RESEARCH ARTICLE

Direct evidence of recombination in the *recA* gene of *Aeromonas bestiarum*

Ariadna Sanglas[1], Vicenta Albarral[1], Maribel Farfán[1,2*], J. Gaspar Lorén[1] and M. Carmen Fusté[1,2]

[1]Departament de Microbiologia i Parasitologia Sanitàries, Facultat de Farmàcia, Universitat de Barcelona. Av. Joan XXIII s/n, 08028 Barcelona, Spain.

[2]Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona. Av. Diagonal 643, 08028 Barcelona, Spain.

*Corresponding author:

Ph.D. Maribel Farfán

Mailing address: Departament de Microbiologia i Parasitologia Sanitàries, Facultat de Farmàcia, Universitat de Barcelona, Av. Joan XXIII s/n, 08028 Barcelona, Spain

Phone: +34 93 402 44 97

Fax: +34 93 402 44 98

E-mail: mfarfan@ub.edu

E-mail address:

Ariadna Sanglas, arisanglas@gmail.com

Vicenta Albarral, valbarral@ub.edu

J. Gaspar Lorén, jgloren@ub.edu

M. Carmen Fusté, mcfuste@ub.edu

Running title:

Recombination in *Aeromonas bestiarum*

ABSTRACT

Two hundred and twenty-one strains representative of all *Aeromonas* species were characterized using the *recA* gene sequence, assessing its potential as a molecular marker for the genus *Aeromonas*. The inter-species distance values obtained demonstrated that *recA* has a high discriminatory power. Phylogenetic analysis, based on full-length gene nucleotide sequences, revealed a robust topology with clearly separated clusters for each species. The maximum likelihood tree showed the *Aeromonas bestiarum* strains in a well-defined cluster, containing a subset of four strains of different geographical origins in a deep internal branch. Data analysis provided strong evidence of recombination at the end of the *recA* sequences in these four strains. Intergenomic recombination corresponding to partial regions of the two adjacent genes *recA* and *recX* (248 bp) was identified between *A. bestiarum* (major parent) and *A. eucrenophila* (minor parent). The low number of recombinant strains detected (1,8 %) suggests that horizontal flow between *recA* sequences is relatively uncommon in this genus. Moreover, only a few nucleotide differences were detected among these fragments, indicating that recombination has occurred recently. Finally, we also determined if the recombinant fragment could have influenced the structure and basic functions of the RecA protein, comparing models reconstructed from the translated amino acid sequences of our *A. bestiarum* strains with known *E.coli* RecA structures.

KEYWORDS

Aeromonas; Aeromonas bestiarum; *recA;* molecular marker; recombination; protein prediction

INTRODUCTION

The main goal of bacterial phylogenetic studies is to reconstruct the correct genealogical relationships among the strains analysed, estimate when their divergence occurred, and determine the sequence of events along the evolutionary lineages. Nevertheless, not all the genes commonly used for this purpose are always suitable, and their properties need to be considered before their application in a phylogenetic study [11]. The gene should be conserved, encode essential cell functions, have only a single copy, and be present in all species of the genus [52].

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 recombination, in contrast with eukaryotes, bacteria have acquired three basic mechanisms by which they can incorporate genes from other bacterial species. Nevertheless, their genomes are not simply arbitrary assortments of genes of mixed heritage. Recombination in bacteria is always restricted to small DNA fragments, is unidirectional and independent of reproduction, and occurs with a relatively low frequency, although genes codifying virulence factors or antibiotic resistance experiment more frequent recombination changes [10, 12].

The impact of recombination on bacterial phylogenies has been the subject of considerable discussion [10, 13, 17, 36, 41, 56]. Recently, with the availability of sequencing techniques and the analytical power of new programs, 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. 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 [12, 27].

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 new methods have been developed to test the presence of recombination, as well as to identify the parental and recombinant individuals or the recombination break-points. Those methods can be classified in different categories: similarity, distance, phylogenetic, compatibility, and nucleotide substitution distribution [23, 40]. Their performance 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 [41].

The genus *Aeromonas* Stanier 1943 belongs to the family *Aeromonadaceae* within the class *Gammaproteobacteria* [29]. 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 are usual microbiota (as well as primary or secondary pathogens) of fish, amphibians and other animals. Some species, mainly *A. caviae*, *A. hydrophila* and *A. veronii* bv. Sobria, are opportunistic pathogens of humans [19]. Several attempts have been made to generate phylogenies using DNA gene sequences to reconstruct the correct genealogical ties among species in *Aeromonas* [9, 11, 30, 46]. However, the genes chosen for this purpose are not always suitable, and do not necessarily give congruent phylogenies [21, 51].

In our study we investigated the discriminatory power of the *recA* gene sequences at inter- and intra-specific levels for application in *Aeromonas* phylogenetic studies. Any cluster showing incongruences was analyzed looking for the presence of potential recombinant fragments in their *recA* gene sequences.

MATERIAL AND METHODS

Data set

221 *Aeromonas* strains were analyzed based on the nucleotide sequences of the *recA* gene: 125 belonged to the "*Aeromonas hydrophila* Species Complex" (study 1), and 150 (54 strains from study 1) represented all the species and subspecies of this genus described to date, including the type strains as well as some strains considered as synonymous (study 2) (Table S1). The strains were obtained from several culture collections

(CECT, Colección Española de Cultivos Tipo, Universitat de València, Valencia, Spain; CIP, Collection de l'Institut Pasteur, Paris, France; JCM, Japan Collection of Microorganisms, RIKEN BioResource Center, Ibaraki, Japan; LMG, Culture Collection of the Laboratorium voor Microbiologie Gent, Universiteit Gent, Ghent, Belgium), kindly supplied by different authors or research groups (Katri Berg, University of Helsinki, Helsinki, Finlandia; Yogesh Shouche, Molecular Biology Laboratory, National Centre for Cell Science, Pune, India; Margarita Gomila, Universitat de les Illes Balears, Palma de Mallorca, Spain; Mª José Figueras, Universitat Rovira i Virgili, Reus, Spain; Antonio Martínez-Murcia, Universidad de Alicante, Spain), or obtained by our group from freshwater and food samplings [35]. Strains were grown aerobically on tryptone soy agar (TSA; Pronadisa, Laboratorios Conda) supplemented with 1% (w/v) NaCl for 24-48h at their optimum temperature, which ranged between 25 and 30°C. For long-term storage, pure cultures were stored frozen at -40 and - 80 ºC in triptone soy broth (TSB; Oxoid, Thermo Fisher Scientific Inc.) containing 20% glycerol and on Protect™ cryobeads (Technical Service Consultants Ltd.). Species affiliation, source, geographical origin and the type of phylogenetic study of all the strains analysed are listed in Table S1.

DNA extraction, primers, and PCR conditions

Genomic DNA was extracted and purified with the REALPURE® Genomic DNA extraction kit (Durviz) and stored at -20°C until use. Primer3 software was used to design PCR and sequencing primers (http://primer3.sourceforge.net/, [55]). Oligonucleotide primers were designed from published genome sequences of *Aeromonas hydrophila* ATCC 7966^T (GenBank accession number: CP000462, [50]) and *Aeromonas salmonicida* subsp. *salmonicida* A449 (GenBank accession number: CP000644, [43]). All primers used in this study are shown in Table S2. Six different primer sets were used to amplify by PCR the complete *recA* gene and its flanking regions (Table S2). Additional internal primers were designed for the sequencing of *recA*. The oligonucleotides were synthesized by Invitrogen™ (Thermo Fisher Scientific). The conditions of amplification by PCR were optimized in a 50 µL final volume reaction, containing 0.5 – 10 µL of genomic DNA as the template, 0.2 – 2 µM each primer, 0.2 mM each dNTP, 0.5 mM $MgCl_2$, 5% dimethyl sulfoxide (DMSO) (optional), 1X Buffer I (10X Buffer I: 100 mM Tris-HCl, pH 8.3, 500 mM KCl, 15 mM $MgCl_2$, 0.01% (w/v) gelatin) and 1.25 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems). Amplifications were performed in Veriti® (Applied Biosystems) and Applied Biosystems® 2720 thermal cyclers using the following program: initial denaturation at 95°C for 5 min, followed by 35 cycles of denaturation at 94°C for 45 s, annealing at 53–58°C for 1 min and elongation at 72°C for 2 min, and a final extension at 72°C for 10 min. The PCR products

were resolved by electrophoresis in 1% agarose gels stained with GelRed™ (Biotium) and visualized by UV transillumination. Amplicons were purified with a MSB® Spin PCRapace kit (STRATEC Molecular). Purified PCR products were directly sequenced on both strands using either the PCR or internal primers. Sequencing reactions were performed with the BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and analyzed on an ABI PRISM® 3700 DNA sequencer (Applied Biosystems) by the Genomics Unit of the Scientific and Technological Centers of the University of Barcelona (CCiTUB).

Phylogenetic and sequence analysis

Multiple sequence alignments were performed using the ClustalW program [22] implemented in MEGA6 software [54]. We also applied the graphical dots plot method for the exploratory sequence analysis. This method was carried out with functions included in *phyclust* [7] and *ape* [38] packages using the R programming language [42]. Maximum likelihood (ML) phylogenetic analyses were performed using the PhyML 3.1 program [16] with 1000 bootstrap replicates to assess tree topology robustness. Phylogenetic trees were reconstructed based on the best fit model of evolution for each dataset estimated in the MEGA6. Phylogenetic trees were visualized using the NJPlot program [39].

Recombination detection methods

To detect potential recombination events, we analysed the incongruences in the tree topology, examined the sequence alignments and the dots plots, determined the G+C content (mol %) with the DnaSP program (version 5, [24]), and calculated the codon usage bias with MEGA6 software. In addition, we evaluated the possible recombination events, and identified potential major and minor parents and the location of possible recombination breakpoints, using seven methods implemented in the RDP4 program [28]: RDP [26], GENECONV [37], BOOTSCAN [47], MAXCHI [34], CHIMAERA [40], SiScan [15] and 3SEQ [5] with their default parameters. Sequences statistically supported by at least two detection methods (P-value < 0.05) were considered as potential recombinants.

A phylogenetic network was constructed with the SplitsTree4 program (version 4.13, [18]), using the median joining method [2] from a multiple sequence alignment. The existence of recombination was also evaluated using the *Phi Test* (Pairwise homoplasy index, [6]), implemented in SplitsTree4, which is significant at a 95% confidence interval (P-value < 0.05).

Protein homology modelling

Protein prediction of the query sequences was carried out by homology modelling [48], using the resolved structure of a known related protein. This method identifies one or more known protein structure related to the target sequence (templates), aligns the template with the translated target sequence, builds a three-dimensional model based on the alignment, and finally, evaluates the model obtained. To find the best template structure, we searched for bacterial RecA proteins available in the RCSB Protein Data Bank (PDB; http://www.rcsb.org/pdb/). Homology modelling was performed using the structure prediction servers Phyre^2 (Protein Homology/analogY Recognition Engine version 2.0, [20]) and SWISS-MODEL [1]. Predicted models were visualized with the PyMol molecular graphics system (version 0.99rc6, Schrödinger, LLC.). Protein disorder prediction was carried out with the DISOPRED server [57], using a method to predict dynamically disordered regions from the amino acid sequence.

RESULTS

Phylogenetic studies

We sequenced the complete *recA* gene in 221 *Aeromonas* strains, including all the species formally accepted to date (Table S1), in order to assess its potential as a molecular marker for the genus *Aeromonas*, particularly to solve incongruences and obtain a better discrimination between closely related species. In the first of two phylogenetic studies based on the full-length nucleotide sequences of the *recA* gene, we analyzed 125 strains belonging to the "*Aeromonas hydrophila* species complex" (AHC) (*A.hydrophila*, *A.dhakensis*, *A.bestiarum*, *A.piscicola*, *A.popoffii, A.salmonicida*) (study 1), a taxonomically complex group that includes closely genetically related species with a difficult phenotypical discrimination. Secondly, we performed a phylogenetic analysis (study 2) with 150 strains representative of all the species of this genus. Sequence length varies between 1,059 (*A. diversa*, *A. popoffii*, *A. schubertii*, *A. simiae*, *A. sobria*, and *A. taiwanensis*) and 1,068 (*A. cavernicola*) bp depending on the species.

The *recA* gene-based ML phylogenetic tree obtained in study 1 (Fig. 1A) shows a robust topology with clearly separated clusters for each species, which is very similar to that previously obtained from the concatenated

sequences of six genes [14]. The ML tree corresponding to study 2 (Fig. 1B) was similarly robust and all the strains were clearly grouped according to the species.

Intra- and inter-specific distance values estimated from sequence data obtained from studies 1 and 2 are shown in Tables S3 and S4. The mean intra- and inter-pairwise distance values for species included in study 1 were 0.0159 ± 0.0061 and 0.0749 ± 0.0134, respectively, and in study 2, 0.0165 ± 0.0082 and 0.1065 ± 0.0190, respectively. Additionally, distance density plots are shown in Fig. S1. In all cases, the inter-specific values obtained allowed a clear separation of species. In species with only one strain (*A. australiensis, A. cavernicola, A. fluvialis, A. sanarellii, A. taiwanensis*), it was obviously not possible to calculate the distance variation values at intra-specific level.

The ML phylogenies showed the strains of the different species studied grouped in coherent clusters. The *A. bestiarum* group appeared clearly separated from other species, with a subset of four strains in a deep internal branch. Notably, these 4 strains were isolated from distinct geographical origins (CECT 5741: environment, Germany; CECT 5742: water, Switzerland; LMG 13667: water, USA; AE147: lake water, Finland) and obtained from different culture collections or authors. We analyzed their *recA* sequences to investigate why these strains appeared separated from the rest.

Recombination in the *recA* gene

We conducted a preliminary exploratory analysis of the *recA* sequences, using graphic methods such as multiple sequence alignments and dots plots. These graphical representations provide a quick and intuitive data analysis, and make the overall trends and local variations easier to visualize. Figure 2 presents the dots plots generated from all the sequences corresponding to the phylogenetic studies 1 (A) and 2 (B). The graph shows the polymorphic (segregating) sites detected along the sequences, showing in a different colour those bases that vary from the consensus sequence (top of the graph). Both plots allowed us to graphically see separated groups of species that were in clear concordance with those obtained in the phylogenetic trees (Fig. 1). Moreover, both graphs detected a particular region with a nearly identical sequence at the end of the *recA* gene in the four strains belonging to the *A. bestiarum* cluster, which was clearly different from the other strains of the species.

In order to analyze the possible exogenous origin of the sequence fragment detected, we applied seven recombination detection methods implemented in the RDP4 program to the data set (Table 1). The analysis provided statistical support for a unique recombination event in the 4 *A. bestiarum* strains, identifying the beginning breakpoint at position 979. The presumptive recombinant region therefore comprises a fragment of 81 bp (alignment positions: 979-1065, excluding the stop codon). The changes detected in this fragment are identical in the 4 strains, except at position 1,014 for the strain LMG 13667. In addition, all methods identified *A. bestiarum* as the major parent (the contributor of the non-recombinant region) with high probability values (> 97%). No minor parent (source of the recombinant region) was identified, although two methods suggested *A. eucrenophila*.

To confirm if the region detected at the end of the *recA* is a fragment acquired by recombination, we determined the G+C content of the total gene and that corresponding to the non-recombinant (positions 1-978), and recombinant (positions 979-1,062) regions (Table 2). The G+C content obtained for the terminal region in the recombinant strains was clearly higher (64%) than in the remaining *A. bestiarum* strains (61.7%), providing evidence of an exogenous origin. The results of a codon usage bias study showed no significant differences among the regions analysed (Table S5).

All these results were confirmed when we generated ML trees from the non-recombinant (Fig. 3A) and recombinant (Fig. 3B) regions of these sequences. The tree constructed from the recombinant fragment sequences (Fig. 3B) showed the four strains grouped in a cluster separated from *A. bestiarum* but closely related to *A. eucrenophila* and *A. tecta*.

In order to determine the recombinant fragment endpoint, we enlarged the sequenced fragment to the adjacent gene, *recX,* in a set of 18 strains, which included the four recombinants and a group of strains representative of *A. bestiarum*, as well as *A. eucrenophila* and *A. tecta* (Table 2), the probable sources of the recombinant fragment. For each strain, we obtained the nucleotide sequence with the full-length *recA* gene, the intergenic region and the partial *recX* gene, which was called *recA-recX*. From a multiple sequence alignment of 18 *recA-recX* sequences (positions 1-1,522), we analyzed the recombination with the RDP4 program. Five methods detected the recombinant fragment (with statistical support) in the same four *A. bestiarum* strains (Table 1; Fig. S2). In this case, we were able to determine the minor parent, an *A.*

eucrenophila strain, with more than 94% similarity. The end point of the recombinant fragment was located at position 1,235. Therefore, this region begins at position 979 of the *recA* gene and ends at position 129 of the *recX* gene, with a total length of 248 bp (Fig. S3), all fragments being nearly identical, differing only in 1 to 8 bp (Fig. S4).

These results were confirmed by the dots plot (Fig. 2B), since the pattern in the recombinant region of the four strains was highly homologous with the *A.eucrenophila* species, and very different from the *A. bestiarum* strains. However, the pattern in the non-recombinant regions of both genes was similar to *A. bestiarum*.

The G+C content (Table 2) and the codon usage determination (Table S5) was also in agreement with these results. When considering the complete *recA* gene and the non-recombinant regions of the *recA-recX* sequences, the values obtained were very similar to *A. bestiarum*, but in the recombinant region, these values were close to those of *A. eucrenophila.*

Finally, a split decomposition analysis of the *recA-recX* sequences gave a reticulated structure (Fig. 4), providing evidence for recombination. This network identified the parent strains (*A. bestiarum* and *A. eucrenophila*) of the recombinant sequences (in red). The *Phi Test* provided additional statistically significant recombination evidence ($P = 1.7 \times 10^{-16}$).

Predicted RecA protein structure

In order to determine if the recombinant region detected could affect the structure or function of the RecA protein, we constructed three-dimensional structures of RecA of the *A. bestiarum* (CECT4227^T) type strain and a recombinant isolate (CECT5741) by homology modelling. Based on these query sequences, we searched for homologous proteins of bacterial RecA in the PDB protein database, obtaining a total of 60 structures that corresponded to the following species: *Mycobacterium smegmatis* (30), *Mycobacterium tuberculosis* (14), *Escherichia coli* (14), *Thermotoga maritima* (1), *Deinococcus radiodurans* (1), and a partial structure of the C-terminal region of *E. coli* K12. All structures were resolved experimentally using different techniques. The search results for template structures using the program SWISS-MODEL identified *E. coli* structures as the best for modelling the *A. bestiarum* RecA protein, with the highest percentages of identity (77.5 - 80.3%), similarity (52 - 53%) and coverage (92-96%). We chose three *E.coli* structures obtained by X-

298 ray diffraction to be used as templates: the first was crystallized without DNA (PDB ID: 2REB, [53]), the
299 second with a single-strand DNA (PDB ID: 3CMW, [8]) and the third with a double strand (PDB ID: 3CMT, [8]).
300 The *E. coli* template without DNA (2REB) was a free monomer, while the templates with DNA were
301 polymerized into filaments, taking an inactive (3CMW) or active state (3CMT).
302
303 Figure S5 compares the deduced amino acid sequences and the predicted protein secondary structure
304 corresponding to the type strain of *A. bestiarum* and the recombinant CECT5741 using *E. coli* 2REB as a
305 template. The amino acid sequences of both *A. bestiarum* strains showed high homology over the entire
306 translated region except at the end, which corresponds to the recombinant region with a high sequence
307 variation. The figure also highlights the differences between both strains in this region. The secondary
308 structure of the RecA protein was identical for both strains, with ten α-helices and twelve β-sheets, but some
309 differences were observed compared with *E. coli*. For example, β0 and β5 sheets were absent in *A.*
310 *bestiarum*, but present in *E. coli*, and two β-sheets (β5 and β6) in the *A. bestiarum* protein before the F helix
311 were not observed in *E. coli* (Fig. S5). However, these differences correspond to predicted regions with a low
312 degree of confidence. Similarly, some small discrepancies were observed when using the 3CMW and 3CMT
313 models as templates (data not shown).
314
315 Three *E. coli* structures (2REB, 3CMW and 3CMT) were used as a template for the construction of three-
316 dimensional models, with identical results for both *A. bestiarum* strains. Figure S6 shows the three-
317 dimensional models obtained (type strain in green and recombinant in violet) using the *E. coli* 2REB as a
318 template, with high similarity for both protein structures. When we overlapped the *E.coli* and *A. bestiarum*
319 structures (Fig. 5), four remarkable differences were detected. First, *E.coli* presented a β0 sheet at the N-
320 terminal region, which was absent in *A. bestiarum*. Second, only *A. bestiarum* presented two β-sheets in
321 Loop-1, which corresponds to one of the DNA binding sites. Third, the β5 sheet of *E.coli* was not present in *A.*
322 *bestiarum*. Lastly, the recombinant region located in the C-terminal end could not be visualized in the
323 predicted structure because it corresponds to a non-crystallized region in the *E.coli* RecA protein.
324
325 It was impossible to predict the tertiary structure of the C-terminal region, which contains the recombinant
326 fragment, because in the *E. coli* structures used as a template this region was either disorderly (2REB) or not

crystallized (3CMW and 3CMT), and therefore without any structure. Currently, there are no available RecA *E. coli* structures with this C-terminal region resolved.

We analyzed the predicted RecA protein disordered profile plot of the recombinant strain (Fig. S7), which showed a higher probability of disorder in the C-terminal region (the recombinant fragment). A detailed analysis of the last 25 amino acids of this region corresponding to the type strain of *A. bestiarum*, the recombinant isolate and the different *E.coli* templates showed that despite the variation in sequences, they have in common a high number of negatively charged amino acids (Fig. S5).

DISCUSSION

In this study, we assessed the suitability of the *recA* gene as a molecular marker for *Aeromonas.* Partial sequences (348-600 bp) of this gene have been previously used for the identification of *Aeromonas* strains. Sepe et al. [49] concluded that *recA* was useful for the construction of reliable phylogenies, but they were unable to develop a simple PCR procedure because the primers used were not sufficiently specific and resulted in multiple bands. These partial sequences of *recA* have also been used in different MLST or MLSA studies [3, 30, 32, 33]. In a phylogenetic study, it is important to analyse the full sequence of the chosen genes because not all the regions in a protein-coding gene are subject to the same selective pressure, some accumulating changes more freely, while others are more conserved. If we analyze only a fragment, it may not be representative of the full-length gene sequence variation. To avoid possible under-estimations of the total changes in the sequence, in our work we sequenced the entire gene. In addition, our analysis included a high number of strains, 125 (study 1) and 150 (study 2), representative of all the species and subspecies of the genus *Aeromonas,* which allowed us to determine the intra-specific distance values. The inter-specific distance values obtained demonstrated that *recA* has a high discriminatory power among the *Aeromonas* species.

Phylogenetic analysis of the genus *Aeromonas* based on *recA* gene sequences confirmed the monophyletic origin of this bacterial group [25]. As in previous work using other molecular markers [14, 30, 32], we obtained a perfect clustering of the strains belonging to the same species, including those considered synonymous.

Moreover, the robustness of the ML trees was statistically supported by high bootstrap values in the majority of the group nodes.

In the cluster corresponding to *A. bestiarum* we detected a group of four strains with a recombinant fragment at the end of the sequence. Enlarging the sequenced fragment to the adjacent gene (*recX*), we were able to determine the end and the total length of the recombinant region, as well as the major and minor parent. The fact that the recombinant sequences in the 4 strains were quite similar (only a maximum of 8 nucleotide differences) indicates that the recombination event is a recent phenomenon.

In the literature, some recombination events in *Aeromonas* linked to the *recA* and other conserved or accessory genes are described. Silver et al. [51] were the first to report an episode of recombination in this genus. In their study, they considered only clinical isolates belonging to the *A.veronii* species complex, in which they found phylogenetic inconsistencies in the trees obtained from different individual genes (*gyrB*, *recA*, *dnaJ*, *chiA*, *ascFG*, *ascV*, *aexT* and *aexU*) compared with the tree constructed from the concatenated sequences. These incongruences were interpreted as a consequence of a possible recombination event later confirmed by two different methods, PhiTest and GARD, which detected recombination in some genes (*ascFG*, *aexU*, *gyrB* and *dnaJ*). Martino et al. [32] also determined the presence of recombination in an MLST study analyzing 6 genes (*gyrB*, *groL*, *gltA, metG*, *ppsA* and *recA*) from a collection of 77 *Aeromonas* strains isolated from fish and shellfish, including all the species type strains. In this case, recombination was determined by generating a split tree network and calculating the *Phi Test* using the SplitTree program, and applying 5 recombination detection methods with the RDP3 program. In two recent studies [12, 45], recombination was suspected from the inconsistencies detected in the phylogenies, and was later confirmed by several programs. In contrast with these studies, we also characterized the recombinant fragment and determined its origin. In all these cases, recombination affects a low number of strains and the acquired region is generally small, so the cohesion of the main groups in the phylogenies is not seriously affected and most bacterial species remain delineable as discrete evolutionary lineages, as in our work.

We tested if the acquisition of the recombinant fragment can affect the structure and function of the RecA protein by constructing three-dimensional protein models for the *A. bestiarum* (CECT 4227^T) type strain and a recombinant strain (CECT 5741) from their *recA* sequences. Despite possible errors in comparative models

(in the packaging or conformation), for example, incorrect alignment of the sequence modelling with known related structures, this methodology has proven very useful [48]. Among the few RecA protein structures of prokaryotic origin available in the databases, those of *E. coli* were selected as a template, based on the identity and similarity percentages obtained after comparison with our translated sequences.

RecA protein has an important biological role, being responsible for homologous recombination processes and the DNA repair system, which activates the emergency or SOS response in prokaryotes. Three structural domains can be distinguished in the protein: N-terminal, central and C-terminal. The amino acid sequence alignment showed differences between the two *A. bestiarum* strains and *E. coli*, especially in the C-terminal fragment, which appears to be a region that accumulates more synonymous and non-synonymous changes. No significant differences were observed between the two *A. bestiarum* sequences in the secondary and tertiary predictions obtained using different crystallized structures, with or without DNA, as a template. Some discrepancies were detected between the three-dimensional structures of *E. coli* and *A. bestiarum*, but they did not affect the major RecA protein functional domains, the ATP (Walker A, Walker B and MAW) and DNA binding sites (Loop-1 and Loop-2), except in the case of the Loop-1, in which *A. bestiarum* showed two β-sheets absent in *E. coli*. Nevertheless, this result should be considered with caution because it could be an artefact of the modelling method.

In summary, we evaluated the usefulness of *recA* as a molecular marker for the genus *Aeromonas*, and the importance of the presence of recombination events and their influence on phylogenies. We characterized the recombinant fragment detected in four *A.bestiarum* strains and determined its origin. Our study reveals that although recombination is present in some *Aeromonas* strains (17,4 % of the *A.bestiarum* species), it is infrequent, and its impact on the phylogenies is low (1,8% of the total strains analyzed). This challenges the postulation that HGT is so common in bacterial populations that it precludes the existence of biological species [44].

ACKNOWLEDGEMENTS

This work was supported by projects from the Ministerio de Educación y Ciencia, Spain (CGL2008-03281/BOS) and the Universitat de Barcelona (ARZ00F01).

APPENDIX A. SUPPLEMENTARY DATA

The following are Supplementary data to this article:

Table S1

Aeromonas strains studied in this work and *recA* sequences GenBank accession numbers.

Table S2

recA primers and PCR settings used in this work.

Table S3

Intra- and inter-specific genetic distances obtained from the *recA* sequences of the *Aeromonas hydrophila* Complex (AHC) species (study 1).

Table S4

Intra- and inter-specific genetic distances determined from the *recA* sequences of all the species of the genus *Aeromonas* (study 2).

Table S5

Comparison of codon usage between recombinant strains and closely related species in non-recombinant and recombinant regions of *recA* and *recX* genes.

Figure S1

Within species (black) and between species (grey) *p-distance* distribution obtained from the *recA* gene sequence data (A: study 1 and B: study 2).

Figure S2

Recombination analysis with the RDP4 program.

Figure S3

Gene cluster and recombinant fragment detected for the four *A. bestiarum* recombinants.

Figure S4

Comparison of nucleotide variation in the complete recombinant fragment *recA-recX* corresponding to the four *A. bestiarum* recombinant strains.

Figure S5

Comparative amino acid sequence analysis and predicted RecA secondary structure of the type strain of *A. bestiarum* and the recombinant CECT5741 (both in green) with *E. coli* (PDB ID: 2REB) as a template (in blue) using the program Phyre[2].

Figure S6

Predicted RecA structures of the *A. bestiarum* type strain (A, green) and the recombinant CECT5741 (B, violet).

Figure S7

Predicted RecA protein disordered profile plot of the recombinant strain CECT5741, generated using the DISOPRED2 server.

File S1

recA gene multiple sequence alignment by ClustalW (MEGA6) of 125 strains belonging to the *Aeromonas hydrophila* species complex (study 1) in FASTA format.

File S2

recA gene multiple sequence alignment by ClustalW (MEGA6) of 150 *Aeromonas* strains, including all the species in the genus (study 2), in FASTA format.

File S3

recA+recX multiple sequence alignment by ClustalW (MEGA6) of 18 *Aeromonas* strains in FASTA format.

REFERENCES

[1] 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.

[2] Bandelt, H.J., Forster, P., Röhl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37-48.

[3] Beaz-Hidalgo, R., Martínez-Murcia, A., Figueras, M.J. (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*. Syst. Appl. Microbiol. 36, 171-176.

[4] Beaz-Hidalgo, R., Martínez-Murcia, A., Figueras, M.J. (2015) *Aeromonas dhakensis* Beaz-Hidalgo et al. 2015. *In* List of new names and new combinations previously effectively, but not validly, published. Validation list no. 161. Int. J. Syst. Evol. Microbiol. 65, 1-4.

[5] Boni, M.F., Posada, D., Feldman, M.W. (2007) An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics. 176, 1035-1047.

[6] Bruen, T.C., Philippe, H., Bryant, D. (2006) A simple and robust statistical test for detecting the presence of recombination. Genetics. 172, 2665-2681.

[7] Chen, W.C. (2010) Phylogenetic clustering with R package phyclust. Phyloclustering-Phylogenetic clustering website: http://thirteen-01.stat.iastate.edu/snoweye/phyclust/

[8] Chen, Z., Yang, H., Pavletich, N.P. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature. 453, 489-494.

[9] Colston, S.M., Fullmer, M.S., Beka, L., Lamy, B., Gogarten, J.P., Graf, J. (2014) Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. mBio. 5, e02136-14.

[10] Didelot, X., Maiden, M.C. (2010) Impact of recombination on bacterial evolution. Trends Microbiol. 18, 315-322.

[11] 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.

[12] Farfán, M., Albarral, V., Sanglas, A., Lorén, J.G., Fusté, M.C. (2013) The effect of recombination in *Aeromonas*. In: Muñoz-Torrero, D., Cortés, A., Mariño, E.L. (Eds.), Recent Advances in Pharmaceutical Sciences III. Transworld Research Network, India, pp. 179-193.

[13] 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) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc. Natl. Acad. Sci. USA. 98, 182-187.

[14] 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, Croatia, pp. 39-54.

[15] Gibbs, M.J., Armstrong, J.S., Gibbs, A.J. (2000) Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics. 16, 573-582.

[16] Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307-321.

[17] Hanage, W.P., Fraser, C., Spratt, B.G. (2005) Fuzzy species among recombinogenic bacteria. BMC Biol. 3, 6.

[18] Huson, D.H., Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254-267.

[19] Janda, J.M, Abbott, S.L. (2010) The genus *Aeromonas*: taxonomy, pathogenicity, and infection. Clin. Microbiol. Rev. 23, 35-73.

[20] Kelley, L.A., Sternberg, M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4, 363-371.

[21] 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.

[22] Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics. 23, 2947-2948.

[23] Lemey, P., Posada, D. (2009) Introduction to recombination detection. In: Vandamme, A.M., Salemi, M., Lemey, P. (Eds.), The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing, 2nd Ed. Cambridge University Press, New York, pp. 493.

[24] Librado, P., Rozas, J. (2009) DnaSp v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25, 1451-1452.

[25] Lorén, J.G., Farfán, M., Fusté, M.C. (2014) Molecular phylogenetics and temporal diversification in the genus *Aeromonas* based on the sequences of five housekeeping genes. PLoS One. 9, e88805.

[26] Martin, D., Rybicki, E. (2000) RDP: detection of recombination amongst aligned sequences. Bioinformatics. 16, 562-563.

[27] Martin, D.P., Lemey, P., Posada, D. (2011) Analysing recombination in nucleotide sequences. Mol. Ecol. Res. 11, 943-955.

[28] Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B. (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, vev003.

[29] Martin-Carnahan, A., Joseph, S.W. (2005) Genus I. *Aeromonas* Stanier 1943, 213AL. In: Garrity, G.M., Brenner, D.J., Krieg, N.R., Staley, J.T. (Eds.), Bergey's Manual of Systematic Bacteriology, Vol. 2, Part B. Springer, NewYork, pp. 557-578.

[30] Martínez-Murcia, A.J., Monera, A., Saavedra, M.J., Oncina, R., Lopez-Alvarez, M., Lara, E., Figueras, M.J. (2011) Multilocus phylogenetic analysis of the genus *Aeromonas*. Syst. Appl. Microbiol. 34, 189-199.

[31] Martínez-Murcia, A., Beaz-Hidalgo, R., Svec, P., Saavedra, M.J., Figueras, M.J., Sedlacek, I. (2013). *Aeromonas cavernicola* sp. nov., isolated from fresh water of a brook in a cavern. Curr. Microbiol. 66, 197-204.

[32] 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

595 sequence typing, phenotype, and presence of putative virulence genes. Appl. Environ. Microbiol. 77,
596 4986-5000.
597
598 [33] Martino, M.E., Fasolato, L., Montemurro, F., Novelli, E., Cardazzo, B. (2014) *Aeromonas* spp.: ubiquitous
599 or specialized bugs? Environ. Microbiol. 16, 1005-1018.
600
601 [34] Maynard Smith, J.M. (1992) Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126-129.
602
603 [35] Miñana-Galbis, D., Farfán, M., Lorén, J.G., Fusté, M.C. (2002) Biochemical identification and numerical
604 taxonomy of *Aeromonas* spp. isolated from environmental and clinical samples in Spain. J. Appl.
605 Microbiol. 93, 420-430.
606
607 [36] Ochman, H., Lerat, E., Daubin, V. (2005) Examining bacterial species under the spectre of gene transfer
608 and exchange. Proc. Natl. Acad. Sci. USA. 102, 6595-6599.
609
610 [37] Padidam, M., Sawyer, S., Fauquet, C.M. (1999) Possible emergence of new geminiviruses by frequent
611 recombination. Virology. 265, 218-225.
612
613 [38] Paradis, E., Claude, J., Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language.
614 Bioinformatics. 20, 289.290.
615
616 [39] Perrière, G., Gouy, M. (1996) WWW-Query: an on-line retrieval system for biological sequence banks.
617 Biochimie. 78, 364-369.
618
619 [40] Posada, D., Crandall, K.A. (2001) Evaluation of methods for detecting recombination from DNA
620 sequences: computer simulations. Proc. Natl. Acad. Sci. USA. 98, 13757-13762.
621
622 [41] Posada, D., Crandall, K.A., Holmes, E.C. (2002) Recombination in evolutionary genomics. Annu. Rev.
623 Genet. 36, 75-97.
624

[42] R Core Team (2013). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www. R-project.org/

[43] 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) The genome of *Aeromonas salmonicida* subsp. *salmonicida* A449: insights into the evolution of a fish pathogen. BMC Genomics. 9, 427.

[44] Riley, M.A., Lizotte-Waniewski, M. (2009) Population genomics and the bacterial species concept. Methods Mol. Biol. 532, 367-377.

[45] Roger, F., Marchandin, H., Jumas-Bilak, E., Kodjo, A., colBVH study group, Lamy, B. (2012) Multilocus genetics to reconstruct aeromonad evolution. BMC Microbiol. 12, 62.

[46] Saavedra, M.J., Figueras, M.J., Martínez-Murcia, A.J. (2006) Updated phylogeny of the genus *Aeromonas*. Int. J. Syst. Evol. Microbiol. 56, 2481-2487.

[47] Salminen, M.O., Carr, J.K., Burke, D.S., McCutchan, F.E. (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by BOOTSCANning. AIDS Res. Hum. Retroviruses. 11, 1423-1425.

[48] Sánchez, R., Sali, A. (1997) Advances in comparative protein-structure modelling. Curr. Opin. Struct. Biol. 7, 206-214.

[49] Sepe, A., Barbieri, P., Peduzzi, R., Demarta, A. (2008) Evaluation of *recA* sequencing for the classification of *Aeromonas* strains at the genotype level. Lett. Appl. Microbiol. 46, 439-444.

[50] 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,

654 J.F. (2006) Genome sequence of *Aeromonas hydrophila* ATCC 7966^T: jack of all trades. J. Bacteriol.
655 188, 8272-8282.
656
657 [51] Silver, A.C, Williams, D., Faucher, J., Horneman, A.J., Gogarten, J.P., Graf, J. (2011). Complex
658 evolutionary history of the *Aeromonas veronii* group revealed by host interaction and DNA sequence
659 data. PLoS One. 6, e16751.
660
661 [52] Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P., Kampfer, P., Maiden, M., Nesme, X.,
662 Rossello-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C., Whitman, W.B. (2002) Report of
663 the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol.
664 Microbiol. 52, 1043-1047.
665
666 [53] Story, R.M., Weber, I.T., Steitz, T.A. (1992) The structure of the *E. coli* RecA protein monomer and
667 polymer. Nature. 355, 318-325.
668
669 [54] Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary
670 Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725-2729.
671
672 [55] Untergrasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G. (2012)
673 Primer3 - new capabilities and interfaces. Nucleic. Acids. Res. 40, e115.
674
675 [56] Vinuesa, P., Silva, C., Werner, D., Martínez-Romero, E. (2005) Population genetics and phylogenetic
676 inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium*
677 species cohesion and delineation. Mol. Phylogenet. Evol. 34, 29-54.
678
679 [57] Ward, J.J., McGuffin, L.J., Bryson, K., Buxton, B.F., Jones, D.T. (2004) The DISOPRED server for the
680 prediction of protein disorder. Bioinformatics. 20, 2138-2139.
681
682

682 **TABLES**

683

684 **Table 1**

685 Recombinant analysis (RDP4). Upper: results obtained with the complete *recA* gene sequences (1,068
686 positions) from 150 strains. Lower: *recA-recX* region sequences (1,522 positions) from 18 strains, including
687 the recombinant strains and those species closely related.

688

Recombinant strains	Detection methods	Events[a]	Breakpoint[b]		Major parent	Minor parent	*P- value*
			Begin	End			
Sequence analysis of *recA* gene							
AE147 CECT5741 CECT5742 LMG13667	RPD	1	899 / 939	Undetermined	*A.bestiarum* (98.5 – 99.0 %)	Unknown	9.6×10^{-7}
	GENECONV	1	899 / 939	Undetermined	*A.bestiarum* (98.1 – 99.9 %)	Unknown	1.9×10^{-5}
	BOOTSCAN	1	899 / 939	Undetermined	*A.bestiarum* (98.1 – 99.9 %)	Unknown	2.6×10^{-7}
	MAXCHI	1	869	Undetermined	*A.bestiarum* (98.1 – 99.9 %)	*A.eucrenophila* (92.9 – 94.7 %)	9.8×10^{-4}
	CHIMAERA	1	899 / 939	Undetermined	*A.bestiarum* (98.1 – 99.9 %)	Unknown	2.9×10^{-5}
	Siscan	1	794 / 833 / 834	Undetermined	*A.bestiarum* (97.6 – 98.0 %)	Unknown	1.8×10^{-8}
	3SEQ	1	869	Undetermined	*A.bestiarum* (98.1 – 99.9 %)	*A.eucrenophila* (94.7 – 95.7 %)	4.5×10^{-7}
Sequence analysis of *recA* - *recX* fragment							
AE147 CECT5741 CECT5742 LMG13667	GENECONV	1	969	1,188 / 1,202 / 1,235	*A.bestiarum* (98.7 – 99.3 %)	*A.eucrenophila* (96.7 – 97.3 %)	1.3×10^{-17}
	BOOTSCAN	1	885 / 986	1,188 / 1,205 / 1,235	*A.bestiarum* (98.8 – 99.1 %)	*A.eucrenophila* (94.9 – 97.5 %)	8.8×10^{-14}
	MAXCHI	1	969	1,202 / 1,214 / 1,235	*A.bestiarum* (98.2 – 99.6 %)	*A.eucrenophila* (96.2 – 97.3 %)	5.4×10^{-13}
	CHIMAERA	1	971	1,188 / 1,202 / 1,235	*A.bestiarum* (98.2 – 99.6 %)	*A.eucrenophila* (96.7 – 97.3 %)	3.1×10^{-12}
	3SEQ	1	969	1,188 / 1,205 / 1,235	*A.bestiarum* (98.7 – 99.3 %)	*A.eucrenophila* (96.7 – 97.3 %)	3.4×10^{-25}

689

690 [a] Recombination events statistically supported by at least two recombinant detection methods (*P-value* < 0.05).

691 [b] Position in alignment

692

693

694

Table 2

Comparison of the G+C content, corresponding to different regions of *recA* and *recX* genes, among recombinant strains and closely related *Aeromonas* species.

***Aeromonas* species**[a]	**Mean G+C content (mol %)**				
	***recA*[b]**			***recX*[b]**	**Recombinant fragment[b] 248 pb**
	Gene	**Positions 1-978**	**Positions 979-1062**	**Positions 1-129**	
A. bestiarum (n = 5)	59.1	58.9	61.7	63.7	60.7
***A. bestiarum* recombinants (n = 4)**	**59.2**	**58.8**	**64.0**	**63.0**	**62.9**
A. eucrenophila (n = 5)	60.2	59.7	66.4	62.9	63.2
A. tecta (n = 4)	59.9	59.3	67.0	61.4	62.5

[a] Reference strains of *A. bestiarum* (CECT4227^T, 112A, 559A, LMG13663, HE73), recombinant strains of *A. bestiarum* (AE147, CECT5741, CECT5742, LMG13667), strains of *A. eucrenophila* (CECT4224^T, CECT4827, CECT4853, CECT4854, CECT4855) and *A. tecta* (MDC91^T, MDC92, MDC93, MDC94); n, number of isolates.

[b] *recA-recX* sequences (GenBank accession numbers: KM260547 – KM260564).

FIGURE LEGENDS

Figure 1

Maximum likelihood phylogenies based on *recA* gene sequences: A (study 1) *Aeromonas hydrophila* Complex strains, including the relative positions of the four recombinant *A. bestiarum* strains (in bold); B (study 2) strains representing all the *Aeromonas* species. Species included in study 1 and 2 are indicated in bold on the right. The *A. bestiarum* cluster, including the four recombinant strains, is shaded in grey. Bootstrap values higher than 70% are given at the corresponding nodes. Scale bar represents the number of substitutions per site.

Figure 2

Dots plots showing the segregating sites of: A) the complete *recA* gene sequence for each strain analysed in the phylogenetic study 1, and B) the *recA-recX* sequences corresponding to 18 strains, including the recombinant (Table 2). The consensus sequence is shown at the top. Polymorphic sites for each sequence are marked with a different colour depending on the base (A, yellow; G, blue; C, green; T, red) in relation to the consensus sequence. The uppercase letters on the left indicate the sequence clusters corresponding to the following species: A, *A.bestiarum*; B, *A.piscicola*; C, *A.popoffii*; D, *A.salmonicida*; E, *A.dhakensis*; F, *A.hydrophila*; G, *A.bestiarum*; H, *A.eucrenophila*; I, *A.tecta*. Recombinant strains were included in the *A.bestiarum* cluster.

Figure 3

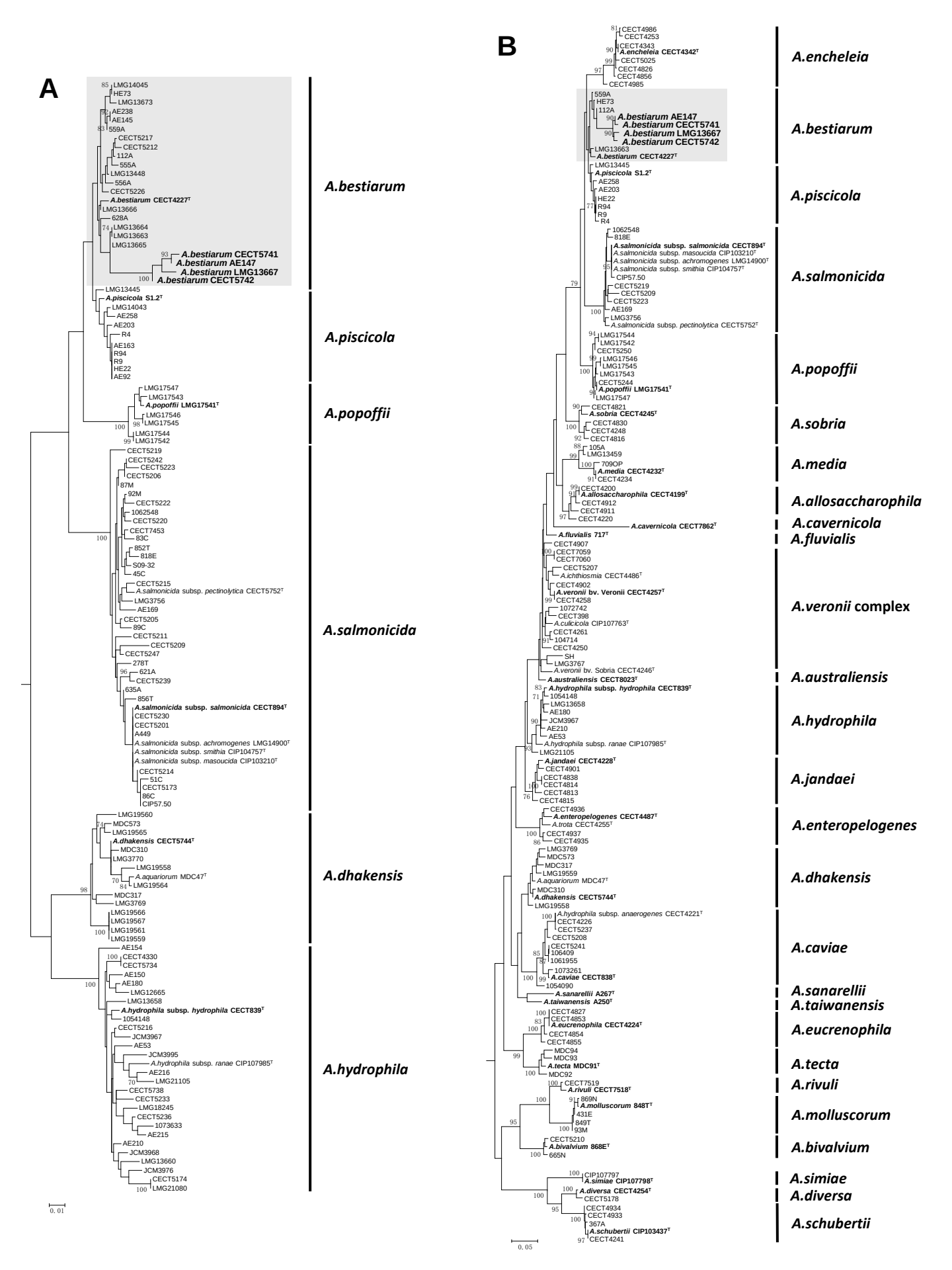
Maximum likelihood phylogenies based on non-recombinant (A, positions 1-978) and recombinant (B, positions 979 to the end) *recA* gene region sequences. The relative positions of the four recombinant *A.bestiarum* strains are in bold. Bootstrap values higher than 70% are given at the corresponding nodes. Scale bar represents the number of substitutions per site.

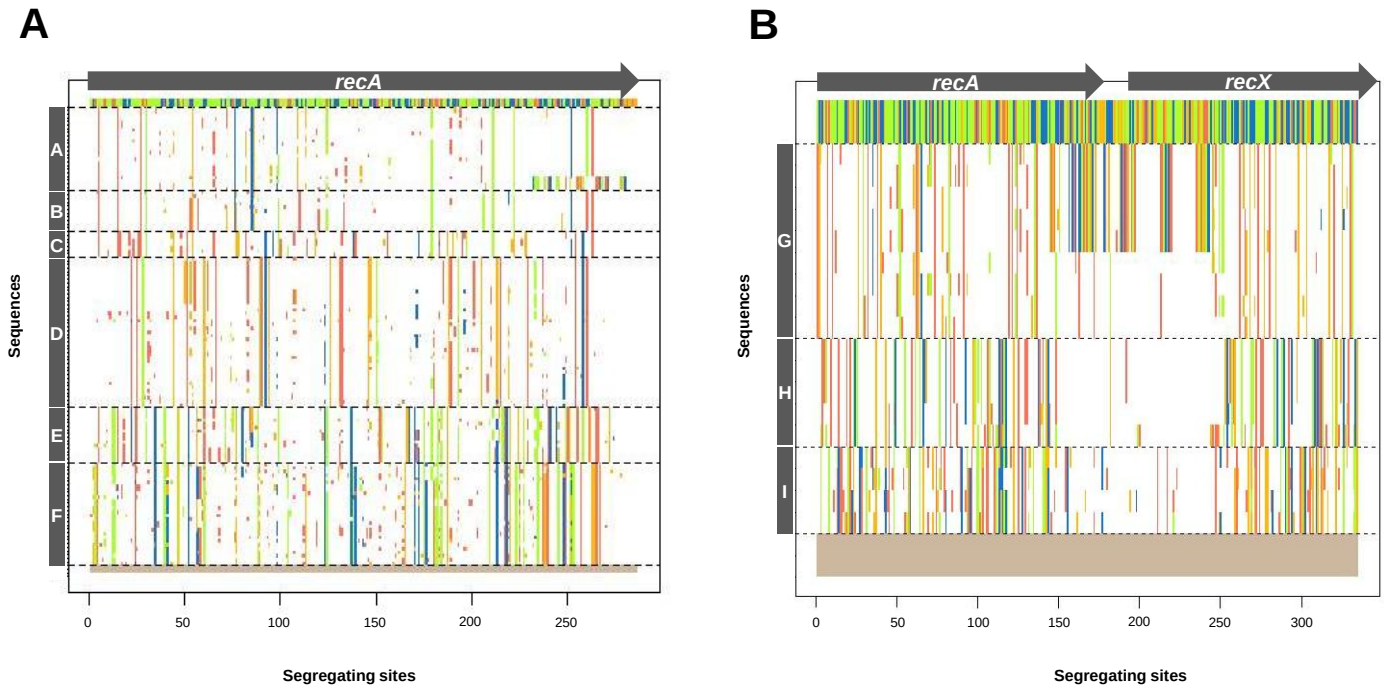
Figure 4

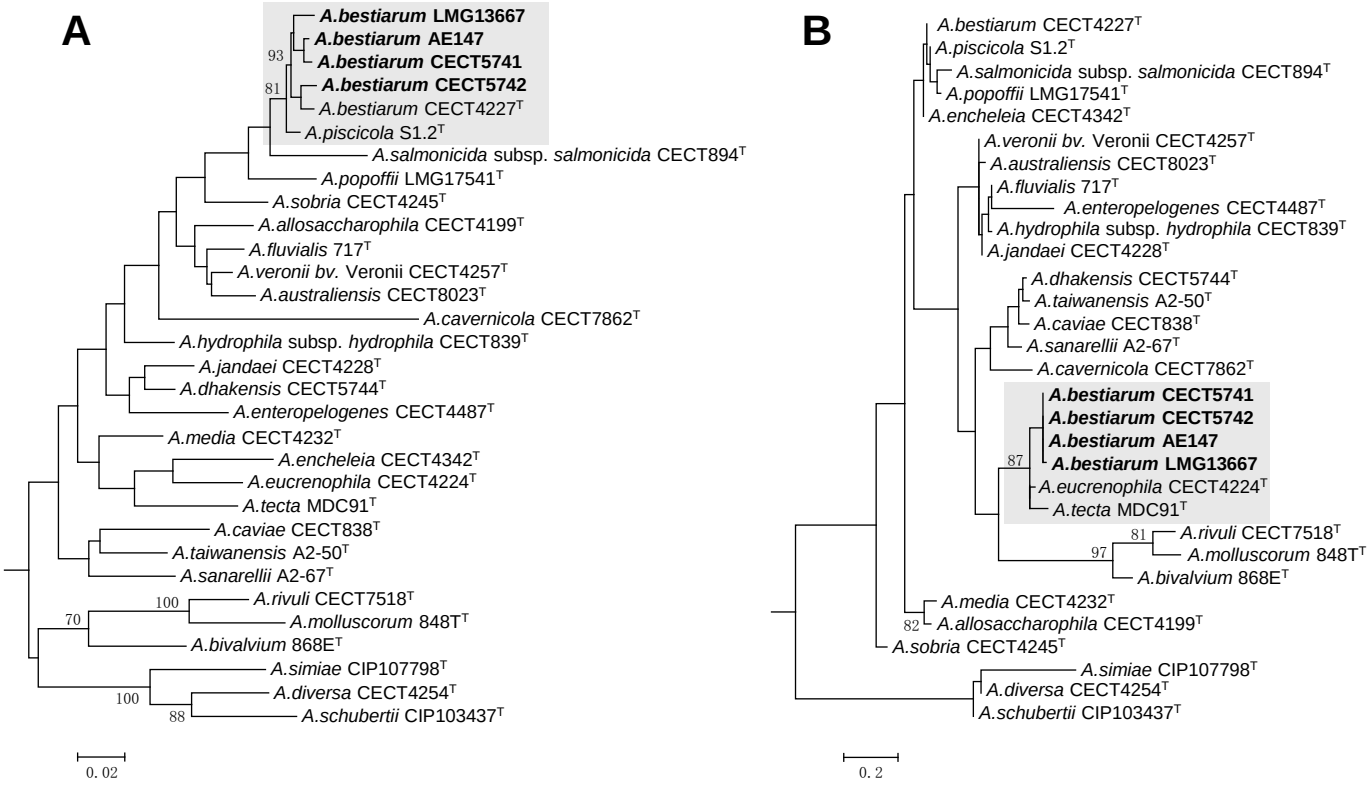
Phylogenetic network constructed with the *recA* gene sequences showing the relationships (in red) of the four recombinant strains with a set of 14 strains corresponding to the most closely related species (*A.bestiarum*, *A.eucrenophila*, *A.tecta*).

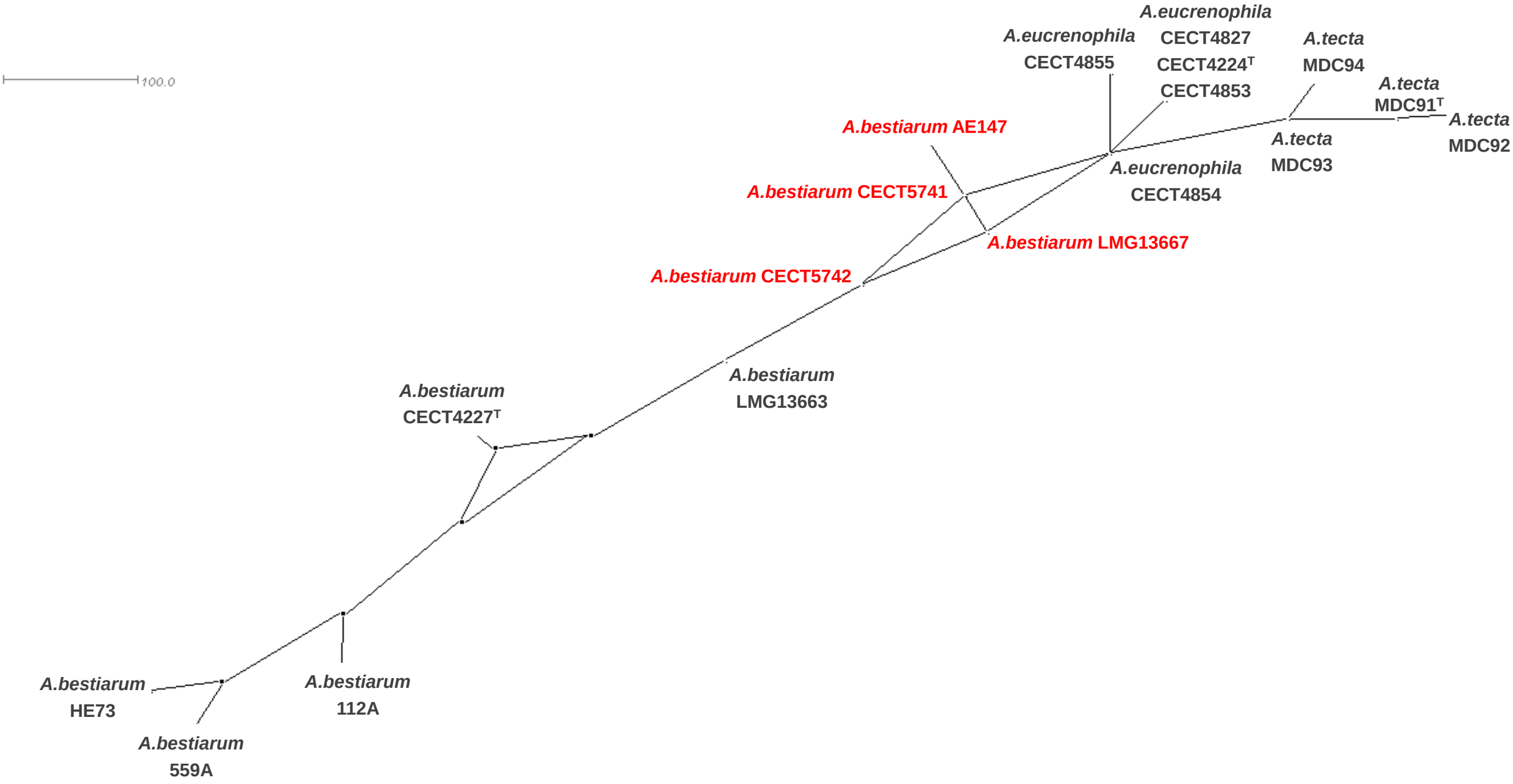
Figure 5

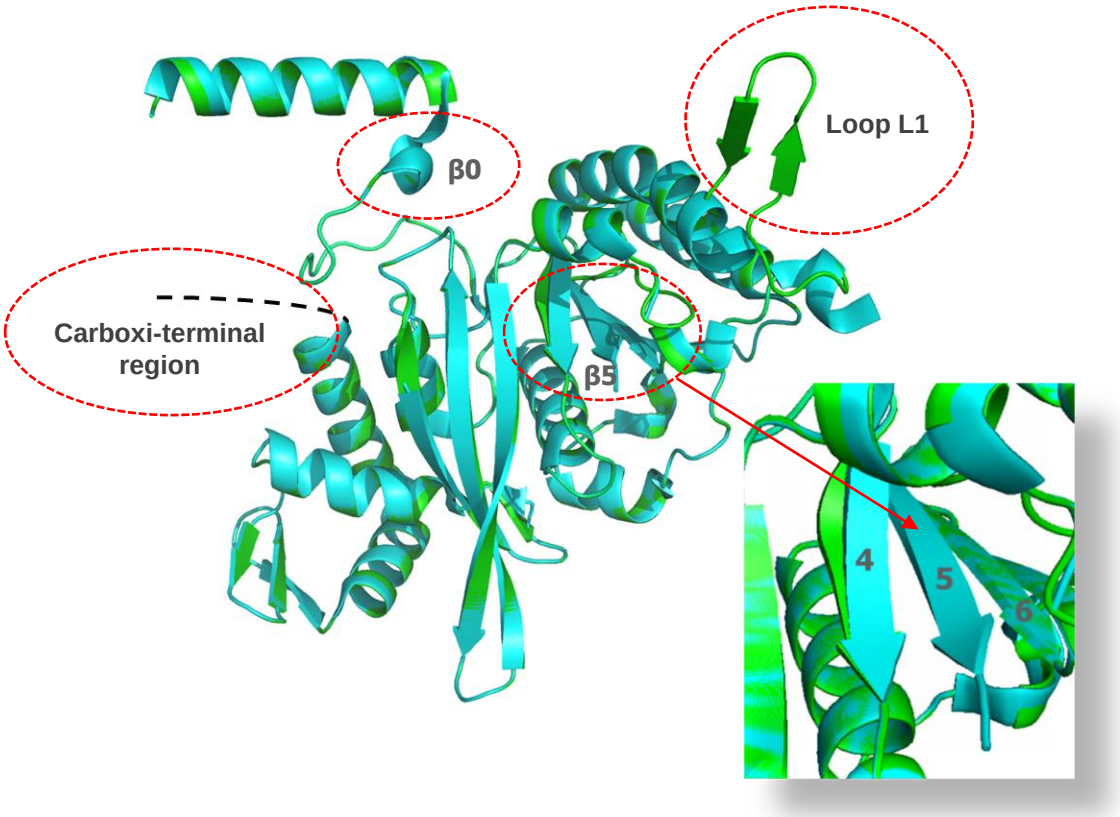
Three-dimensional *E. coli* RecA protein (PDB: 2REB; blue) and *A. bestiarum* CECT4227^T (green) structural alignment predicted using the PyMOL program. Differences are marked with red circles.


A
LMG14045
HE73
LMG13673
AE238
AE145
559A
CECT5217
CECT5212
112A
555A
LMG13448
556A
CECT5226
A.bestiarum CECT4227T
LMG13666
628A
LMG13664
LMG13663
LMG13665
A.bestiarum CECT5741
A.bestiarum AE147
A.bestiarum LMG13667
A.bestiarum CECT5742
LMG13445
A.piscicola S1.2T
LMG14043
AE258
AE203
R4
AE163
R94
R9
HE22
AE92
LMG17547
LMG17543
A.popoffii LMG17541T
LMG17546
LMG17545
LMG17544
LMG17542
CECT5219
CECT5242
CECT5223
CECT5206
87M
92M
CECT5222
1062548
CECT5220
CECT7453
83C
852T
818E
S09-32
45C
CECT5215
A.salmonicida subsp. pectinolytica CECT5752T
LMG3756
AE169
CECT5205
89C
CECT5211
CECT5209
CECT5247
278T
621A
CECT5239
635A
856T
A.salmonicida subsp. salmonicida CECT894T
CECT5230
CECT5201
A449
A.salmonicida subsp. achromogenes LMG14900T
A.salmonicida subsp. smithia CIP104757T
A.salmonicida subsp. masoucida CIP103210T
CECT5214
51C
CECT5173
86C
CIP57.50
LMG19560
MDC573
LMG19565
A.dhakensis CECT5744T
MDC310
LMG3770
LMG19558
A.aquariorum MDC47T
LMG19564
MDC317
LMG3769
LMG19566
LMG19567
LMG19561
LMG19559
AE154
CECT4330
CECT5734
AE150
AE180
LMG12665
LMG13658
A.hydrophila subsp. hydrophila CECT839T
1054148
CECT5216
JCM3967
AE53
JCM3995
A.hydrophila subsp. ranae CIP107985T
AE216
LMG21105
CECT5738
CECT5233
LMG18245
CECT5236
1073633
AE215
AE210
JCM3968
LMG13660
JCM3976
CECT5174
LMG21080
0. 01
A.bestiarum
A.piscicola
A.popoffii
A.salmonicida
A.dhakensis
A.hydrophila
B
CECT4986
CECT4253
CECT4343
A.enchelei CECT4342T
CECT5025
CECT4826
CECT4856
CECT4985
559A
HE73
112A
A.bestiarum AE147
A.bestiarum CECT5741
A.bestiarum LMG13667
A.bestiarum CECT5742
LMG13663
A.bestiarum CECT4227T
LMG13445
A.piscicola S1.2T
AE258
AE203
HE22
R94
R9
R4
1062548
818E
A.salmonicida subsp. salmonicida CECT894T
A.salmonicida subsp. masoucida CIP103210T
A.salmonicida subsp. achromogenes LMG14900T
A.salmonicida subsp. smithia CIP104757T
CIP57.50
CECT5219
CECT5209
CECT5223
AE169
LMG3756
A.salmonicida subsp. pectinolytica CECT5752T
LMG17544
LMG17542
CECT5250
LMG17546
LMG17545
LMG17543
CECT5244
A.popoffii LMG17541T
LMG17547
CECT4821
A.sobria CECT4245T
CECT4830
CECT4248
CECT4816
105A
LMG13459
709OP
A.media CECT4232T
CECT4234
CECT4200
A.allosaccharophila CECT4199T
CECT4912
CECT4911
CECT4220
A.cavernicola CECT7862T
A.fluvialis 717T
CECT4907
CECT7059
CECT7060
CECT5207
A.ichthiosmia CECT4486T
CECT4902
A.veronii bv. Veronii CECT4257T
CECT4258
1072742
CECT398
A.culicicola CIP107763T
CECT4261
104714
CECT4250
SH
LMG3767
A.veronii bv. Sobria CECT4246T
A.australiensis CECT8023T
A.hydrophila subsp. hydrophila CECT839T
1054148
LMG13658
AE180
JCM3967
AE210
AE53
A.hydrophila subsp. ranae CIP107985T
LMG21105
A.jandaei CECT4228T
CECT4901
CECT4838
CECT4814
CECT4813
CECT4815
CECT4936
A.enteropelogenes CECT4487T
A.trota CECT4255T
CECT4937
CECT4935
LMG3769
MDC573
MDC317
LMG19559
A.aquariorum MDC47T
MDC310
A.dhakensis CECT5744T
LMG19558
A.hydrophila subsp. anaerogenes CECT4221T
CECT4226
CECT5237
CECT5208
CECT5241
106409
1061955
1073261
A.caviae CECT838T
1054090
A.sanarellii A267T
A.taiwanensis A250T
CECT4827
CECT4853
A.eucrenophila CECT4224T
CECT4854
CECT4855
MDC94
MDC93
A.tecta MDC91T
MDC92
CECT7519
A.rivuli CECT7518T
869N
A.molluscorum 848T
431E
849T
93M
CECT5210
A.bivalvium 868ET
665N
CIP107797
A.simiae CIP107798T
A.diversa CECT4254T
CECT5178
CECT4934
CECT4933
367A
A.schubertii CIP103437T
CECT4241
0. 05
A.encheleia
A.bestiarum
A.piscicola
A.salmonicida
A.popoffii
A.sobria
A.media
A.allosaccharophila
A.cavernicola
A.fluvialis
A.veronii complex
A.australiensis
A.hydrophila
A.jandaei
A.enteropelogenes
A.dhakensis
A.caviae
A.sanarellii
A.taiwanensis
A.eucrenophila
A.tecta
A.rivuli
A.molluscorum
A.bivalvium
A.simiae
A.diversa
A.schubertii


A
recA
A
B
C
D
E
F
Sequences
0
50
100
150
200
250
Segregating sites
B
recA
recX
G
H
I
Sequences
0
50
100
150
200
250
300
Segregating sites


A
A.bestiarum LMG13667
A.bestiarum AE147
A.bestiarum CECT5741
A.bestiarum CECT5742
A.bestiarum CECT4227T
A.piscicola S1.2T
A.salmonicida subsp. salmonicida CECT894T
A.popoffii LMG17541T
A.sobria CECT4245T
A.allosaccharophila CECT4199T
A.fluvialis 717T
A.veronii bv. Veronii CECT4257T
A.australiensis CECT8023T
A.cavernicola CECT7862T
A.hydrophila subsp. hydrophila CECT839T
A.jandaei CECT4228T
A.dhakensis CECT5744T
A.enteropelogenes CECT4487T
A.media CECT4232T
A.encheleia CECT4342T
A.eucrenophila CECT4224T
A.tecta MDC91T
A.caviae CECT838T
A.taiwanensis A2-50T
A.sanarellii A2-67T
A.rivuli CECT7518T
A.molluscorum 848TT
A.bivalvium 868ET
A.simiae CIP107798T
A.diversa CECT4254T
A.schubertii CIP103437T
93
81
100
70
100
88
0. 02
B
A.bestiarum CECT4227T
A.piscicola S1.2T
A.salmonicida subsp. salmonicida CECT894T
A.popoffii LMG17541T
A.encheleia CECT4342T
A.veronii bv. Veronii CECT4257T
A.australiensis CECT8023T
A.fluvialis 717T
A.enteropelogenes CECT4487T
A.hydrophila subsp. hydrophila CECT839T
A.jandaei CECT4228T
A.dhakensis CECT5744T
A.taiwanensis A2-50T
A.caviae CECT838T
A.sanarellii A2-67T
A.cavernicola CECT7862T
A.bestiarum CECT5741
A.bestiarum CECT5742
A.bestiarum AE147
A.bestiarum LMG13667
A.eucrenophila CECT4224T
A.tecta MDC91T
A.rivuli CECT7518T
A.molluscorum 848TT
A.bivalvium 868ET
A.media CECT4232T
A.allosaccharophila CECT4199T
A.sobria CECT4245T
A.simiae CIP107798T
A.diversa CECT4254T
A.schubertii CIP103437T
87
81
97
82
0. 2


100.0
A.eucrenophila
CECT4855
A.eucrenophila
CECT4827
CECT4224T
CECT4853
A.tecta
MDC94
A.tecta
MDC91T
A.tecta
MDC92
A.bestiarum AE147
A.tecta
MDC93
A.eucrenophila
CECT4854
A.bestiarum CECT5741
A.bestiarum LMG13667
A.bestiarum CECT5742
A.bestiarum
LMG13663
A.bestiarum
CECT4227T
A.bestiarum
HE73
A.bestiarum
112A
A.bestiarum
559A


Loop L1
β0
Carboxi-terminal
region
β5
4
5
6