



Species Delimitation, Phylogenetic Relationships, and Temporal Divergence Model in the Genus *Aeromonas*

J. G. Lorén¹, Maribel Farfán^{1,2*} and M. C. Fusté^{1,2}

¹ Departament de Biologia, Sanitat i Medi Ambient, Secció de Microbiologia, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona, Spain, ² Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain

OPEN ACCESS

Edited by:

Jesus L. Romalde,
Universidade de Santiago
de Compostela, Spain

Reviewed by:

Tomochika Fujisawa,
Institut Pasteur, France
David R. Arahal,
Universitat de València, Spain
Antony T. Vincent,
Institut de Biologie Intégrative et des
Systèmes, Canada

*Correspondence:

Maribel Farfán
mfarfan@ub.edu

Specialty section:

This article was submitted to
Evolutionary and Genomic
Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 12 January 2018

Accepted: 05 April 2018

Published: 20 April 2018

Citation:

Lorén JG, Farfán M and Fusté MC
(2018) Species Delimitation,
Phylogenetic Relationships,
and Temporal Divergence Model
in the Genus *Aeromonas*.
Front. Microbiol. 9:770.
doi: 10.3389/fmicb.2018.00770

The definition of species boundaries constitutes an important challenge in biodiversity studies. In this work we applied the Generalized Mixed Yule Coalescent (GMYC) method, which determines a divergence threshold to delimit species in a phylogenetic tree. Based on the tree branching pattern, the analysis fixes the transition threshold between speciation and the coalescent process associated with the intra-species diversification. This approach has been widely used to delineate eukaryote species and establish their diversification process from sequence data. Nevertheless, there are few examples in which this analysis has been applied to a bacterial population. Although the GMYC method was originally designed to assume a constant (Yule) model of diversification at between-species level, it was later evaluated simulating other conditions. Our aim was therefore to determine the species delineation in *Aeromonas* using the GMYC method and assess which model best explains the speciation process in this bacterial genus. The application of the GMYC method allowed us to clearly delineate the *Aeromonas* species boundaries, even in the controversial groups, such as the *A. veronii* or *A. media* species complexes.

Keywords: species delimitation, GMYC, diversification model, *Aeromonas*, *mdh*, *recA*

INTRODUCTION

As in other organisms, diversification in bacteria leads to entities that we call species, which group together organisms that have evolved separately from others. Species can be differentiated from populations of individual strains by a maximum likelihood method that determines the point of transition of the evolutionary processes from the level of species (speciation and extinction) to the population (coalescence). The entities determined in this way maintain the biological properties and the levels of sequence divergence of the traditionally defined species.

Several methods have been proposed for species delimitation in bacteria. The classical approach uses a sequence divergence of 3 per cent pairwise distance in 16S rRNA (Schloss and Handelsman, 2006) or the 1 percent recommended by Acinas et al. (2004) as a threshold for separating species. However, rates of substitution may vary among lineages, as can the levels of variation within and between species, which makes it difficult to assume a universal threshold. Other widely applied methods for species delimitation are based on multilocus sequences, or more recently whole

genome sequencing of several individuals has been used to determine the average nucleotide identity (ANI) (Richter and Roselló-Mora, 2009) and *in silico* DNA-DNA hybridization (*isDDH*) (Meier-Kolthoff et al., 2013) values to separate species. When comparing these methods with traditional phenotypic approaches, discrepancies can sometimes arise.

The study of evolutionary patterns in prokaryotes is hampered by the lack of a reliable fossil record, limited morphological differentiation and frequently complex taxonomic relationships. The few studies in this field suggest a constant rate of diversification in free-living or symbiotic bacteria (Martin et al., 2004; Vinuesa et al., 2005; Sanglas et al., 2017), while in some pathogens, such as *Borrelia burgdorferi*, it seems to follow a radiation pattern (Morlon et al., 2012). Differences in the cladogenesis of *B. burgdorferi* could be explained by its association with vertebrate and arthropod hosts, which could restrict the gene flow between populations. Understanding the evolution of prokaryote biological diversity therefore remains a significant challenge for biologists (Barraclough et al., 2009).

The main objective of our research in the last years has been the study of diversification in the genus *Aeromonas*, a γ -*Proteobacteria* that comprises a group of Gram-negative, rod-shaped bacteria, which are found in aquatic environments worldwide and are members of the microbiota (as well as primary or secondary pathogens) of fish, amphibians and other animals (Martin-Carnahan and Joseph, 2005; Janda and Abbott, 2010). *Aeromonas* is an ideal genus to study the diversification processes in bacteria because its species, a combination of free-living bacteria and host-associated strains, can be isolated from a wide variety of habitats. We began by investigating the rate and pattern of cladogenesis in this bacterial genus from a collection of strains including the type strains of all the *Aeromonas* species, using the sequences of five housekeeping genes (Lorén et al., 2014). However, the frequently high intra-specific diversity in bacteria is not always adequately reflected by the type strain.

We therefore performed a second analysis using molecular data from the sequences of two housekeeping genes (*mdh* and *recA*) obtained from 150 strains belonging to 27 species of *Aeromonas* (Sanglas et al., 2017). Phylogenies that mix variation between and within species often need to be reduced to trees with only one sequence per species to avoid invalidating the results (Fontaneto et al., 2012). To fulfill these conditions we constructed a tree with the consensus sequence for each species and used the BEAST program to obtain the species tree (Sanglas et al., 2017). In both cases the results of the analysis allowed us to determine that the process of speciation in *Aeromonas* follows a constant model of diversification.

The genus *Aeromonas*, includes 30 taxonomic species, many of them described recently. However, some descriptions were based on only one strain, or using techniques such as the 16S rRNA gene sequences, which have been demonstrated to have a low discriminatory power for delimiting species in *Aeromonas* (Figueras et al., 2000). Consequently, the improper characterization of some *Aeromonas* species has led to several reclassifications, as in the case of *A. aquariorum* (Beaz-Hidalgo et al., 2013), *A. ichthiosmia* (Huys et al., 2001), *A. culicicola* (Huys et al., 2005), among others. In addition, the genus includes several

species complexes, such as *A. hydrophila* (Martin-Carnahan and Joseph, 2005), *A. veronii* (Silver et al., 2011) or *A. media* (Talagrand-Reboul et al., 2017), which are constituted by groups of strains with high intra-specific heterogeneity, giving rise to controversies about the species delineation.

In this study, we aimed to determine the species delineation and corroborate the diversification process in *Aeromonas* by applying a sequence-based method that allows the use of several strains per species, the Generalized Mixed Yule Coalescent (GMYC) method. This approach has been widely used to delineate eukaryote species and establish the diversification process from sequence data (Pons et al., 2006; Fontaneto et al., 2007; Lahaye et al., 2008). Nevertheless, there are only a few previous examples in which this analysis has been applied to a bacterial population (Barraclough et al., 2009; Powell, 2012). The method determines a divergence threshold to delimit species in a phylogenetic tree. Based on the tree branching pattern, the analysis fixes the transition threshold between speciation and the coalescent process associated with the intra-species diversification. The pruned tree provided by the threshold allows the use of classical methods of diversification analysis to determine the speciation process followed by the species in the analyzed population.

It is presupposed that branch lengths between species are derived from speciation and extinction rates, whereas branch lengths within a species reflect a coalescence process at the population level (Pons et al., 2006). The method determines the locations of ancestral nodes that define putative species and applies a likelihood ratio test to assess the fit of the branch lengths to a mixed lineage birth-population coalescent model (Pons et al., 2006). As the analysis to determine the transition threshold is based on the tree branching pattern, the tree has to accomplish certain requirements, for example, be ultrametric, fully resolved (without multifurcations) and reliable (with a high support in the nodes).

Our aim was therefore to determine the species delineation in *Aeromonas* based on the GMYC method. Although initially this method assumed diversification at between-species level follows a constant (Yule) model (Pons et al., 2006; Fontaneto et al., 2007), it was later evaluated simulating other diversification models (Fujisawa and Barraclough, 2013). Despite previous results indicate that *Aeromonas* follows a constant (Yule) model of diversification (Lorén et al., 2014; Sanglas et al., 2017), we also determined which model best explains the speciation process in this bacterial genus.

MATERIALS AND METHODS

Gene Sequences

A collection of 147 *Aeromonas* strains, representative of the 30 species recognized up to August 2017, was selected for the study. Two housekeeping genes (*mdh* and *recA*) were chosen for the analysis; for each strain, the full-length sequences for both genes were previously obtained (Sanglas et al., 2017). In the case of the more recently accepted *Aeromonas* species: *A. lacus*, *A. aquatica* and *A. finlandiensis*, sequences were obtained from genomes

available in the GenBank¹. The strains and sequences are listed in Supplementary Table S1, including the GenBank accession numbers.

Data Sets

Phylogenetic reconstruction was carried out from the concatenated sequences of *mdh* and *recA* genes. For each gene, the translated sequences were aligned using the ClustalW program implemented in MEGA6 (Tamura et al., 2013) and translated back to obtain the nucleotide alignments. Both alignments were concatenated with the DAMBE program (v5.3.10; Xia, 2013) and later the sequences were checked for the presence of incongruences or gaps. Divergent and ambiguously aligned blocks were also removed using the Gblocks program (Castresana, 2000).

Sequence Analysis

To ascertain if the sequences used allowed a good species discrimination we determined the intra- and inter-specific distances among the different *Aeromonas* species. Data were graphically depicted with box plots obtained using the R package ggplot2 (Wickham, 2016). Sequences were also analyzed with different distance models that consider the influence of saturation, base frequency, transition–transversions or multiple substitutions. All the analyses were conducted with the function *dist.dna* implemented in the R package *ape* (Paradis et al., 2004). To determine the polymorphic sites along the sequences we construct a dots plot graph with the R package *phyclust* (Chen, 2010, 2011). The R package *NbClust* (Charrad et al., 2014) was used to determine the number of clusters from the *p* distances of the alienated sequences (Supplementary Table S2).

Phylogenetic Reconstruction

Bayesian phylogenetic trees were reconstructed with the BEAST program (v1.8.1; Drummond and Rambaut, 2007; Drummond et al., 2012) from the data sets. The model of evolution for the each gene was determined using the jModelTest 2 program (Darriba et al., 2012). The general time-reversible model with discrete gamma distribution and invariant sites (GTR + G + I) was selected as the best-fit model of nucleotide substitution. The Bayesian analysis was performed using a GTR model with four gamma categories, a Yule process of speciation, and a constant clock model of evolution as the tree priors, as well as other default parameters. We used the divergence time between *Escherichia coli* and *Salmonella enterica* estimated by Ochman and Wilson as the calibration point (Ochman and Wilson, 1987a,b). Accordingly, we calibrated the divergence of *Aeromonas* with a normally distributed prior with a mean of 140 Ma and a standard deviation of 10 Ma. We performed three independent Markov Chain Monte Carlo (MCMC) runs 20 million generations, sampling every 2,000 generations. Posterior distributions for parameter estimates and likelihood scores to approximate convergence were visualized with the Tracer program (v1.6.0; Rambaut et al., 2014). Visual inspection of traces within and across runs, as well as the effective sample sizes (ESS)

¹<http://www.ncbi.nlm.nih.gov/genbank/>

of each parameter (>200), allowed us to confirm that the analysis was adequately sampled. A maximum clade credibility (MCC) tree was chosen by TreeAnnotator (v1.8.1; Drummond et al., 2012) from the combined output of the three MCMC runs using the LogCombiner program² after the removal of the initial trees (20–25%) as burn-in. The MCC tree was visualized with the program FigTree (v1.4.2)³.

GMYC Analysis

The GMYC analysis uses the information contained in a tree to delimit species and determine the diversification model. To analyze the influence of the priors chosen, we constructed 3 different trees from the same DNA sequence alignment, changing the model used to express the branching pattern of the tree (the Yule or a coalescent model) and the rate of molecular evolution, considering a constant or a relaxed clock, as recommended by Michonneau (unpublished). The topology of the trees would be almost identical in all cases; however, the branch lengths would vary.

The analysis was conducted using the function *gmyc* available in the *splits* package implemented in R (Ezard et al., 2014), with the single threshold option as recommended by Fujisawa and Barraclough (2013). The method compares the null (no threshold) and alternative hypothesis (one threshold) and infers the number of genetic entities. Branching events between species are modeled with a Yule model (Barraclough and Nee, 2001), while branching events within species are adjusted to a neutral coalescent process (Hudson, 1991).

²<http://beast.bio.ed.ac.uk/logcombiner>

³<http://beast.bio.ed.ac.uk/figtree>

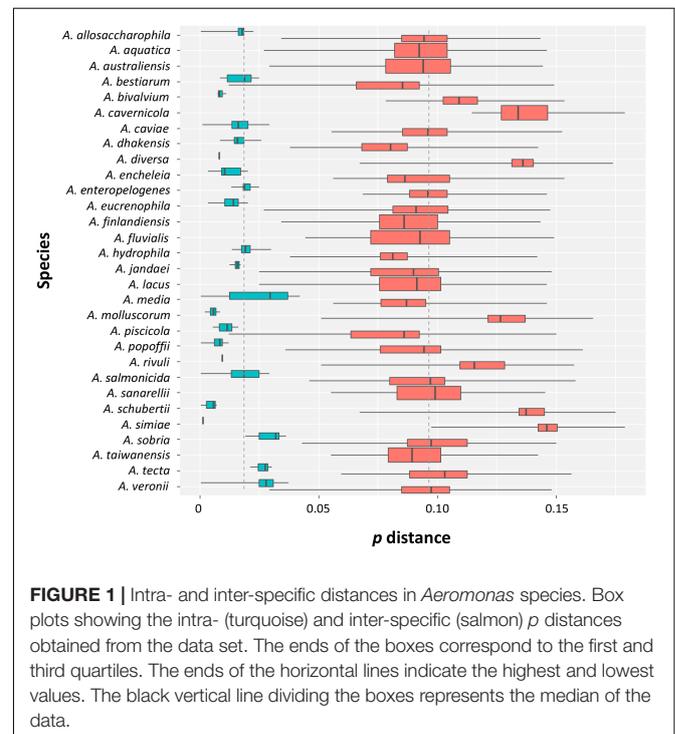


FIGURE 1 | Intra- and inter-specific distances in *Aeromonas* species. Box plots showing the intra- (turquoise) and inter-specific (salmon) *p* distances obtained from the data set. The ends of the boxes correspond to the first and third quartiles. The ends of the horizontal lines indicate the highest and lowest values. The black vertical line dividing the boxes represents the median of the data.

Results are evaluated by a likelihood ratio test (LRT) between the null hypothesis, which considers that all strains analyzed constitute a single species, and the alternative hypothesis (GMYC model) that assumes the existence of a species-delimiting threshold. The LRT significance was calculated using a chi-square test with 2 degrees of freedom.

The gmyc function gives a likelihood score for the model that considers all sequences belonging to the same species, and a likelihood score considering the sequences split in different species. The output also lists how many clusters and entities are associated with the highest likelihood score with the corresponding confidence intervals (CI) and the estimated threshold time when there was a transition between the speciation- and coalescent-level events. The R package also contains functions that plot (1) the number of lineages-through-time (LTT), with the inferred position of the threshold (red vertical line); (2) the likelihood profile through time; (3) the tree with the clusters highlighted in red. Additionally, the “support” for the delineated species can be plotted, indicating whether the results are reliable or not.

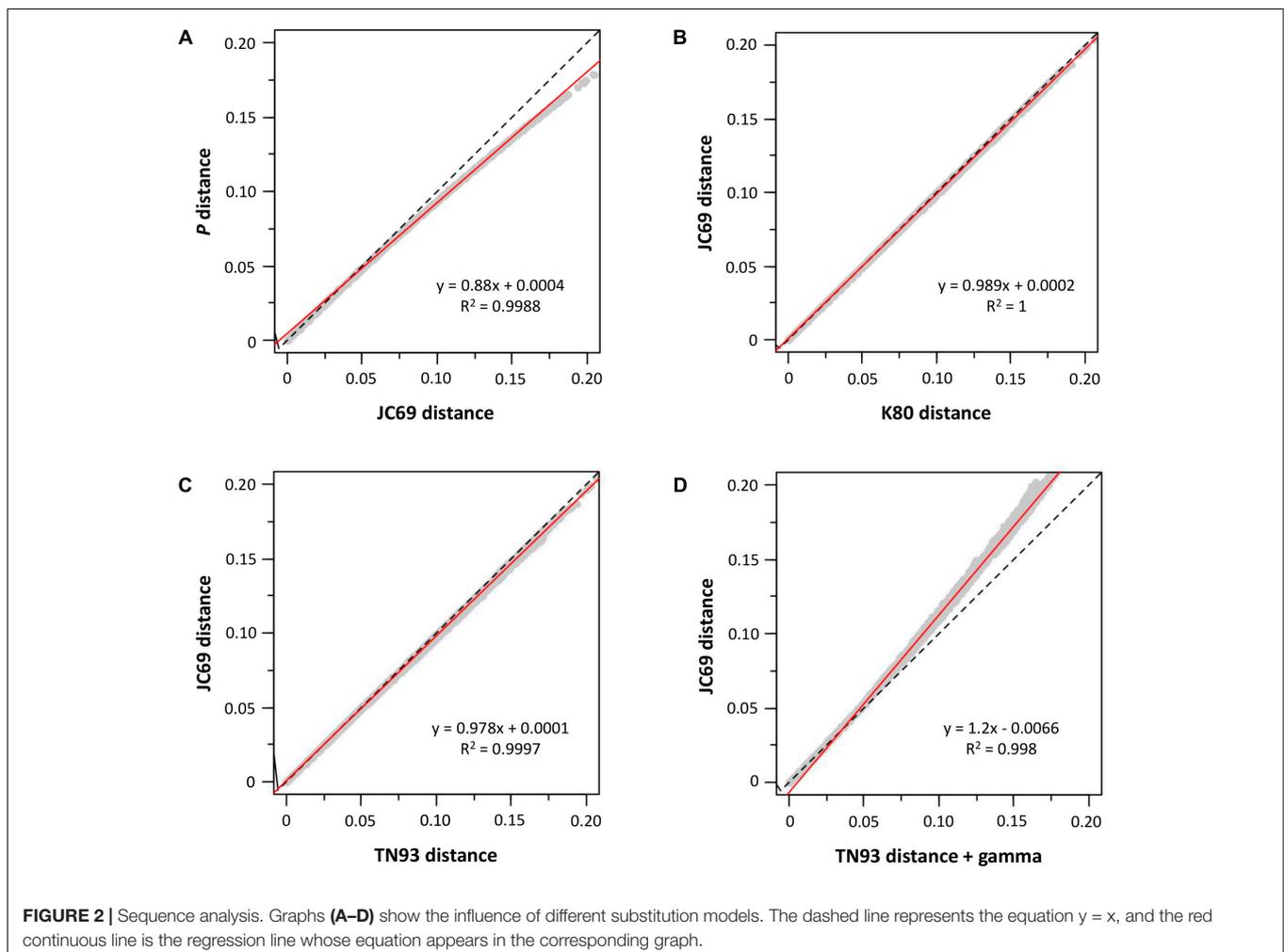
The use of only one indirect calibration point in the construction of the phylogeny, due to the absence of more reliable

calibration data, can be a source of uncertainty. As the GMYC method relies on relative rather than absolute branch lengths, we repeated the analysis with the same alignment but removing the outgroup, obtaining the corresponding tree. Following the suggestion of Fujisawa and Barraclough (2013), we scaled branch lengths in the tree to have a root age of 1.0 before running the analysis.

Diversification Analyses

A standard lineage-through-time (LTT) plot was constructed using the R package ape (Paradis et al., 2004) to graphically visualize and evaluate the temporal pattern of lineage diversification in *Aeromonas*.

We used the birth–death likelihood (BDL) tests implemented in laser (Rabosky, 2009) to detect the diversification model and the speciation and extinction rates (λ and μ) from the pruned tree obtained by cutting the original tree at the threshold given by the GMYC analysis. To test the null hypothesis of no-rate change versus variable-rate change in diversification, we have applied the ML approach of Rabosky, the test ΔAIC_{RC} (Rabosky, 2009). This statistic is calculated as: $\Delta AIC_{RC} = AIC_{RC} - AIC_{RV}$, where AIC_{RC} is the Akaike information criterion (AIC) score for the best fitting



rate-constant diversification model, and AIC_{RV} is the AIC for the best fitting variable-rate diversification model. Thus, a positive value for AIC_{RV} indicates that the data are best approximated with a rate-variable model, while a negative AIC_{RV} value suggests a rate-constant model of diversification. We tested eight different models, two of which were rate-constant (pure-birth or Yule and birth-death) and six were rate-variable (DDL, DDX, Yule 2-,3-,4- and 5- rates) (Lorén et al., 2014).

We calculated the gamma (γ) statistic (Pybus and Harvey, 2000; Fordyce, 2010) and its significance by simulating 5,000 phylogenies, as described previously (Lorén et al., 2014). This statistic compares the relative node positions in a phylogeny with those expected under a constant diversification rate model, in which the statistic follows a standard normal distribution. Positive γ values evidence that nodes are closer to the tips than expected under the constant rate model. When γ is negative, the internal nodes are closer to the root than expected under a constant model, indicating a decrease in diversification through

time. In addition, we compared the observed empirical gamma value with a distribution of the gamma statistics obtained by simulation.

RESULTS

Sequence Analysis

The analysis involved 147 *Aeromonas* strains. The number of total positions analyzed was 1,979 bp. All positions containing gaps and missing data were eliminated in the construction of the phylogenetic tree.

The intra- and inter-specific distances determined from the sequences (Supplementary Figure S1 and Supplementary Table S3), with the exception of the conflicting *A. veronii* group and the *A. bestiarum*/*A. piscicola* cluster, allowed a good discrimination of the species with no overlap among the distances (Figure 1).

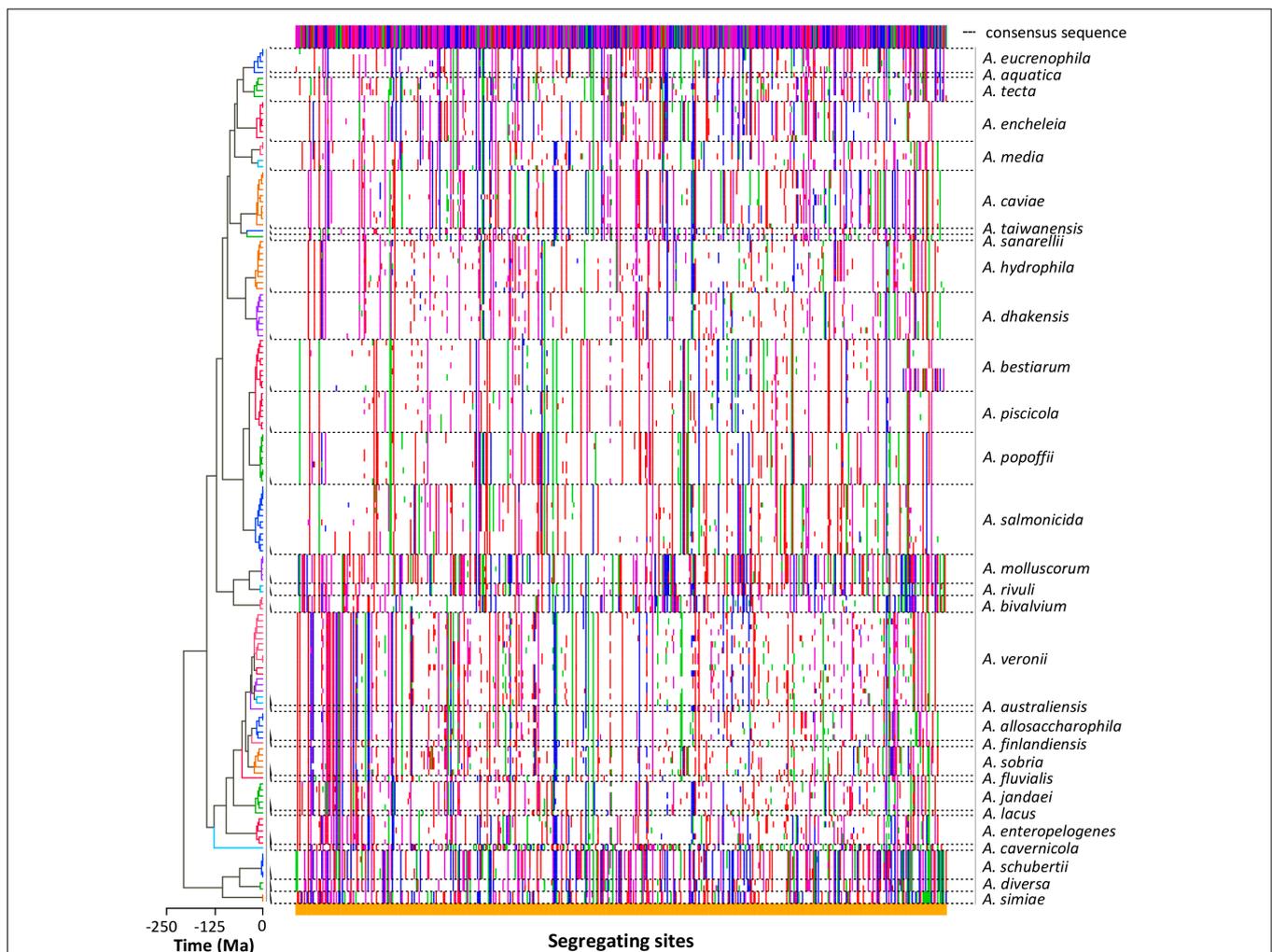


FIGURE 3 | Sequence polymorphism. The dots plot shows the segregating sites determined in each sequence arranged in the same order as in the tree (left). Polymorphic differences between the sequences and the consensus sequence (top) are marked in colors depending on the base (A green, G blue, C purple, T red). Dashed lines delimit the sequences belonging to the same *Aeromonas* species.

We also determined the sequence evolution of the concatenated sequences of *mdh* and *recA*. As can be seen in **Figure 2**, we analyzed our sequences for the influence of saturation (**Figure 2A**), base frequencies (**Figure 2B**), transitions and transversions (**Figure 2C**), and the gamma correction for multiple substitutions (**Figure 2D**). The sequences showed no influence of saturation, base frequency bias or the transition-transversion ratio; only the heterogeneity in the inter-site substitution rate could have some effect (**Figure 2D**).

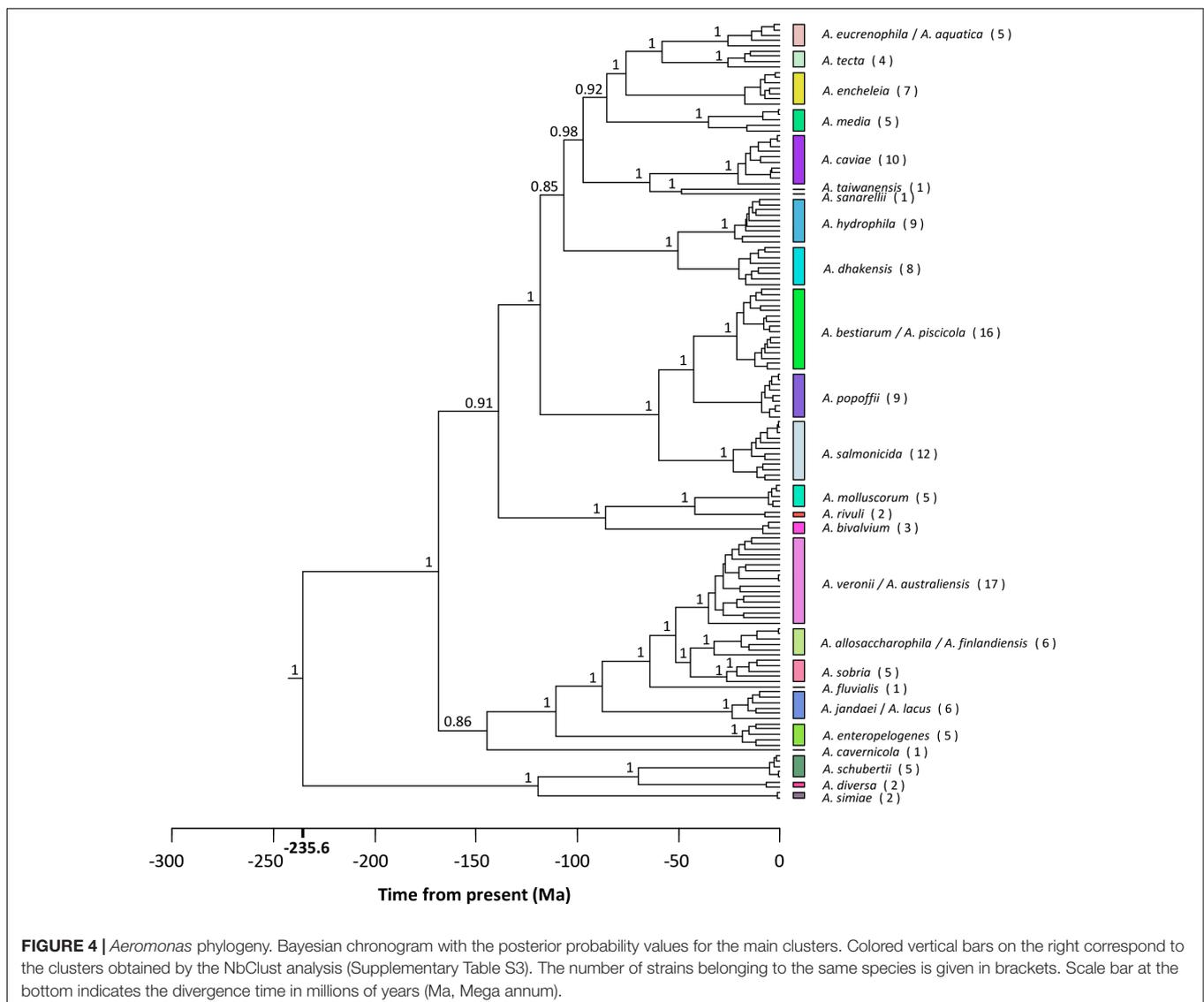
The nucleotide substitution analysis (**Figure 3**) showed a clear diversity among the sequences of the strains belonging to the same species. Nevertheless, the different species exhibited segregating site patterns that allowed them to be separated.

Phylogenetic Reconstruction

The best-fit models of sequence evolution were implemented according to the Akaike Information Criterion (AIC) scores for the substitution models evaluated, using jModeltest.

The general time reversible (GTR) model was selected as the best model of evolution for the concatenated sequences using a discrete gamma distribution and a fraction of invariable sites (GTR + G + I). **Figure 4** shows the *Aeromonas* Bayesian phylogeny, all the strains belonging to the same species clustering together in the same group. The posterior values obtained for each node were close to 1 for the majority of the main clades. The figure also shows the clusters obtained with the *p* distance analysis (Supplementary Table S2), which fully matched those established from the phylogeny.

Figure 5 depicts the phylogeny obtained with the corresponding lineage-through-time plot (LTT plot). The LTT plot (**Figure 5B**) clearly shows two different linear relationships with a sudden change in the slopes at approximately -25 Ma. This breakpoint roughly matches the region of the tree between -25 Ma and the present (**Figure 5A**), which accumulates the majority of the nodes (75%). We applied a linear regression



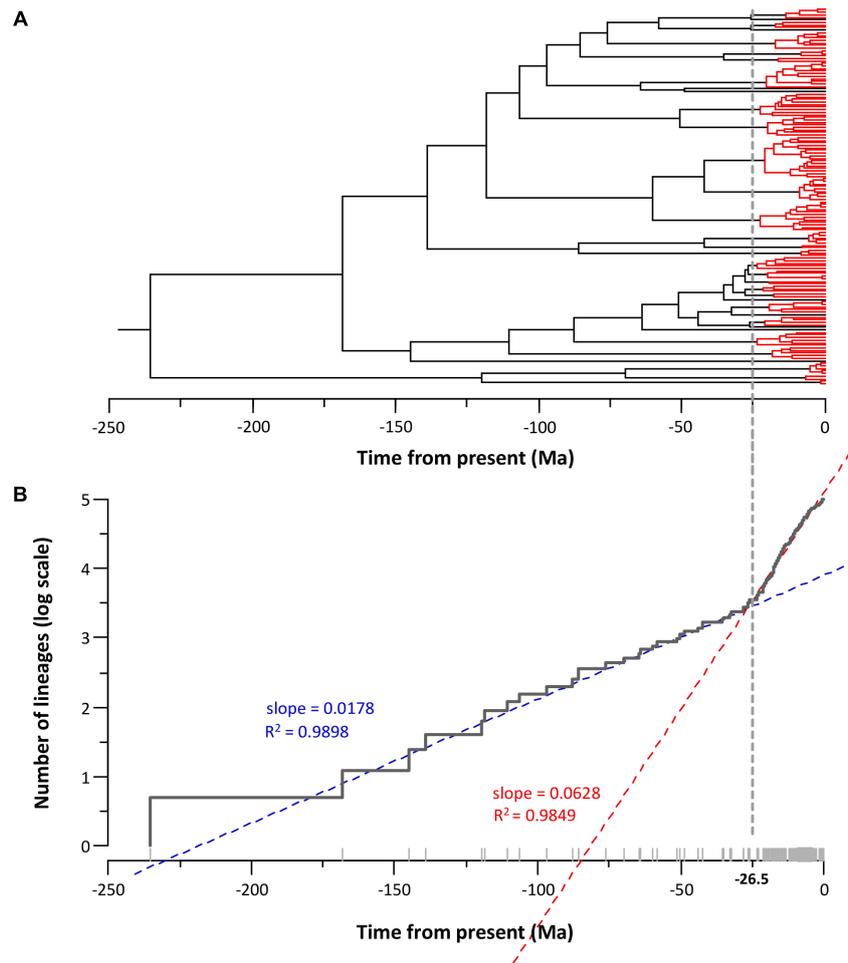


FIGURE 5 | *Aeromonas* diversification. **(A)** *Aeromonas* chronogram showing the temporal node distribution. Nodes posterior to -25 Ma (breakpoint) are marked in red. **(B)** LTT plot corresponding to the chronogram showing the breakpoint and the linear adjustment of the two segments (before and after the breakpoint).

model to the LTT plot points between the crown age (-235.6 Ma) and the breakpoint (-25 Ma) (segment I), and those obtained from the breakpoint to the present (segment II). The fit of the two segments to a linear model was very good, with R^2 values of 0.9898 and 0.9849 for segments I and II, respectively. The two regression lines intersected at -26.5 Ma, and the slopes for segment I and II were 0.0178 and 0.0628 (Figure 5B). Using the Chow test (Chow, 1960), we verified that the two linear regression slopes differed significantly (F test = 1715.8, P value = $1.12e^{-100}$).

Generalized Mixed Yule Coalescent

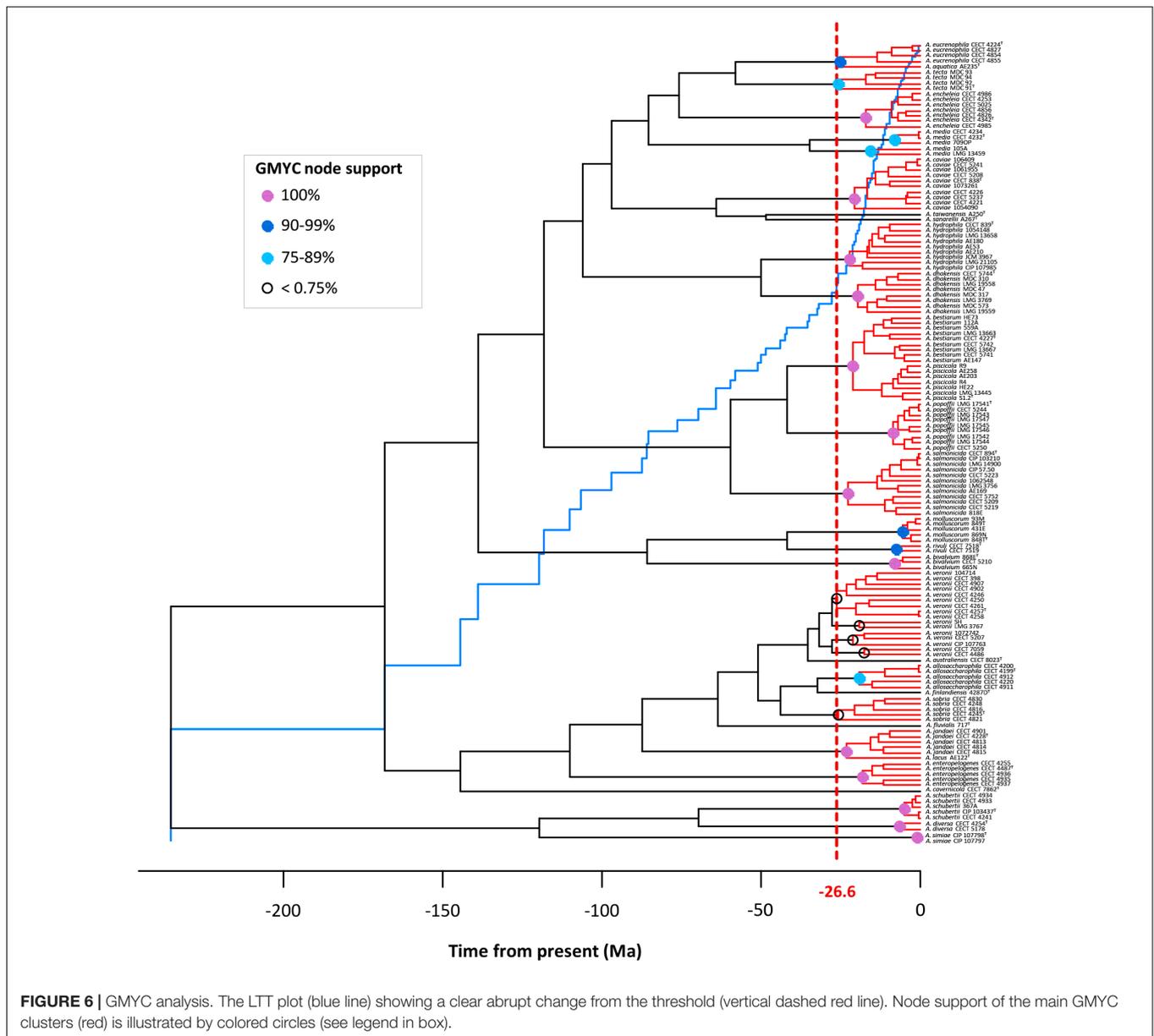
The existence of distinct *Aeromonas* lineages was confirmed by the branch length analysis. The resulting LTT plot showed a steep upturn in branching rates toward the present, marking the transition from the between- to within-species rate of lineage branching (Figure 6). The GMYC model fitted a transition in branching rate occurring at -26.6 Ma. The support in the majority of the GMYC clusters was high, confirming the reliability of the results (Figure 6).

The number of ML entities determined was 31 with a confident interval (CI) of 24–36 (Table 1). The λ values for the diversification and coalescent processes were 0.0092 and 0.2779, respectively. The scaling parameter p for the diversification process was close to 1 (1.154), which indicates a constant speciation rate model with no extinction (Yule model), while the value for the coalescent process was clearly below 1 ($6.186e^{-08}$) indicating a deficit of recent coalescent events (Fujisawa and Barraclough, 2013). No substantial differences were detected between Bayesian models using a coalescent or Yule prior when constructing the tree (Table 1).

The calibration point used to date the phylogeny also had no influence. Identical results were obtained when the dated tree was compared with the one obtained by repeating the analysis with the same alignment but removing the outgroup.

Diversification Analysis

As suggested by Rabosky (2006), we calculated the significance of ΔAIC_{RC} for the set of analyzed models by using the Yule model



to simulate 5,000 phylogenies of the same size and diversification rate as those obtained from our data, and determined the P value from the resulting distributions. As can be seen in **Table 2**, we cannot reject the null hypothesis of a Yule model to a level of significance of $\alpha = 0.05$, which means that the diversification in *Aeromonas* is constant.

For further corroboration, we determined the gamma statistic of Pybus and Harvey, a powerful tool principally used for comparing models of decreasing speciation rate through time and a constant rate of diversification. The estimated γ value from the chronogram (pruned tree) was 0.798 with a P value of 0.425, indicating that we cannot reject the null hypothesis of constant diversification for our phylogeny. Data obtained from the simulation also corroborate a constant diversification process in *Aeromonas* (**Figure 7**).

DISCUSSION

The choice of appropriate genes is crucial for the reconstruction of reliable molecular phylogenies. They should be housekeeping genes, evolving at a constant rate (good molecular clocks), and have unsaturated synonymous and non-synonymous substitutions. Unlike the majority of phylogenetic studies, in our analysis we checked our sequences to ensure that the genes used in this work fulfilled the above conditions and allowed a good species separation.

The phylogeny constructed from the concatenated sequences corroborates the monophyletic origin of this group of bacteria. In the chronogram obtained, the majority of the nodes were strongly supported, with posterior values close to 1. In addition, the main clade distribution was in

TABLE 1 | GMYC analysis.

	GMYC	
	Yule Constant clock	Coalescent Constant clock
Number of ML entities (CI) ^a	31 (24 – 36)	31 (25 – 38)
Threshold time	–26.60	–24.60
pdiv value ^b	1.1548	1.2170
pcoal value ^c	6.1860e-08	2.0497e-09
λ_{div} ^b	0.0092	0.0080
λ_{coal} ^c	0.2779	0.3026

^aML, maximum likelihood; CI, confidence interval. ^bdiv, diversification process. ^ccoal, coalescent process.

agreement with previously published phylogenies (Roger et al., 2012; Colston et al., 2014; Lorén et al., 2014; Sanglas et al., 2017). We obtained a perfect clustering of the strains belonging to the same species, including those considered synonymous.

The analyses based on the LTT plot detected an increase in the diversification rates from a certain point in time. As **Figure 5** illustrates, a high percentage of nodes accumulate close to the present due to the greater similarity of the sequences belonging to the same species.

The LTT plot constructed from the phylogenetic tree clearly shows two different linear relationships with a sudden change in the slopes at –26.5 Ma. The fit of the two segments to a linear model was very good, with high *R* squared values. The Chow test verified that the two linear regression slopes differed significantly (P value = $1.12e^{-100}$), which indicates a change in the rate of lineage branching.

In this work we carried out a quantitative analysis of sequence data to delimit putative *Aeromonas* species by detecting shifts in the rate of lineage branching. The branch-length analysis is based on a probabilistic model that distinguishes between species diversification and coalescent processes. This approach compensates for undefined species limits by including confidence intervals when allocating species-defining nodes and does not rely on population limits; thus, species represented by a single strain can be included (Pons et al., 2006).

Some studies recommend the use of BEAST for the tree construction as input for the GMYC analysis (Monaghan et al., 2009; Tang et al., 2014). From a set of given data (sequences)

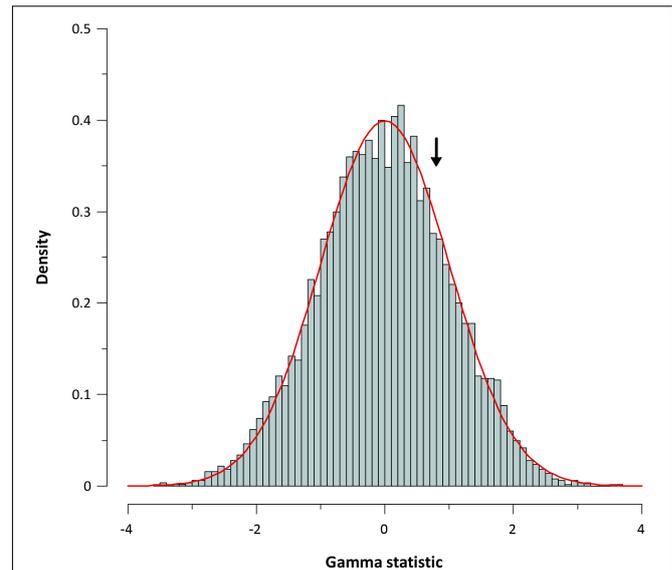


FIGURE 7 | Gamma statistic distribution. Gamma statistic distribution obtained by simulating 5,000 phylogenies under the Yule model. The arrow indicates the empirical gamma (γ) value obtained.

and a substitution model, the Bayesian Inference performs a probability analysis in search of the best set of trees that maximize the posterior probability. From an initial tree, it uses the Markov Chains Monte Carlo to evaluate the posterior probability of the different states proposed. After generating a large number of trees (frequently around 10,000), it uses the subsequent probability to ascertain how many times a node is repeated in each tree and this is done for each node. Subsequently, the Bayes theorem is used to build a single phylogenetic tree (maximum clade credibility, MCC). The same method provides an estimate of the reliability of the results obtained through the posterior probability values of each node, without the need to evaluate it *a posteriori*.

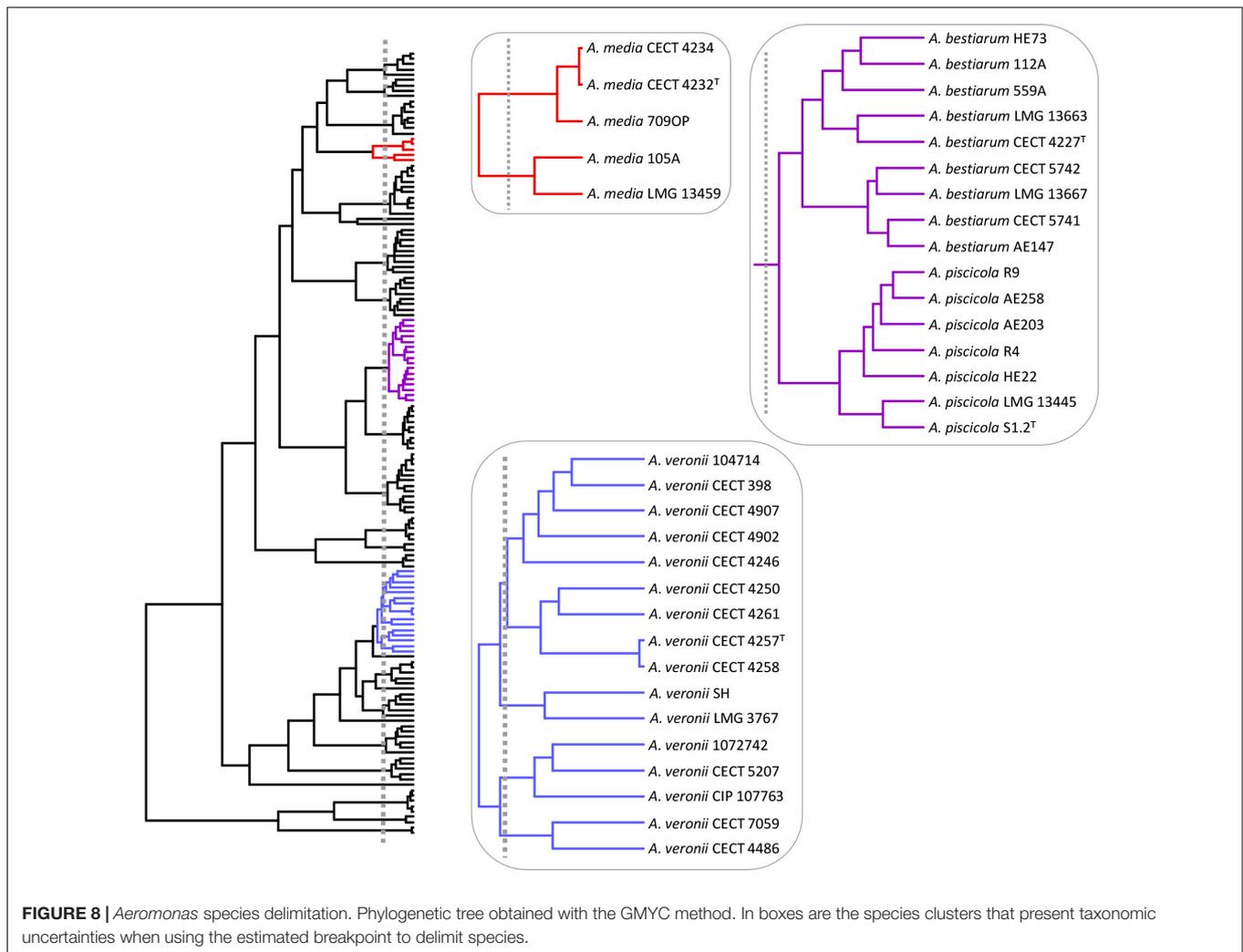
Although Monaghan et al. (2009) recommended the use of the coalescent tree prior instead of the Yule prior for constructing the tree for the GMYC analysis, in our work, in agreement with Talavera et al. (2013), we obtained identical results with the different BEAST options.

Barracough et al. (2009) used the GMYC method to delimit bacterial species based on 16S rRNA sequences, concluding that

TABLE 2 | Diversification models.

Models	AIC ^a							
	Pure birth	Birth-death	DDX	DDL	Yule-2-rates	Yule-3-rates	Yule-4-rates	Yule-5-rates
	150.55	151.83	152.21	152.55	152.80	152.30	153.02	155.56
ΔAIC_{RC} test ^b	Best constant model	Best variable model	ΔAIC_{RC}	<i>P</i> value ^c	Best model		λ (ML)	
	Pure birth	DDX	–1.6569	0.753	Pure birth (Yule)		0.017	

^a AIC, Akaike Information Criterion. ^b ΔAIC_{RC} test (Rabosky, 2006). ^c *P* value obtained by simulation of 5,000 phylogenies.



the 16S rRNA is a rather conservative molecule for surveying bacterial diversity and cannot be used to distinguish putative species. They suggest that studies using sequences of multiple genes or whole genomes of individuals would be more useful. Accordingly, we applied the GMYC method to delimit species in the genus *Aeromonas*, and the results corroborate the species previously described with other methods (phenotypic analysis, MLSA, genome sequence analysis) for this bacterial genus. The resulting tree exhibited the branching rate pattern of the species-to-population transition. Only the controversial *A. veronii* and *A. media* species complexes, the *A. piscicola*/*A. bestiarum* group, and two of the most recently proposed species were not entirely resolved.

The *A. veronii* group includes two biovars, *A. veronii* Veronii and *A. veronii* Sobria, as well as two other *Aeromonas* species considered as synonymous, *A. culicicola* (Huys et al., 2005) and *A. ichthiosmia* (Huys et al., 2001). When we analyzed the *A. veronii* species complex, four groups of strains were revealed: a main cluster consisting of 9 strains, including the two biovars, *A. veronii* Veronii and *A. veronii* Sobria; a second group with two strains, including *A. ichthiosmia*; a third comprising three

strains, including *A. culicicola*; and the fourth with two strains (Figure 8).

Considering the GMYC results, it would be interesting to review the inclusion of *A. culicicola* and *A. ichthiosmia* in the already controversial *A. veronii* group (Huys et al., 2001; Esteve et al., 2003). In addition, Colston et al. (2014) determined the ANI and *is*DDH values from the genomes of different *Aeromonas* species. When the authors compared the type strain of *A. culicicola* and *A. ichthiosmia* with all the strains defined as *A. veronii*, the *is*DDH values obtained were below the threshold of 70% (69.1–69.6% for *A. culicicola* and 67.4–68.2% for *A. ichthiosmia*) and ANI values close to 96%, suggesting these species are different from *A. veronii*.

Aeromonas media described by Allen et al. (1983) traditionally comprised two hybridization groups: *A. media* HG5A and HG5B (Hänninen and Siitonen, 1995), with the strains LMG 13459 (CDC 0862-83) and CECT 4232 (ATCC 33907) representing both groups (Popoff et al., 1981). This was later corroborated by Küpfer et al. (2006) and more recently by Talagrand-Reboul et al. (2017). Our results reveal the existence of two groups of strains outside the GMYC threshold that determine the

species in *Aeromonas* (Figure 8), which is in accordance with Talagrand-Reboul et al. (2017), who analyzed the relationships of 40 *A. media* strains with the recently proposed species of *A. rivipollensis*. These authors also identified two groups of strains inside the *A. media* cluster with ANIb (ANI calculator and EzGenome) and *isDDH* values between groups of $\leq 94.6\%$ and $\leq 61.2\%$, respectively. These values, which are below the considered threshold for isolates of the same species, support the existence of two different species within this group.

Another controversial GMYC clade grouped together *A. bestiarum/A. piscicola* (Figure 8). Colston et al. (2014) determined 0.04 substitutions per site for this pair of strains, which was clearly lower than those determined for other *Aeromonas* species, such as *A. eucrenophila/A. tecta* (1.0) or *A. schubertii/A. diversa* (0.9). Furthermore, the ANI or *isDDH* values for *A. bestiarum/A. piscicola* were close to the threshold to be considered as members of the same species.

Recently, three new *Aeromonas* species have been formally accepted: *A. finlandensis*, *A. lacus* and *A. aquatica* (Beaz-Hidalgo et al., 2015). Nevertheless, our analysis showed that *A. lacus* and *A. aquatica* should probably not be considered new species (Figure 6). In the case of *A. lacus*, the results obtained are in agreement with the ANI values calculated from the genomes in the original description of these species. When *A. jandaei* was compared with *A. lacus*, the ANI value obtained was 95.38%, similar to the species cutoff of 96%.

The GMYC analysis also reveals that diversification in *Aeromonas* is a constant process and follows a Yule model. This result is independent of the number of sequences per species used, and agrees with a study of macroevolutionary dynamic models in a huge number of eukaryote and prokaryote taxa (Maruvka et al., 2013). Prokaryote populations are less affected by extinction and founder effects than larger less abundant organisms (Lynch and Conery, 2003; Butterfield, 2011). Bacterial species may be described as metapopulations extending over time and evolving separately from other species (Cohan, 2002a,b; Achtman and Wagner, 2008). Although microbial dispersion can be limited by geographical barriers and the resulting physical isolation may influence microbial evolution, due to their small size bacteria generally have an unrestricted capacity for dispersion (Finlay and Clarke, 1999) via a variety of passive mechanisms and over long distances (Fierer, 2008).

REFERENCES

- Achtman, M., and Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* 6, 431–440. doi: 10.1038/nrmicro1872
- Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, L., Distel, D. L., et al. (2004). Fine scale phylogenetic architecture of a complex bacterial community. *Nature* 430, 551–554. doi: 10.1038/nature02649
- Allen, D. A., Austin, B., and Colwell, R. R. (1983). *Aeromonas media*, a new species isolated from river water. *Int. J. Syst. Bacteriol.* 33, 599–604. doi: 10.1099/00207713-33-3-599
- Barracough, T. G., Hughes, M., Ashford-Hodges, N., and Fujisawa, T. (2009). Inferring evolutionary significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biol. Lett.* 5, 425–428. doi: 10.1098/rsbl.2009.0091

CONCLUSION

The GMYC method clearly delineated the species boundaries in *Aeromonas*, even in the controversial groups, such as the *A. veronii* or *A. media* species complexes. Some of these taxonomic uncertainties are due to the use of inappropriate methods for defining species [biochemical tests with miniaturized methods (API) or 16S rRNA sequences]. It would therefore be interesting to revise bacterial species separation using the emerging new analyses based on genome sequences (ANI, *isDDH*). The method presented here, widely used to delimit species in eukaryotes, offers an alternative that is easy to use and fully practicable for all laboratories. In addition, the GMYC method has promising potential for phylogenetic community ecology studies.

AUTHOR CONTRIBUTIONS

JGL and MCF conceived and designed the research. JGL and MF performed the computations. JGL, MCF, and MF analyzed the data and discussed the results. All authors contributed to and revised the final manuscript version. They all approved the final version to be published.

FUNDING

This work was supported by a grant from the University of Barcelona for Open Access Publishing in scientific journals.

ACKNOWLEDGMENTS

We thank Dr. Ariadna Sanglas for her invaluable help and contribution of sequences to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.00770/full#supplementary-material>

- Barracough, T. G., and Nee, S. (2001). Phylogenetics and speciation. *Trends Ecol. Evol.* 16, 391–399. doi: 10.1016/S0169-5347(01)02161-9
- Beaz-Hidalgo, R., Latif-Eugenin, F., Hossain, M. J., Berg, K., Niemi, R. M., Rapala, J., et al. (2015). *Aeromonas aquatica* sp. nov., *Aeromonas finlandensis* sp. nov. and *Aeromonas lacus* sp. nov. isolated from Finnish waters associated with cyanobacterial blooms. *Syst. Appl. Microbiol.* 38, 161–168. doi: 10.1016/j.syapm.2015.02.005
- Beaz-Hidalgo, R., Martínez-Murcia, A., and 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. doi: 10.1016/j.syapm.2012.12.007
- Butterfield, N. J. (2011). Animals and the invention of the Phanerozoic Earth system. *Trends Ecol. Evol.* 26, 81–87. doi: 10.1016/j.tree.2010.11.012

- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334
- Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61, 1–36. doi: 10.18637/jss.v061.i06
- Chen, W. C. (2010). *Phylogenetic Clustering with R Package Phyclust. Phyloclustering-Phylogenetic Clustering Website*. Available at: <https://CRAN.R-project.org/package=phyclust>
- Chen, W. C. (2011). *Overlapping Codon Model, Phylogenetic Clustering, and Alternative Partial Expectation Conditional Maximization Algorithm*. Ph.D. dissertation, Iowa State University, Ames, IA.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regression. *Econometrica* 28, 591–605. doi: 10.2307/1910133
- Cohan, F. M. (2002a). What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487. doi: 10.1146/annurev.micro.56.012302.160634
- Cohan, F. M. (2002b). Sexual isolation and speciation in bacteria. *Genetica* 116, 359–370. doi: 10.1023/A:1021232409545
- Colston, S. M., Fullmer, M. S., Beka, L., Lamy, B., Gogarten, J. P., and Graf, J. (2014). Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. *mBio* 5:e02136. doi: 10.1128/mBio.02136-14
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109
- Drummond, A. J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214. doi: 10.1186/1471-2148-7-214
- Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973. doi: 10.1093/molbev/mss075
- Esteve, C., Valera, L., Gutiérrez, C., and Ventosa, A. (2003). Taxonomic study of sucrose-positive *Aeromonas jandaei*-like isolates from faeces, water and eels: emendation of *A. jandaei* Carnahan et al. 1992. *Int. J. Syst. Evol. Microbiol.* 53, 1411–1419. doi: 10.1099/ijs.0.02504-0
- Ezard, T., Fujisawa, T., and Barraclough, T. (2014). *Splits: Species' Limits by Threshold Statistics. R Package Version 1.0–19*. Available at: <http://R-Forge.R-project.org/projects/splits/>
- Fierer, N. (2008). “Microbial biogeography: patterns in microbial diversity across space and time,” in *Accessing Uncultivated Microorganisms: From the Environment to Organisms and Genomes and Back*, ed. K. Zengler (Washington DC: ASM Press), 95.
- Figueras, M. J., Soler, L., Chacón, M. R., Guarro, J., and Martínez-Murcia, A. J. (2000). Extended method for discrimination of *Aeromonas* spp. by 16S rDNA RFLP analysis. *Int. J. Syst. Evol. Microbiol.* 50, 2069–2073. doi: 10.1099/00207713-50-6-2069
- Finlay, B. J., and Clarke, K. J. (1999). Ubiquitous dispersal of microbial species. *Nature* 400:828.
- Fontaneto, D., Herniou, E. A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., et al. (2007). Independently evolving species in asexual bdelloid rotifers. *PLoS Biol.* 5:e87. doi: 10.1371/journal.pbio.0050087
- Fontaneto, D., Tang, C. Q., Obertegger, U., Leasi, F., and Barraclough, T. G. (2012). Different diversification rates between sexual and asexual organisms. *Evol. Biol.* 39, 262–270. doi: 10.1007/s11692-012-9161-z
- Fordyce, J. A. (2010). Interpreting the gamma statistic in phylogenetic diversification rate studies: a rate decrease does not necessarily indicate an early burst. *PLoS One* 5:e11781. doi: 10.1371/journal.pone.0011781
- Fujisawa, T., and Barraclough, T. G. (2013). Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Syst. Biol.* 62, 707–724. doi: 10.1093/sysbio/syt033
- Hänninen, M. L., and Siitonen, A. (1995). Distribution of *Aeromonas* phenospecies and genospecies among strains isolated from water, foods or from human clinical samples. *Epidem. Infect.* 115, 39–50. doi: 10.1017/S0950268800058106
- Hudson, R. R. (1991). Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* 7, 1–44.
- Huys, G., Cnockaert, M., and Swings, J. (2005). *Aeromonas culicicola* Pidiyar et al. 2002 is a later subjective synonym of *Aeromonas veronii* Hickman-Brenner et al. 1987. *Syst. Appl. Microbiol.* 28, 604–609. doi: 10.1016/j.syapm.2005.03.012
- Huys, G., Kämpfer, P., and 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. *Syst. Appl. Microbiol.* 24, 177–182. doi: 10.1078/0723-2020-00038
- Janda, J. M., and Abbott, S. L. (2010). The genus *Aeromonas*: taxonomy, pathogenicity, and infection. *Clin. Microbiol. Rev.* 23, 35–73. doi: 10.1128/CMR.00039-09
- Küpfer, M., Kuhnert, P., Korczak, B. M., Peduzzi, R., and Demarta, A. (2006). Genetic relationships of *Aeromonas* strains inferred from 16S rRNA, gyrB and rpoD gene sequences. *Int. J. Syst. Evol. Microbiol.* 28, 604–609.
- Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G., et al. (2008). DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2923–2928. doi: 10.1073/pnas.0709936105
- Lorén, J. G., Farfán, M., and 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. doi: 10.1371/journal.pone.0088805
- Lynch, M., and Conery, J. S. (2003). The origins of genome complexity. *Science* 302, 1401–1404. doi: 10.1126/science.1089370
- Martin, A. P., Costello, E. K., Meyer, A. F., Nemergut, D. R., and Schmidt, S. K. (2004). The rate and pattern of cladogenesis in microbes. *Evolution* 58, 946–955. doi: 10.1111/j.0014-3820.2004.tb00429.x
- Martin-Carnahan, A., and Joseph, S. W. (2005). “Genus I. *Aeromonas* stanier 1943, 213AL,” in *Bergey's Manual of Systematic Bacteriology*, Vol. 2, eds G. M. Garrity, D. J. Brenner, N. R. Krieg, and J. T. Staley (New York, NY: Springer), 557–578.
- Maruvka, Y. E., Shnerb, N. M., Kessler, D. A., and Ricklefs, R. E. (2013). Model for macroevolutionary dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2460–E2469. doi: 10.1073/pnas.1220014110
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequences-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60
- Monaghan, M. T., Wild, R., Elliot, M., Fujisawa, T., Balke, M., Inward, D. J. G., et al. (2009). Accelerated species inventory on Madagascar using coalescent-based models of species delimitation. *Syst. Biol.* 58, 298–311. doi: 10.1093/sysbio/syp027
- Morlon, H., Kems, B. D., Plotkin, J. B., and Brisson, D. (2012). Explosive radiation of a bacterial species group. *Evolution* 66, 2577–2586. doi: 10.1111/j.1558-5646.2012.01598.x
- Ochman, H., and Wilson, A. C. (1987a). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* 26, 74–86.
- Ochman, H., and Wilson, A. C. (1987b). “Evolutionary history of enteric bacteria,” in *Escherichia coli and Salmonella typhimurium: Molecular and Cellular Aspects*, eds F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (Washington, DC: ASM Publications), 1649–1654.
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., et al. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55, 595–609. doi: 10.1080/10635150600852011
- Popoff, M. Y., Coynault, C., Kiredjian, M., and Lemelin, M. (1981). Polynucleotide sequence relatedness among motile *Aeromonas* species. *Curr. Microbiol.* 5, 109–114. doi: 10.1007/BF01567430
- Powell, J. R. (2012). Accounting for uncertainty in species delineation during the analysis of environmental DNA sequence data. *Methods Ecol. Evol.* 30, 1–11. doi: 10.1111/j.2041-210X.2011.00122.x
- Pybus, O. G., and Harvey, P. H. (2000). Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. Biol. Sci.* 267, 2267–2272. doi: 10.1098/rspb.2000.1278
- Rabosky, D. (2009). *LASER: Likelihood Analysis of Speciation/Extinction Rates from Phylogenies. R Package Version 2.3*. Available at: <https://cran.r-project.org/src/contrib/Archive/laser>

- Rabosky, D. L. (2006). Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60, 1152–1164. doi: 10.1111/j.0014-3820.2006.tb01194.x
- Rambaut, A., Suchard, M. A., Xie, D., and Drummond, A. J. (2014). *Tracer v1.6*. Available at: <http://beast.bio.ed.ac.uk/Tracer>
- Richter, M., and Roselló-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Roger, F., Marchandin, H., Jumas-Bilak, E., Kodjo, A., ColBVH study group, and Lamy, B. (2012). Multilocus genetics to reconstruct aeromonad evolution. *BMC Microbiol.* 12:62. doi: 10.1186/1471-2180-12-62
- Sanglas, A., Albarral, V., Farfán, M., Lorén, J. G., and Fusté, M. C. (2017). Evolutionary roots and diversification of the genus *Aeromonas*. *Front. Microbiol.* 8:127. doi: 10.3389/fmicb.2017.00127
- Schloss, P. D., and Handelsman, J. (2006). Toward a census of bacteria in soils. *PLoS Comput. Biol.* 2:e92. doi: 10.1371/journal.pcbi.0020092
- Silver, A. C., Williams, D., Faucher, J., Horneman, A. J., Gogarten, J. P., and Graf, J. (2011). Complex evolutionary history of the *Aeromonas veronii* group revealed by host interaction and DNA sequence data. *PLoS One* 6:e16751. doi: 10.1371/journal.pone.0016751
- Talagrand-Reboul, E., Roger, F., Kimper, J. L., Colston, S. M., Graf, J., Ltif-Eugenin, F., et al. (2017). Delineation of taxonomic species within complex of species: *Aeromonas media* and related species as a test case. *Front. Microbiol.* 8:621. doi: 10.3389/fmicb.2017.00621
- Talavera, G., Dinca, V., and Vila, R. (2013). Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods Ecol. Evol.* 4, 1101–1110. doi: 10.1111/2041-210X.12107
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tang, C. Q., Humphreys, A. M., Fontaneto, D., and Barraclough, T. (2014). Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods Ecol. Evol.* 5, 1086–1094. doi: 10.1111/2041-210X.12246
- Vinuesa, P., Silva, C., Werner, D., and 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. *Mol. Phylogenet. Evol.* 34, 29–54. doi: 10.1016/j.ympev.2004.08.020
- Wickham, H. (2016). *ggplot: Elegant Graphics for Data Analysis*, 2nd Edn. New York, NY: Springer-Verlag. doi: 10.1007/978-3-319-24277-4
- Xia, X. (2013). DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* 30, 1720–1728. doi: 10.1093/molbev/mst064

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Lorén, Farfán and Fusté. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.