestiarum*, *A.eucrenophila* i *A.tecta*, considerant les seqüències del gen *recA* complet i de regions intragèniques (no recombinant i recombinant). Els resultats de les soques recombinants, d'*A.eucrenophila* i d'*A.tecta* es comparen amb els obtinguts pel clade *A.bestiarum*, de manera que si el codó utilitzat per codificar l'aminoàcid és el mateix s'indica amb un punt (·), si és diferent s'especifica el codó, i en absència de l'aminoàcid s'anota amb una barra inclinada (/).

<i>recA</i> - Gen complet																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	GCC	TGC	GAC	GAA	TTC	GGC	CAC	ATC	AAG	CTG	ATG	AAC	CCG	CAG	CGT	TCC	ACC	GTC	TGG	TAC
Recombinants	·	TGC TGT	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
<i>A.eucrenophila</i>	·	TGT	·	GAA GAG	·	·	·	·	·	·	·	·	·	·	CGC	·	·	·	·	·
<i>A.tecta</i>	·	·	·	GAG	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
<i>recA</i> - Regió no recombinant (1-978)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	GCC	TGC	GAC	GAA	TTC	GGC	CAC	ATC	AAG	CTG	ATG	AAC	CCG	CAG	CGT	TCC	ACC	GTC	TGG	TAC
Recombinants	·	TGC TGT	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
<i>A.eucrenophila</i>	·	TGT	·	·	·	·	·	·	·	·	·	·	·	·	CGC	·	·	·	·	·
<i>A.tecta</i>	·	·	·	GAG	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
<i>recA</i> - Regió recombinant (979-final)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	GCC GCG	/	GAC	GAA	TTC	GGT	/	/	AAG	CTC	/	/	CCG	CAA	/	TCC AGC	/	GTG	/	/
Recombinants	GCC	/	·	GAG	TTT	GGC	/	ATC	·	·	/	/	CCT	/	/	TCG	/	·	/	/
<i>A.eucrenophila</i>	GCC	/	·	GAG	·	GGC	/	ATC	·	·	/	/	CCT	/	/	TCG	/	·	/	/
<i>A.tecta</i>	GCC	/	·	GAG	·	GGC	/	ATC	·	·	/	/	CCT	/	/	TCG	ACC	·	/	/

Abreviatures: Ala, alanina; Cys, cisteïna; Asp, àcid aspàrtic; Glu, àcid glutàmic; Gly, glicina; His, histidina; Ile, isoleucina; Lys, lisina; Leu, leucina; Met, metionina; Asn, asparagina; Pro, prolina; Gln, glutamina; Arg, arginina; Ser, serina; Thr, treonina; Val, valina; Trp, triptòfan; Tyr, tirosina.

Annex VII. Composició d'aminoàcids del gen *recX*

Freqüències d'aminoàcids expressades en percentatges per les 4 soques recombinants d'*A.bestiarum* (en gris) i un grup representatiu de soques de les espècies *A.bestiarum*, *A.eucrenophila* i *A.tecta*, considerant les seqüències del gen *recX* i de regions intragèniques (recombinant i no recombinant).

<i>recX</i> (1-416)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	15,3	0,8	7,2	11,5	5,2	5,3	0,0	3,1	3,2	10,8	2,3	2,7	1,5	4,6	11,3	5,3	1,5	4,6	1,5	2,3
Recombinants	14,3	0,7	7,1	11,6	5,8	5,1	0,0	2,9	2,9	8,0	3,6	2,4	5,1	3,6	10,9	5,8	1,6	5,1	1,4	2,2
<i>A.eucrenophila</i>	13,6	0,7	5,9	12,9	5,1	5,8	1,4	2,9	3,0	8,1	3,6	1,3	5,1	2,9	11,4	7,2	0,9	4,3	1,4	2,2
<i>A.tecta</i>	12,8	1,5	7,5	10,6	6,1	6,1	2,2	4,4	3,1	8,6	3,7	1,5	3,9	2,0	10,8	7,5	0,7	3,3	1,5	2,2

<i>recX</i> - Regió recombinant (1-129)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	22,2	0,0	5,6	19,4	5,6	0,0	0,0	0,0	0,0	11,1	5,6	0,0	0,0	2,8	11,1	11,1	2,8	2,8	0,0	0,0
Recombinants	18,6	0,0	4,7	18,6	7,0	0,0	0,0	0,0	0,0	2,3	9,3	0,0	11,6	0,0	9,3	11,6	2,3	4,7	0,0	0,0
<i>A.eucrenophila</i>	18,6	0,0	4,7	18,6	7,0	0,0	0,0	0,0	0,0	2,8	9,3	0,0	11,6	0,0	9,3	11,2	2,3	4,7	0,0	0,0
<i>A.tecta</i>	19,4	0,0	4,8	16,4	7,9	0,0	0,0	2,4	0,0	3,6	9,7	0,0	7,9	0,0	9,7	14,5	0,0	3,6	0,0	0,0

<i>recX</i> - Regió no recombinant (130-416)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	12,6	1,1	7,8	8,4	5,1	7,4	0,0	4,2	4,4	10,7	1,1	3,8	2,1	5,3	11,4	3,2	1,1	5,3	2,1	3,2
Recombinants	12,4	1,1	8,2	8,4	5,3	7,4	0,0	4,2	4,2	10,5	1,1	3,4	2,1	5,3	11,6	3,2	1,3	5,3	2,1	3,2
<i>A.eucrenophila</i>	11,4	1,1	6,5	10,3	4,2	8,4	2,1	4,2	4,4	10,5	1,1	1,9	2,1	4,2	12,4	5,5	0,2	4,2	2,1	3,2
<i>A.tecta</i>	10,0	2,1	8,7	8,2	5,3	8,7	3,2	5,3	4,5	10,8	1,1	2,1	2,1	2,9	11,3	4,5	1,1	3,2	2,1	3,2

Abreviatures: Ala, alanina; Cys, cisteïna, Asp, àcid aspàrtic; Glu, àcid glutàmic; Phe, fenilalanina; Gly, glicina; His, histidina; Ile, isoleucina; Lys, lisina; Leu, leucina; Met, metionina; Asn, asparragina; Pro, prolina; Gln, glutamina; Arg, arginina; Ser, serina; Thr, treonina; Val, valina; Trp, triptòfan i Tyr, tirosina.

Annex VIII. Ús de codons del gen *recX*

Codons utilitzats per les 4 soques recombinants d'*A.bestiarum* (en gris) i un grup representatiu de soques de les espècies *A.bestiarum*, *A.eucrenophila* i *A.tecta*, considerant les seqüències del gen *recX* i de regions intragèniques (recombinant i no recombinant). Els resultats de les soques recombinants, d'*A.eucrenophila* i d'*A.tecta* es comparen amb els obtinguts pel clade *A.bestiarum*, de manera que si el codó utilitzat per codificar l'aminoàcid és el mateix s'indica amb un punt (·), si és diferent s'especifica el codó, i en absència de l'aminoàcid s'anota amb una barra inclinada (/).

<i>recA</i> - Gen complet																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	GCC	TGC	GAC	GAA	TTC	GGC	CAC	ATC	AAG	CTG	ATG	AAC	CCG	CAG	CGT	TCC	ACC	GTC	TGG	TAC
Recombinants	·	TGC TGT	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
<i>A.eucrenophila</i>	·	TGT	·	GAA GAG	·	·	·	·	·	·	·	·	·	·	CGC	·	·	·	·	·
<i>A.tecta</i>	·	·	·	GAG	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
<i>recA</i> - Regió no recombinant (1-978)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	GCC	TGC	GAC	GAA	TTC	GGC	CAC	ATC	AAG	CTG	ATG	AAC	CCG	CAG	CGT	TCC	ACC	GTC	TGG	TAC
Recombinants	·	TGC TGT	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
<i>A.eucrenophila</i>	·	TGT	·	·	·	·	·	·	·	·	·	·	·	·	CGC	·	·	·	·	·
<i>A.tecta</i>	·	·	·	GAG	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
<i>recA</i> - Regió recombinant (979-final)																				
Espècie	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
<i>A.bestiarum</i>	GCT GCC GCG	/	GAC	GAA	TTC	GGT	/	/	AAG	CTC	/	/	CCG	CAA	/	TCC AGC	/	GTG	/	/
Recombinants	GCC	/	·	GAG	TTT	GGC	/	ATC	·	·	/	/	CCT	/	/	TCG	/	·	/	/
<i>A.eucrenophila</i>	GCC	/	·	GAG	·	GGC	/	ATC	·	·	/	/	CCT	/	/	TCG	/	·	/	/
<i>A.tecta</i>	GCC	/	·	GAG	·	GGC	/	ATC	·	·	/	/	CCT	/	/	TCG	ACC	·	/	/

Abreviatures: Ala, alanina; Cys, cisteïna; Asp, àcid aspàrtic; Glu, àcid glutàmic; Phe, fenilalanina; Gly, glicina; His, histidina; Ile, isoleucina; Lys, lisina; Leu, leucina; Met, metionina; Asn, asparagina; Pro, prolina; Gln, glutamina; Arg, arginina; Ser, serina; Thr, treonina; Val, valina; Trp, triptòfan i Tyr, tirosina.

Annex IX. Alineament de seqüències de la proteïna RecA de diferents espècies bacterianes

Alineament de les seqüències aminoacídiques de RecA de la soca tipus d'*A.bestiarum* CECT4227^T i de l'aïllat recombinant CECT5741, juntament amb les seqüències d'altres 5 espècies que tenen l'estructura resolta i està dipositada a la base de dades PDB (2REB, *E.coli*; 1UBG, *Mycobacterium smegmatis*; 1MO4, *Mycobacterium tuberculosis*; 3HR8, *Thermotoga maritima* i 1XP8 *Deinococcus radiodurans*).

```

CECT4227T -----MDQNKQKALAAALGQIEKQFGKGSIMRLGDSKT-MDIEAISTGS
CECT5741 -----MDQNKQKALAAALGQIEKQFGKGSIMRLGDSKT-MDIEAISTGS
2REB -----AIDENKQKALAAALGQIEKQFGKGSIMRLGEDRS-MDVIETISTGS
1UBG -----MAQQAPDREKALELAMAQIDKNFGKGSVMRLGEEVR-QPISVIPTGS
1MO4 -----MTQTPDREKALELAVAQIEKSYGKGSVMRLGDEAR-QPISVIPTGS
3HR8 -----MPEEKQKKSIVLEKALKRIEENFGKGSIMILGDETQVPVEVIPTGS
1XP8  GSHMSKDATKEISAPTDAKERSKAIETAMSQIEKAFGKGSIMKLGAESEK-LDVQVSTGS
      :.:.:.: * : :.:.: :****:* * * . :.:.:****
      :.:.:.: * : :.:.: :****:* * * . :.:.:****

CECT4227T LSLDVALGIGGLPCGRIVEIYGPESSGKTTLTQLVIAEAQKKGKVCFAIDAEHALDPIYA
CECT5741  LSLDVALGIGGLPCGRIVEIYGPESSGKTTLTQLVIAEAQKKGKVCFAIDAEHALDPIYA
2REB    LSLDIALGAGGLPMGRIVEIYGPESSGKTTLTQLVIAAQRREGKTCFAIDAEHALDPIYA
1UBG    ISLDVALGIGGLPRGRVIEIYGPESSGKTTVALHAVANAQAAGGIAAFIDAEHALDPEYA
1MO4    IALDVALGIGGLPRGRVIEIYGPESSGKTTVALHAVANAQAAGGVAAAFIDAEHALDPDYA
3HR8    LAIDIATGVGGYPRGRIVEIFGQESSGKTTLALHAIIEAQKMGVAAAFIDAEHALDPVYA
1XP8    LSLDLALGVGGIPRGRITEIYGPESSGKTTLALAIVAQAQKAGGTCAFIDAEHALDPVYA
      :*: * * * * * : * : * * * * * : * : * * * * * : * : * * * * * : * : * * * * *

CECT4227T AKLGVNVDLLISQPDTEQALEICDMLVRSNAVDVIIVDSVAALTPKAEIEGEMGDSHV
CECT5741  AKLGVNVDLLISQPDTEQALEICDMLVRSNAVDVIIVDSVAALTPKAEIEGEMGDSHV
2REB    RKLGVVDIDNLLCSQPDTEQALEICDALARSGAVDVIIVDSVAALTPKAEIEGEGIDSHM
1UBG    KKLGVDTDSLVSQPDTEQALEIADMLVRSALDIIVIDSVAALVPRAEIEGEMGDSHV
1MO4    KKLGVDTDSLVSQPDTEQALEIADMLIRSGALDIVVIDSVAALVPRAEIEGEMGDSHV
3HR8    KNLGVDLKSLLSQPDHGEQALEIIVDELVRSVVDLIVVDSVAALVPRAEIEGAMQDMQV
1XP8    RALGVNTDELLVSQPDNGEQALEIMELLVRSGAIDVVVDSVAALVPRAEIEGDMGDSL
      ***: ..** ***** : * **..:****:*****.***:*** **

CECT4227T GLQARLMSQALRKLTANIKNANCLCIFINQIRMKIGVMFGSPETTGGNALKFYASVRLD
CECT5741  GLQARLMSQALRKLTANIKNANCLCIFINQIRMKIGVMFGSPETTGGNALKFYASVRLD
2REB    GLAARMMSQAMRKLKAGNLKQSNLLIFINQIRMKIGVMFGNPETTGGNALKFYASVRLD
1UBG    GLQARLMSQALRKMTGALNNSGTTAIFINQLREKIGVMFGSPETTGGKALKFYASVRLD
1MO4    GLQARLMSQALRKMTGALNNSGTTAIFINQLRDKIGVMFGSPETTGGKALKFYASVRMD
3HR8    GLQARLMSQALRKIAGSVNKSVAIVFTNQIRMKIGVMFGSPETTGGALKFYATMRME
1XP8    GLQARLMSQALRKLTAILSKTGTAIFINQVREKIGVMYGNPETTGGRALKFYASVRLD
      ** **:*****:***:.. :.:.: ** **:* ****:*:***** *****:***:

CECT4227T IRRI-GAIKEGDEVVGNETRVKVVKNKVAPPFKQAEFQIFYGVGISKEGELVDLVGKHKL
CECT5741  IRRI-GAIKEGDEVVGNETRVKVVKNKVAPPFKQAEFQIFYGVGISKEGELVDLVGKHKL
2REB    IRRI-GAVKEGENVVGSETRVKVVKNKIAAPPFKQAEFQIYEGEINIFYGELVDLVGKKEKL
1UBG    VRRI-ETLKDGTDAVGNRTRVKVVKNKVSPPFKQAEFDILYGGISREGSLIDMGVEHGF
1MO4    VRRV-ETLKDGTNAVGNRTRVKVVKNKCLAPPFKQAEFDILYGKGISREGSLIDMGVDQGL
3HR8    VRRG-EPIKEGKDVIGNVSVKIVKNKVAPPFKTAQTYIIYGKIDREYELFNIAVNEGI
1XP8    VRKIGQPTKVGNDAVANTVKIKTVKNKVAAPFKEVELALVYKGFQDQLSDLVGLAAMDID
      **: . * * :.:. . * **** .*** .: :.:. * * :.:. :.:.:

CECT4227T IDKAGAWYSYN---GE--KIGQGKANVMKLFSENKAMAAEVEARLRELLSGDAPA-EKP
CECT5741  IDKAGAWYSYN---GE--KIGQGKANVMKLFSENKAMAAEVEARLRELLSGAAPVDDKP
2REB    IEKAGAWYSYK---GE--KIGQGKANATAWLKDNPETAKEIEKKVRELLS--NPNSTPD
1UBG    IRKSGSWFTYE---GE--QLGQKENARKFLENTDVANEIEKKIKEKLGIGAVVTAEA-
1MO4    IRKSGAWFTYE---GE--QLGQKENARNFLVENADVADEIEKKIKEKLGIGAVVTDPS
3HR8    VDRKGSWYTYTTLKGEVSLGQGSNAVQFLKDNPEIAGEIERRIREKYGLLSVEKEEQR
1XP8    IKKAGSFYSYG---DE--RIGQGKEKTIAERPEMEQEIRDRVMAAIRAGNAGEAPAL
      : : * : : * . * : * * . : . : * : . :

CECT4227T VAADADELEAESEQEF
CECT5741  AAIEADEFEPEGE-ELE
2REB    FSVDDSEGVAETNEDF-
1UBG    -DDVLPAPVDF-----
1MO4    NDGVLPAVDF-----
3HR8    KEKSSGEEAS-----
1XP8    APAPAAPEAAEA-----

```

Abreviatures: (*) indica que totes les posicions presenten el mateix residu; (:) que les posicions que han patit canvis conservatius i (.) indica canvis no conservatius.

10. BIBLIOGRAFIA

10. BIBLIOGRAFIA

- Abbott SL, Cheung WKW, Janda JM (2003) The Genus *Aeromonas*: biochemical characteristics, atypical reactions, and phenotypic identification schemes. *Journal of Clinical Microbiology*, **41**, 2348-2357.
- Achtman M, Wagner M (2008) Microbial diversity and the genetic nature of microbial species. *Nature reviews. Microbiology*, **6**, 431-440.
- Aihara H, Ito Y, Kurumizaka H *et al.* (1997) An interaction between a specified surface of the C-terminal domain of RecA protein and double-stranded DNA for homologous pairing. *Journal of Molecular Biology*, **274**, 213-221.
- Alavandi S, Ananthan S (2003) Biochemical characteristics, serogroups, and virulence factors of *Aeromonas* species isolated from cases of diarrhoea and domestic water samples in Chennai. *Indian Journal of Medical Microbiology*, **21**, 233-238.
- Altwegg M, Steigerwalt AG, Altwegg-Bissig R, Lüthy-Hottenstein J, Brenner DJ (1990) Biochemical identification of *Aeromonas* genospecies isolated from humans. *Journal of Clinical Microbiology*, **28**, 258-264.
- Arbogast BS, Edwards SV., Wakeley J, Beerli P, Slowinski JB (2002) Estimating Divergence Times From Molecular Data on Phylogenetic and Population Genetic Timescales. *Annual Review of Ecology and Systematics*, **33**, 707-740.
- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195-201.
- Barraclough TG, Nee S (2001) Phylogenetics and speciation. *Trends in Ecology & Evolution*, **16**, 391-399.
- Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH (2007) Phylogenetic reconstruction. In *Evolution*, Cold Spring Harbor Laboratory Press.
- Battistuzzi FU, Feijao A, Hedges SB (2004) A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evolutionary Biology*, **4**, 44.
- Battistuzzi FU, Filipowski A, Hedges SB, Kumar S (2010) Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Molecular Biology and Evolution*, **27**, 1289-1300.
- Bauer A, Kirby W, Sherris J, Turck M (1966) Antibiotic susceptibility testing by a standardized single disk method. *American Journal of Clinical Pathology*, **45**.
- Beaz-Hidalgo R, Latif-Eugenín F, Hossain MJ (2015) *Aeromonas aquatica* sp. nov., *Aeromonas finlandiensis* sp. nov. and *Aeromonas lacus* sp. nov. isolated from Finnish waters associated with cyanobacterial blooms. *Systematic and Applied Microbiology*, **38**, 161-168.

- Beaz-Hidalgo R, Martínez-Murcia A, Figueras MJ (2013) Reclassification of *Aeromonas hydrophila* subsp. *dhakensis* (Huys et al. 2002) and *Aeromonas aquariorum* (Martínez-Murcia et al. 2008) as *Aeromonas dhakensis* sp. nov. comb nov. and emendation of the species *Aeromonas hydrophila*. *Systematic and Applied Microbiology*, **36**, 171-176.
- Bell CE (2005) Structure and mechanism of *Escherichia coli* RecA ATPase. *Molecular Microbiology*, **58**, 358-66.
- Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, **27**, 343-50.
- Benkert P, Künzli M, Schwede T (2009) QMEAN server for protein model quality estimation. *Nucleic Acids Research*, **37**, 510-514.
- Bennett-Lovsey RM, Herbert AD, Sternberg MJE, Kelley LA (2007) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins*, **70**, 611-625.
- Bergey H, Breed S (1966) *Bergey's Manual of Determinative Bacteriology*. Ed. Baltimore: Williams & Wilkins.
- Bergey D, Buchanan R, Gibbons N (1974) *Bergey's Manual of Determinative Bacteriology*. Ed. Baltimore: Williams & Wilkins.
- Berman HM, Kleywegt GJ, Nakamura H, Markley JL (2013) The future of the protein data bank. *Biopolymers*, **99**, 218-222.
- Berman HM, Westbrook J, Feng Z *et al.* (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235-242.
- Biasini M, Bienert S, Waterhouse A *et al.* (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, **42**, 252-258.
- Boni MF, Posada D, Feldman MW (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, **176**, 1035-1047.
- Bradley P, Misura KMS, Baker D (2005) Toward high-resolution de novo structure prediction for small proteins. *Science*, **309**, 1868-1871.
- Brien TFO (2002) Emergence, spread, and environmental effect of antimicrobial resistance: How use of an antimicrobial anywhere can increase resistance to any antimicrobial anywhere else. *Clinical Infectious Diseases*, **34**, 78-84.
- Brocks JJ, Buick R, Summons RE, Logan G a. (2003) A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochimica et Cosmochimica Acta*, **67**, 4321-4335.
- Bromham L, Penny D (2003) The modern molecular clock. *Nature Reviews. Genetics*, **4**, 216-224.

- Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. *Genetics*, **172**, 2665-2681.
- Buchan DW a, Minneci F, Nugent TCO, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research*, **41**, 349-357.
- Campbell MJ, Davis RW (1999) Toxic mutations in the recA gene of E. coli prevent proper chromosome segregation. *Journal of Molecular Biology*, **286**, 417-435.
- Cardozo T, Totrov M, Abagyan R (1995) Homology modeling by the ICM method. *Proteins*, **23**, 403-414.
- Caro-Quintero A, Konstantinidis KT (2012) Bacterial species may exist, metagenomics reveal. *Environmental Microbiology*, **14**, 347-355.
- Carrine E (2004) Evolutionary timing of the origins of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic. *Geobiology*, **2**, 1-20.
- Chai B, Wang H, Chen X (2012) Draft genome sequence of high-melanin-yielding *Aeromonas media* strain WS. *Journal of Bacteriology*, **194**, 6693-6694.
- Chao CM, Lai CC, Tang HJ, Ko WC, Hsueh P-R (2013) Skin and soft-tissue infections caused by *Aeromonas* species. *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology*, **32**, 543-547.
- Charif D, Lobry JR (2007) Seqin{R} 1.0-2: a contributed package to the {R} project for statistical computing devoted to biological sequences retrieval and analysis. In: *Structural approaches to sequence evolution: molecules, networks, populations* (eds Bastolla M, Porto M, Roman HE, Vendruscolo M), 207-232. Springer Verlag.
- Carnahan AM, Behram S, Joseph, SW (1991) Aerokey II: A flexible key for identifying clinical *Aeromonas* species. *Journal of Clinical Microbiology*, **29**, 2843-2849.
- Chen Z, Yang H, Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature*, **453**, 489-484.
- Cohan FM (2002) What are bacterial species? *Annual Review of Microbiology*, **56**, 457-487.
- Cohan FM, Koeppel AF (2008) The origins of ecological diversity in prokaryotes. *Current Biology: CB*, **18**, 1024-1034.
- Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Research*, **36**, 197-201.
- Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science : a publication of the Protein Society*, **2**, 1511-1519.
- Colston SM, Fullmer MS, Beka L, Lamy B, Gogarten JP (2014) Bioinformatic Genome Comparisons for Taxonomic and Phylogenetic Assignments Using *Aeromonas* as a Test Case. *Molecular Biology*, **5**, 1-13.

- Cordaux R, Michel-Salzat A, Bouchon D (2001) *Wolbachia* infection in crustaceans: novel hosts and potential routes for horizontal transmission. *Journal of Evolutionary Biology*, **14**, 237-243.
- Cox MM (2007a) Regulation of bacterial RecA protein function. *Critical Reviews in Biochemistry and Molecular Biology*, **42**, 41-63.
- Cox MM (2007b) Motoring along with the bacterial RecA protein. *Nature Reviews. Molecular Cell Biology*, **8**, 127-138.
- Davison J (1999) Genetic exchange between bacteria in the environment. *Plasmid*, **42**, 73-91.
- Degnan P, Lazarus A, Brock C, Wernegreen J (2004) Host symbiont stability and fast evolutionary rates in bacterium association: cospeciation of *Camponotus* species and their endosymbionts, *Candidatus Blochmannia*. *Systematic Biology*, **53**, 95-110.
- Denamur E, Matic I (2006) Evolution of mutation rates in bacteria. *Molecular Microbiology*, **60**, 820-827.
- Doolittle RF, Feng D, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science*, **271**, 470-477.
- Drees JC, Lusetti SL, Cox MM (2004) Inhibition of RecA protein by the *Escherichia coli* RecX protein: modulation by the RecA C terminus and filament functional state. *The Journal of Biological Chemistry*, **279**, 52991-52997.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biology*, **4**.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Drummond AJ, Suchard M a, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969-1973.
- Duploux A, Iturbe-Ormaetxe I, Beatson S a *et al.* (2013) Draft genome sequence of the male-killing *Wolbachia* strain wBol1 reveals recent horizontal gene transfers from diverse sources. *BMC Genomics*, **14**, 20.
- Egelman EH, Stasiak A (1993) Electron microscopy of RecA-DNA complexes: Two different states, their functional significance and relation to the solved crystal structure. *Micron*, **24**, 309-324.
- Eggler AL, Lusetti SL, Cox MM (2003) The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. *The Journal of Biological Chemistry*, **278**, 16389-16396.
- Etherington GJ, Dicks J, Roberts IN (2005) Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics*, **21**, 278-281.

-
- Euzéby J (2015) List of prokaryotic names with standing in nomenclature. <http://www.bacterio.net/>
- Farfán M, Miñana-Galbis D, Fusté MC, Lorén JG (2009) Divergent evolution and purifying selection of the *flaA* gene sequences in *Aeromonas*. *Biology Direct*, **4**, 23.
- Farfán M, Miñana-Galbis D, Garreta A, Lorén JG, Fusté MC (2010) Malate dehydrogenase: a useful phylogenetic marker for the genus *Aeromonas*. *Systematic and Applied Microbiology*, **33**, 427-435.
- Farfán M, Albarral V, Sanglas A, Lorén JG, Fusté MC (2013a) 11 . The effect of recombination in *Aeromonas*. Ed. Recent Advances in Pharmaceutical Sciences III, 179-193.
- Farfán M, Spataro N, Sanglas A, Albarral V, Lorén JG, Bosch, E Fusté MC (2013b) Draft Genome Sequence of the *Aeromonas diversa* Type Strain. *Genome Announcements*, **1**, 6-7.
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783-791.
- Felsenstein J (1988) Phylogenies from molecular sequence data: inference and reliability. *Annual Review of Genetics*, **22**, 521-565.
- Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, Sunderland, Massachusetts.
- Feng D, Cho G, Doolittle R (1997) Determining divergence times with a protein clock: Update and reevaluation. *Proceedings of the National Academy of Sciences*, **94**, 13028-13033.
- Fontaneto D, Tang CQ, Obertegger U, Leasi F, Barraclough TG (2012) Different diversification rates between sexual and asexual organisms. *Evolutionary Biology*, **39**, 262-270.
- Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science*, **323**, 741-746.
- Fraser C, Hanage WP, Spratt BG (2007) Recombination and the nature of bacterial speciation. *Science*, **315**, 476-480.
- Fusté M, Farfán M, Miñana-Galbis D, Albarral V, Sanglas A, Lorén JG (2012) Population Genetics of the “*Aeromonas hydrophila* Species Complex.” In: *Studies in Population Genetics*, 39-54.
- Galtier N, Daubin V (2008) Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **363**, 4023-4029.
- Garcia-Vallvé S, Romeu A, Palau J (2000) Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome Research*, **10**, 1719-1725.
- Gasteiger E, Hoogland C, Gattiker A, Dubaud S, Wilkins, MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: *The proteomics*, 571–608. The Proteomics Protocols Handbook, Humana Press.

- Gerhardt P, Murray R, Wodd W, KRIEG N (1994) *Methods for general and molecular bacteriology*. American Society for Microbiology, Washington, D.C.
- Gibbs MJ, Armstrong JS, Gibbs a. J (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, **16**, 573-582.
- Goward CR, Nicholls DJ (1994) Malate dehydrogenase: a model for structure, evolution, and catalysis. *Protein Science*, **3**, 1883-1888.
- Von Graevenitz A (2007) The role of *Aeromonas* in diarrhea: a review. *Infection*, **35**, 59-64.
- Gruber T, Eisen J, Gish K, Bryant D (2011) The phylogenetic relationship of *Chlorobium tepidum* and *Chloroflexus aurantiacus* based upon their RecA sequences. *FEMS Microbiology Letters*, **162**, 53-60.
- Guex N, Peitsch M (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714-2723.
- Guex N, Peitsch MC, Schwede T (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis*, **30 Suppl 1**, 162-173.
- Guindon S, Dufayard JF, Lefort V *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.
- Guindon S, Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, **52**, 696-704.
- Haas J, Roth S, Arnold K *et al.* (2013) The Protein Model Portal: a comprehensive resource for protein structure and model information. *Database: The Journal of Biological Databases and Curation*, **2013**.
- Han H-J, Kim D-Y, Kim W-S *et al.* (2011) Atypical *Aeromonas salmonicida* infection in the black rockfish, *Sebastes schlegeli Hilgendorf*, in Korea. *Journal of Fish Diseases*, **34**, 47-55.
- Hastings W (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97.
- Hedges SB, Chen H, Kumar S *et al.* (2001) A genomic timescale for the origin of eukaryotes. *BMC Evolutionary Biology*, **1**.
- Hedges SB, Kumar S (2004) Precision of molecular time estimates. *Trends in Genetics*, **20**, 242-247.
- Hein J (1990) Reconstructing evolution of sequences subject to recombination using parsimony. *Molecular Tools for Screening Biodiversity*, **200**, 185-200.
- Heled J, Bouckaert R, Drummond AJ, Xie W (2013) * BEAST in BEAST 2.0 Estimating species trees from multilocus data, 1-18.

- Heled J, Drummond AJ (2010) Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, **27**, 570-580.
- Hinrichs K-U (2002) Microbial fixation of methane carbon at 2.7 Ga: Was an anaerobic mechanism possible? *Geochemistry Geophysics Geosystems*, **3**.
- Ho SYW, Phillips MJ, Drummond AJ, Cooper A (2005) Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution*, **22**, 1355-1363.
- Holmes DE, Nevin KP, Lovley DR (2004) Comparison of 16S rRNA, *nifD*, *recA*, *gyrB*, *rpoB* and *fusA* genes within the family *Geobacteraceae* fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, **54**, 1591-1599.
- Holmes EC, Worobey M, Rambaut A (1999) Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution*, **16**, 405-409.
- Holt J, Krieg N, Sneath P, Staley J, Williams S (1994) *Bergey's Manual of Determinative Bacteriology*. (1994) *Bergey's Manual of Determinative Bacteriology*.
- Hooft R, Vriend G, Sander C, Abola E (1996) Errors in protein structures. *Nature*, **381**, 272.
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310-2314.
- Hug L a., Roger AJ (2007) The impact of fossils and taxon sampling on ancient molecular dating analyses. *Molecular Biology and Evolution*, **24**, 1889-1897.
- Husmeier D, Wright F (2001) Detection of recombination in DNA multiple alignments with hidden Markov models. *Journal of Computational Biology: a Journal of Computational Molecular Cell Biology*, **8**, 401-427.
- Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68-73.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254-267.
- Huson DH, Bryant D (2013) User manual for SplitsTree4, 1-57.
- Huson DH, Scornavacca C (2011) A survey of combinatorial methods for phylogenetic networks. *Genome Biology and Evolution*, **3**, 23-35.
- Huys G, Cnockaert M, Swings J (2005) *Aeromonas culicicola* (Pidiyar et al. 2002) is a later subjective synonym of *Aeromonas veronii* (Hickman-Brenner et al. 1987). *Systematic and Applied Microbiology*, **28**, 604-609.
- Huys G, Denys R, Swings J (2002) DNA-DNA reassociation and phenotypic data indicate synonymy between *Aeromonas enteropelogenes* (Schubert et al . 1990) and *Aeromonas trota* (Carnahan et al. 1991). *International Journal of Systematic and Evolutionary Microbiology*, **52**, 1969-1972.

- Huys G, Kämpfer P, Swings J (2001) New DNA-DNA hybridization and phenotypic data on the species *Aeromonas ichthiosmia* and *Aeromonas allosaccharophila*: *A.ichthiosmia* (Schubert et al. 1990) is a later synonym of *A. veronii* (Hickman-Brenner et al. 1987). *Systematic and Applied Microbiology*, **24**, 177-82.
- Igbinosa IH, Igumbor EU, Aghdasi F, Tom M, Okoh AI (2012) Emerging *Aeromonas* species infections and their significance in public health. *The Scientific World Journal*, **2012**.
- Isonhood JH, Drake M (2002) *Aeromonas* species in Foods. *Journal of Food Protection*, **3**, 447-582.
- Jakobsen IB, Wilson SR, Eastep S (1997) The partition matrix: exploring variable phylogenetic signals along nucleotide sequence alignments. *Molecular Biology and Evolution*, **14**, 474-484.
- Janda JM (1991) Recent advances in the study of the taxonomy, pathogenicity, and infectious syndromes associated with the genus *Aeromonas*. *Clinical Microbiology Reviews*, **4**, 397-410.
- Janda JM, Abbott SL (1998) Evolving concepts regarding the genus *Aeromonas*: an expanding Panorama of species, disease presentations, and unanswered questions. *Clinical Infectious Diseases: an Official Publication of the Infectious Diseases Society of America*, **27**, 332-344.
- Janda JM, Abbott SL (2010) The genus *Aeromonas*: taxonomy, pathogenicity, and infection. *Clinical Microbiology Reviews*, **23**, 35-73.
- Janda JM, Duffey PS (1988) Mesophilic aeromonads in human disease: current taxonomy, laboratory identification, and infectious disease spectrum. *Clinical Infectious Diseases: an Official Publication of the Infectious Diseases Society of America*, **10**, 980-997.
- Jaroszewski L, Rychlewski L, Li Z (2005) FFAS03: a server for profile-profile sequence alignments. *Nucleic acids*, **33**, 284-288.
- Jones DT (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *Journal of Molecular Biology*, **287**, 797-815.
- Karlin S, Brocchieri L (1996) Evolutionary conservation of RecA genes in relation to protein structure and function. *Journal of Bacteriology*, **178**, 1881-94.
- Karplus K (2009) SAM-T08, HMM-based protein structure prediction. *Nucleic Acids Research*, **37**, 492-497.
- Kass R, Raftery A (1995) Bayes Factor. *Journal of the American Statistical Association*, **90**, 773-795.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, **10**, 845-858.
- Kelley L a, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*, **4**, 363-371.

-
- Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, **32**, 526-531.
- Kishino H, Thorne JL, Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution*, **18**, 352-361.
- Kober B (2000) HIV signature and sequence variation analysis. In: *Computational Analysis of HIV Molecular Sequences* (eds Rodrigo A, Learn G), 55-72. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Kopp J, Schwede T (2004) Automated protein structure homology modeling: a progress report. *Pharmacogenomics*, **5**, 405-416.
- Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics*, **23**, 1289-1291.
- Kosakovskiy SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) GARD: a genetic algorithm for recombination detection. *Bioinformatics*, **22**, 3096-3098.
- Kuo C-H, Ochman H (2009) Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biology Direct*, **4**, 35.
- Küpfer M, Kuhnert P, Korczak BM, Peduzzi R, Demarta A (2006) Genetic relationships of *Aeromonas* strains inferred from 16S rRNA, *gyrB* and *rpoB* gene sequences. *International Journal of Systematic and Evolutionary Microbiology*, **56**, 2743-2751.
- Kurumizaka H, Aihara H, Ikawa S *et al.* (1996) A Possible Role of the C-terminal Domain of the RecA Protein. *The Journal of Biological Biochemistry*, **271**, 33515-33524.
- Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, **157**, 105-132.
- Laskowski R a., MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, **26**, 283-291.
- Leaché AD, Rannala B (2011) The accuracy of species tree estimation under simulation: a comparison of methods. *Systematic Biology*, **60**, 126-137.
- Lemey P, Lott M, Martin DP, Moulton V (2009a) Identifying recombinants in human and primate immunodeficiency virus sequence alignments using quartet scanning. *BMC Bioinformatics*, **10**, 126.
- Lemey P, Salemi M, Vandamme A (2009b) *The Phylogenetic Handbook*.
- Lepage T, Bryant D, Philippe H, Lartillot N (2007) A general comparison of relaxed molecular clock models. *Molecular Biology and Evolution*, **24**, 2669-2680.

- Levin B (1981) Periodic selection, infectious gene exchange and the genetic structure of *E.coli* populations. *Genetics*, 1-23.
- Lewis-Rogers N (2004) Evolutionary analyses of genetic recombination. *Signpost, Kerala, India*, **661**, 49-78.
- Li Y, Liu Y, Zhou Z *et al.* (2011) Complete genome sequence of *Aeromonas veronii* strain B565. *Journal of Bacteriology*, **193**, 3389-3390.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451-1452.
- Linder HP, Hardy CR, Rutschmann F (2005) Taxon sampling effects in molecular clock dating: An example from the African Restionaceae. *Molecular Phylogenetics and Evolution*, **35**, 569-582.
- Liu L, Pearl D (2007) Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology*, 504-514.
- Lorén JG, Farfán M, Fusté MC (2014) Molecular phylogenetics and temporal diversification in the genus *Aeromonas* based on the sequences of five housekeeping genes. *PLoS one*, **9**.
- Lusetti SL, Cox MM (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annual Review of Biochemistry*, **71**, 71-100.
- Lusetti SL, Shaw JJ, Cox MM (2003a) Magnesium ion-dependent activation of the RecA protein involves the C terminus. *The Journal of Biological Chemistry*, **278**, 16381-16388.
- Lusetti SL, Wood EA, Fleming CD *et al.* (2003b) C-terminal deletions of the *Escherichia coli* RecA protein. Characterization of in vivo and in vitro effects. *The Journal of Biological Chemistry*, **278**, 16372-16380.
- Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83-85.
- Maddison WP (1997) Gene Trees in Species Trees. *Systematic Biology*, **46**, 523.
- Maiden MCJ, Bygraves JA, Feil E (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*, **95**, 3140-3145.
- Majewski J, Zawadzki P, Pickerill P *et al.* (2000) Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *Journal of Bacteriology*, **182**, 1016-1023.
- Martin AP, Costello EK, Meyer AF, Nemergut DR, Schmidt SK (2004) The rate and pattern of cladogenesis in microbes. *Evolution*, **58**, 946-955.
- Martin DP, Lemey P, Lott M *et al.* (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, **26**, 2462-2463.

-
- Martin DP, Lemey P, Posada D (2011) Analysing recombination in nucleotide sequences. *Molecular Ecology Resources*, **11**, 943-955.
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, **1**.
- Martin D, Posada D (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research & Human Retroviruses*, **21**, 98-102.
- Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics*, **16**, 562-563.
- Martin-Carnahan A, Joseph S (2005) Order XII. Aeromonadales ord. nov. In: *Bergey's Manual of Systematic Bacteriology*, 556-578.
- Martinez-Murcia AJ, Benlloch S, Collins MD (1992a) Phylogenetic interrelationships of members of the genera *Aeromonas* and *Plesiomonas* as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA-DNA hybridizations. *International journal of Systematic Bacteriology*, **42**, 412-421.
- Masui R, Mikawa T, Kuramitsu S (1997) Local folding of the N-terminal domain of *Escherichia coli* RecA controls protein-protein interaction. *Journal of Biological Chemistry*, **272**, 27707-27715.
- Maynard Smith J (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*, **34**, 126-129.
- McGrew DA, Knight KL (2003) Molecular design and functional organization of the RecA protein. *Critical Reviews in Biochemistry and Molecular Biology*, **38**, 385-432.
- McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404-405.
- McGuire G, Wright F (2000) TOPAL 2.0: improved detection of mosaic sequences within multiple alignments. *Bioinformatics*, **16**, 130-134.
- McMahon M, Wilson I (2001) The occurrence of enteric pathogens and *Aeromonas* species in organic vegetables. *International Journal of Food Microbiology*, **70**, 55-62.
- Mcvean G, Awadalla P, Fearnhead P (2002) A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics*, **1241**, 1231-1241.
- Minárik P, Tomásková N, Kollárová M, Antalík M (2002) Malate dehydrogenases-structure and function. *General Physiology and Biophysics*, **21**, 257-265.
- Miñana-Galbis D, Farfán M, Fusté MC, Lorén JG (2004) *Aeromonas molluscorum* sp. nov., isolated from bivalve molluscs. *International Journal of Systematic and Evolutionary Microbiology*, **54**, 2073-2078.

- Miñana-Galbis D, Farfán M, Lorén JG, Fusté MC (2002) Biochemical identification and numerical taxonomy of *Aeromonas* spp. isolated from environmental and clinical samples in Spain. *Journal of Applied Microbiology*, **93**, 420-430.
- Miñana-Galbis D, Urbizu-Serrano A, Farfán M, Fusté MC, Lorén JG (2009) Phylogenetic analysis and identification of *Aeromonas* species based on sequencing of the cpn60 universal target. *International Journal of Systematic and Evolutionary Microbiology*, **59**, 1976-1983.
- Montgomerie S, Cruz J a, Shrivastava S *et al.* (2008) PROTEUS2: a web server for comprehensive protein structure prediction and structure-based annotation. *Nucleic Acids Research*, **36**, 202-209.
- Morandi A, Zhaxybayeva O, Gogarten JP, Graf J (2005) Evolutionary and diagnostic implications of intragenomic heterogeneity in the 16S rRNA gene in *Aeromonas* strains. *Journal of Bacteriology*, **187**, 6561-6564.
- Morlon H, Kempes BD, Plotkin JB, Brisson D (2013) Explosive radiation of a bacterial species group. *Evolution*, **66**, 2577-2586.
- Morlon H, Parsons TL, Plotkin JB (2011) Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences*, **108**, 16327-16332.
- Morlon H, Potts MD, Plotkin JB (2010) Inferring the dynamics of diversification: A coalescent approach. *PLoS Biology*, **8**.
- Nakhleh L, Ruths D Gene Trees , Species Trees , and Species Networks. 1-27.
- Nalin R, Gandon C, Cvejic JH, Bruneteau M, Berry A (2000) NOTE A recA gene phylogenetic analysis confirms the close proximity of *Frankia* to *Acidothermus*. *International Journal of Systematic and Evolutionary Microbiology*, **50**, 781-785.
- Nee S (2001) Inferring speciation rates from phylogenies. *Evolution*, **55**, 661-668.
- Nee S (2006) Birth-Death models in macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 1-17.
- Nee S, May RM, Harvey PH (1994) The reconstructed evolutionary process. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **344**, 305-311.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**, 418-426.
- Nhung PH, Hata H, Ohkusu K *et al.* (2007) Use of the novel phylogenetic marker dnaJ and DNA-DNA hybridization to clarify interrelationships within the genus *Aeromonas*. *International Journal of Systematic and Evolutionary Microbiology*, **57**, 1232-1237.
- Nielsen M, Lundegaard C, Lund O, Petersen TN (2010) CPHmodels-3.0-remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Research*, **38**, 576-581.
- Ochman H, Davalos LM (2006) The nature and dynamics of bacterial genomes. *Science*, **311**, 1730-1733.

- Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*, **405**, 299-304.
- Ochman H, Wilson AC (1987) Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *Journal of Molecular Evolution*, **26**, 74-86.
- Ohta T, Basten CJ (1992) Gene conversion generates hypervariability at the variable regions of kallikreins and their inhibitors. *Molecular Phylogenetics and Evolution*, **1**, 87-90.
- Oren A, Garrity GM (2014) List of new names and new combinations previously effectively, but not validly, published. *International Journal of Systematic and Evolutionary Microbiology*, **64**, 2184-2187.
- Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology*, **265**, 218-225.
- Paradis E (2003) Analysis of diversification: combining phylogenetic and taxonomic data. *Proceedings. Biological sciences / The Royal Society*, **270**, 2499-2505.
- Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289-290.
- Parker JL, Shaw JG (2011) *Aeromonas* spp. clinical microbiology and disease. *The Journal of Infection*, **62**, 109-118.
- Pascual J, Macián MC, Arahal DR, Garay E, Pujalte MJ (2010) Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, *recA*, *pyrH*, *rpoD*, *gyrB*, *rctB* and *toxR* genes. *International Journal of Systematic and Evolutionary Microbiology*, **60**, 154-165.
- Payne GW, Vandamme P, Morgan SH *et al.* (2005) Development of a *recA* gene-based identification approach for the entire *Burkholderia* genus. *Applied and Environmental Microbiology*, **71**, 3917-3927.
- Pérez-Losada M, Browne EB, Madsen A *et al.* (2006) Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary genetics in Infectious Diseases*, **6**, 97-112.
- Perriere G, Gouy M (1996) WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie*, 364-369.
- Popoff MY, Coynault C, Kiredjian M, Lemelin M (1981) Polynucleotide sequence relatedness among motile *Aeromonas* species. *Current Microbiology*, **5**, 109-114.
- Popoff M, Lallier R (1984) Biochemical and serological characteristics of *Aeromonas*. In: *Methods in Microbiology* (ed Bergan T), 127-145. Academic Press.
- Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Molecular Biology and Evolution*, **19**, 708-717.

- Posada D, Crandall K a (2001) Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 13757-13762.
- Posada D, Crandall KA (2002) The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution*, **54**, 396-402.
- Posada D, Crandall K a, Holmes EC (2002) Recombination in evolutionary genomics. *Annual Review of Genetics*, **36**, 75-97.
- Pybus OG, Harvey PH (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proceedings. Biological sciences / The Royal Society*, **267**, 2267-2272.
- Rabosky DL (2006a) LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. *Evolutionary Bioinformatics Online*, **2**, 273-276.
- Rabosky DL (2006b) Likelihood methods for detecting temporal shifts in diversification rates. *Evolution*, **60**, 1152-1164.
- Rabosky DL, Grundler M, Anderson C *et al.* (2014) BAMMtools: An R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution*, **5**, 701-707.
- Rabosky DL, Lovette IJ (2008) Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? *Evolution; International Journal of organic evolution*, **62**, 1866-1875.
- Reddy C, Beveridge T, Breznak J *et al.* (2007) *Methods for General and Molecular Microbiology*. ASM Press, Washington, D.C.
- Reith ME, Singh RK, Curtis B *et al.* (2008) The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. *BMC Genomics*, **9**, 427.
- Rest JS, Mindell DP (2003) Retroids in archaea: phylogeny and lateral origins. *Molecular Biology and Evolution*, **20**, 1134-1142.
- Richard SB, Madern D, Garcin E, Zaccai G (2000) Halophilic adaptation: novel solvent protein interactions observed in the 2.9 and 2.6 Å resolution structures of the wild type and a mutant of malate dehydrogenase from *Haloarcula marismortui*. *Biochemistry*, **39**, 992-1000.
- Ricklefs RE (2007) Estimating diversification rates from phylogenetic information. *Trends in Ecology & Evolution*, **22**, 601-610.
- Rizzo J, Rouchka EC (2007) Review of phylogenetic tree construction. *Bioinformatics Laboratory Technical Report Series*, 1-7.
- Roca A, Cox M (1997) RecA protein: structure, function, and role in recombinational DNA repair. *Progress in Nucleic Acid Research and Molecular Biology*, **56**, 129-223.

- Roche D, Buenavista M (2011) The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand. *Nucleic Acids*, **39**, 171-176.
- Roger F, Marchandin H, Jumas-Bilak E, Kodjo A, and Lamy B (2012) Multilocus genetics to reconstruct aeromonad evolution. *BMC Microbiology*, **12**, 62.
- Rohde R, Muller R (2005) Cycles in fossil diversity. *Nature*, **434**, 9-11.
- Rossello-Mora R, Amann R (2001) The species concept for procaryotes. *FEMS Microbiology Ecology*, **25**, 39-67.
- Rossi F, Dellaglio F, Torriani S (2006) Evaluation of *recA* gene as a phylogenetic marker in the classification of dairy propionibacteria. *Systematic and Applied Microbiology*, **29**, 463-469.
- Rutschmann F (2006) Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Distributions*, **12**, 35-48.
- Saavedra MJ, Figueras MJ, Martínez-murcia AJ (2006) Updated phylogeny of the genus *Aeromonas*. *International Journal of Systematic and Evolutionary Microbiology*, **56**, 2481-2487.
- Sahney S, Benton MJ (2008) Recovery from the most profound mass extinction of all time. *Proceedings. Biological sciences / The Royal Society*, **275**, 759-765.
- Sali A (1995) Comparative protein modeling by satisfaction of spatial restraints. *Molecular Medicine Today*, **2**, 270-277.
- Salminen MO, Carr JK, Burke DS, McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Research and Human Retroviruses*, **11**, 1423-1425.
- Sanderson MJ (1996) A nonparametric approach of rate constancy to estimating divergence times in the absence. *Methods*, 1218-1231.
- Sanderson M, Donoghue MJ (1996) Reconstructing shifts in diversification rates on phylogenetic trees. *Tree*, **27**, 10153-10157.
- Schmidt T, Bergner A, Schwede T (2013) Modelling three-dimensional protein structures for applications in drug design. *Drug Discovery Today*, **19**, 890-897.
- Schwede T, Jürgen K, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, **31**, 3381-3385.
- Sen K (2005) Development of a rapid identification method for *Aeromonas* species by multiplex-PCR. *Canadian Journal of Microbiology*, **51**, 957-966.
- Sepe A, Barbieri P, Peduzzi R, Demarta A (2008) Evaluation of *recA* sequencing for the classification of *Aeromonas* strains at the genotype level. *Letters in Applied Microbiology*, **46**, 439-444.

- Sepkoski J (2002) *A compendium of fossil marine animal genera*. *Bull Am Paleontol.* 363, 1-560.
- Seshadri R, Joseph SW, Chopra AK (2006) Genome sequence of *Aeromonas hydrophila* ATCC 7966T: jack of all trades. *Journal of Bacteriology*, **188**, 8272-8282.
- Shapiro B, Friedman J, Cordero O *et al.* (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science*, **336**, 48-51.
- Sheridan PP, Freeman KH, Brenchley JE (2003) Estimated minimal divergence times of the major bacterial and archaeal phyla. *Geomicrobiology Journal*, **20**, 1-14.
- Sneath PHA (1993) Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences. *International Journal of Systematic Bacteriology*, **43**, 626-629.
- Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, **33**, 244-248.
- Soler L, Yáñez M, Chacon M (2004) Phylogenetic analysis of the genus *Aeromonas* based on two housekeeping genes. *International Journal of Systematic and Evolutionary Microbiology*, **54**, 1511-1519.
- Spataro N, Farfán M, Albarral V, Sanglas A, Lorén JG, Fusté MC, Bosch E (2013) Draft Genome Sequence of *Aeromonas molluscorum* Strain 848T^T. *Genome Announcements*, **2**, 3-4.
- Stackebrandt E, Frederiksen W, Garrity GM (2002) TAXONOMIC Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology, 1043-1047.
- Stadler T (2010a) Tree Sim R-Simulating trees under the birth-death model.
- Stadler T (2010b) Sampling-through-time in birth-death trees. *Journal of Theoretical Biology*, **267**, 396-404.
- Stine OC, Sozhamannan S, Gou Q *et al.* (2000) Phylogeny of *Vibrio cholerae* based on recA sequence. *Infection and Immunity*, **68**, 7180-7185.
- Story RM, Steitz TA (1992) Structure of the recA protein-ADP complex. *Nature*, **355**, 374-376.
- Sullivan C, Diggle M, Clarke S (2005) Multilocus sequence typing: Data analysis in clinical microbiology and public health. *Molecular Biotechnology*, **29**, 245-254.
- Summons RE, Jahnke LL, Hope JM, Logan GA (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature*, **400**, 554-557.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585-595.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512-526.

-
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725-2729.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876-4882.
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673-4680.
- Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution*, **15**, 1647-1657.
- Tomás JM (2012) The main *Aeromonas* pathogenic factors. *ISRN Microbiology*, **2012**, 256-261.
- Torriani S, Felis GE, Dellaglio F (2001) Differentiation of *Lactobacillus plantarum*, *L.pentosus*, and *L.paraplantarum* by *recA* gene sequence analysis and multiplex PCR assay with *recA* gene-derived primers. *Applied and Environmental Microbiology*, **67**, 3450-3454.
- Untergasser A, Cutcutache I, Koressaar T *et al.* (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Research*, **40**, 115.
- Urwin R, Maiden MCJ (2003) Multi-locus sequence typing: a tool for global epidemiology. *Trends in Microbiology*, **11**, 479-487.
- Vanloock MS, Yu X, Yang S *et al.* (2003a) ATP-mediated conformational changes in the RecA filament. *Structure*, **11**, 187-196.
- Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*.
- Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the alpha- and beta- subunits of ATP synthase, myosin, kinases and other ATP requiring enzymes and a common nucleotide binding fold. *The EMBO Journal*, **1**, 945-951.
- Wallace R, Shaffer J, Murphy R. (1979) Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: The effect of single base pair mismatch. *Nucleic Acids Research*, **6**, 3543-3558.
- Wallner B, Elofsson A (2003) Can correct protein models be identified? *Protein Science*, **12**, 1073-1086.
- Wallner B, Fang H, Elofsson A (2003) Automatic consensus-based fold recognition using Pcons, ProQ, and Pmodeller. *Proteins*, **53**, Suppl 6, 534-541.
- Wang T-F, Chen L-T, Wang AH-J (2008) Right or left turn? RecA family protein filaments promote homologous recombination through clockwise axial rotation. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, **30**, 48-56.
- Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT (2004) The DISOPRED server for the prediction of protein disorder. *Bioinformatics*, **20**, 2138-2139.

- Wass MN, Kelley L a, Sternberg MJE (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Research*, **38**, 469-473.
- Wei-Chen C (2011) *Overlapping codon model, phylogenetic clustering, and alternative partial expectation conditional maximization algorithm*.
- Weiller GF (1998) Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Molecular Biology and Evolution*, **15**, 326-335.
- Welch JJ, Bromham L (2005) Molecular dating when rates vary. *Trends in Ecology & Evolution*, **20**, 320-327.
- Weng FY, Chiou CS, Lin PHP, Yang SS (2009) Application of recA and rpoB sequence analysis on phylogeny and molecular identification of *Geobacillus* species. *Journal of Applied Microbiology*, **107**, 452-464.
- Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews*, **35**, 957-976.
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, **35**, 407-410.
- Wiuf C, Christensen T, Hein J (2001) A simulation study of the reliability of recombination detection methods. *Molecular Biology and Evolution*, **18**, 1929-1939.
- Wollenberg K, Arnold J, Avise JC (1996) Recognizing the forest for the trees: testing temporal patterns of cladogenesis using a null model of stochastic diversification. *Molecular Biology and Evolution*, **13**, 833-849.
- Xia X (2013) DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, **30**, 1720-1728.
- Xing X, Bell CE (2004a) Crystal structures of *Escherichia coli* RecA in a compressed helical filament. *Journal of Molecular Biology*, **342**, 1471-1485.
- Xing X, Bell CE (2004b) Crystal structures of *Escherichia coli* RecA in complex with MgADP and MnAMP-PNP. *Biochemistry*, **43**, 16142-16152.
- Yáñez M, Catalán V, Apráiz D, Figueras M, Martínez-Murcia A (2003) Phylogenetic analysis of members of the genus *Aeromonas* based on *gyrB* gene sequences. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 875-883.
- Yang Z (2006) *Computational Molecular Evolution*. Oxford Series in Ecology and Evolution.
- Yang Z, Bielawski JR (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution*, **15**, 496-503.
- Yoder AD, Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. *Molecular Biology and Evolution*, **17**, 1081-1090.

- Yu X, Egelman EH (1997) The RecA hexamer is a structural homologue of ring helicases. *Nature*, **4**, 101-104.
- Yu X, Jacobs S a, West SC, Ogawa T, Egelman EH (2001) Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 8419-8424.
- Zbinden A, Köhler N, Bloemberg G V. (2011) recA-based PCR assay for accurate differentiation of *Streptococcus pneumoniae* from other viridans streptococci. *Journal of Clinical Microbiology*, **49**, 523-527.
- Zha X, Zhang W, Zhou C, Zhang L (2014) Detection and characterization of *Wolbachia* infection in silkworm. *Genetics and Molecular Biology*, **37**, 573-580.
- Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.
- Zhou J, Bowler LD, Spratt BG (1997) Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of *Neisseria meningitidis* and commensal *Neisseria* species. *Molecular Microbiology*, **23**, 799-812.
- Zuckerlandl E, Pauling L (1962) *Molecular Disease, evolution and genic heterogeneity*. Ed. M Kasha M, Pullman P. Academic Press.

11. PUBLICACIONES

Population Genetics of the “*Aeromonas hydrophila* Species Complex”

M^a Carmen Fusté, Maribel Farfán, David Miñana-Galbis,
Vicenta Albarral, Ariadna Sanglas and José Gaspar Lorén
*Department of Health Microbiology and Parasitology, Faculty of Pharmacy,
University of Barcelona, Barcelona,
Spain*

1. Introduction

Population genetics studies the genetic variability of individuals in a population based on the allele frequencies at several genes or loci and tries to explain this variability in terms of mutation, selection or genetic recombination. The statistical analysis of these frequencies allows models of evolution to be established, which will help us to understand and predict the past and present gene flow in the population (Maynard-Smith, 1991). For the most part population genetics has been designed for diploid organisms with sexual reproduction. In the words of Bruce Levin, “the genetic theory of adaptive evolution was developed by sexually reproducing eukaryotes, for sexually reproducing eukaryotes” (Levin & Bergstrom, 2000). As a consequence, before being applied to prokaryotes, population genetics needs to be adapted.

In theory the haploid nature of bacteria should simplify their analysis, since dominance or over-dominance is not an issue and the genotype can usually be deduced directly from the phenotype. However, central to classical population genetics are infinite population size, random mating, and free recombination. Consequently, as expressed by Maynard-Smith, “the alleles present at one locus are independent of those at other loci. Changes in the frequency of an allele at one locus, therefore, are independent of what is happening elsewhere in the genome: each locus can be treated individually” (Maynard-Smith, 1995). It is true that the size of bacterial populations can be practically infinite but recombination occurs extremely rarely so that changes affecting one locus can lead to the modification of others. In the succinct words of Maynard-Smith, “the genome should be treated as an inter-related whole, and not as a set of independently changing genes”. The crux of the problem is knowing the exact level of recombination in bacterial populations, since “it is considerably more challenging to elaborate a theory for a population with little recombination than for one with no recombination, or a lot” (Maynard-Smith, 1995). In bacterial population genetics, sometimes we detect a degree of recombination that is too high for a pure phylogenetic approach, but too low for assessing a random interchange.

Stronger evidence for restricted recombination comes from measurements of linkage disequilibrium: that is, the tendency for particular alleles at different loci to co-occur

(Maynard-Smith et al., 1993; Haubold et al., 1998). Linkage disequilibrium (and the too-frequent occurrence of a particular combination of alleles, which is a manifestation of such disequilibrium) shows that recombination is restricted, but not absent. The determination of the relative importance of mutation in comparison with recombination is central to bacterial population genetics (Feil et al., 1999). Previous studies have demonstrated a wide variety of situations among bacterial species ranging from the clonal diversification of *Salmonella* (Selander et al., 1990) or *Escherichia coli* (Orskov et al., 1990), which are mainly due to mutation, to the frequent recombination found in *Neisseria gonorrhoeae* (O'Rourke & Stevens, 1993) or *Helicobacter pylori* (Salaun et al., 1998). Most of the population studies done with bacterial species suggest that recombination occurs in nature, and indeed may be highly important in generating variation, but that it is infrequent compared to mutation. Consequently, bacterial populations consist largely of independent clonal lineages.

The development of protein electrophoresis was a breakthrough for the study of bacterial population genetics. A pioneer in the field, Milkman (1973) used the methodology to study whether electrophoretic variation is selective or neutral. As described by Selander et al., Multilocus Enzyme Electrophoresis (MLEE) "has long been a standard method in eukaryotic population genetics and systematics" before it was applied "for studying the genetic diversity and structure in natural populations of bacteria. This research established basic population frameworks for the analysis of variation in serotypes and other phenotypic characters and has provided extensive data for systematics and useful marker systems for epidemiology" (Selander et al., 1986).

In 1998, Maiden et al. introduced Multilocus Sequence Typing (MLST), an extension of MLEE based on nucleotide sequencing that is able to determine higher levels of discrimination (more alleles per locus). In MLST "alleles are identified directly from the nucleotide sequences of internal fragments of genes rather than by comparing the electrophoretic mobilities of the enzymes they encode" (Maiden et al., 1998). In addition, this method is fully portable between laboratories and data can be stored in a single multilocus sequence database accessible via the internet (<http://www.mlst.net>). This approach has given "a new dimension to the elucidation of genomic relatedness at the inter- and intraspecific level by sequence analysis of housekeeping genes subject to stabilising selection. This technique has been mainly used in epidemiology, but it offers the opportunity to incorporate the insights available from population genetics and phylogenetic approaches into bacterial systematics" (Stackebrandt et al., 2002; Pérez-Losada et al., 2005; Robinson et al., 2010).

The results obtained using MLST indicate that despite the high diversity observed in bacteria, it is possible to recognize clusters with a lower degree of variation. The number of sequence types (STs) obtained is less than expected if we consider the product of the individual allelic frequencies. Nevertheless, most of the strains belong to one or a few allelic profiles, whereas most of the STs are represented by one or few strains. In addition, most STs cluster in clonal complexes constituted by closely related genotypes. Typically, each clonal complex consists of a predominant ST and a varied group of less common STs, which have different alleles only in one or two loci (Feil et al., 2004). These differences could correspond to initial stages of clonal divergence from an ancestral genotype that was the origin of the clone (Vogel et al., 2010; Willems, 2010).

2. The genus *Aeromonas*

The genus *Aeromonas* Stanier 1943 belongs to the family *Aeromonadaceae* within the class *Gammaproteobacteria* (Martin-Carnahan & Joseph, 2005). *Aeromonads* are autochthonous inhabitants of aquatic environments including chlorinated and polluted waters, although they can also be isolated from a wide variety of environmental and clinical sources. They cause infections in vertebrates and invertebrates, such as frogs, birds, various fish species and domestic animals. In recent years, some authors have considered *Aeromonas* as an emergent pathogen in humans, producing intestinal and extraintestinal diseases. *Aeromonads* are facultative anaerobic chemoorganotrophs capable of anaerobic nitrate respiration and dissimilatory metal reduction (Martin-Carnahan & Joseph, 2005).

The interest in the taxonomy of the genus *Aeromonas* has increased markedly in recent years (Janda & Abbott, 2010), and its classification, with 25 species currently recognized, remains challenging. Novel species are continuously being described, strains and species described so far are being rearranged, and DNA-DNA hybridization studies have observed discrepancies (Janda & Abbott, 2010). Historically, the genus *Aeromonas* has been divided into two groups: nonmotile, psychrophilic species, best represented by *A. salmonicida*, which are generally associated with fish diseases and motile mesophilic species associated with human diseases, including *A. hydrophila*, *A. veronii* and *A. caviae* (Martin-Carnahan & Joseph, 2005). More species have since been described and genealogies have to be adapted accordingly, although this is not always straightforward.

Bacterial species are formally defined as a group of strains that share several phenotypical characteristics and show values of DNA-DNA hybridization $\geq 70\%$ and a 16S rRNA sequence similarity $\geq 97\%$ with their close relatives (Stackebrandt et al., 2002). Indeed, there are hardly any examples in which strains with divergence in the 16S rRNA sequence $\leq 97\%$ are defined as one species (Roselló-Mora & Amann, 2001). In *Aeromonas* 16S rRNA gene sequences are highly similar, being identical in some close related species such as *A. salmonicida*, *A. bestiarum*, *A. popoffii* and *A. piscicola*, which hampers their utility in defining species in this genus (Figueras et al., 2000; Beaz-Hidalgo et al., 2009).

Several attempts have been made to generate phylogenies using DNA gene sequences to reconstruct the correct genealogical ties among species in *Aeromonas* (Küpfer et al., 2006; Saavedra et al., 2006; Miñana-Galbis et al., 2009), but the genes chosen for this purpose are not always suitable and do not give congruent phylogenies (Farfán et al., 2010; Silver et al., 2011). Recently, two papers presenting MLST schemes for *Aeromonas* have been published (Martínez-Murcia et al., 2011; Martino et al., 2011) and there is an online MLST database of the genus *Aeromonas*, managed by Keith Jolley and curated by Barbara Cardazzo (<http://pubmlst.org/aeromonas>). All this accumulated data should help to establish a reliable clustering of the *Aeromonas* species and elucidate their exact boundaries.

Finally, the availability of complete genomes of different species is also useful in this task, but unfortunately in the case of *Aeromonas* only three genomes have been completed, one corresponding to the type strain of *A. hydrophila* subsp. *hydrophila* ATCC 7966 isolated from a tin of milk (Seshadri et al., 2006), the second to a fish pathogen, the strain A449 of *A. salmonicida* subsp. *salmonicida* (Reith et al., 2008) and the third to *A. veronii* isolated from an aquaculture pond sediment (Li et al., 2011). The information given by the genomes of *A. hydrophila* and *A. salmonicida* indicate that while they are of identical size (4,7Mb) and share

multiple housekeeping and virulence genes, *A. salmonicida* has acquired several mobile genetic elements, and undergone genome rearrangements and loss of genes in the process of adapting to a specific host. The genome of *A. veronii* is smaller (4,3Mb) and contains fewer virulence genes than the others.

3. “*Aeromonas hydrophila* species complex”

An example of the taxonomic complexity of the genus *Aeromonas* is the difficulty in discriminating between the phenotypically and genetically closely related species belonging to the “*Aeromonas hydrophila* species complex” (AHC), which includes: *A. hydrophila*, composed by three subspecies: *A. hydrophila* subsp. *hydrophila*, *A. hydrophila* subsp. *ranae* and *A. hydrophila* subsp. *dhakensis*, *A. bestiarum*, *A. popoffii* and *A. salmonicida*, divided in five subspecies: *A. salmonicida* subsp. *salmonicida*, *A. salmonicida* subsp. *masoucida*, *A. salmonicida* subsp. *achromogenes*, *A. salmonicida* subsp. *pectinolytica* and *A. salmonicida* subsp. *smithia* (Miñana-Galbis et al., 2002; Martin-Carnahan & Joseph, 2005). Recently, two additional species have been described in this group, *A. aquariorum* and *A. piscicola* (Martínez-Murcia et al., 2008; Beaz-Hidalgo et al., 2009). Members of the AHC were first described as strains producing the enzymes elastase, lecithinase or stapholysin (Abbott et al., 2003). They are genetically closely related and share multiple phenotypic characteristics, which makes discrimination among the species included in this group extremely difficult (Miñana-Galbis et al., 2002).

Several approaches have been used to discriminate among the AHC species: Amplified Fragment Length Polymorphisms (AFLP) (Huys et al., 1997), fluorescent AFLP (FAFLP) (Huys et al., 2002), MLEE (Miñana-Galbis et al., 2004), Random Amplified Polymorphic DNA (RAPD) and MALDI-TOF MS analysis (Martínez-Murcia et al., 2008). Although the results obtained with these methods have been useful for the taxonomy and phylogeny of the AHC, providing a hypothesis for the genealogy of strains and detailing their patterns of descent and degree of genetic variation accumulated over time, only the MLEE study has been used to elucidate their population genetic structure.

Previous studies based on the sequence analysis of several housekeeping genes have demonstrated that the AHC is not monophyletic (Soler et al., 2004; Küpfer et al., 2006; Saavedra et al., 2006; Nhung et al., 2007; Beaz-Hidalgo et al., 2009; Miñana-Galbis et al., 2009). Nevertheless, controversially, other studies have shown the monophyly of this group (Martínez-Murcia et al., 2005; Farfán et al., 2010). This conflict could be due to the incongruence of phylogenies derived from distinct gene sequence analysis.

In order to establish the population structure and divergence of the species included in this group of *Aeromonas* we studied a set of strains representative of the AHC, in which we analyzed the nucleotide sequences (total or partial) of 6 housekeeping genes: *cpn60* (555 bp), *dnaJ* (891 bp), *gyrB* (1089 bp), *mdh* (936 bp), *recA* (1065 bp), *rpoD* (843 bp), giving a total fragment length of 5379 bp.

4. Phylogenetic analysis

The relationships among the analyzed *Aeromonas* isolates are represented as a genealogical tree (Fig. 1). The tree reveals that the AHC splits into three main clusters separated by a

mean genetic distance of 0.071 (Table 1). The *A. salmonicida* cluster, although comprising five subspecies, constitutes a homogeneous clade with the lowest mean genetic distance of 0.016 (Table 1). The closest relative to *A. salmonicida*, *A. bestiarum*, has a different population structure. Rather than being a single group, it is divided in three clades, one constituted by the great majority of the *A. bestiarum* strains, a second corresponding to *A. popoffii* and the third including some *A. bestiarum* strains together with isolates of *A. piscicola*. The mean genetic distance of the *A. bestiarum* cluster is higher than in *A. salmonicida* 0.029 (Table 1). The *A. hydrophila* group, which exhibits the highest genetic distance 0.037 (Table 1), clearly separates in two clades, one including the *A. hydrophila* subsp. *hydrophila* and *A. hydrophila* subsp. *ranae* and the second constituted by *A. hydrophila* subsp. *dhakensis* and *A. aquariorum*. The presence of different clades in some of these species poses questions about the clade-species relationships.

	TN93	Standard error
All sequences	0.0717	0.0024
<i>A. bestiarum</i>	0.0286	0.0014
<i>A. bestiarum</i> (clade 1)	0.0157	0.0009
<i>A. bestiarum</i> (clade 2)	0.0084	0.0008
<i>A. bestiarum</i> (clade 3)	0.0100	0.0008
<i>A. hydrophila</i>	0.0370	0.0016
<i>A. hydrophila</i> (clade 1)	0.0219	0.0010
<i>A. hydrophila</i> (clade 2)	0.0158	0.0012
<i>A. salmonicida</i>	0.0163	0.0009

Table 1. Mean genetic distances for all sequences and clusters calculated using the Tamura Nei (1993) distance. Standard error estimates were obtained by the bootstrap method with 500 replicates.

	TN93	Standard error
<i>A. bestiarum</i> vs. <i>A. salmonicida</i>	0.0655	0.0033
<i>A. bestiarum</i> vs. <i>A. hydrophila</i>	0.1042	0.0043
<i>A. hydrophila</i> vs. <i>A. salmonicida</i>	0.1101	0.0049
<i>A. bestiarum</i> (clade 1) vs. <i>A. bestiarum</i> (clade 2)	0.0507	0.0027
<i>A. bestiarum</i> (clade 1) vs. <i>A. bestiarum</i> (clade 3)	0.0279	0.0021
<i>A. bestiarum</i> (clade 2) vs. <i>A. bestiarum</i> (clade 3)	0.0526	0.0031
<i>A. hydrophila</i> (clade 1) vs. <i>A. hydrophila</i> (clade 2)	0.0558	0.0030

Table 2. Mean genetic distances among different clusters obtained using the Tamura Nei (1993) distance. Standard error estimates were obtained by the bootstrap method with 500 replicates.

Estimation of distance frequency within and between clusters from multilocus sequence data provides interesting insights into the population structure of these groups. The frequency distribution of the pairwise genetic distances clearly identifies the AHC species phylogenetic structure (Fig. 2). Within the strains of the *A. bestiarum* group (Fig. 2A), the plot shows three peaks with distance values ranging from 0 to 0.063. The lowest values correspond to pairwise comparisons among isolates within clades (1, 2 and 3) and the highest to those between clades. Distances between groups allowed a clear separation of *A.*

bestiarum from the other species groups. When we plotted the pairwise distance distribution within the *A. hydrophila* group (Fig. 2B) two peaks were shown and again the lowest values are those within the clades and the highest between the clades (0-0.062). *A. hydrophila* is clearly separated from *A. salmonicida* and *A. bestiarum* despite the overlap in the graph (Fig. 2B), which is a consequence of the similar distance values between *A. hydrophila* and *A. salmonicida* (0.110) and *A. hydrophila* and *A. bestiarum* (0.104). Otherwise, species boundaries are objectively defined in the phylogenetic tree (Fig. 1). In the *A. salmonicida* group, the within distance distribution (0-0.030) appears as a single peak, as does the interspecies distance.

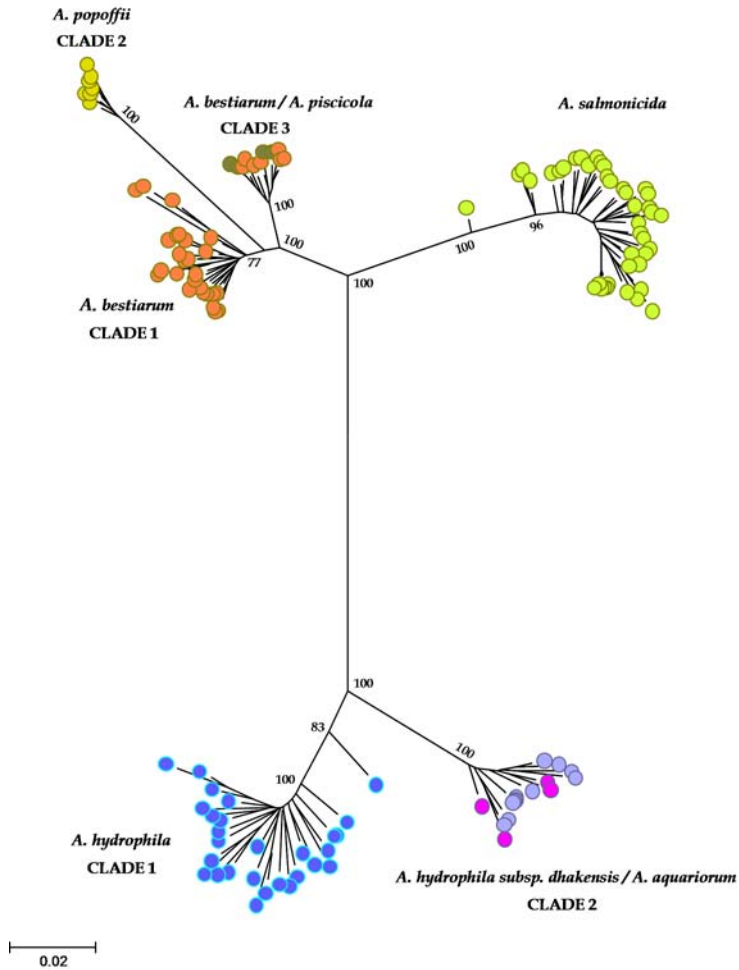


Fig. 1. Maximum likelihood tree inferred from the concatenated nucleotide sequences generated with MEGA5 software (Tamura et al., 2011). The tree was constructed using the Tamura Nei distance (TN93) and the rate of variation among sites was modelled with a gamma distribution (shape parameter = 0.5723) assuming heterogeneity of invariant sites

(best model for the data). Numbers at the branch nodes represent bootstrap values (500 replicates). The scale bar indicates the number of nucleotide substitutions per site. The GenBank/EMBL/DDBJ accession numbers of the nucleotide sequences used in this study are EU306796-EU306797, EU306804-EU306806, EU306814-EU306820, EU306822-EU306824, EU306826-EU306829, EU741625, EU741635-EU741636, EU741642, FJ936120, FJ936135, GU062399, JN711508-JN711610 (*cpn60* gene); FJ936122, FJ936136, JN215529-JN215534, JN711611-JN711731 (*dnaJ* gene); FJ936137, JN215535, JN711732-JN711858 (*gyrB* gene); HM163293-HM163294, HM163305-HM163307, HM163312-HM163318, JN660159-JN660273 (*mdh* gene); JN660274-JN660400 (*recA* gene) and EF465509-EF465510, FJ936132, FJ936138, JN215536-JN215541, JN712315-JN712433 (*rpoD* gene).

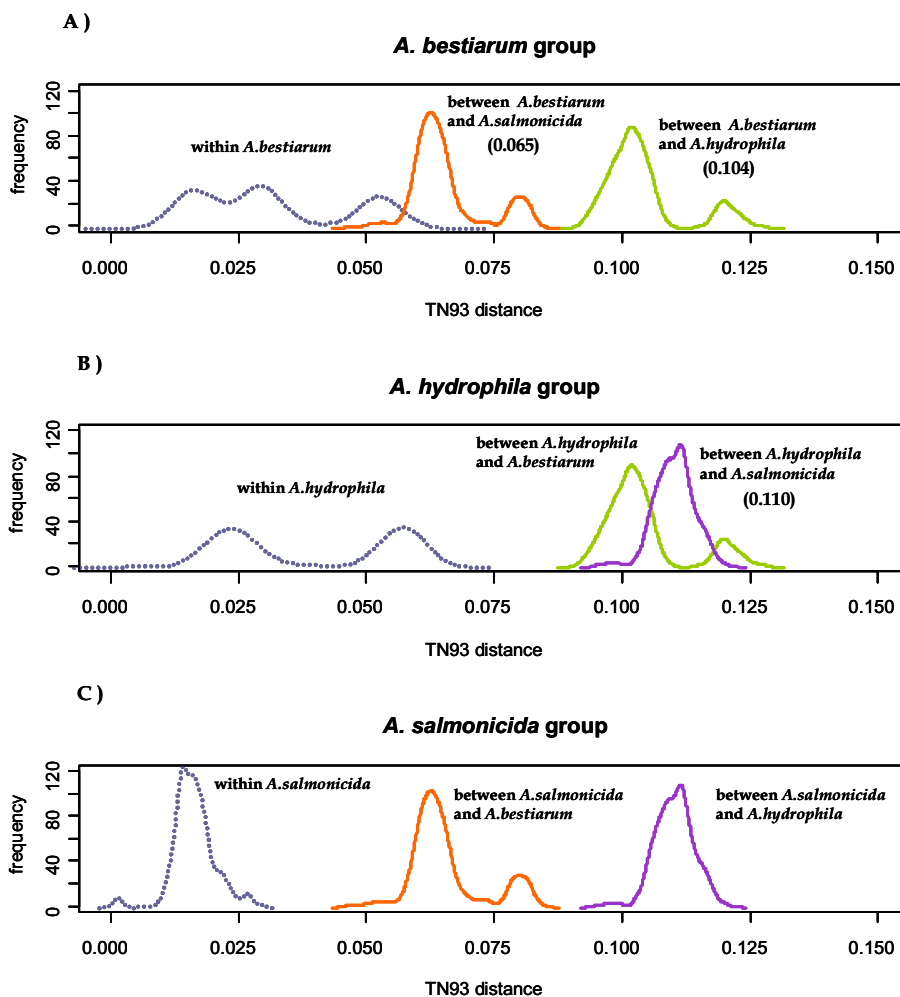


Fig. 2. Distribution of TN93 distances within and between clusters determined by pairwise comparisons. Mean distances are indicated within brackets.

Although to our knowledge there is no function that could be used to define a level of divergence for distinguishing species, it seems clear that, in bacteria, there is always an exponential relationship between interspecies recombination and sequence divergence (Roberts & Cohan, 1993; Zawadzki et al., 1995; Vulic et al., 1997). Values of divergence among clades (Table 2) seem to correspond to those of a genetically isolated biological species. The lowest value corresponds to *A. bestiarum* clade 3, indicating that its isolates are undergoing speciation, but species boundaries are not as well-defined as in the other clades.

5. Linkage disequilibrium and population structure

Linkage equilibrium is characterized by the statistical independence of alleles at all loci. A common method used in bacterial population genetics to quantify the degree of linkage between a set of loci in MLSA data is to use the index of association (I_A) (Brown et al., 1980; Maynard-Smith et al., 1993). This index compares the ratio of variance calculated from the pairwise distribution of allele mismatches in the data (V_O) and the ratio expected under a null hypothesis of linkage equilibrium (V_E). In this case

$$E(V_O) = \sum h_i(1-h_i),$$

where h_i is an unbiased estimator of the population genetic diversity equivalent to heterozygosity in diploid organisms (Table 4). The index of association (I_A) originally used by Brown et al. (1980) is computed as

$$I_A = V_O/V_E - 1,$$

which would give a value of zero if there is no association between loci (Maynard-Smith et al., 1993). Values of this index significantly different from zero reflect strong linkage disequilibrium (lower rates of recombination). The value of I_A depends on the number of loci analyzed. Haubold & Hudson (2000) subsequently proposed a standardized index of association (I_A^S) defined as

$$I_A^S = (1/l-1) I_A,$$

where l is the number of loci studied. This index has the advantage of being comparable between studies as long as it can be assumed that the neutral mutation parameter $\theta = 2Ne\mu$ is constant (Hudson, 1994). Two methods are commonly used for determining whether the computed I_A represents a significant deviation from linkage equilibrium ($I_A \sim 0$): resampling from randomized data sets using Monte Carlo simulations or using the parametric method described by Haubold et al. (1998). Both methods are perfectly good alternatives.

The genetic structure of the AHC, which has been previously determined using enzyme electrophoresis (MLEE), has revealed a clear clonal structure with strong linkage disequilibrium among 15 different protein loci (Miñana-Galbis et al., 2004). Additionally, this study demonstrated the usefulness of MLEE for separating the strains belonging to *A. bestiarum* and *A. salmonicida*, which are almost indistinguishable by phenotypic characteristics and 16S rRNA sequence analysis and present borderline DNA-DNA homology values.

In a multilocus sequence analysis (using six housekeeping genes) of a new set of AHC strains in which we included representatives of *A. piscicola* and *A. aquariorum*, we obtained

I_A^S values different from 0 in all cases, indicating the absence of recombination and again revealing strong linkage disequilibrium when considering both the total population and the different groups of species (Table 5). This is in spite of the high number of alleles per locus and polymorphic sites (Table 3) and huge genetic diversity (Table 4). Values of V_O , V_E , I_A^S and the 5% critical values as determined by the Monte Carlo process (L_{MC}) and from the parametric approach (L_{para}) of Haubold et al. (1998) are shown in Table 5.

Gene	Number of alleles	Number of polymorphic sites	$\pi \pm \text{s.e.}^{**}$
<i>cpn60</i>	95	144	0.075 \pm 0.009
<i>dnaJ</i>	102	237	0.082 \pm 0.007
<i>gyrB</i>	108	266	0.063 \pm 0.005
<i>mdh</i> *	113	225	0.059 \pm 0.009
<i>recA</i> *	106	282	0.071 \pm 0.009
<i>rpoD</i>	110	212	0.090 \pm 0.010

* full-length sequence gene; ** nucleotide diversity (π) \pm standard error (s.e.)

Calculated by using MEGA5 software (Tamura et al., 2011).

Table 3. Sequence variation at six loci. Standard error estimates were obtained by the bootstrap method with 500 replicates.

	STs	h_i^*						$H \pm \text{s.e.}^{**}$
		<i>cpn60</i>	<i>dnaJ</i>	<i>gyrB</i>	<i>mdh</i>	<i>recA</i>	<i>rpoD</i>	
All STs	127	0.9931	0.9959	0.9966	0.9978	0.9945	0.9973	0.9959 \pm 0.0007
<i>A. bestiarum</i>	44	0.9704	0.9863	0.9863	0.9905	0.9852	0.9873	0.9843 \pm 0.0029
<i>A. hydrophila</i>	42	0.9930	0.9884	0.9954	1.0000	0.9930	0.9930	0.9938 \pm 0.0015
<i>A. salmonicida</i>	41	0.9744	0.9902	0.9878	0.9890	0.9707	0.9951	0.9846 \pm 0.0039

* genetic diversity at individual loci (h_i); ** mean genetic diversity (H) \pm standard error (s.e.)

$h_j = (n/n-1) \sum p_{ij}^2$, where p_{ij} is the frequency of the i th allele at the j th locus and n the number of loci.

Data were calculated by using R statistical software (R Development Core Team, 2010).

Table 4. Genetic diversity (h) at six loci for all STs and major species sets.

	V_O	V_E	I_A^S	L_{para}	L_{MC}	P_{MC}
All STs	0.0635	0.0248	0.3131	0.0253	0.0255	< 1.00 \times 10 ⁻⁰⁴
<i>A. bestiarum</i>	0.2314	0.0924	0.3010	0.0986	0.1001	< 1.00 \times 10 ⁻⁰⁴
<i>A. hydrophila</i>	0.1382	0.0369	0.5495	0.0395	0.0405	< 1.00 \times 10 ⁻⁰⁴
<i>A. salmonicida</i>	0.1916	0.0908	0.2222	0.0978	0.0988	< 1.00 \times 10 ⁻⁰⁴

For the meaning of acronyms, see text. The L_{MC} and P_{MC} results were obtained from 10000 resamplings. Data were calculated by using R statistical software except for L_{para} , which was determined with the LIAN 3.5 program (Haubold & Hudson, 2000).

Table 5. Multilocus linkage disequilibrium analysis of the AHC.

During the last years, with the availability of the first DNA sequence data of individual genes, evidence of recombination at the molecular level has accumulated for *Aeromonas* in genes such as *dnaJ*, *gyrB* and *recA* (Silver et al., 2011). Incongruence between trees reconstructed from individual genes appeared as further proof of recombination at the gene level. Nevertheless, multilocus analysis with gene sequences also revealed a clear clonal

structure in this bacterial group (Table 5). The question is that although bacteria are capable of accumulating gene fragments from other bacterial species or mutations, the recombinant segments are not long enough to break the clonal structure of the population. While the absence of linkage ($I_A^S \sim 0$) is difficult to explain without assuming high levels of recombination, linkage disequilibrium does not exclude the presence of significant levels of recombination (Touchon et al., 2009).

In our study we have also determined the presence of recombinant fragments in the *recA* (in four *A. bestiarum* strains) and *dnaJ* genes (five strains, 2 *A. bestiarum*, 2 *A. hydrophila* and 2 *A. salmonicida*). However, although these strains cluster separately when the corresponding tree is constructed, revealing the different origin of the gene fragments, they group together with the other strains when a concatenated tree is generated. This confirms that recombination is not sufficient to break the genetic cohesion of this group.

6. Gene flow and divergence

The existence of barriers to gene flow such as geographical separation, ecological adaptation or the accumulation of genetic differences ultimately leads to distinct lineages. These processes are usually more complicated in prokaryotes, since the boundaries of their species are sometimes distorted by gene transfer between divergent organisms. In addition, the mechanisms that can contribute to the cohesion of groups are very different in bacteria. We have determined the divergence among the different AHC clades using the nucleotide fixation index, N_{ST} (Nei & Kumar, 2000). The use of the N_{ST} as a measure of population differentiation under certain circumstances has been criticized (Jost, 2008), but it is still used for describing the average amount of such differentiation observed from multiple locus data (Ryman & Leimar, 2009). In our study, the determination of the N_{ST} values indicated a high level of interclade genetic differentiation ($N_{ST} = 0.8025$). On average, most of the 80% of the total variance of nucleotide diversity was attributable to genetic differentiation among clades, whereas about 20% was found within populations. High levels of diversification were also found among most of the AHC groups of species (Table 6), with values always higher than 0.7 except in the case of *A. bestiarum* clade 3 / *A. bestiarum* clade 1 ($N_{ST} = 0.5$) and *A. hydrophila* clade 2 / *A. hydrophila* clade 1 ($N_{ST} = 0.6$). The clear divergence between the different clades described suggests they form coherent groups in which the phenomenon of recombination, if present, fails to break this consistency.

In bacterial populations it seems reasonable to equate the effect of lateral gene transfer (LTG) from other species with the product Nm (effective population size \times migration rate) determined from N_{ST} . Indeed, assuming the limitations of the Wright island model (Wright, 1940), the value of N_{ST} at the equilibrium ($1/(2Nm+1)$) allows Nm to be calculated. The values obtained for all AHC clades ($N_{ST} = 0.8$, $Nm \sim 0.13$) again suggest that gene flow (in this case, the lateral gene transfer) is insufficient to counteract their genetic differentiation.

The McDonald-Kreitman Test (MKT, McDonald & Kreitman, 1991) was applied to our data sets to detect signs of selection (Table 7). Under neutrality, the ratio of nonsynonymous-to-synonymous fixed substitutions (between species) should be the same as the ratio of nonsynonymous-to-synonymous polymorphism (within species). We observed a high level of fixed replacements between most species studied. In all comparisons the ratio of nonsynonymous-to-synonymous substitutions was higher for fixed differences than for

polymorphisms. This result agrees with the presence of mutations under positive selection spreading quickly through a population. These changes do not contribute to polymorphism but have a cumulative effect on divergence and the fixed NS/S is consequently greater than polymorphic NS/S (Egea et al., 2008). Two of the McDonald-Kreitman tests were not significant, *A. bestiarum* versus *A. salmonicida* and *A. piscicola* (*A. bestiarum* clade 3) versus the other strains of the *A. bestiarum* group.

	<i>A. best</i> clade 1	<i>A. best</i> clade 2	<i>A. best</i> clade 3	<i>A. hyd</i> clade 1	<i>A. hyd</i> clade 2	<i>A. salm</i>
<i>A. bestiarum</i> clade 1	---	0.1703	0.4532	0.1366	0.1068	0.1907
<i>A. bestiarum</i> clade 2	0.7460	---	0.1174	0.0863	0.0683	0.1051
<i>A. bestiarum</i> clade 3	0.5245	0.8099	---	0.1077	0.0833	0.1480
<i>A. hydrophila</i> clade 1	0.7854	0.8529	0.8228	---	0.1264	0.1264
<i>A. hydrophila</i> clade 2	0.8240	0.8798	0.8571	0.6428	---	0.0993
<i>A. salmonicida</i>	0.7239	0.8264	0.7716	0.7982	0.8343	---

Calculated using DnaSP v5.10 software (Librado & Rozas, 2009).

Table 6. Pairwise estimates of population differentiation, N_{ST} (lower-left) and gene flow, Nm (upper-right).

	Fixed		Polymorphic		NI	P
	S	NS	S	NS		
<i>A. bestiarum</i> vs. <i>A. hydrophila</i>	83	26	1235	88	0.227	0.000000 ***
<i>A. bestiarum</i> vs. <i>A. salmonicida</i>	57	7	964	77	0.650	0.325391 ns
<i>A. hydrophila</i> vs. <i>A. salmonicida</i>	118	36	1141	86	0.247	0.000000 ***
<i>A. bestiarum</i> (clades 1 and 3) vs. <i>A. bestiarum</i> (clade 2)	88	15	596	43	0.423	0.015757 *
<i>A. bestiarum</i> (clade 1 and 2) vs. <i>A. bestiarum</i> (clade 3)	21	3	659	48	0.510	0.402179 ns
<i>A. hydrophila</i> (clade 1) vs. <i>A. hydrophila</i> (clade 2)	56	16	803	48	0.209	0.000009 ***

Acronyms are for synonymous substitutions (S); nonsynonymous substitutions (NS); neutrality index (NI); P -value from Fisher's exact test (P). * $0.01 < P < 0.05$; ** $0.001 < P < 0.01$; *** $P < 0.001$; not significant (ns). Calculated using DnaSP v5.10 software (Librado & Rozas, 2009).

Table 7. McDonald-Kreitman Test for molecular evidence of selection.

7. Conclusions

Developments in gene sequence analysis have greatly enhanced the study of bacterial population genetics. Gene-wide approaches to mapping bacterial diversity, which have already proved effective for gaining fresh insight into bacterial evolution, have the potential to reveal the phenotypic basis of genetic diversity in the AHC and to investigate the dynamics of this complex bacterial community. The general objective of the work described in this chapter has been to evaluate the suitability of combining population genetics and phylogenetic approaches for the delineation of bacterial species in the AHC, considered by many specialists a taxonomically tangled group.

The results obtained from the linkage disequilibrium analysis and sequence divergence show that the AHC is composed of four robust groups that basically correspond with the

phenotypically described species *A. hydrophila*, *A. bestiarum*, *A. popoffii* and *A. salmonicida*. The average divergence between these clusters seems to exclude a significant influence of recombination in the genetic structure of this bacterial group and therefore they are valid taxonomic units, despite the extensive variability within some of them. Phenotypic characteristics lead to the differentiation of five *A. salmonicida* subspecies, but the lack of a consistent signal in our multilocus sequence analysis only allowed the possible differentiation of *A. salmonicida* subsp. *pectinolytica*. Similarly, it is impossible to differentiate between *A. hydrophila* subsp. *hydrophila* and *A. hydrophila* subsp. *ranae*. These results are in agreement with those obtained in a previous study with isolates belonging to the *A. veronii* Group, in which the authors failed to achieve the differentiation of biovars within these *Aeromonas* species (Silver et al., 2011). Nevertheless, *A. hydrophila* subsp. *dhakensis* strains, which cluster together with *A. aquariorum* isolates, exhibited the divergence levels of a biological species and hence deserve full species status. Consequently, the *A. aquariorum* isolates should be reclassified. Finally, in the *A. bestiarum* group we distinguished a clade (clade 3) that includes *A. piscicola* isolates as well as several strains probably misclassified as *A. bestiarum*. This clade seems to constitute an incipient new species with low values of differentiation and species boundaries less well defined than in *A. bestiarum* (clade 1) or *A. popoffii* (clade 2).

It has been frequently postulated that in bacterial populations, lateral gene transfer is so common that it precludes the existence of true biological species. One of the aims of this study has been to verify if this hypothesis is applicable to our AHC data. Three lines of evidence suggest the contrary. First of all, using the Tamura Nei model, which best fits our data, we found considerable interspecific nucleotide diversity, suggesting a high degree of divergence that hampers recombination among AHC species. Secondly, the linkage disequilibrium analysis of six loci reveals a strong disequilibrium with I_A^S values, suggesting little or null influence of recombination in the genetic structure of AHC species. Thirdly, the N_{ST} values obtained reflect a high degree of differentiation between clades. In short, the genetic structure of the AHC appears to confirm that the entities phenotypically described as species form cohesive groups in which genetic recombination plays a limited role in reducing genetic variation and can be defined as biological species.

Like other authors (Lan & Reeves, 2001; Vinuesa et al., 2005), we agree that a combination of phylogenetic and population genetic studies is currently the best theoretical and practical approach to delineate species as natural and discrete lineages in the bacterial world.

8. Acknowledgments

We are very grateful to Katri Berg (University of Helsinki, Finland), Margarita Gomila (Universitat de les Illes Balears, Mallorca, Spain) and Carmen Gallegos (Hospital Sant Llàtzer, Mallorca, Spain) for kindly providing some of the strains used in this study. This research was supported by the project CGL-2008-03281/BOS from the Ministerio de Educación y Ciencia, Spain.

9. References

- Abbott, L.S.; Cheung, W.K.W. & Janda, M.J. (2003). The genus *Aeromonas*: Biochemical characteristics, atypical reactions, and phenotypic identification schemes. *Journal of Clinical Microbiology*, Vol. 41, No. 6, pp. 2348-2357.

- Beaz-Hidalgo, R.; Alperi, A.; Figueras, M.J. & Romalde, J.L. (2009). *Aeromonas piscicola* sp. nov., isolated from diseased fish. *Systematic and Applied Microbiology*, Vol. 32, No. 7, pp. 471-479.
- Brown, A.D.; Feldman, M.W. & Nevo, E. (1980). Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, Vol. 96, No. 2, pp. 523-536.
- Egea, R.; Casillas, S. & Barbadilla, A. (2008). Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research*, Vol. 36, No. 2, pp. W157-W162.
- Farfán, M.; Miñana-Galbis, D.; Garreta, A.; Lorén, J.G. & Fusté, M.C. (2010). Malate dehydrogenase: a useful phylogenetic marker for the genus *Aeromonas*. *Systematic and Applied Microbiology*, Vol. 33, No. 8, pp. 427-435.
- Feil, E.J.; Maiden, M.C.J.; Achtman, M. & Spratt, B.G. (1999). The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Molecular Biology and Evolution*, Vol. 16, No. 11, pp. 1496-1502.
- Feil, E.J.; LI, B.C.; Aanensen, D.M.; Hanage, W.P. & Spratt, B.G. (2004). eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus typing data. *Journal of Bacteriology*, Vol. 186, No. 5, pp. 1518-1530.
- Figueras, M.J.; Soler, L.; Chacón, M.R.; Guarro, J. & Martínez-Murcia, A.J. (2000). Extended method for discrimination of *Aeromonas* spp. by 16S rDNA RFLP analysis. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 50, No. 6, pp. 2069-2073.
- Janda, J.M. & Abbott, S. (2010). The genus *Aeromonas*: taxonomy, pathogenicity, and infection. *Clinical Microbiology Reviews*, Vol. 23, No. 1, pp. 35-73.
- Jost, L. (2008). G_{ST} and its relatives do not measure differentiation. *Molecular Evolution*, Vol. 17, No. 18, pp. 4015-4026.
- Haubold, B.; Travisano, M.; Rainey, P.B. & Hudson, R.R. (1998). Detecting linkage disequilibrium in bacterial populations. *Genetics*, Vol. 150, No. 4, pp. 1341-1348.
- Haubold, B. & Hudson, R.R. (2000). LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics*, Vol. 16, No. 9, pp. 847-848.
- Hudson, R.R. (1994). Analytical results concerning linkage disequilibrium in models with genetic transformation and conjugation. *Journal of Evolutionary Biology*, Vol. 7, No. 5, pp. 535-548.
- Huys, G.; Kämpfer, P.; Altwegg, M.; Kersters, I.; Lamb, A.; Coopman, R.; Lüthy-Hottenstein, J.; Vancanneyt, M.; Janssen, P. & Kersters, K. (1997). *Aeromonas popoffii* sp. nov., a mesophilic bacterium isolated from drinking water production plants and reservoirs. *International Journal of Systematic Bacteriology*, Vol. 47, No. 4, pp. 1165-1171.
- Huys, G.; Kämpfer, P.; Albert, M.J.; Kühn, I.; Denys, R. & Swings, J. (2002). *Aeromonas hydrophila* subsp. *dhakensis* subsp. nov., isolated from children with diarrhoea in Bangladesh, and extended description of *Aeromonas hydrophila* subsp. *hydrophila* (Chester 1901) Stanier 1943 (Approved Lists 1980). *International Journal of Systematic Bacteriology*, Vol. 52, No. 3, pp. 705-712.
- Küpfer, M.; Kuhnert, P.; Korczyk, B.M.; Peduzzi, R. & Demarta, A. (2006). Genetic relationships of *Aeromonas* strains inferred from 16S rRNA, *gyrB* and *rpoB* gene sequences. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 56, No. 12, pp. 2743-2751.
- Lan, R. & Reeves, P.R. (2001). When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends in Microbiology*, Vol. 9, No. 9, pp. 419-424.
- Levin, B.R. & Bergstrom, C.T. (2000). Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in

- prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 97, No. 13, pp. 6981-6985.
- Li, Y.; Liu, Y.; Zhou, Z.; Huang, H.; Ren, Y.; Zhang, Y.; Li, G.; Zhou, Z. & Wang, L. (2011). Complete genome sequence of *Aeromonas veronii* strain B565. *Journal of Bacteriology*, Vol. 193, No. 13, pp. 3389-3390.
- Librado, P. & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, Vol. 25, No. 11, pp. 1451-1452.
- Maiden, M.C.; Bygraves, J.A.; Feil, E.; Morelli, G.; Russell, J.E.; Urwin, R.; Zhang, Q.; Zhou, J.; Zurth, K.; Caugant, D.A.; Feavers, I.M.; Achtman, M. & Spratt, B.G. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 95, No. 6, pp. 3140-3145.
- Martin-Carnahan, A. & Joseph, S.W. (2005). Genus I. *Aeromonas* Stanier 1943, 213^{AL}. In: *Bergey's Manual of Systematic Bacteriology*, Garrity, G.M.; Brenner, D.J.; Krieg, N.R. & Staley, J.T. (eds.), Vol. 2, part B, pp. 557-578, Springer, New York.
- Martínez-Murcia, A.J.; Soler, L.; Saavedra, M.J.; Chacón, M.R.; Guarro, J.; Stackebrandt, E. & Figueras, M.J. (2005). Phenotypic, genotypic, and phylogenetic discrepancies to differentiate *Aeromonas salmonicida* from *Aeromonas bestiarum*. *International Microbiology*, Vol. 8, No. 4, pp. 259-269.
- Martínez-Murcia, A.J.; Saavedra, M.J.; Mota, V.R.; Maier, T.; Stackebrandt, E. & Cousin, S. (2008). *Aeromonas aquariorum* sp. nov., isolated from aquaria of ornamental fish. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 58, No. 5, pp. 1169-1175.
- Martínez-Murcia, A.J.; Monera, A.; Saavedra, M.J.; Oncina, R.; López-Alvarez, M.; Lara, E. & Figueras, M.J. (2011). Multilocus phylogenetic analysis of the genus *Aeromonas*. *Systematic and Applied Microbiology*, Vol. 34, No. 3, pp. 189-199.
- Martino, M.E.; Fasolato, L.; Montemurro, F.; Rosteghin, M.; Manfrin, A.; Patarnello, T.; Novelli, E. & Cardazzo, B. (2011). Determination of microbial diversity of *Aeromonas* strains on the basis of multilocus sequence typing, phenotype, and presence of putative virulence genes. *Applied and Environmental Microbiology*, Vol. 77, No. 14, pp. 4986-5000.
- Maynard-Smith, J. (1991). The population genetics of bacteria. *Proceedings of the Royal Society B Biological Sciences*, Vol. 245, No. 1312, pp. 37-41.
- Maynard-Smith, J.; Smith, N.H.; O'Rourke, M. & Spratt, B.G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 90, No. 10, pp. 4384-4388.
- Maynard-Smith, J. (1995). Do bacteria have population genetics?, In: *Population Genetics of Bacteria*, Baumberg, S.; Young, J.P.W.; Wellington, E.M.H. & Saunders, J.R. (eds.), pp. 1-12, Cambridge University Press, Cambridge.
- McDonald, J.H. & Kreitman, M. (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, Vol. 351, pp. 652-654.
- Milkman, R. (1973). Electrophoretic variation in *Escherichia coli* from natural sources. *Science*, Vol. 182, No. 116, pp. 1024-1026.
- Miñana-Galbis, D.; Farfán, M.; Lorén, J.G. & Fusté, M.C. (2002). Biochemical identification and numerical taxonomy of *Aeromonas* spp. isolated from environmental and clinical samples in Spain. *Journal of Applied Microbiology*, Vol. 93, No. 3, pp. 420-430.
- Miñana-Galbis, D.; Farfán, M.; Fusté, M.C. & Lorén, J.G. (2004). Genetic diversity and population structure of *Aeromonas hydrophila*, *Aer. bestiarum*, *Aer. salmonicida* and

- Aer. popoffii* by multilocus enzyme electrophoresis (MLEE). *Environmental Microbiology*, Vol. 6, No. 3, pp. 198-208.
- Miñana-Galbis, D.; Urbizu-Serrano, A.; Farfán, M.; Fusté, M.C. & Lorén, J.G. (2009). Phylogenetic analysis and identification of *Aeromonas* species based on sequencing of the *cpn60* universal target. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 59, No. 8, pp. 1976-1983.
- Nei, M., & Kumar, S. (2000). Genetic polymorphism and evolution, In: *Molecular Evolution and Phylogenetics*, pp. 231-264, Oxford University Press, New York.
- Nhung, P.H.; Hata, H.; Ohkusu, K.; Noda, M.; Shah, M.M.; Goto, K. & Ezaki, T. (2007). Use of the novel phylogenetic marker *dnaJ* and DNA-DNA hybridization to clarify interrelationships within the genus *Aeromonas*. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 57, No. 6, pp. 1232-1237.
- O'Rourke, M. & Stevens, E. (1993). Genetic structure of *Neisseria gonorrhoeae* populations: a non-clonal pathogen. *Journal of General Microbiology*, Vol. 139, No. 11, pp. 2603-2611.
- Orskov, F.; Whittam, T.S.; Cravioto, A. & Orskov, I. (1990). Clonal relationships among classic enteropathogenic *Escherichia coli* (EPEC) belonging to different O groups. *The Journal of Infectious Diseases*, Vol. 162, No. 1, pp. 76-81.
- Pérez-Losada, M.; Brown, E.B.; Madsen, A.; Wirth, T.; Viscidi, R.P. & Crandall, K.A. (2006). Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infection, Genetics and Evolution*, Vol. 6, No. 2, pp. 97-112.
- R Development Core Team (2010). *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.r-project.org/>.
- Reith, M.E.; Singh, R.K.; Curtis, B.; Boyd, J.M.; Bouevitch, A.; Kimball, J.; Munholland, J.; Murphy, C.; Sarty, D.; Williams, J.; Nash, J.H.E.; Johnson, S.C. & Brown, L.L. (2008). The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. *BMC Genomics*, Vol. 9, No. 427.
- Roberts, M.S. & Cohan, F.M. (1993). The effect of DNA sequence divergence on sexual isolation in *Bacillus*. *Genetics*, Vol. 134, No. 2, pp. 401-408.
- Robinson, A.; Falush, D. & Feil, E.J. (eds.) (2010). *Bacterial Population Genetics in Infectious Disease*, Wiley & Sons, ISBN 978-0-470-42474-2, New Jersey.
- Roselló-Mora, R. & Amann, R. (2001). The species concept for prokaryotes. *FEMS Microbiology Reviews*, Vol. 25, No. 1, pp. 39-67.
- Ryman, N. & Leimar, O. (2009). G_{ST} is still a useful measure of genetic differentiation - a comment on Jost's D. *Molecular Ecology*. Vol. 18, No. 10, pp. 2084-2087.
- Saavedra, M.J.; Figueras, M.J. & Martínez-Murcia, A.J. (2006). Updated phylogeny of the genus *Aeromonas*. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 56, No. 10, pp. 2481-2487.
- Salaun, L.; Audibert, C.; Le Lay, G.; Burocoa, C.; Fauchère, J.L. & Picard, B. (1998). Panmictic structure of *Helicobacter pylori* demonstrated by the comparative study of six genetic markers. *FEMS Microbiology Letters*, Vol. 161, No. 2, pp. 231-239.
- Seshadri, R.; Joseph, S.W.; Chopra, A.K.; Sha, J.; Shaw, J.; Graf, J.; Haft, D.; Wu, M.; Ren, Q.; Rosovitz, M.J.; Madupu, R.; Tallon, L.; Kim, M.; Jin, S.; Vuong, H.; Stine, O.C.; Ali, A.; Horneman, A.J. & Heidelberg, J.F. (2006). Genome sequence of *Aeromonas hydrophila* ATCC 7966: jack of all trades. *Journal of Bacteriology*, Vol. 188, No. 23, pp. 8272-8282.
- Selander, R.K.; Caugant, D.A.; Ochman, H.; Musser, J.M.; Gilmour, M.N. & Whittam, T.S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population

- genetics and systematics. *Applied and Environmental Microbiology*, Vol. 51, No. 5, pp. 873-884.
- Selander, R.K.; Beltran, P.; Smith, N.H.; Helmuth, R.; Rubin, F.A.; Kopecko, D.J.; Ferris, K.; Tall, B.D.; Cravioto, A. & Musser, J.M. (1990). Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. *Infection & Immunity*, Vol. 58, No. 7, pp. 2262-2275.
- Silver, A.C.; Williams, D.; Faucher, J.; Horneman, A.J.; Gogarten, J.P. & Graf, J. (2011). Complex evolutionary history of the *Aeromonas veronii* group revealed by host interaction and DNA sequence data. *PLoS ONE*, Vol. 6, No. 2, e16751.
- Soler, L.; Yañez, M.A.; Chacón, M.R.; Aguilera-Arreola, M.G.; Catalán, V.; Figueras, M.J. & Martínez-Murcia, A.J. (2004). Phylogenetic analysis of the genus *Aeromonas* based on two housekeeping genes. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 54, No. 5, 1511-1519.
- Stackebrandt, E.; Frederiksen, W.; Garrity, G.M.; Grimont, P.A.D.; Kämpfer, P.; Maiden, M.C.J.; Nesme, X.; Rosselló-Mora, R.; Swings, J.; Trüper, H.G.; Vauterin, L.; Ward, A.C. & Whitman, W.B. (2002). Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, Vol. 52, No. 3, 1043-1047.
- Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M. & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, DOI: 10.1093/molbev/msr121.
- Touchon, M.; Hoede, C.; Tenaillon, O.; Barbe, V.; Baeriswy, S.; Bidet, P.; Bingen, E.; Bonacorsi, S.; Bouchier, C.; Bouvet, O.; Calteau, A.; Chiapello, H.; Clermont, O.; Cruveiller, S.; Danchin, A.; Diard, M.; Dossat, C.; Karoui, M.E.; Frapy, E.; Garry, L.; Ghigo, J.M.; Gilles, A.M.; Johnson, J.; LeBouguéneq, C.; Lescat, M.; Mangenot, S.; Martínez-Jéhanne, V.; Matic, I.; Nassif, X.; Oztas, S.; Petit, M.A.; Pichon, C.; Rouy, Z.; Ruf, C.S.; Schneider, D.; Tourret, J.; Vacherie, B.; Vallenet, D.; Médigue, C.; Rocha, E.P.C. & Denamur, E. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genetics*, Vol. 5, No. 1, e1000344.
- Vinuesa, P.; Silva, C.; Werner, D. & Martínez-Romero, E. (2005). Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. *Molecular Phylogenetics and Evolution*, Vol. 34, No. 1, pp. 29-54.
- Vogel, U., Schoen, C. & Elias, J. (2010). Population genetics of *Neisseria meningitidis*. In: *Bacterial Population Genetics in Infectious Disease*, Robinson, A.; Falush, D. & Feil, E.J. (eds.), pp. 247-267, Wiley & Sons, ISBN 978-0-470-42474-2, New Jersey.
- Vulic, M.; Dionisio, F.; Taddei, F. & Radman, M. (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 94, No. 18, pp. 9763-9767.
- Willems, R.J. (2010). Population genetics of *Enterococcus*. In: *Bacterial Population Genetics in Infectious Disease*, Robinson, A.; Falush, D. & Feil, E.J. (eds.), pp. 195-216, Wiley & Sons, ISBN 978-0-470-42474-2, New Jersey.
- Wright, S. (1940). Breeding structure of populations in relation to speciation. *American Naturalist*, Vol. 74, No. 752, pp. 232-248.
- Zawadzki, P.; Roberts, M.S. & Cohan, F.M. (1995). The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. *Genetics*, Vol. 140, No. 3, pp. 917-932.



Transworld Research Network
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

Recent Advances in Pharmaceutical Sciences III, 2013: 179-193 ISBN: 978-81-7895-605-3
Editors: Diego Muñoz-Torrero, Amparo Cortés and Eduardo L. Mariño

11. The effect of recombination in *Aeromonas*

Maribel Farfán^{1,2}, Vicenta Albarral^{1,2}, Ariadna Sanglas^{1,2}, J. Gaspar Lorén¹
and M. Carmen Fusté^{1,2}

¹Departament de Microbiologia i Parasitologia Sanitàries, Facultat de Farmàcia, Universitat de Barcelona. Av. Joan XXIII s/n, 08028 Barcelona, Spain; ²Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona. Av. Diagonal 643, 08028 Barcelona, Spain

Abstract. Although several approaches have been attempted, the estimation of recombination frequencies in natural populations of bacteria remains challenging. Previous studies have demonstrated a wide variety of situations among bacterial species, ranging from the clonal diversification of *Salmonella* or *Escherichia coli*, which are mainly due to mutation, to the frequent recombination found in *Neisseria gonorrhoeae* or *Helicobacter pylori*. Most of the population studies done with bacterial species suggest that recombination occurs in nature but that it is infrequent compared to mutation. Consequently, bacterial populations consist largely of independent clonal lineages. Our research suggests little or null influence of recombination in the genetic structure of '*Aeromonas hydrophila* Species Complex', despite the presence of some strains with recombinant gene fragments.

Introduction

Bacteria reproduce asexually, giving two identical individuals after their division, with the exception of changes produced by mutation or recombination. Although this reproduction process is not associated with

Correspondence/Reprint request: Dr. Maribel Farfán, Departament de Microbiologia i Parasitologia Sanitàries Facultat de Farmàcia, Universitat de Barcelona. Av. Joan XXIII s/n, 08028 Barcelona, Spain
E-mail: mfarfan@ub.edu

recombination, in contrast with eukaryotes, bacteria have acquired three basic mechanisms by which they can incorporate genes from other bacterial species: transformation, conjugation, and transduction. Nevertheless, although bacteria might incorporate foreign genes from other species, their genomes are not simply arbitrary assortments of genes of mixed heritage. Once the bacterial cells acquire genes by means of one of these mechanisms, they have to be incorporated into a replicon; if not, the introduced genes would become diluted in the population or degraded by the restriction endonucleases. Bacterial interchange promotes the acquisition of novel genetic elements, the impact of which has been extensively studied in human and animal pathogens and commensals, where they are often associated with the emergence of new phenotypes [1]. Recombination in bacteria is: always restricted to small DNA fragments, unidirectional, independent of reproduction, and occurs with a relatively low frequency, although genes codifying virulence factors or antibiotic resistance experiment more frequent recombination changes. The incorporation of genes or parts of genes through recombination always results in mosaic genomes that are composed of regions with different evolutionary histories [2]. Homologous recombination is widespread in the genomes of many bacteria, and is usually a consequence of *recA* mediated homology-dependant recombination. When the incorporated fragment replaces an identical DNA sequence its effect cannot be detected, although the process seems to be very frequent when the two bacteria involved in this interchange are closely related. Homologous recombination in this case might play a crucial role in DNA repair [3]. On the contrary, if recombination has a measurable effect on the genome of the recipient, it is considered as an effective recombination event.

The impact of recombination on bacterial phylogenies has been the subject of considerable discussion [4-9]. Recently, with the availability of sequencing techniques and the analytical power of new programmes, the detection of recombination events has increased dramatically. This has led to the questioning of existing phylogenies and the methods used for their construction, such as Maximum Likelihood (ML) and Maximum Parsimony (MP), which assume that the analyzed sequences have the same evolutionary history. The presence of recombination would break this assumption, since in this case sequences would have different underlying phylogenies that are more easily envisaged as a network rather than a tree. Due to the importance of recombination in evolutionary analysis, it is essential to be able to identify whether a given set of sequences has undergone recombination events, define the boundaries of the recombinational units, and evaluate the impact of recombination on our ability to reconstruct evolutionary histories and estimate population genetic parameters.

To investigate genetic exchange in bacteria, large data sources have been used, such as those deposited in the Multi Locus Sequence Typing (MLST) databases at www.mlst.net [10]. Multilocus data allow us to determine the degree of recombination based on the type of population structure. If the population shows a clonal structure (linkage disequilibrium), then recombination is absent and the accumulated genetic changes will be a consequence of mutation. However, in the case of bacteria, this assumption is not always true and the clonal structure is not always broken, even if a certain degree of recombination is present [11, 12].

The increasing availability of whole genome sequences has enabled a more complete study of the recombination process in bacteria. The genes sequenced in a MLST study (usually six or seven) are not always representative of the entire genome, and can give biased results. The analyzed sequences do not correspond to the complete gene, only fragments of about 400-500 bp, so the changes determined cannot be representative of the full gene, particularly in the case of a protein-codifying gene. Unfortunately, full genome sequences that could obviate all these questions are still usually limited to a few isolates of each species and frequently have been chosen for specific reasons (clinical, environmental or industrial, etc.), so are not representative of the entire population.

In this study we will consider the impact of recombination on bacterial phylogenies and the consequences of inaccurate approaches to inferring phylogenetic relationships. Traditionally, recombination in a given set of sequences has been identified by the incongruence of the different gene trees analyzed, the presence of mosaic structures, and variations in the G+C content or the codon bias. Several methods have been developed to test the presence of recombination in a given set of sequences, and to identify the parental and recombinant individuals or the recombinational break-points. Those methods can be classified in different categories: similarity, distance, phylogenetic, compatibility, and nucleotide substitution distribution [13, 14]. The performance of these methods varies depending on the level of recombination, but in general most of them are efficient, and although they can have trouble in detecting recombination when the level of divergence is low, their discriminatory power increases when the level of recombination is high [5]. In addition, methods that use the substitution patterns or incompatibility among sites seem to be more powerful than those based on phylogenetic incongruence. This might be partially explained by the fact that, in general, phylogenetic methods can only detect recombination events that change the topology of the tree, and at high recombination rates there should be many such events [5]. What is important is to increase the chances of detecting recombination while minimizing the detection of false positives, so

the chosen method will depend on the level of divergence among the sequences analysed [15, 16]. In either case, the best option is not to rely on a single method to detect recombination [13, 17].

1. Recombination in bacteria

Although several approaches have been attempted, the estimation of recombination frequencies in natural populations of bacteria remains challenging. One of the parameters commonly used is the determination of the rate of recombination relative to mutation [18], but this is not always easy to calculate, except in the case of the most recent events [4], which might not be representative of the entire history of the population.

The determination of the relative importance of mutation in comparison with recombination is central to bacterial population genetics [18]. Previous studies have demonstrated a wide variety of situations among bacterial species, ranging from the clonal diversification of *Salmonella* [19] or *Escherichia coli* [20], which are mainly due to mutation, to the frequent recombination found in *Neisseria gonorrhoeae* [21] or *Helicobacter pylori* [22]. Most of the population studies done with bacterial species suggest that recombination occurs in nature, and indeed may be highly important in generating variation, but that it is infrequent compared to mutation. Consequently, bacterial populations consist largely of independent clonal lineages.

Comparison of results from analyses performed with different methodologies is problematic; nevertheless, studies using the same methods across different genera have suggested wide variation in recombination rates with value differences of two or three orders of magnitude [23]. In addition, conflicting levels of recombination have been obtained for concrete bacterial species, such as *Escherichia coli*, *Staphylococcus aureus*, *Bacillus cereus* or *Haemophilus influenzae* [24], depending on the sampling and analytical methodologies used. The analyzed isolates have to be sampled carefully in order to be informative about the underlying recombination process. If they are not representative of the whole population, an important bias might be introduced. Unfortunately, in most of the populations studied, particularly in the case of pathogen bacteria, samples are not fully representative, usually with an overrepresentation of virulent strains, which are frequently under higher selective forces than the corresponding non-virulent strains.

Recombination studies using whole genome data have contributed to a better understanding of recombination in bacteria. Several studies have reported differences in the prevalence of recombination at different regions of the same bacterial genome [25], being apparently higher in those genes under

positive selection [25, 26]. Genomic regions encoding proteins with a role in pathogenicity are often under positive selection and frequently exhibit high rates of recombination, even in the case of bacteria in which recombination is relatively rare [28, 29]. A possible explanation for the relative prevalence of recombination in positive selected regions of the bacterial genome is that the only observable recombination is likely to be the one that unites beneficial mutations and removes deleterious ones.

2. The genus *Aeromonas*

The genus *Aeromonas* Stanier 1943 belongs to the family *Aeromonadaceae* within the class *Gammaproteobacteria* [30]. Aeromonads are autochthonous inhabitants of aquatic environments, including chlorinated and polluted waters, although they can also be isolated from a wide variety of environmental and clinical sources. They cause infections in vertebrates and invertebrates, such as frogs, birds, various fish species, and domestic animals. In recent years, some authors have considered *Aeromonas* as an emergent pathogen in humans, producing intestinal and extraintestinal diseases. Aeromonads are facultative anaerobic chemoorganotrophs capable of anaerobic nitrate respiration and dissimilatory metal reduction [30].

Several attempts have been made to generate phylogenies using DNA gene sequences to reconstruct the correct genealogical ties among species in *Aeromonas* [31-33], but the genes chosen for this purpose are not always suitable, and do not necessarily give congruent phylogenies [34, 35]. Recently, two papers presenting MLST schemes for *Aeromonas* have been published [10, 36], and there is an online MLST database for the genus *Aeromonas*, managed by Keith Jolley and curated by Barbara Cardazzo (<http://pubmlst.org/aeromonas>). All this accumulated data should help to establish a reliable clustering of the *Aeromonas* species and elucidate their exact boundaries.

Finally, the availability of complete genomes of different species is also useful in this task, but unfortunately, in the case of *Aeromonas*, only six genomes have been completed to date, corresponding to: the type strain of *A. hydrophila* ATCC 7966, isolated from a tin of milk [37]; the strain A449 of *A. salmonicida*, a fish pathogen described by Reith *et al.* ([38], Fig. 1); an *A. caviae* strain Ae398 isolate from a stool sample [39]; an *A. veronii* strain B565 isolated from an aquaculture pond sediment [40]; and more recently, an *A. aquariorum* strain AAK1 isolated from blood [41] and the highly melanin-yielding *A. media* strain WS [42]. The information given by the genomes of *A. hydrophila* and *A. salmonicida* indicates that while they are of identical size (4.7Mb) and share multiple housekeeping and virulence genes, *A. salmonicida*

has acquired several mobile genetic elements, and undergone genome rearrangements and loss of genes in the process of adapting to a specific host. The genome of *A. veronii* is smaller (4.3Mb) and contains fewer virulence genes. Similarly, *A. caviae* presents a small genome (4.43Mb), but in contrast to *A. veronii* several putative virulence genes have been identified, as in *A. aquariorum*, which has the biggest genome reported for *Aeromonas* (4.81Mb). The last genome completed corresponds to an *A. media* strain (4.3Mb) in which no virulence genes have been reported.

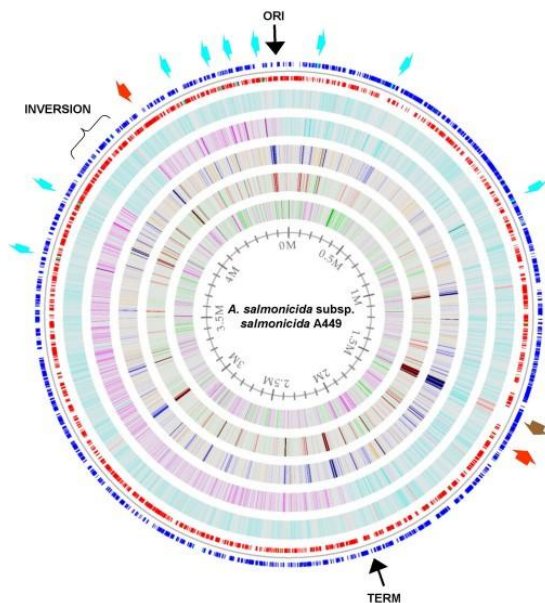


Figure 1. A genome atlas representation of the *A. salmonicida* subsp. *salmonicida* A449 chromosome (ref. [38]).

3. The ‘*Aeromonas hydrophila* species complex’

An example of the taxonomic complexity of the genus *Aeromonas* is the difficulty in discriminating between the phenotypically and genetically closely related species belonging to the “*Aeromonas hydrophila* species complex” (AHC), which includes: *A. hydrophila*, composed of three subspecies: *A. hydrophila* subsp. *hydrophila*, *A. hydrophila* subsp. *ranae* and *A. hydrophila* subsp. *dhakensis*, *A. bestiarum*, *A. popoffii*, and *A. salmonicida*,

divided in five subspecies: *A. salmonicida* subsp. *salmonicida*, *A. salmonicida* subsp. *masoucida*, *A. salmonicida* subsp. *achromogenes*, *A. salmonicida* subsp. *pectinolytica*, and *A. salmonicida* subsp. *smithia* [30, 43]. Recently, two additional species have been described in this group, *A. aquariorum* and *A. piscicola* [44, 45]. Members of the AHC were first described as strains producing the enzymes elastase, lecitinase or stapholysin [46]. They are genetically closely related and share multiple phenotypic characteristics, which makes discrimination among the species included in this group extremely difficult [43].

In order to establish the population structure and divergence of the species included in the AHC group, Fusté *et al.* [12] studied a set of representative strains, in which they analyzed the nucleotide sequences (total or partial) of 6 housekeeping genes. The authors concluded, from the linkage disequilibrium analysis and sequence divergence results, that the AHC is composed of four robust groups that basically correspond with the phenotypically described species *A. hydrophila*, *A. bestiarum*, *A. popoffii*, and *A. salmonicida*, in which recombination, if present, does not break their clonal structure.

4. Population structure and recombination in *Aeromonas*

The few references in the bibliography dealing with recombination in *Aeromonas* reach similar conclusions about its low incidence, with the exception of the study by Silver *et al.* [35], which reports a notable effect of recombination in the “*A. veronii* species complex”. In this study, the strains were obtained from patients, veterinary samples, and medicinal leeches. The aligned sequences were investigated for evidence of recombination because the maximum-likelihood inferred phylogenies for each gene family showed low bootstrap support for most clades. Applying two tests for recombination, employing a variety of approaches, it was demonstrated that at least for some strains, horizontal gene transfer occurs at a sufficient frequency to blur the signal from vertically inherited genes, despite strains being adapted to distinct niches.

Martino *et al.* [11] analyzed a collection of *Aeromonas* including 23 type and reference strains, and 77 strains isolated from fish, crustaceans and molluscs in a MLST study using 6 housekeeping genes. Based on an eBURST analysis, the authors determined the recombination/mutation (r/m) ratio for the entire population and for the three major groups identified. The r/m values obtained (ranging between 0.07 and 0.13) suggest a reduced rate of recombination. Analysis with the SplitsTree program revealed that most of the genes, although showing a reticulate network, were not significantly affected by intragenic recombination with the exception of *recA*. When a set

of programmes included in the RDP3 package were applied to the *Aeromonas* sequences, several recombination events were identified supported by at least three of those programmes. Nevertheless, they concluded that, in the case of *Aeromonas*, the impact of recombination may be negligible, based on the very similar topologies of the phylogenetic trees, the low *r/m* rates, and the reduced network structure determined by the split decomposition analysis.

Roger *et al.* [47], in a MLST study that includes isolates from different origins but with a particularly high representation of clinical strains, determined that the standardized I_A (I_A^S) values showed the existence of significant linkage disequilibrium, indicative of a clonal population structure. When using at least four methods of the RDP software, they detected some recombination among the population in all but two of the seven loci analyzed. In addition, the use of split decomposition determined that most of the sequence types (STs) were not affected by recombination, even though more recombination events were found within the clonal complexes, particularly for the STs in the *A. caviae* clade. Differences in branching order were observed in both distance and ML trees when gene and concatenated sequence trees were compared, suggesting the occurrence of recombination. The authors conclude that recombination is present in the genus *Aeromonas*, at least in some strains, but at a relatively low frequency.

Finally, our group determined the genetic population structure of a group of *Aeromonas* corresponding to the AHC [12], which had been previously analyzed by enzyme electrophoresis (MLEE), revealing a clear clonal structure with strong linkage disequilibrium among 15 different protein loci [48]. We used a higher number of AHC isolates including representatives of *A. piscicola* [45] and *A. aquariorum* [44], two more recently described strains grouped within the AHC. The I_A^S values obtained in this study were different from 0 in all cases, indicating the absence of recombination and again, revealing strong linkage disequilibrium when considering both the total population and the different sets of species. This is in spite of the high number of alleles per locus and polymorphic sites and huge genetic diversity.

During the last years, with the availability of the first DNA sequence data of individual genes, evidence of recombination at the molecular level has accumulated for *Aeromonas* in housekeeping genes such as *dnaJ*, *gyrB* and *recA* [11, 35, 47] or structural and accessory genes [35]. In our study we have also determined the presence of potential recombinant fragments in the *recA* and *dnaJ* genes of some strains. However, although these strains cluster separately when the corresponding gene tree is constructed, revealing a possible different origin of the gene fragments, they group together with the other strains of the same species when a concatenated tree is generated. This confirms that recombination is not sufficient to break the genetic cohesion of this group.

5. Intragenic recombination in the *dnaJ* gene of the ‘*Aeromonas hydrophila* species complex’

We have recently studied the possible existence of recombination in the *dnaJ* gene (Fig. 2) in an AHC group with 90 strains (87% of environmental and 13% of clinical origin). Group I included 29 *A. salmonicida* strains (5 subspecies), Group II 31 strains (22 *A. bestiarum*, 5 *A. popoffii* and 4 *A. piscicola*) and Group III 30 strains (26 *A. hydrophila* and 4 *A. aquariorum*). Gene sequences used were obtained from our previous work [12]. We added a new strain *A. hydrophila* JCM 3968 (GenBank accession numbers: JN711671 (*dnaJ*), JN711562 (*cpn60*), JN711795 (*gyrB*), KC525968 (*mdh*), KC525969 (*recA*) and JN712375 (*rpoD*)). We determined the presence of mosaic structures, different G+C content and codon usage bias in the sequences but none of the results were conclusive.

In our study we also detected incongruences in the phylogenies when the *dnaJ* gene and the concatenated trees were compared (Fig. 3). Five strains, 2 *A. bestiarum* (orange), 1 *A. hydrophila* (blue), and 2 *A. salmonicida* (green), clustered out of the corresponding species group in the *dnaJ* gene tree (Fig. 3a), revealing a possible different origin of the gene fragments. Nevertheless, they grouped together with the other strains of the same species when a concatenated tree was generated (Figs. 3b and 4).

Consequently, we also analyzed our sequences using six recombination detection programmes in the RDP4 package, which significantly detected possible recombination events in one (*A. salmonicida* CECT 5214) out of the five strains, when all strains or the corresponding species subgroups were analyzed (Table 1). The investigation also provided well-supported evidence for recombination in two *A. hydrophila* strains (CECT 4330 and CECT 5734), which clustered among the other *A. hydrophila* isolates in the *dnaJ* and the concatenated trees, although they were separated in a deeper branch. No recombination events were detected among the *A. bestiarum* isolates (Group II).

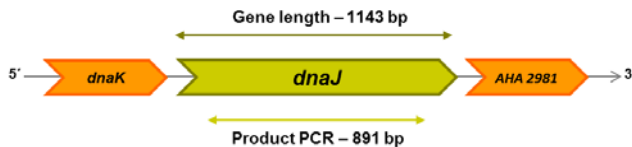


Figure 2. Schematic representation of the *dnaJ* locus and its flanking regions for *Aeromonas*, based on the whole genome sequence of *Aeromonas hydrophila* ATCC 7966^T (GenBank accession number CP000462, [37]). Partial sequences for *dnaJ* (891 bp) were obtained from the GenBank database or determined as previously described [49].

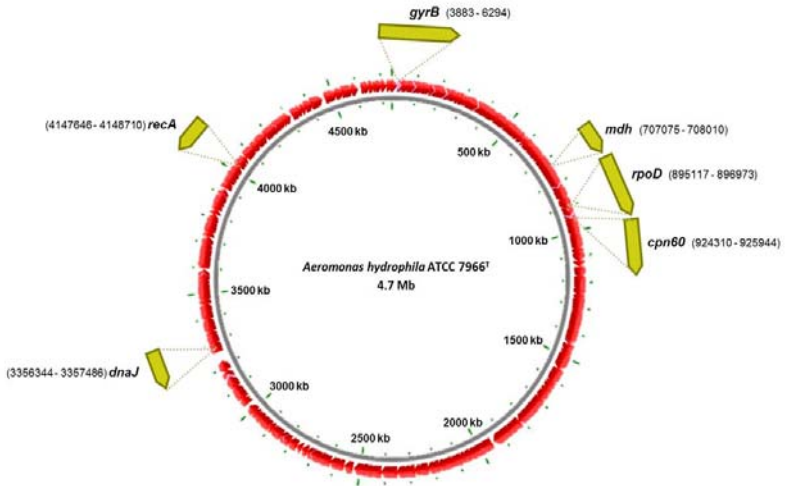


Figure 4. Distribution of six genes sequenced in the circular map of the genome of *A. hydrophila* ATCC 7966^T (GenBank accession number CP000462, [37]). Genes are shown outside the circle and have standard abbreviations. Arrows indicate direction of transcription. Detailed genomic position is listed in parentheses after each gene name.

Table 1. Recombination analysis summary.

	Recombination sequence	Number of methods detecting recombination events with statistical significance	Recombination detection methods						Events	
			1	2	3	4	5	6		
<i>All strains</i> N=90	<i>A. salmonicida</i> CECT 5214	3	█		█				█	1
Group I <i>A. salmonicida</i> N=29	<i>A. salmonicida</i> CECT 5214	4		█	█	█			█	1 or 2
	<i>A. salmonicida</i> CECT 5219	1				█				1
Group III <i>A. hydrophila</i> N=30	<i>A. hydrophila</i> CECT 4330	2				█			█	1
	<i>A. hydrophila</i> CECT 5734	2							█	1

Automated screening for recombination from multiple alignment of *dnaJ* sequences was performed using programme RDP4 (<http://darwin.uvigo.es/rdp/rdp.html>, [51]), which used six recombination detection programmes: RDP (1), GENECONV (2), BootScan (3), MaxChi (4), Chimaera (5) and 3Seq (6), with their default parameters. Sequences statistically supported by at least two recombination detection programmes (*P*-value <0.05) were considered as possible recombinants.

6. Conclusions

Developments in gene sequence analysis have greatly enhanced the study of recombination in bacterial populations. Gene-wide approaches to mapping bacterial diversity, which have already proved effective for gaining insight into bacterial evolution, have the potential to reveal the phenotypic basis of genetic diversity in *Aeromonas*, and to investigate the dynamics of this complex bacterial community. The objective of the work described in this chapter has been to evaluate the importance of the presence of recombination events and their influence on phylogenies, as it has been frequently postulated that in bacterial populations, horizontal gene transfer (HGT) is so common that it precludes the existence of biological species. Our research suggests little or null influence of recombination in the genetic structure of AHC species, despite the existence of some strains with recombinant gene fragments.

Assuming that the cohesion of major phylogenetic groups within the prokaryotes is due to vertical transmission and common ancestry rather than preferential HGT, it is possible to construct robust phylogenies reflecting the evolutionary history of bacteria, using a sufficient number of orthologous housekeeping genes (concatenated trees). In these phylogenies, bacterial species are delineable as ‘classical Darwinian’ evolutionary lineages [52-55].

The foregoing consideration does not exclude the existence of horizontal gene transfer, which in fact occurs, and has important evolutionary consequences, but it is doubtful that HGT is the essence of modern genome phylogeny [53]. Moreover, as demonstrated in *Salmonella*, *Streptococcus*, and *Bacillus*, homologous recombination decays exponentially with sequence divergence; in other words, a sequence divergence between two strains of 10% suppresses the recombination rate between them by a factor of about 100 [56, 57]. We are currently pursuing recombination studies in other genes of this *Aeromonas* group.

Acknowledgements

This research was supported by projects from the Ministerio de Educación y Ciencia, Spain (CGL2008-03281/BOS), and the Universitat de Barcelona (ARZ00F01).

References

1. Hacker, J., Carniel, E. 2001. *EMBO Rep.*, 2, 376.
2. Spratt, B. G., Bowler, L. D., Zhang, Q. Y., Zhou, J., Smith, J. M. 1992. *J. Mol. Evol.*, 34, 115.

3. Michod, R. E., Bernstein, H., Nedelcu, A. M. 2008. *Infect. Genet. Evol.*, 8, 267.
4. Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M. S., Day, N. P., Enright, M. C., Goldstein, R., Hood, D. W., Kalia, A., Moore, C. E., Zhou, J., Spratt, B. G. 2001. *Proc. Natl. Acad. Sci. USA*, 98, 182.
5. Posada, D., Crandall, K. A., Holmes, E. C. 2002. *Annu. Rev. Genet.*, 36, 75.
6. Hanage, W. P., Fraser, C., Spratt, B. G. 2005. *BMC Biol.*, 3, 6.
7. Vinuesa, P., Silva, C., Werner, D., Martínez-Romero, E. 2005. *Mol. Phylogenet. Evol.*, 34, 29.
8. Ochman, H., Lerat, E., Daubin, V. 2005. *Proc. Natl. Acad. Sci. USA*, 102, 6595.
9. Didelot, X., Maiden, M. C. 2010. *Trends Microbiol.*, 18, 315.
10. Aanensen, D. M., Spratt, B. G. 2005. *Nucleic. Acids. Res.*, 33, W728.
11. Martino, M. E., Fasolato, L., Montemurro, F., Rosteghin, M., Manfrin, A., Patarnello, T., Novelli, E., Cardazzo, B. 2011. *Appl. Environ. Microbiol.*, 77, 4986.
12. Fusté, M. C., Farfán, M., Miñana-Galbis, D., Albarral, V., Sanglas, A., Lorén, J. G. 2012. Population genetics of the '*Aeromonas hydrophila* Species Complex'. In *Studies in Population Genetics*, M. C. Fusté (Ed.), InTech, Croatia, 39.
13. Posada, D., Crandall, K. A. 2001. *Proc. Natl. Acad. Sci. USA*, 98, 13757.
14. Lemey, P., Posada, D. 2009. Introduction to recombination detection. In *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd Ed., A. M. Vandamme, M. Salemi, P. Lemey (Eds.), Cambridge University Press, New York, 493.
15. Sawyer, S. A. 1999. GENECONV: A computer package for the statistical detection of gene conversion. Distributed by the author, *Department of Mathematics, Washington University in St. Louis*, available at <http://www.math.wustl.edu/~sawyer>.
16. Martin, D., Rybicki, E. 2000. *Bioinformatics*, 16, 562.
17. Wiuf, C., Christensen, T., Hein, J. 2001. *Mol. Biol. Evol.*, 18, 1929.
18. Milkman, R., Bridges, M. M. 1990. *Genetics*, 126, 505.
19. Feil, E. J., Maiden, M. C., Achtman, M., Spratt, B. G. 1999. *Mol. Biol. Evol.*, 16, 1496.
20. Selander, R. K., Beltran, P., Smith, N. H., Helmuth, R., Rubin, F. A., Kopecko, D. J., Ferris, K., Tall, B. D., Cravioto, A., Musser, J. M. 1990. *Infect. Immun.*, 58, 2262.
21. Orskov, F., Whittam, T. S., Cravioto, A., Orskov, I. 1990. *J. Infect. Dis.*, 162, 76.
22. O'Rourke, M., Stevens, E. 1993. *J. Gen. Microbiol.*, 139, 2603.
23. Salaun, L., Audibert, C., Le Lay, G., Burucoa, C., Fauchère, J. L., Picard, B. 1998. *FEMS Microbiol. Lett.*, 161, 231.
24. Pérez-Losada, M., Browne, E. B., Madsen, A., Wirth, T., Viscidi, R. P., Crandall, K. A. 2006. *Infect. Genet. Evol.*, 6, 97.
25. Denamur, E., Lecomte, G., Darlu, P., Tenaillon, O., Acquaviva, C., Sayada, C., Sunjevaric, I., Rothstein, R., Elion, J., Taddei, F., Radman, M., Matic, I. 2000. *Cell*, 103, 711.
26. Lefébure, T., Stanhope, M. J. 2007. *Genome Biol.*, 8, R71.
27. Orsi, R. H., Sun, Q., Wiedmann, M. 2008. *BMC Evol. Biol.*, 8, 233.
28. Liu, X., Gutacker, M. M., Musser, J. M., Fu, Y. X. 2006. *J. Bacteriol.*, 188, 8169.
29. Gomes, J. P., Bruno, W. J., Nunes, A., Santos, N., Florindo, C., Borrego, M. J., Dean, D. 2007. *Genome Res.*, 17, 50.

30. Martin-Carnahan, A., Joseph, S. W. 2005. Genus I. *Aeromonas* Stanier 1943, 213^{AL}. In *Bergey's Manual of Systematic Bacteriology*, G. M. Garrity, D. J. Brenner, N. R. Krieg, J. T. Staley (Eds.), Springer, New York, Vol. 2. Part B, 557.
31. Saavedra, M. J., Figueras, M. J., Martínez-Murcia, A. J. 2006. *Int. J. Syst. Evol. Microbiol.*, 56, 2481.
32. Küpfer, M., Kuhnert, P., Korczak, B. M., Peduzzi, R., Demarta, A. 2006. *Int. J. Syst. Evol. Microbiol.*, 56, 2743.
33. Miñana-Galbis, D., Urbizu-Serrano, A., Farfán, M., Fusté, M. C., Lorén, J. G. 2009. *Int. J. Syst. Evol. Microbiol.*, 59, 1976.
34. Farfán, M., Miñana-Galbis, D., Garreta, A., Lorén, J. G., Fusté, M. C. 2010. *Syst. Appl. Microbiol.*, 33, 427.
35. Silver, A. C., Williams, D., Faucher, J., Horneman, A. J., Gogarten, J. P., Graf, J. 2011. *PLoS One*, 6, e16751.
36. Martínez-Murcia, A. J., Monera, A., Saavedra, M. J., Oncina, R., López-Alvarez, M., Lara, E., Figueras, M. J. 2011. *Syst. Appl. Microbiol.*, 34, 189.
37. Seshadri, R., Joseph, S. W., Chopra, A. K., Sha, J., Shaw, J., Graf, J., Haft, D., Wu, M., Ren, Q., Rosovitz, M. J., Madupu, R., Tallon, L., Kim, M., Jin, S., Vuong, H., Stine, O. C., Ali, A., Horneman, A. J., Heidelberg, J. F. 2006. *J. Bacteriol.*, 188, 8272.
38. Reith, M. E., Singh, R. K., Curtis, B., Boyd, J. M., Bouevitch, A., Kimball, J., Munholland, J., Murphy, C., Sarty, D., Williams, J., Nash, J. H., Johnson, S. C., Brown, L. L. 2008. *BMC Genomics.*, 9, 427.
39. Beatson, S. A., das Graças de Luna, M., Bachmann, N. L., Alikhan, N. F., Hanks, K. R., Sullivan, M. J., Wee, B. A., Freitas-Almeida, A. C., Dos Santos, P. A., de Melo, J. T., Squire, D. J., Cunningham, A. F., Fitzgerald, J. R., Henderson, I. R. 2011. *J. Bacteriol.*, 193, 1286.
40. Li, Y., Liu, Y., Zhou, Z., Huang, H., Ren, Y., Zhang, Y., Li, G., Zhou, Z., Wang, L. 2011. *J. Bacteriol.*, 193, 3389.
41. Wu, C. J., Wang, H. C., Chen, C. S., Shu, H. Y., Kao, A. W., Chen, P. L., Ko, W. C. 2012. *J. Bacteriol.*, 194, 4114.
42. Chai, B., Wang, H., Chen, X. 2012. *J. Bacteriol.*, 194, 6693.
43. Miñana-Galbis, D., Farfán, M., Lorén, J. G., Fusté, M. C. 2002. *J. Appl. Microbiol.*, 93, 420.
44. Martínez-Murcia, A. J., Saavedra, M. J., Mota, V. R., Maier, T., Stackebrandt, E., Cousin, S. 2008. *Int. J. Syst. Evol. Microbiol.*, 58, 1169.
45. Beaz-Hidalgo, R., Alperi, A., Figueras, M. J., Romalde, J. L. 2009. *Syst. Appl. Microbiol.*, 32, 471.
46. Abbott, S. L., Cheung, W. K., Janda, J. M. 2003. *J. Clin. Microbiol.*, 41, 2348.
47. Roger, F., Marchandin, H., Jumas-Bilak, E., Kodjo, A., BVH, C., Lamy, B. 2012, *BMC Microbiol.*, 12, 62.
48. Miñana-Galbis, D., Farfán, M., Fusté, M. C., Lorén, J. G. 2004. *Environ. Microbiol.*, 6, 198.
49. Nhung, P. H., Hata, H., Ohkusu, K., Noda, M., Shah, M. M., Goto, K., Ezaki, T. 2007. *Int. J. Syst. Evol. Microbiol.*, 57, 1232.

50. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. 2011. *Mol. Biol. Evol.*, 28, 2731.
51. Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefevre, P. 2010. *Bioinformatics*, 26, 2462.
52. Daubin, V., Gouy, M., Perrière, G. 2002. *Genome Res.*, 12, 1080.
53. Kurland, C. G., Canback, B., Berg, O. G. 2003. *Proc. Natl. Acad. Sci. USA.*, 100, 9658.
54. Lerat, E., Daubin, V., Moran, N. A. 2003. *PLoS Biol.*, 1, E19.
55. Vinuesa, P., Silva, C., Werner, D., Martínez-Romero, E. 2005. *Mol. Phylogenet. Evol.*, 34, 29.
56. Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F. M., Dowson, C. G. 2000. *J. Bacteriol.*, 182, 1016.
57. Fraser, C., Hanage, W. P., Spratt, B. G. 2007, *Science*, 315, 4.



Short communication

Reclassification of *Aeromonas hydrophila* subspecies *anaerogenes*

David Miñana-Galbis*, Maribel Farfán, Vicenta Albarral, Ariadna Sanglas, J. Gaspar Lorén, M. Carmen Fusté

Departament de Microbiologia i Parasitologia Sanitàries, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain

ARTICLE INFO

Article history:

Received 15 November 2012

Received in revised form 9 April 2013

Accepted 15 April 2013

Keywords:

*Aeromonas**Aeromonas hydrophila* subspecies *anaerogenes**Aeromonas caviae**Aeromonas hydrophila*

Taxonomy

Reclassification

Housekeeping genes

ABSTRACT

Technological advances together with the continuous description of new taxa have led to frequent reclassifications in bacterial taxonomy. In this study, an extensive bibliographic revision, as well as a sequence analysis of nine housekeeping genes (*cpn60*, *dnaJ*, *dnaX*, *gyrA*, *gyrB*, *mdh*, *recA*, *rpoB* and *rpoD*) and a phenotypic identification of *Aeromonas hydrophila* subspecies *anaerogenes* were performed. All data obtained from previous physiological, phylogenetic, and DNA–DNA hybridization studies together with those presented in this study suggested that *A. hydrophila* subspecies *anaerogenes* belonged to the species *Aeromonas caviae* rather than *A. hydrophila*. Therefore, the inclusion of *A. hydrophila* subsp. *anaerogenes* in the species *A. caviae* is proposed.

© 2013 Elsevier GmbH. All rights reserved.

Bacterial taxonomy is constantly changing due to the large number of reclassifications and the description of novel taxa. DNA–DNA hybridization remains the gold standard technique for defining bacterial species, particularly when 16S rRNA gene sequence similarity values are higher than 97–98% [36,38]. However, the lack of comparative DNA–DNA hybridization studies on species belonging to highly related genera or the publication of conflicting data has promoted several species reassignments [1,3,18,23]. In contrast, sequence analysis of housekeeping genes provides higher discriminatory power at the interspecific and intraspecific levels than 16S rRNA gene sequence analysis. As an excellent alternative to DNA–DNA reassociation, the advantages of the housekeeping gene technique include the availability for most laboratories, high reproducibility, low cost and speed, and the open access of the resulting sequences in public databases. Analysis of a set of five, or even fewer, housekeeping genes, with a multilocus sequence analysis (MLSA) approach, is useful for species delineation, phylogenetic analysis and population genetic studies [4,9,15,17,21,34,39] and, therefore, this technique could supplant DNA–DNA hybridization.

Taxonomic reclassifications have also resulted from the application of new techniques to classical taxa and the inclusion of data from the description of novel taxa. In the present study, the taxonomic position of *Aeromonas hydrophila* subspecies *anaerogenes*

was re-evaluated and its formal inclusion in *Aeromonas caviae* is proposed.

Anaerogenic strains of *A. hydrophila* were described by Schubert in 1964 [30] under the designation *A. hydrophila* subspecies *anaerogenes*, with strain ATCC 15467^T (=544^T = CECT 4221^T = CIP 76.15^T) being designated as the type strain. This taxon was included in the description of the genus *Aeromonas* in the 8th edition of *Bergey's Manual of Determinative Bacteriology* [31]. In a later taxonomic study, Popoff and Véron [27] showed that *A. hydrophila* and *Aeromonas punctata* were grouped into the same cluster in which aerogenic and anaerogenic strains could be divided into two biovars, *A. hydrophila* biovar *hydrophila* and *A. hydrophila* biovar *anaerogenes*, respectively. In this study, Popoff and Véron proposed the emendation of the *A. hydrophila* description with the consideration of *A. punctata* as an illegitimate synonym of *A. hydrophila*, the formal description of *Aeromonas sobria* sp. nov., and the exclusion of *A. hydrophila* subsp. *proteolytica* from the genus *Aeromonas*. Nevertheless, *A. hydrophila* with three subspecies (*A. hydrophila* subsp. *hydrophila*, *A. hydrophila* subsp. *anaerogenes*, and *A. hydrophila* subsp. *proteolytica*) and *A. punctata* with two subspecies (*A. punctata* subsp. *punctata*, and *A. punctata* subsp. *caviae*) were included in the *Approved Lists of Bacterial Names* [32], whereas the name *A. sobria* was not validly published until 1981 [20,28], and *A. hydrophila* subsp. *proteolytica* was reclassified as *Vibrio proteolyticus* in 1980 and validated in 1982 [2,20].

On the basis of DNA–DNA hybridization experiments, Popoff et al. [26] showed that *A. hydrophila* biovar *hydrophila* included

* Corresponding author. Tel.: +34 93 402 44 97; fax: +34 93 402 44 98.
E-mail address: davidminyana@ub.edu (D. Miñana-Galbis).

three hybridization groups (HGs), *A. hydrophila* biovar *anaerogenes* had up to another three HGs and *A. sobria* two HGs. These HGs correspond to the current species *A. hydrophila* (HG1), *Aeromonas bestiarum* (HG2), *Aeromonas salmonicida* (HG3), *Aeromonas caviae* (HG4), *Aeromonas media* (HG5), *Aeromonas eucrenophila* (HG6), *A. sobria* (HG7), and *Aeromonas veronii* biovar *sobria* (HG8) [13]. In the same study by Popoff et al., the type strain of *A. hydrophila* subsp. *anaerogenes* (strain 544^T) belonged to the same HG of the type strain of *A. caviae* (strain 545^T = ATCC 15468^T = CECT 838^T), which was an *Aeromonas* species described in the first edition of *Bergey's Manual of Systematic Bacteriology* [25] and validly published in the same year [20]. For this reason, *A. hydrophila* subsp. *anaerogenes* has not been included in the list of species and subspecies of the genus *Aeromonas* described in the last two editions of *Bergey's Manual of Systematic Bacteriology* [13,25] or *The Prokaryotes* [7,8]. In 2002, Huys et al. [10] emended the description of *A. hydrophila* subsp. *hydrophila* and proposed *A. hydrophila* subsp. *dhakensis* as a new *A. hydrophila* subspecies. They also considered the taxonomical status of other *A. hydrophila* subspecies, such as *A. hydrophila* subsp. *anaerogenes* and *A. hydrophila* subsp. *proteolytica*, indicating that, following the description of the genus *Aeromonas* by Popoff [25], *A. hydrophila* subsp. *anaerogenes* should not be considered as a subspecies of *A. hydrophila*. Popoff established that *A. hydrophila* subsp. *anaerogenes* comprised three DNA–DNA hybridization groups, which were subsequently placed in *A. caviae* (groups 239 and 545) and *A. eucrenophila* (strain 546), as already mentioned above. However, as *A. hydrophila* subsp. *anaerogenes* was not validly reclassified, this *Aeromonas* subspecies is still a validated subspecies name [5].

Other evidence has suggested that *A. hydrophila* subsp. *anaerogenes* and *A. caviae* belong to the same taxon. A multilocus enzyme electrophoresis (MLEE) study of *A. hydrophila*, *A. caviae* and *A. sobria* showed that strains 544 (*A. hydrophila* subsp. *anaerogenes*) and 545 (*A. caviae*) belonged to the same zymotype (C₂), and strain 543 (=ATCC 7966^T = CECT 839^T), the type strain of *A. hydrophila*, constituted a distinct zymotype (H₃) [24]. In a phylogenetic study of the genera *Aeromonas* and *Plesiomonas* [14], the 16S rRNA gene sequence of the type strain of *A. hydrophila* subsp. *anaerogenes* (GenBank accession no. X60409) was reported to exhibit 2 and 16 nucleotide differences with those of the type strains of *A. caviae* (GenBank accession no. X60408) and *A. hydrophila* subsp. *hydrophila* (GenBank accession no. X60404), respectively. More recently, different phylogenetic studies based on sequence analysis of housekeeping genes [6,11,12,33], such as *gyrB*, *mdh*, *rpoB* and *rpoD*, and the multilocus sequence analysis (MLSA) published by Martínez-Murcia et al. [15] and Roger et al. [29], have shown that the type strain of *A. hydrophila* subsp. *anaerogenes* clusters with the *A. caviae* type strain at sequence distances within the ranges of intraspecific values, and they should be considered as synonyms.

In this current study, sequences of nine genes (*cpn60*, *dnaj*, *dnaX*, *gyrA*, *gyrB*, *mdh*, *recA*, *rpoB* and *rpoD*) were analyzed from the type strains of *A. hydrophila* subsp. *anaerogenes*, *A. caviae* and *A. hydrophila* subsp. *hydrophila* taken from either GenBank or those previously obtained by our group (Supplementary material in SAM Online). Sequence alignments and Jukes-Cantor distances were obtained using the MEGA5 software [35]. The distance values obtained by comparing the gene sequences of *A. hydrophila* subsp. *anaerogenes* CECT 4221^T and *A. caviae* CECT 838^T were <3%, which was below the interspecific threshold values [6,11,15,19]. In contrast, *A. hydrophila* subsp. *anaerogenes* CECT 4221^T and *A. hydrophila* subsp. *hydrophila* CECT 839^T showed distance values higher than 4.5%, and therefore should be considered as different species (Table 1). Distance values from the concatenated sequences were 1.96% between *A. hydrophila* subsp. *anaerogenes* and *A. caviae*, and 7.41% between *A. hydrophila* subsp. *anaerogenes* and *A. hydrophila*

Table 1

Sequence distances (%) and nucleotide differences between *A. hydrophila* subsp. *anaerogenes* CECT 4221^T and the type strains of *A. caviae* and *A. hydrophila*.

Genes (num. of nt analyzed)	<i>A. caviae</i>	<i>A. hydrophila</i>
Concatenated (6642 nt)		
nt differences	127	464
Jukes-Cantor distances	1.96%	7.41%
<i>cpn60</i> (555 nt)		
nt differences	3	39
Jukes-Cantor distances	0.54%	7.38%
<i>dnaj</i> (843 nt)		
nt differences	15	78
Jukes-Cantor distances	1.80%	9.88%
<i>dnaX</i> (492 nt)		
nt differences	11	42
Jukes-Cantor distances	2.27%	9.06%
<i>gyrA</i> (705 nt)		
nt differences	4	34
Jukes-Cantor distances	0.57%	4.98%
<i>gyrB</i> (1113 nt)		
nt differences	28	56
Jukes-Cantor distances	2.56%	5.21%
<i>mdh</i> (936 nt)		
nt differences	22	68
Jukes-Cantor distances	2.39%	7.64%
<i>recA</i> (597 nt)		
nt differences	17	40
Jukes-Cantor distances	2.90%	7.02%
<i>rpoB</i> (516 nt)		
nt differences	11	24
Jukes-Cantor distances	2.16%	4.80%
<i>rpoD</i> (816 nt)		
nt differences	16	83
Jukes-Cantor distances	1.99%	10.93%

subsp. *hydrophila*. All gene sequence alignments are available as Supplementary material in SAM Online.

As the type strain of *A. hydrophila* subsp. *anaerogenes* is considered to belong to the species *A. caviae*, few phenotypic studies of the genus *Aeromonas* have included this strain in recent years [22,37]. For this reason, strains *A. hydrophila* subsp. *anaerogenes* CECT 4221^T, *A. caviae* CECT 838^T and *A. hydrophila* subsp. *hydrophila* CECT 839^T were analyzed by 16 key phenotypic tests useful for the phenotypic differentiation of *Aeromonas* species [16]. The results obtained for the type strains of *A. caviae* and *A. hydrophila* subsp. *hydrophila* were identical to those previously described [16], and strain CECT 4221^T (*A. hydrophila* subsp. *anaerogenes*) showed identical phenotypic features to those of CECT 838^T (*A. caviae*). Five key tests allowed a clear phenotypic differentiation between *A. hydrophila* subsp. *hydrophila* and *A. caviae*/*A. hydrophila* subsp. *anaerogenes* type strains, whereas, in contrast to *A. hydrophila* subsp. *anaerogenes* CECT 4221^T and *A. caviae* CECT 838^T, *A. hydrophila* subsp. *hydrophila* CECT 839^T was positive for the lysine decarboxylase and Voges–Proskauer tests and the production of gas from D-glucose, as well as hydrogen sulfide from cysteine and elastase.

On the basis of bibliographic data from previous phylogenetic, MLSA, MLEE and DNA–DNA hybridization studies and new evidence from physiological tests and distance values obtained from the housekeeping gene analysis performed in this study, the reclassification of *A. hydrophila* subsp. *anaerogenes* to *A. caviae* is proposed, with the consequent loss of its taxonomic rank as a subspecies.

Acknowledgements

We would like to thank M. Palau for her contribution to this study. This work has been supported by a project from the Ministerio de Educación y Ciencia, Spain (CGL2008-03281/BOS).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.syapm.2013.04.006>.

References

- [1] Ara, I., Tsetseg, B., Daram, D., Suto, M., Ando, K. (2012) *Herbidospira mongolien-sis* sp. nov., isolated from soil, and reclassification of *Herbidospira osyris* and *Streptosporangium claviforme* as synonyms of *Herbidospira cretacea*. *Int. J. Syst. Evol. Microbiol.* 62, 2322–2329.
- [2] Baumann, P., Baumann, L., Bang, S.S., Woolkalis, M.J. (1980) Reevaluation of the taxonomy of *Vibrio*, *Beneckeia*, and *Photobacterium*: abolition of the genus *Beneckeia*. *Curr. Microbiol.* 4, 127–132.
- [3] Coorevits, A., Dinsdale, A.E., Halket, G., Lebbe, L., De Vos, P., Van Landschoot, A., Logan, N.A. (2012) Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (nom. corr., formerly 'thermoglucosidiasius'); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus* comb. nov.; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* comb. nov.; and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. *Int. J. Syst. Evol. Microbiol.* 62, 1470–1485.
- [4] de la Haba, R.R., Márquez, M.C., Papke, R.T., Ventosa, A. (2012) Multilocus sequence analysis of the family *Halomonadaceae*. *Int. J. Syst. Evol. Microbiol.* 62, 520–538.
- [5] Euzéby, J.P. 2012 List of Prokaryotic Names with Standing in Nomenclature <http://www.bacterio.net>
- [6] Farfán, M., Miñana-Galbis, D., Garreta, A., Lorén, J.G., Fusté, M.C. (2010) Malate dehydrogenase: a useful phylogenetic marker for the genus *Aeromonas*. *Syst. Appl. Microbiol.* 33, 427–435.
- [7] Farmer, J.J., III, Arduino, M.J., Hickman-Brenner, F.W. (1992) The genera *Aeromonas* and *Plesiomonas*. In: Balows, A., Trüper, H.G., Dworkin, M., Harder, W., Schleifer, K.H. (Eds.), *The Prokaryotes: A Handbook on the Biology of Bacteria*, 2nd ed., Springer-Verlag, New York, pp. 3012–3045.
- [8] Farmer, J.J., III, Arduino, M.J., Hickman-Brenner, F.W. (2006) The genera *Aeromonas* and *Plesiomonas*. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., Stackebrandt, E. (Eds.), *The Prokaryotes: A Handbook on the Biology of Bacteria*, vol. 6, 3rd ed., Springer, New York, pp. 564–596.
- [9] Fusté, M.C., Farfán, M., Miñana-Galbis, D., Albarral, V., Sanglas, A., Lorén, J.G. (2012) Population genetics of the “*Aeromonas hydrophila* species complex”. In: Fusté, M.C. (Ed.), *Studies in Population Genetics*, InTech, Rijeka, pp. 39–54.
- [10] Huys, G., Kämpfer, P., Albert, M.J., Kühn, I., Denys, R., Swings, J. (2002) *Aeromonas hydrophila* subsp. *dhakensis* subsp. nov., isolated from children with diarrhoea in Bangladesh, and extended description of *Aeromonas hydrophila* subsp. *hydrophila* (Chester 1901) Stanier 1943 (approved lists 1980). *Int. J. Syst. Evol. Microbiol.* 52, 705–712.
- [11] Küpfer, M., Kuhnert, P., Korczak, B.M., Peduzzi, R., Demarta, A. (2006) Genetic relationships of *Aeromonas* strains inferred from 16S rRNA, *gyrB* and *rpoB* gene sequences. *Int. J. Syst. Evol. Microbiol.* 56, 2743–2751.
- [12] Lamy, B., Laurent, F., Kodjo, A. (2010) Validation of a partial *rpoB* gene sequence as a tool for phylogenetic identification of aeromonads isolated from environmental sources. *Can. J. Microbiol.* 56, 217–228.
- [13] Martin-Carnahan, A., Joseph, S.W. (2005) Genus I. *Aeromonas* Stanier 1943, 213^{AL}, in: Garrity, G.M., Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), *Bergey's Manual of Systematic Bacteriology: Part B*, vol. 2, 2nd ed., Springer, New York, pp. 557–578.
- [14] Martínez-Murcia, A.J., Benlloch, S., Collins, M.D. (1992) Phylogenetic interrelationships of members of the genera *Aeromonas* and *Plesiomonas* as determined by 16S ribosomal DNA sequencing: lack of congruence with results of DNA–DNA hybridizations. *Int. J. Syst. Bacteriol.* 42, 412–421.
- [15] Martínez-Murcia, A.J., Monera, A., Saavedra, M.J., Oncina, R., López-Alvárez, M., Lara, E., Figueras, M.J. (2011) Multilocus phylogenetic analysis of the genus *Aeromonas*. *Syst. Appl. Microbiol.* 34, 189–199.
- [16] Miñana-Galbis, D., Farfán, M., Lorén, J.G., Fusté, M.C. (2010) Proposal to assign *Aeromonas diversa* sp. nov. as a novel species designation for *Aeromonas* group 501. *Syst. Appl. Microbiol.* 33, 15–19.
- [17] Miñana-Galbis, D., Farfán, M., Lorén, J.G., Fusté, M.C. (2010) The reference strain *Aeromonas hydrophila* CIP 57.50 should be reclassified as *Aeromonas salmonicida* CIP 57.50. *Int. J. Syst. Evol. Microbiol.* 60, 715–717.
- [18] Miñana-Galbis, D., Pinzón, D.L., Lorén, J.G., Manresa, À., Oliart-Ros, R.M. (2010) Reclassification of *Geobacillus pallidus* (Scholz et al. 1988) Banat et al. 2004 as *Aeribacillus pallidus* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 60, 1600–1604.
- [19] Miñana-Galbis, D., Urbizu-Serrano, A., Farfán, M., Fusté, M.C., Lorén, J.G. (2009) Phylogenetic analysis and identification of *Aeromonas* species based on sequencing of the *cpn60* universal target. *Int. J. Syst. Evol. Microbiol.* 59, 1976–1983.
- [20] Moore, W.E.C., Cato, E.P., Moore, L.V.H. (1985) Index of the bacterial and yeast nomenclatural changes published in the *International Journal of Systematic Bacteriology* since the 1980 Approved Lists of Bacterial Names (1 January 1980 to 1 January 1985). *Int. J. Syst. Bacteriol.* 35, 382–407.
- [21] Mulet, M., Lalucat, J., García-Valdés, E. (2010) DNA sequence-based analysis of the *Pseudomonas* species. *Environ. Microbiol.* 12, 1513–1530.
- [22] Noterdaeme, L., Bigawa, S., Steigerwalt, A.G., Brenner, D.J., Ollevier, F. (1996) Numerical taxonomy and biochemical identification of fish-associated motile *Aeromonas* spp. *Syst. Appl. Microbiol.* 19, 624–633.
- [23] Pang, H., Kitahara, M., Tan, Z., Wang, Y., Qin, G., Ohkuma, M., Cai, Y. (2012) Reclassification of *Lactobacillus kimchii* and *Lactobacillus bobalius* as later subjective synonyms of *Lactobacillus paralimentarius*. *Int. J. Syst. Evol. Microbiol.* 62, 2383–2387.
- [24] Picard, B., Goulet, P. (1985) Comparative electrophoretic profiles of esterases, and of glutamate, lactate and malate dehydrogenases, from *Aeromonas hydrophila*, *A. caviae* and *A. sobria*. *J. Gen. Microbiol.* 131, 3385–3391.
- [25] Popoff, M. (1984) Genus III *Aeromonas* Kluyver and Van Niel 1936, 398^{AL}, in: Krieg, N.R., Holt, J.G. (Eds.), *Bergey's Manual of Systematic Bacteriology*, vol. 1, 1st ed., The Williams & Wilkins Co., Baltimore, pp. 545–548.
- [26] Popoff, M., Coynault, C., Kiredjian, M., Lemelin, M. (1981) Polynucleotide sequence relatedness among motile *Aeromonas* species. *Curr. Microbiol.* 5, 109–114.
- [27] Popoff, M., Véron, M. (1976) A taxonomic study of the *Aeromonas hydrophila*–*Aeromonas punctata* group. *J. Gen. Microbiol.* 94, 11–22.
- [28] Popoff, M., Véron, M. (1981) Validation of the publication of new names and new combinations previously effectively published outside the IJSB, list No. 6. *Int. J. Syst. Bacteriol.* 31, 215–218.
- [29] Roger, F., Marchandin, H., Jumas-Bilak, E., Kodjo, A., the colBVH study group, Lamy, B. (2012) Multilocus genetics to reconstruct aeromonad evolution. *BMC Microbiol.* 12, 62.
- [30] Schubert, R.H.W. (1964) Zur Taxonomie der anaerogenen Aeromonaden. *Zentralbl. Bakteriol. Parasitenk. Infektionskr. Hyg. Abt. I Orig.* 193, 343–352.
- [31] Schubert, R.H.W. (1974) Genus II. *Aeromonas* Kluyver and Van Niel, 1936, 398. In: Buchanan, R.E., Gibbons, N.E. (Eds.), *Bergey's Manual of Determinative Bacteriology*, 8th ed., The Williams & Wilkins Co., Baltimore, pp. 345–348.
- [32] Skerman, V.B.D., McGowan, V., Sneath, P.H.A. (1980) Approved lists of bacterial names. *Int. J. Syst. Bacteriol.* 30, 225–420.
- [33] Soler, L., Yáñez, M.A., Chacón, M.R., Aguilera-Arreola, M.G., Catalán, V., Figueras, M.J., Martínez-Murcia, A.J. (2004) Phylogenetic analysis of the genus *Aeromonas* based on two housekeeping genes. *Int. J. Syst. Evol. Microbiol.* 54, 1511–1519.
- [34] Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A.D., Kämpfer, P., Maiden, M.C.J., Nesme, X., Rosselló-Móra, R., Swings, J., Trüper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B. (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047.
- [35] Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- [36] Tindall, B.J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W., Kämpfer, P. (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266.
- [37] Valera, L., Esteve, C. (2002) Phenotypic study by numerical taxonomy of strains belonging to the genus *Aeromonas*. *J. Appl. Microbiol.* 93, 77–95.
- [38] Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.H., Glöckner, F.O., Rosselló-Móra, R. (2010) Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. *Syst. Appl. Microbiol.* 33, 291–299.
- [39] Zeigler, D. (2003) Gene sequences useful for predicting relatedness of whole genomes in bacteria. *Int. J. Syst. Evol. Microbiol.* 53, 1893–1900.

Draft Genome Sequence of the *Aeromonas diversa* Type Strain

Maribel Farfán,^{a,b} Nino Spataro,^c Ariadna Sanglas,^a Vicenta Albarral,^a J. Gaspar Lorén,^a Elena Bosch,^c M. Carmen Fusté^{a,b}

Departament de Microbiologia i Parasitologia Sanitàries, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain^a; Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain^c; Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain^b

M.F. and N.S. are co-first authors.

We present here the first genome sequence of the *Aeromonas diversa* type strain (CECT 4254^T). This strain was isolated from the leg wound of a patient in New Orleans (Louisiana) and was originally described as enteric group 501 and distinguished from *A. schubertii* by DNA-DNA hybridization and phenotypical characterization.

Received 22 April 2013 Accepted 26 April 2013 Published 6 June 2013

Citation Farfán M, Spataro N, Sanglas A, Albarral V, Lorén JG, Bosch E, Fusté MC. 2013. Draft genome sequence of the *Aeromonas diversa* type strain. *Genome Announc.* 1(3): e00330-13. doi:10.1128/genomeA.00330-13.

Copyright © 2013 Farfán et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to M. Carmen Fusté, mcfuste@ub.edu, or Elena Bosch, elena.bosch@upf.edu.

The genus *Aeromonas* Stanier 1943, 213^{AL}, belongs to the family *Aeromonadaceae* within the class *Gammaproteobacteria* (1). *Aeromonads* are autochthonous inhabitants of aquatic environments, including chlorinated and polluted waters, although they can also be isolated from a wide variety of environmental and clinical sources. They cause infections in vertebrates and invertebrates, such as frogs, birds, various fish species, and domestic animals. In recent years, some authors have considered *Aeromonas* as an emergent pathogen in humans, producing intestinal and extra-intestinal diseases. *Aeromonads* are facultative anaerobic chemoorganotrophs capable of anaerobic nitrate respiration and dissimilatory metal reduction (1). *Aeromonas* sp. 2478-85 Hickman-Brenner et al. 1988 (CDC 2478-85, ATCC 43946, CECT 4254) is the type strain of *Aeromonas diversa* (2). This strain was isolated from the leg wound of a patient of the Charity Hospital in New Orleans (Louisiana), initially designated enteric group 501 (3) and then in 1991 proposed to be DNA hybridization group 13 (HG13) (4).

The draft genome sequence of the *A. diversa* type strain was obtained with a shotgun strategy using Roche 454 sequencing technology. A total of 158,564 reads with an average length of 442 nucleotides (15× coverage) were *de novo* assembled by a combined strategy (Newbler *de novo* and Velvet *de novo*) using the AMOS package 3.1.0. A total of 104 contigs, 101 of >1 kb in length, were constructed, with an N₅₀ of 64,780 bp; the largest contig assembled measured 227,398 bp and the calculated draft genome size was 4.02 Mb, which is slightly smaller than the other *Aeromonas* genomes reported to date (ranging from 4.3 to 4.97 Mb) (5–13). The G+C mole percentage was 61.7.

The gene prediction and protein annotation were performed by applying the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (<http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html>) on the assembled contigs. A total of 3,709 protein-coding sequences were identified, together with 68 tRNA genes and 6 rRNA genes. Protein annotation using the VFDB database (<http://www.mgc.ac.cn/VFs/>) of virulent factors for bacterial pathogens detected seven putative virulence factors, including a gene involved in ferric uptake (*hemE*), a fosfoheptose isomerase

gene related to lipopolysaccharide (LPS) biosynthesis (*gmhA*), three genes (*cheW*, *fliN* and *cheY*) involved in flagellar motor switch component signal transmission, and two type III secretion system genes (transport H⁺ ATPase gene, *yscR*). Further studies on these virulence-associated proteins will enhance our understanding of *Aeromonas* infections in humans.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [APVG00000000](https://www.ncbi.nlm.nih.gov/nuccore/APVG00000000). The version described in this paper is the first version, [APVG01000000](https://www.ncbi.nlm.nih.gov/nuccore/APVG01000000).

ACKNOWLEDGMENTS

We thank Roger Anglada and Nuria Bonet from the UPF Genomics Core Facility, as well as the National Institute of Bioinformatics, and particularly Angel Carreño and Txema Heredia, for computational support.

This work was partially funded by grant SAF2011-29239 from Ministerio de Ciencia e Innovación, Spanish Government, grant 2009SGR-1101 from Direcció General de Recerca, Generalitat de Catalunya, and grant ARZ00F01 from the University of Barcelona.

REFERENCES

1. Martin-Carnahan A, Joseph SW. 2005. Genus I. *Aeromonas* Stanier 1943, 213^{AL}, p 557–578. In Garrity GM, Brenner DJ, Krieg NR, Staley JT (ed), *Bergey's manual of systematic bacteriology*, vol 2, part B. Springer Verlag, New York, NY.
2. Miñana-Galbis D, Farfán M, Gaspar Lorén J, Carmen Fusté MC. 2010. Proposal to assign *Aeromonas diversa* sp. nov. As a novel species designation for *Aeromonas* group 501. *Syst. Appl. Microbiol.* 33:15–19.
3. Hickman-Brenner FW, Fanning GR, Arduino MJ, Brenner DJ, Farme JJ, III. 1988. *Aeromonas schubertii*, a new manitol-negative species found in human clinical specimens. *J. Clin. Microbiol.* 26:1561–1564.
4. Carnahan AM, Behram S, Joseph SW. 1991. Aerokey II: a flexible key for identifying clinical *Aeromonas* species. *J. Clin. Microbiol.* 29:2843–2849.
5. Seshadri R, Joseph SW, Chopra AK, Sha J, Shaw J, Graf J, Haft D, Wu M, Ren Q, Rosovitz MJ, Madupu R, Tallon L, Kim M, Jin S, Vuong H, Stine OC, Ali A, Horneman AJ, Heidelberg JF. 2006. Genome sequence of *Aeromonas hydrophila* ATCC 7966T: jack of all trades. *J. Bacteriol.* 188:8272–8282.
6. Reith ME, Singh RK, Curtis B, Boyd JM, Bouevitch A, Kimball J, Munnholland J, Murphy C, Sarty D, Williams J, Nash JH, Johnson SC, Brown

- LL. 2008. The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. *BMC Genomics* 9:427.
7. Beatson SA, das Graças de Luna M, Bachmann NL, Alikhan NF, Hanks KR, Sullivan MJ, Wee BA, Freitas-Almeida AC, Dos Santos PA, de Melo JT, Squire DJ, Cunningham AF, Fitzgerald JR, Henderson IR. 2011. Genome sequence of the emerging pathogen *Aeromonas caviae*. *J. Bacteriol.* 193:1286–1287.
 8. Li Y, Liu Y, Zhou Z, Huang H, Ren Y, Zhang Y, Li G, Zhou Z, Wang L. 2011. Complete genome sequence of *Aeromonas veronii* strain B565. *J. Bacteriol.* 193:3389–3390.
 9. Charette SJ, Brochu F, Boyle B, Filion G, Tanaka KH, Derome N. 2012. Draft genome sequence of the virulent strain 01-B526 of the fish pathogen *Aeromonas salmonicida*. *J. Bacteriol.* 194:722–723.
 10. Wu CJ, Wang HC, Chen CS, Shu HY, Kao AW, Chen PL, Ko WC. 2012. Genome sequence of a novel human pathogen, *Aeromonas aquariorum*. *J. Bacteriol.* 194:4114–4115.
 11. Chan XY, Chua KH, Puthucheary SD, Yin WF, Chan KG. 2012. Draft genome sequence of an *Aeromonas* sp. strain 159 clinical isolate that shows quorum-sensing activity. *J. Bacteriol.* 194:6350.
 12. Chai B, Wang H, Chen X. 2012. Draft genome sequence of high-melanin-yielding *Aeromonas media* strain WS. *J. Bacteriol.* 194:6693–6694.
 13. Han JE, Kim JH, Choresca C, Shin SP, Jun JW, Park SC. 2013. Draft genome sequence of a clinical isolate, *Aeromonas hydrophila* SNUFPC-A8, from a moribund cherry salmon (*Oncorhynchus masou masou*). *Genome Announc.* 1(1):e00133-12. doi:10.1128/genomeA.00133-12.

Draft Genome Sequence of *Aeromonas molluscorum* Strain 848T^T, Isolated from Bivalve Molluscs

Nino Spataro,^a Maribel Farfán,^{b,c} Vicenta Albarral,^b Ariadna Sanglas,^b J. Gaspar Lorén,^b M. Carmen Fusté,^{b,c} Elena Bosch^a

Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain^a; Departament de Microbiologia i Parasitologia Sanitàries, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain^b; Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain^c

N. S. and M. F. are co-first authors.

We report here the draft genome sequence of *Aeromonas molluscorum* 848T, the type strain of this *Aeromonas* species, which was isolated from wedge shells (*Donax trunculus*) obtained from a retail market in Barcelona, Spain, in 1997.

Received 6 May 2013 Accepted 13 May 2013 Published 20 June 2013

Citation Spataro N, Farfán M, Albarral V, Sanglas A, Lorén JG, Fusté MC, Bosch E. 2013. Draft genome sequence of *Aeromonas molluscorum* strain 848T, isolated from bivalve molluscs. *Genome Announc.* 1(3):e00382-13. doi:10.1128/genomeA.00382-13.

Copyright © 2013 Spataro et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Elena Bosch, elena.bosch@upf.edu, or M. Carmen Fusté, mcfuste@ub.edu.

The genus *Aeromonas* Stanier 1943, 213^{AL}, comprises a collection of Gram-negative, rod-shaped, non-spore-forming, oxidase- and catalase-positive, glucose-fermenting, facultatively anaerobic bacteria that are resistant to vibriostatic agent O/129 and generally motile by means of polar flagella (1). The genus *Aeromonas* belongs to the family *Aeromonadaceae* within the *Gammaproteobacteria*. *Aeromonads* are autochthonous to aquatic environments worldwide and are usual microbiota (as well as primary or secondary pathogens) of fish, amphibians, and other animals. Some motile species (mainly *Aeromonas caviae*, *Aeromonas hydrophila*, and *Aeromonas veronii* bv. Sobria) are opportunistic pathogens of humans (2). *Aeromonas molluscorum* was defined on the basis of a group of five strains that were isolated from bivalve molluscs obtained from retail markets in Barcelona in 1997 (3), which clustered together as a separate phenon in a phenotypic study (4). The type strain of this new *Aeromonas* species is *A. molluscorum* 848T (CECT 5864^T, LMG 22214^T). Recently, in 2010, a new strain of *A. molluscorum* was isolated from a tributyltin (TBT)-contaminated sediment in Ria de Aveiro, Portugal (5).

The draft genome sequence of the *A. molluscorum* type strain was obtained with a shotgun strategy using Roche 454 sequencing technology. A total of 122,746 reads with an average length of 404 nucleotides (9× coverage) were *de novo* assembled using a combined strategy (Newbler *de novo* and Velvet *de novo*). A total of 309 contigs, 304 of >1 kb in length, were constructed, with an N_{50} of 21,565 bp; the largest contig assembled measured 138,647 bp and the calculated genome size was 4.24 Mb, which is slightly smaller than the other *Aeromonas* genomes reported to date (ranging from 4.3 to 4.97 Mb) (6–14). The G+C mole percentage was 59.2.

The gene prediction and protein annotation were performed by applying the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (<http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html>) on the contigs assembled. A total of 3,946 protein-coding sequences were identified, together with 70 tRNA genes and 3 rRNA genes. Protein annotation using the VFDB database (<http://www.mgc.ac.cn/VF/>) of virulence factors for bac-

terial pathogens detected six putative virulence factors, including a gene involved in ferric uptake (*hemE*), a gene encoding a phosphoheptose isomerase associated with lipopolysaccharide (LPS) biosynthesis (*gmhA*), two genes (*cheW* and *cheY*) involved in signal transmission to the flagellar motor switch component, and a regulator gene responsible for epithelial cell invasion (*csrA*). We detected a 36.4-kb putative prophage showing high identity to *phiO18P*, a bacteriophage isolated from *A. media* (15) and also detected in the genome sequence of *A. caviae* Ae398 (8). In addition to the complete sequence of *phiO18P*, 3 partial bacteriophage sequences were also detected. Two copies of an insertion sequence (IS) with high homology to ISAs4 (88% identity) were discovered using the server ISFinder (<http://www-is.biotoul.fr/>). This IS element was previously reported in atypical strains of *A. salmonicida* subsp. *salmonicida* (16).

Nucleotide sequence accession numbers. This Whole-Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [AQGQ00000000](https://www.ncbi.nlm.nih.gov/nuclink/AQGQ00000000) (BioProject PRJNA183610). The version described in this paper is the first version, [AQGQ01000000](https://www.ncbi.nlm.nih.gov/nuclink/AQGQ01000000).

ACKNOWLEDGMENTS

We thank Roger Anglada and Nuria Bonet from the UPF Genomics Core Facility, as well as the National Institute of Bioinformatics (<http://www.inab.org>), and particularly Angel Carreño and Txema Heredia, for computational support.

This work was partially funded by grant SAF2011-29239 from the Ministerio de Ciencia e Innovación, Spanish Government; grant 2009SGR-1101 from Direcció General de Recerca, Generalitat de Catalunya; and grant ARZ00F01 from the University of Barcelona.

REFERENCES

- Martin-Carnahan A, Joseph SW. 2005. Genus I. *Aeromonas* Stanier 1943, 213^{AL}, p 557–578. In Garrity GM, Brenner DJ, Krieg NR, Staley JT (ed.), *Bergey's manual of systematic bacteriology*, vol 2, part B. Springer Verlag, New York, NY.
- Janda JM, Abbott SL. 2010. The genus *Aeromonas*: taxonomy, pathogenicity, and infection. *Clin. Microbiol. Rev.* 23:35–73.

3. Miñana-Galbis D, Farfán M, Fusté MC, Lorén JG. 2004. *Aeromonas molluscorum* sp. nov., isolated from bivalve molluscs. *Int. J. Syst. Evol. Microbiol.* 54:2073–2078.
4. Miñana-Galbis D, Farfán M, Lorén JG, Fusté MC. 2002. Biochemical identification and numerical taxonomy of *Aeromonas* spp. isolated from environmental clinical samples in Spain. *J. Appl. Microbiol.* 93:420–430.
5. Cruz A, Henriques I, Correia A, Suzuki S, Mendo S. 2010. *Aeromonas molluscorum* Av27: a potential natural tool for TBT decontamination, p 37–46. In Hamamura N, Suzuki S, Mendo S, Barroso CM, Iwata H, Tanabe S (ed), *Interdisciplinary studies on environmental chemistry—biological responses to contaminants*. Terra Publishing, Tokyo, Japan.
6. Seshadri R, Joseph SW, Chopra AK, Sha J, Shaw J, Graf J, Haft D, Wu M, Ren Q, Rosovitz MJ, Madupu R, Tallon L, Kim M, Jin S, Vuong H, Stine OC, Ali A, Horneman AJ, Heidelberg JF. 2006. Genome sequence of *Aeromonas hydrophila* ATCC 7966^T: jack of all trades. *J. Bacteriol.* 188: 8272–8282.
7. Reith ME, Singh RK, Curtis B, Boyd JM, Bouevitch A, Kimball J, Munholland J, Murphy C, Sarty D, Williams J, Nash JH, Johnson SC, Brown LL. 2008. The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. *BMC Genomics* 9:427.
8. Beatson SA, das Graças de Luna M, Bachmann NL, Alikhan NF, Hanks KR, Sullivan MJ, Wee BA, Freitas-Almeida AC, Dos Santos PA, de Melo JT, Squire DJ, Cunningham AF, Fitzgerald JR, Henderson IR. 2011. Genome sequence of the emerging pathogen *Aeromonas caviae*. *J. Bacteriol.* 193:1286–1287.
9. Li Y, Liu Y, Zhou Z, Huang H, Ren Y, Zhang Y, Li G, Zhou Z, Wang L. 2011. Complete genome sequence of *Aeromonas veronii* strain B565. *J. Bacteriol.* 193:3389–3390.
10. Charette SJ, Brochu F, Boyle B, Fillion G, Tanaka KH, Derome N. 2012. Draft genome sequence of the virulent strain 01-B526 of the fish pathogen *Aeromonas salmonicida*. *J. Bacteriol.* 194:722–723.
11. Wu CJ, Wang HC, Chen CS, Shu HY, Kao AW, Chen PL, Ko WC. 2012. Genome sequence of a novel human pathogen, *Aeromonas aquariorum*. *J. Bacteriol.* 194:4114–4115.
12. Chan XY, Chua KH, Puthucheary SD, Yin WF, Chan KG. 2012. Draft genome sequence of an *Aeromonas* sp. strain 159 clinical isolate that shows quorum-sensing activity. *J. Bacteriol.* 194:6350.
13. Chai B, Wang H, Chen X. 2012. Draft genome sequence of high-melanin-yielding *Aeromonas media* strain WS. *J. Bacteriol.* 194: 6693–6694.
14. Han JE, Kim JH, Choresca C, Shin SP, Jun JW, Park SC. 2013. Draft genome sequence of a clinical isolate, *Aeromonas hydrophila* SNUFPC-A8, from a moribund cherry salmon (*Oncorhynchus masou masou*). *Genome Announc.* 1(1):e00133-12. doi:10.1128/genomeA.00133-12.
15. Beilstein F, Dreiseikelmann B. 2008. Temperate bacteriophage *phi*O18P from an *Aeromonas media* isolate: characterization and complete genome sequence. *Virology* 373:25–29.
16. Nilsson WB, Gudkovs N, Strom MS. 2006. Atypical strains of *Aeromonas salmonicida* contain multiple copies of insertion element ISAsa4, useful as a genetic marker and a target for PCR assay. *Dis. Aquat. Organ.* 70: 209–217.