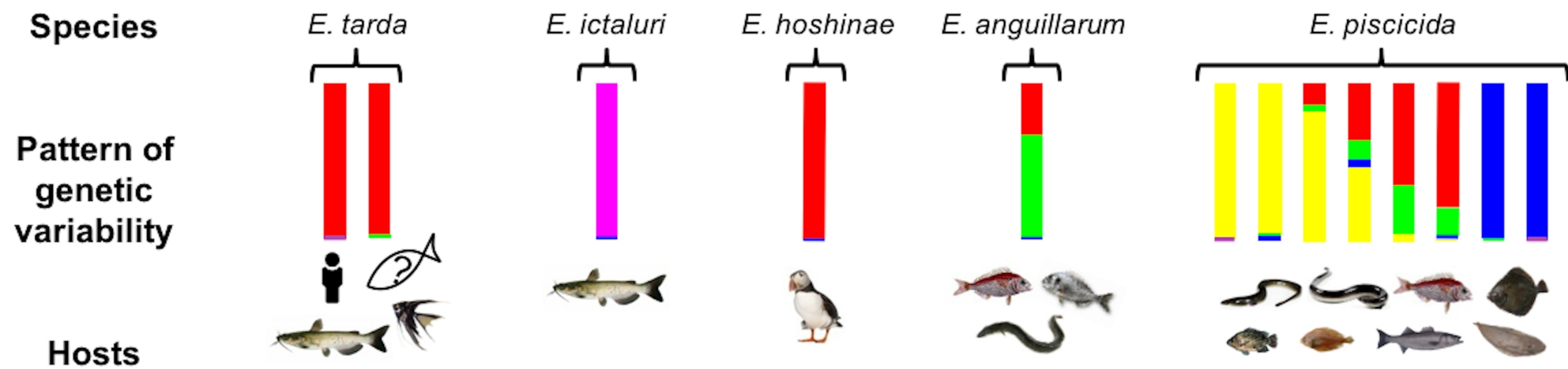
Highlights:

- Analyses performed by ClonalFrame showed strong relation among *Edwardsiella* species.
- *Edwardsiella* population presents a panmictic structure.
- The mutation is more frequent than recombination in the *Edwardsiella* population.
- Population demographic studies of *Edwardsiella* showed a possible recent expansion.



Species
E. tarda
E. ictaluri
E. hoshinae
E. anguillarum
E. piscicida
Pattern of genetic variability
Hosts

Population Genetic and Evolution Analysis of Controversial Genus *Edwardsiella* by Multilocus Sequence Typing

Noemí Buján*, Sabela Balboa, Jesús L. Romalde, Alicia E. Toranzo, Beatriz Magariños

Departamento de Microbioloxía e Parasitoloxía, CIBUS-Facultade de Bioloxía and Instituto de Acuicultura. Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

Corresponding autor:
Departamento de Microbioloxía e Parasitoloxía, CIBUS-Facultade de Bioloxía, Universidade de Santiago de Compostela, Campus Vida, 15782, Santiago de Compostela, Spain. Tel: +34881816910; Fax: +34881816966; E-mail: noemi.bujan@usc.es

Keywords: Pathogen, *Edwardsiella*, MultiLocus, Evolution, Population.

Abstract
At present, the genus *Edwardsiella* compiles five species: *E. tarda*, *E. hoshinae*, *E. ictaluri*, *E. piscicida* and *E. anguillarum*. Some species of this genus such us *E. ictaluri* and *E. piscicida* are important pathogens of numerous fish species. With the description of the two latter species, the phylogeny of *Edwardsiella* became more complicated. With the aim to clarify the relationships among all species in the genus, a multilocus sequence typing (MLST) approach was developed and applied to characterize 56 isolates and 6 reference strains belonging to the five *Edwardsiella* species. Moreover, several analyses based on the MLST scheme were performed to investigate the evolution within the genus, as well as the influence of recombination and mutation in the speciation. *Edwardsiella* isolates presented a high genetic variability reflected in the fourteen sequence types (ST) represented by a single isolates out of eighteen total ST. Mutation events were considerably more frequent than recombination, although both approximately equal influenced the genetic diversification. However, the speciation among species occurred mostly by recombination. *Edwardsiella* genus displays a non-clonal population structure with some degree of geographical isolation followed by a population expansion of *E. piscicida*. A database from this study was created and hosted on pubmlst.org (http://pubmlst.org/edwardsiella/).

1. Introduction
The genus *Edwardsiella* was described in 1965 by Ewing *et al.* to define a distinct taxon within the family *Enterobacteriaceae* (Abbott and Janda, 2006). The species belonging to this genus can be found in many different environmental niches and associated to a high variety of animals including fish, reptiles, amphibians, chickens and other warm-blooded animals (Abbott and Janda, 2006; Mohanty and Sahoo, 2007). *Edwardsiella* is also responsible for some infections of humans causing gastroenteritis presented as acute watery diarrhea (Leung et al., 2012). The most common cause for this zoonosis is the exposition to the causal agent during the fishing or working offshore (Diaz, 2014) since water environment is a common niche of members of *Edwardsiella* (Du, 2007). At present, the genus compiles 5 species: *Edwardsiella tarda*, *Edwardsiella hoshinae*, *Edwardsiella ictaluri*, *Edwardsiella piscicida* and *Edwardsiella anguillarum*. *E. tarda* (Ewing et al., 1965), the best studied taxon in this genus, exhibits a broad geographical distribution and host range. *E. hoshinae* (Grimon et al., 1980) was associated to birds

51 and reptiles and *E. ictaluri* (Hawke et al., 1981) is an important pathogen of cultured
52 channel catfish (*Ictalurus punctatus*). The description of *E. piscicida* (Abayneh et al.,
53 2012) and *E. anguillarum* (Shao et al., 2015) resulted from a reclassification of diverse
54 isolates previously identified as *E. tarda*. *E. piscicida* is, at present, the responsible of
55 major mortalities in different fish species (Camus et al., 2016; Fogelson et al., 2016;
56 Reichley et al., 2016; Shafiei et al., 2016).
57 Molecular typing techniques as Random Amplification of Polymorphic DNA (RAPD),
58 pulsed-field gel electrophoresis (PFGE), Enterobacterial Repetitive Intergenic
59 Consensus (ERIC-PCR), Repetitive Extragenic Palindromic sequence (REP-PCR) and
60 Amplified Fragment Length Polymorphism (AFLP) have been used to study the genetic
61 variability of the *Edwardsiella* species (Castro et al., 2011b; Griffin et al., 2013; Sakai
62 et al., 2009; Maiti et al., 2009; Yang et al., 2013). These techniques allowed the
63 discrimination of the different species of the genus and have shown a low intraspecific
64 genetic diversity among isolates of *Edwardsiella*.
65 To understand the evolutionary history, epidemiology and population dynamics of
66 pathogens is important to study the changes at genetic level. The MultiLocus Sequence
67 Typing (MLST) is considered as a highly sensitive and reproducible method in
68 epidemiological studies of pathogenic bacteria and is widely used as genotyping
69 procedure (Maiden, 2006). With regard to other molecular techniques as RAPD or
70 PFGE, MLST has the advantage of relying on sequence data that can easily be stored in
71 databases, compared across experiments, and progressively enriched by the addition of
72 new isolates. Moreover, MLST provides knowledge about the diversity and the
73 evolutionary relationships among the members of a population. Previously, MLST
74 analysis was performed to investigate the genetic diversity of *E. tarda* isolates obtained
75 from different geographical regions and hosts in China (Yang et al., 2013) using seven
76 housekeeping genes with approximately the same evolutionary rate. Due to this fact, in
77 this work we developed a new MLST scheme, on the basis of sequence polymorphisms
78 within internal fragments of ten housekeeping genes with different evolutionary rates
79 and encompassing representatives of the five *Edwardsiella* species, with the aim to
80 investigate the evolutionary relationships among members of the genus.
81
82 **2. Material and methods**
83 **2.1. Bacterial isolates**
84 A total of 56 isolates and 6 reference strains of *Edwarsiella*, including representative
85 isolates belonging to the five species of this genus were used in this work (Table 1).
86 These isolates were isolated from animals, except the *E. tarda* type strain that was
87 isolated from human feces, and were temporally and geographically diverse. Bacteria
88 were grown in trypticase soy agar supplemented with 1% NaCl (TSA-1, Pronadisa) at
89 25ºC for 24h. All isolates were previously identified by classical biochemical tests and
90 by 16S rRNA gene sequencing (Buján et al., 2017; Castro et al., 2006, 2012). All gene
91 sequences were retrieved from the complete genomes of FL6-60 (NC_017309), C07-
92 087 (NC_020796), EIB202 (NC_013508) and 93-146 (NC_012779.2) strains housed in
93 the GenBank database.
94
95 **2.2. DNA purification, gene amplification and sequencing**
96 Total bacterial DNA was extracted from pure cultures using the Insta-gene matrix (Bio-
97 Rad, Madrid, Spain) following the manufacturer's instructions. Ten genes were
98 selected, based on previous works referenced in Table 2, for MLST analysis, including
99 *adk* (adenylate kinase), *atp*D (β subunit of ATPase), *dna*J (heat shock protein 40), *gap*A
100 (glyceraldehyde-3-phosphodehydrgenase), *gln*A (glutamine synthetase), *hsp*60 (60-KDs

101 heat shock protein), *pho*R (phosphate regulon sensor protein), *pyr*G (cytidine
102 triphosphate synthase), *rpo*A (α subunit of RNA polymerase) and *tuf* (elongation factor
103 Tu). A final concentration of 0.25 μM was used for each primer in the reaction mixture.
104 The PCR reaction mixture consisted in 12,5 μl of PCR Master Mix (2X) (Thermo
105 Scientific), 9,5 μl of nuclease-free water, 1 μl of each couple of primers and 1 μl of
106 template DNA (1000 ng/μl). The primer sequences are described in Table 2 and PCR
107 general conditions were as follows: 94ºC for five min, followed by 35 amplification
108 cycles, 94ºC for 45 s, annealing at changes temperatures (from 45ºC to 60ºC) for 45 s,
109 72ºC for one min, and a final elongation at 72ºC for five min. The purified amplified
110 fragments were direct Sanger sequenced by Stabvida Lcd. (Portugal).
111

112 **2.3. Genetic analyses and population structure**

113 According to MLST standards (Maiden et al., 1998; Urwin and Maiden, 2003),
114 particular allele sequences at particular locus were denominated allele type (AT), and
115 each different combination of ATs was assigned to a distinct sequence type (ST).
116 A descriptive analysis of diversity was performed for the individual genes and
117 concatenated sequences (*adk*, *atp*A, *dna*J, *gap*A, *gln*A, Y-*hsp*60, *pho*R, *pyr*G, *rpo*A,
118 *tuf*). G+C content, number of polymorphic sites, index of association (I_A) and the ratio
119 mean non-synonymous substitutions per non-synonymous site/mean synonymous
120 substitutions per synonymous site (d_N/d_S ratio) were calculated using the software
121 START2 (http://pubmlst.org/software/analysis/start2/). The standardized index of
122 association (I^S_A) from the allelic profile dataset of the whole collection was calculated
123 (Maynard-Smith et al., 1993).
124 The clonal structure of the population was established using eBURST v3.0 program
125 (http://eburst.mlst.net/v3/instructions/). ClonalFrame v1.1 (Didelot and Falush, 2007)
126 was used to construct a dot graph using a 50% majority rule consensus tree with ten
127 runs of MCMC (Markov Chain Monte Carlo). For phylogenetic analysis, concatenated
128 sequences were aligned and trees were performed for the whole collection with the
129 program MEGA version 7 (Kumar et al., 2016) by Neighbor-joining (NJ) method using
130 Kimura-2-parameters method and by Maximum Likelihood (ML) algorithm using
131 Tamura-Nei model, with 1000 bootstrap pseudoreplicates in both cases. For the ML
132 reconstruction, optimal models of evolution were estimated from nucleotide data using
133 MEGA7 software considering 24 substitution types. The best model was selected using
134 Bayesian Information Criterion (BIC).
135 A network constructed by Median Joining method using the program NETWORK 4.1
136 (Fluxus-engineering) (Bandelt et al., 1999) was displayed to establish the relationship
137 among haplotypes.
138

139 **2.4. Recombination analyses**

140 DnaSP5 software (Librado and Rozas, 2009) was employed to calculate the minimal
141 number of recombination events (R_{min}). Moreover, Pairwise Homoplasy Index (PHI)
142 test for recombination was calculated and Split-trees for each individual locus as well as
143 concatenated sequences were generated by SplitsTree4 (Huson and Bryant, 2006).
144 Structure software (Falush et al., 2003, 2007; Hubisz et al., 2009; Pritchard et al., 2000)
145 was used to identify groups with distinct allele frequencies and to detect strains carrying
146 foreign DNA among isolates. The number of clusters (K) was set from three to seven
147 and all runs were replicated five times. The best model of probability, estimated using a
148 burn-in period of 100 000 interactions and a MCMC=50 000, was at a five groups
149 (K=5).
150

151 **2.5. Clonal genealogy**

152 The clonal genealogy was inferred with program ClonalFrame v1.1 (Didelot and Falush,
153 2007) using 100 000 MCMC. The estimated genealogies were represented in a 50%
154 majority rule consensus tree and the evolutionary events at each node were studied.
155 Different population evolutionary parameters as ρ/θ and r/m were calculated by
156 ClonalFrame v1.1. The parameter ρ/θ indicates the ratio at which recombination and
157 mutation occur (Milkman and Bridges, 1990). The relationship of recombination and
158 mutation (r/m) is the ratio of probabilities that an individual nucleotide was altered
159 either through recombination or point mutation and indicates the relative influence of
160 recombination and mutation on clonal divergence (Guttman and Dykhuizen, 1994).
161

162 **2.6. Demographic history and gene flow**

163 DnaSP5 (Librado and Roza, 2009) and Arlequin v3.5 (Excoffier and Lische, 2010) were
164 employed to calculate the average number of nucleotides per site (π) and the number of
165 segregating sites (θ). In addition, the Tajima´s test (D) (Tajima, 1989), and Fu and Li
166 statistics (F* and D*) (Fu and L, 1993) were used to confirm the hypothesis that
167 mutations are selectively neutral (Kimura, 1983).
168 The haplotype structure based on haplotype frequency distribution was evaluated by
169 Fu´s (F'_S) (Fu, 1997) and Strobeck's (*S*) (Strobeck, 1987) statistics. Ramos-Onsis R_2
170 test (Ramos-Onsis and Rozas, 2002) assesses on the differences among the number of
171 singleton mutations and the average number of nucleotide differences in the population
172 and was used to evaluate the demographic expansion.
173 The population differentiation was analyzed on the basis of differences in allele
174 frequencies using Nei´s G_{ST} (Nei, 1978) by Arlequin v3.5 (Excoffier and Lischer, 2010).
175 The same software was used to determinate the gene flow among the five species by
176 fixation index (F_{ST}) (Wright, 1951) and the estimated migrant number among
177 populations (N_M) based in G_{ST}. Program IBDWS (Jensen et al., 2005) was used to
178 estimate the associations between genetic and geographical distances by isolation by
179 distance (IBD) test. These associations were tested with Mantel test (Mantel, 1967) and
180 the significance of the test was determinated by a permutation test of *n*=1000.
181 Geographical distances were measured using geographical coordinates.
182

183 **3. Results**

184 **3.1. Genetic diversity, clustering and phylogenetic analysis**

185 The 56 isolates and 6 reference strains (Table 1) and ten loci selected (Table 2) to
186 performed the MLST analysis resulted in a concatenated sequence of 6629 bp. Allele
187 size used for this analysis are between 470 (*adk*) and 801 (*tuf*) bp. The number of alleles
188 observed for each locus ranged between 5 (*rpo*A) and 14 (*pyr*G) (Table 3). The locus
189 with more polymorphic sites was *dna*J (131 biallelic, 24 triallelic and 3 tetrallelic sites),
190 in contrast with *rpo*A (30 biallelic and 1 triallelic). The most frequent allele found in all
191 genes, except in *adk* (22.6%), was the allele 3 with a percentage ranging between 58%
192 (*gln*A, *hsp*60, *pho*R and *pyr*G) and 91.7% (*rpo*A). The G+C content mean of these
193 genes varied from 52.5% (*rpo*A) to 62.3% (*pho*R). The concatenate sequences were
194 constituted by 95 different alleles (Table 3), 50 of which were sinapomorfic (i.e., shared
195 and derived) and their average G+C %mol content for each species was 58.14% for *E.*
196 *piscicida*, 57.18% for *E. tarda*, 57.34% for *E. ictaluri*, 56% for *E. hoshinae* and 58.6%
197 for *E. anguillarum*.
198 The nucleotide diversity per site (π) and the number of segregation sites (θ) for the
199 concatenated sequences were 0.046 and 0.047 respectively. Values of these estimators
200 for each individual gene in the whole population are shown in Table 3. Synonymous

201 substitution (d_S) occurred more frequently than non-synonymous substitution (d_N),
202 consequently the ratio d_N/d_S for the concatenated sequence was <1, indicating that the
203 population was subjected to purifying selection.
204 The 95 alleles identified from the 56 isolates and 6 reference strains by MLST (Table 1
205 and 3) were grouped in 18 sequence types (STs). Fourteen STs were represented by a
206 single isolate (singletons), while the other four included 2, 2, 8 and 36 isolates. Two
207 clonal complexes were identified by eBURST (CC1 and CC2) using the most stringent
208 criteria (9/10 shared alleles) but it was not possible to assign an ancestral ST (Fig. 1).
209 Both CC compiled only *E. piscicida* isolates isolated from several hosts and from
210 different geographical areas, being the association among STs and host fish not clear.
211 CC1 comprised 10 isolates obtained from different diseased fish in China, Japan and
212 USA, which were distributed in two STs (ST4 and ST11) while the CC2 was only
213 composed by ST7 and ST8 with one strain each, both isolated from tilapia in Japan. It is
214 interesting to note that all *E. piscicida* isolates from turbot in Europe were comprised in
215 a unique ST (ST3), together with one European isolate from sole. Each isolate of *E.*
216 *anguillarum* constituted a singleton (ST13, ST12 and ST18), establishing a DLV
217 (Double-Locus Variant) relation between ST12 and ST13. On the other hand, the *E.*
218 *ictaluri* isolates were grouped in one ST (ST16). Furthermore, a high number of
219 singletons (14) were identified, which reflects some genetic discontinuity within the
220 population. A dot graph constructed employing a 50% majority rule consensus tree with
221 the STs (see Supplementary Figure 1), suggests the existence of missing intermediates
222 and a strong relation among the different species forming only one putative clonal
223 complex. The existence of these hypothetical intermediates was evidenced by the
224 construction of a Median Joining algorithm network (Fig. 2).
225 The phylogenetic relationship of species within the genus *Edwardsiella* was studied
226 performing a NJ and ML tree (Fig. 3). The tree clearly showed the separation of the five
227 species in two branches. Thus *E. tarda* isolates and the unique isolate of *E. hoshinae*
228 were arranged in an independent branch of the tree. The other branch showed three
229 differentiated clusters, each one corresponding with *E. piscicida*, *E. anguillarum* and *E.*
230 *ictaluri* (Fig. 3).
231
232 **3.2. Recombination and mutation**
233 Evolutionary events among *Edwardsiella* species were analyzed. The PHI test did not
234 indicate significant statistical evidences for recombination in any gene. However, the
235 linkage disequilibrium estimated from allelic profiles for the whole isolates collection
236 with I_A parameter (6.298 (P=0.000)) and standardized I^S_A (0.700 (P=0.000)) statistics
237 were significantly different from zero, indicating that some recombination may also
238 occur. The split graphs based on each gene (see Supplementary Figure 2) or the
239 concatenated sequences (Fig. 4) displayed interconnection network structures
240 suggesting the presence of homoplasies. Moreover, some putative recombination events
241 were detected using DNASP5 software with the minimal recombination events (R_{min})
242 (Table 3). ClonalFrame analysis of the concatenated sequences estimated that mutation
243 was considerably more frequent than recombination (ρ/θ=0.035, 95% CI: 0.014-0.068)
244 but their influence for the genetic diversification were approximately equal (r/m=1.09,
245 95% CI: 0.598-1.741). The clonal genealogy showed that the sequence variation on the
246 oldest branch of the tree (node A) arises by mutation events exclusively (see
247 Supplementary Figure 3), diverging two big groups, *E. tarda*-*E. hoshinae* and *E.*
248 *piscicida*-*E. anguillarum*-*E. ictaluri*. The Median joining network (Fig. 2) showing a
249 high number of mutations between the two mentioned groups, corroborated the
250 importance of mutation events in the node A (see Supplementary Figure 3). The

251 evolutionary force responsible for the divergence among species according to
252 ClonalFrame analyses, is the recombination (intermediate nodes B, C and D), although
253 in the intermediate nodes E and F and in the youngest branches, both kinds of event
254 were present (see Supplementary Figure 3).
255 Five genetic ancestors were identified by Structure software. For each isolate, the
256 average proportion of genetic material derived from each genetic ancestor is shown in
257 Fig. 3. The 62 isolates resulted in 14 different patterns classified into five lineages based
258 on which is the major contributor ancestor (≥65%) (Fig. 3). The species *E. piscicida* is
259 the most diverse encompassing three lineages and the admixture isolate *E. piscicida*
260 NCIMB 14824^T. All isolates from European turbot and one isolate from sole formed
261 one tight genetic group belonging to lineage one. Lineage two contained 11 isolates
262 with different host and geographical origin. On the other hand, two strains from USA
263 (C07-087 and 9.8) and two from Japan (07BS and KGE7901) obtained from distinct
264 host (catfish, striped bass and tilapia) are comprised in the lineage five which also
265 included the strain of *E. hoshinae* and the four strains of *E. tarda*. Lineage three
266 comprised exclusively the three isolates of *E. anguillarum* with the same degree of
267 admixture. The two *E. ictaluri* isolates were grouped into lineage four.
268
269 **3.3. Gene flow and demographic trends**
270 The gene flow obtained among the isolates (F_{ST}=0.94, *P*=0) showed an independent
271 evolution due to geographic barriers. Moreover, this hypothesis was confirmed by a N_M
272 value of 0.02 indicating that the genetic exchange among subpopulations is almost non-
273 existent.
274 Tajima´s D, Fu and Li´s (F* and D*), and the Ramos-Onsis R_2 neutrality test for the
275 overall population rejected the neutral theory, although some incongruences were
276 detected (Table 3). Tajima´s D and F* and D* were negative for concatenated
277 sequences reflecting the presence of an excess of rare polymorphisms, which is typical
278 on a population in expansion. Tajima´s D showed the same result in each individual
279 genes but, D* was positive in *adk*, *atp*D, *gln*A, *pho*R, *pyr*G and *tuf*, meaning that the
280 population can be suffering a bottleneck process. However, only four genes had
281 significant values to confirm the results obtained. The deficiency of alleles suggested by
282 positive values of Fu´s F_S and low values of Strobeck´s S´statistic, supported the
283 existence of a bottleneck event in the population. On the other hand, the Ramos-Onsis
284 R_2 values close to zero showed a population in recent expansion.
285 All isolates were divided in three ranges of distances, 0-10 00 km, 500-2500 km and
286 5000-7000 km with different correlation trends. Spatial dependence haplotype
287 frequencies were only detected in related isolates (0-1000 km) (*P*=0.030) indicating that
288 the allele frequencies are more different at this geographical distance, and suggesting
289 isolation by distance. The correlogram showed that in the second distance range the
290 correlation of strains decreased with the distance, and in the third length the correlation
291 of isolates slightly increased with the distance (see Supplementary Figure 4). On the
292 other hand, the Mantel test (*M*) showed non-significant but negative correlation for the
293 full dataset (Z=15344290.2316, r=-0.0314, p=0.3320), which can be indicative of
294 isolation by distance.
295
296 **4. Discussion**
297 MLST is one of the most useful methods to establish the population structure and
298 evolution of microorganisms. With the aim to infer the population structure and
299 evolutionary aspects of the genus *Edwardsiella*, a total of 62 isolates from diverse host
300 and geographical origins were used to develop a MLST scheme. The high sequence

diversity reflected in π and θ values as well as in a high number of polymorphic sites, is supported by the high number of singletons (14 of 18 STs) into the population obtained by the algorithm eBURST (Ramo et al., 2016). The population presented only 107 non-synonymous changes of the 956 polymorphic sites, indicating a strong selection against amino acid changes, a typical feature of housekeeping genes (Pérez-Losada et al., 2006).

The high number of singletons defined by the eBurst scheme also reflects some genetic discontinuity among the strains. This fact may be due to the stringent algorithm used to define clonal complexes rather than to an incongruence of the data (Didelot and Falush, 2007). More complex analyses performed by the construction of the dot graph, showed a tight relation among the species of the genus *Edwardsiella* despite of the existence of missing intermediates in the population structure. On the other hand, it is important to emphasize that the number of isolates available for the recently described species *E. piscicida* (Abayneh et al., 2012) and *E. anguillarum* (Shao et al., 2015) is limited, and this could cause a perturbation in the genetic structure (Balboa et al., 2014). The two clonal complexes of *Edwardsiella* described in this study were composed only by Asiatic *E. piscicida* isolates. The ancestral ST could not be defined, however, hypothetically and based on our studies, the first description of edwardsiellosis caused by *E. piscicida* would have been in 1981 in Japan with the isolation of the isolates EDK1 and E-11-2 (originally identified as *E. tarda*) from diseased eels. Later the infection would disseminate to China being the outbreaks very common in the 90´s causing high number of mortalities in various species of fish (Mohanty and Sahoo, 2007).

The phylogenetic analysis, on the basis of the concatenated sequences, showed each species in a separated monophyletic branch, which was consistent with a recent taxonomic study on the genus *Edwardsiella* (Buján et al., 2017). NJ tree showed two big groups differentiated within the species *E. piscicida*, one of them grouping European turbot isolates and sole, and the other including the rest of *E. piscicida* isolates suggesting that the changes in the sequences can be caused by geographical isolation. Abayneh et al. (2012) reported that the MultiLocus Sequence Analysis (MLSA) approach is capable to resolved isolates according their geographical origin or host. However, ours results by MLST analysis did not allow establish any association.

The bacteria belonging to the family *Enterobacteriaceae* present a high level of genotypic diversity supported by a high recombination/mutation rate (Feil et al., 2001; Salerno et al., 2007). In this study, the r/m index of the population showed that the influence of the recombination and the mutation is roughly equal important in the generation of allelic diversity within a locus. The study of nodes of clonal genealogy indicated that the mutational events increased in the youngest branches supporting evidences of genetic discontinuity among the strains. The same analysis recognized the recombination as the main force responsible of speciation in nodes B, C and D. The construction of median joining network revealed a high number of mutations to separate all species, therefore, the high number of mutations together with the I^S_A value showed that, although the impact of recombination is an important evolution factor, it is not enough for allele random association (De las Rivas et al., 2006).

Structure analysis was performed to examine any possible influence of recombination among the species of the population. The data obtained suggested the existence of five genetic ancestors and all isolates presented some degree of admixture, indicating the import of gene sequences from other species. In the genus *Aeromonas*, the different genetic ancestors were associated with each different species (Martino et al., 2011), but

351 in this study, *E. piscicida* comprised isolates from different lineages (one, two and five)
352 and the admixture strain (*E. piscicida* type strain) whilst the lineage 5 was shared by *E.*
353 *tarda*, *E. hoshinae* and *E. piscicida*. The non-association between genetic ancestors and
354 *Edwardsiella* species can be explained by the high number of recombination events
355 showed in the tree generated by ClonalFrame. On the other hand, the pattern of *E.*
356 *anguillarum* is formed by a mixture between two genetic ancestors demonstrating that,
357 the recombination is the main force to create this new species. The high value of I_A and
358 the considerable genetic admixture between species indicate that the structure of
359 population formed by the five species is panmictic (Maynard-Smith et al., 1993).
360 Despite of these results, the presence of clones within the population or the
361 disequilibrium among the alleles of some genes showed an intermediate panmictic
362 population. Our results demonstrated that the clonal and epidemiologic relation among
363 isolates belonging to different species is a complex work and the real impact of the
364 recombination is not easy to evaluate (Didelot and Maiden, 2010).
365 The demographic results and the structure of the population obtained could be affected
366 by the high number of European turbot isolates that were selected in order to be
367 representative of the fast propagation of edwardsiellosis in Europe (Mohanty and Sahoo,
368 2007).
369 The different information reflected by the indexes used for the study of the
370 demographical history of population, showed a controversial model for the evolution
371 (Bastardo et al., 2015). The positive *Fs* and Strobeck's values which were very close to
372 zero, determined that the population suffered a bottleneck. On the other hand, Tajima's
373 *D* and Fu and Li's F* and D* neutrality test for the overall population resulted in
374 negative values, allowing the rejection of the neutral theory in *Edwardsiella* population
375 as a result of relatively recent population expansion. Moreover, Ramos-Onsis R_2
376 statistic suggested also a possible population expansion (Ramos-Onsis and Rozas, 2002)
377 of *Edwardsiella* with values close to zero. This finding could be supported if some
378 groups were in epidemiological expansion inducing the spreading of the more virulent
379 members in the population. Such hypothesis could be connected with the last outbreaks
380 of *E. piscicida* in Europe (Castro et al., 2011a; Castro et al., 2012; Shafiei et al., 2016).
381 These data are reflected in the independent evolution of the two lineage of *E. piscicida*
382 composed by isolates from Europe and Southeast Asia. *Fs* is a more sensitive indicator
383 for population expansion than *D**, but R_2 statistic is particularly suited for small size
384 samples with recombination (Balboa et al., 2014). Moreover, some genes showed a
385 deviation from the general trend of population indicating unique pressures, spurious
386 results, or a mix of factors. These contradictory results made impossible to distinguish
387 between expansion and balancing selection probably due to the different demographic
388 histories among the five species, the high number of isolates of *E. piscicida* belonging
389 to ST3 and the possibility that the population suffered a bottleneck and a more recent
390 expansion. In addition, the linkage disequilibrium of population together with the high
391 number of mutational events in more recent branches supported this hypothesis and
392 indicated that the genetic hitch-hiking process is probably the reason for this recent
393 expansion (Barton 2000).
394 The geographical isolation plays an important role in the evolution of the population.
395 No recurrent gene flow (Nm<1) and the F_{ST} index close to one, explained the bottleneck
396 occurred in the genus and the elevated number of mutations, possibly caused by the
397 geographical isolation. Moran´s index performed to measure the global spatial
398 autocorrelation showed three patterns: from 0 to 1700 km the isolates present a
399 significant random spatial pattern, from 1700 to 2500 km the values suggested that the
400 geographical distance is a factor to increase the genetic distance but is not significant,

and finally, with more than 2500 km, the genetic distance decreases and isolates are more similar. The genetic similarity among isolates from USA and Asiatic countries suggests that these isolates can share an ancestral origin. Due to the high distance among countries, the more reliable explanation is the commerce flow and the acquisition of fish fry in different countries as occurred with other pathogens (Bastardo et al., 2015).

5. Conclusions

The MLST studies support the existence of five species into *Edwardsiella* genus. The basic knowledge of the population indicates that the structure is panmictic with traces of epidemiological expansion. Moreover, the recombination events are more important in some speciation than the mutation indicating the complexities in the study of this genus. The tools employed provide new information on the genetic mechanisms behind the emergence of the five species.

A database from this study was created and was hosted on pubmlst.org (Jolley and Maiden 2010) and is freely available in Internet (http://pubmlst.org/edwardsiella/). This database will be useful in future works studying evolutionary and epidemiological relationships between strains and for comparison of data generate from worldwide laboratories.

Funding

The work was supported by Ministerio de Economía y Competitividad (Spain) [AGL2012-31049, BES-2010-037413] and Xunta de Galicia (Spain) [GRC-2014/007].

References

Abayneh, T., Colquhoun, D.J., Sørum, H., 2012. Multi-locus Sequence Analysis (MLSA) of *Edwardsiella tarda* isolates from fish. Vet. Microbiol. 158, 367-375.

Abbott, S.L., Janda, J.M., 2006. The Genus *Edwardsiella*. Prokaryotes 6, 72-89.

Balboa, S., Bastardo, A., Romalde, J.L., 2014. Disentangling the population structure and evolution of the clam pathogen *Vibrio tapetis*. Microb. Ecol. 67, 145-154.

Bandelt, H.J., Foster, P., Rhöl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16**,** 37-48.

Barton, N.H., 2000. Genetic hitchhiking. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 355, 1553-1562.

Bastardo, A., Ravelo, C., Romalde, J.L., 2012. Multilocus sequence typing reveals high genetic diversity and epidemic population structure for the fish pathogen *Yersinia ruckeri*. Environ. Microbiol. 14, 1888-1897.

Bastardo, A., Ravelo, C., Romalde, J.L., 2015. Phylogeography of *Yersinia ruckeri* reveals effects of past evolutionary events on the current strain distribution and explains

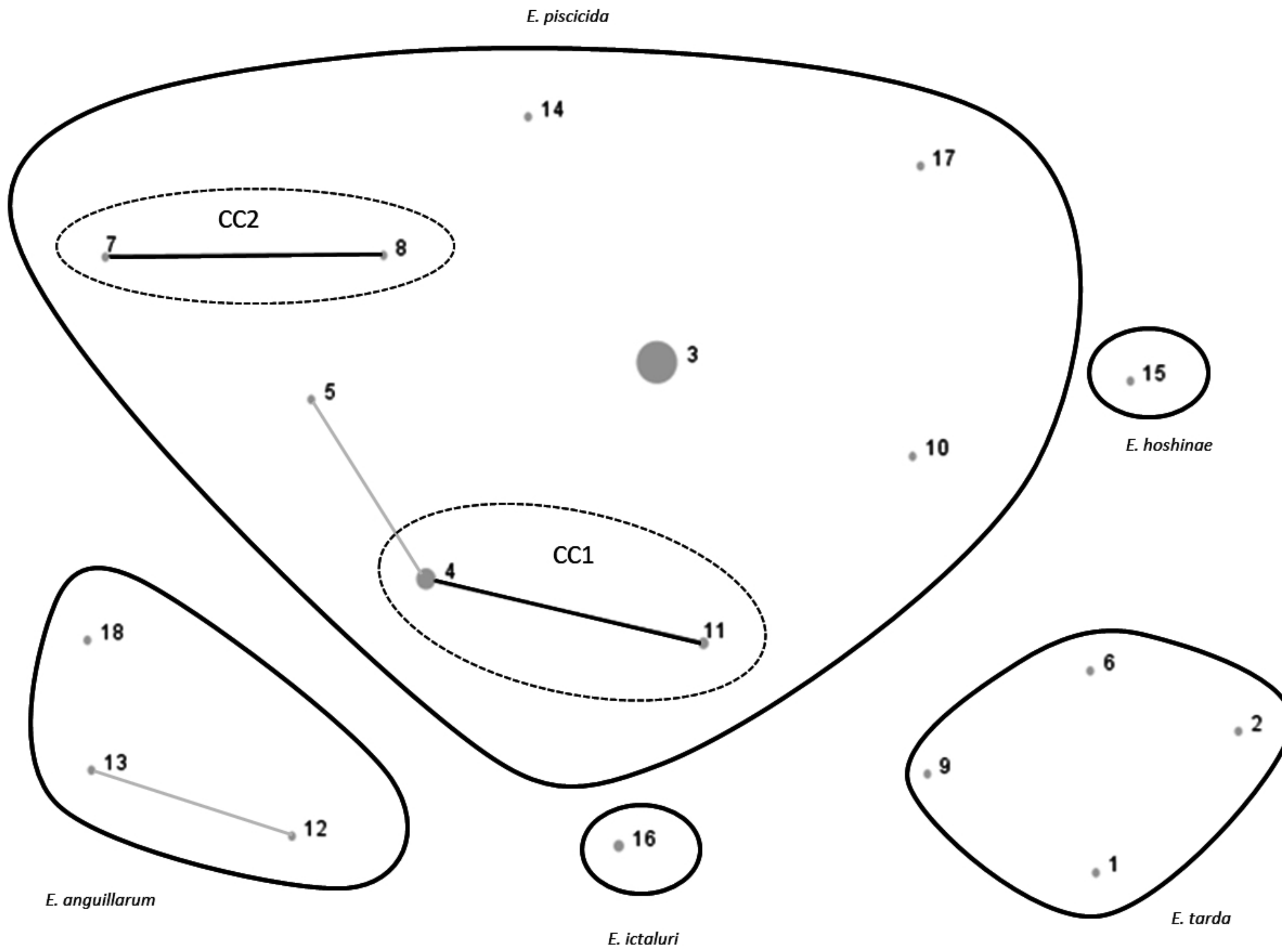
451 variations in the global transmission of enteric redmouth (ERM) disease. Front.
452 Microbiol. DOI: 10.3389/fmicb.2015.01198.
453
454 Buján, N., Mohamed, H., Balboa, S. et al., 2017. Genetic studies to re-affiliate
455 *Edwardsiella tarda* fish isolates to *Edwardsiella piscicida* and *Edwardsiella*
456 *anguillarum* species. Syst. Appl. Microbiol. DOI: 0.1016/j.syapm.2017.09.004.
457
458 Camus, A., Dill, J., McDermott, A. et al., 2016. *Edwardsiella piscicida* associated
459 septicaemia in a blotched fantail stingray *Taeniura meyeni* (Mueller & Henle). J. Fish
460 Dis. 39, 1125-1131.
461
462 Castro, N., Magariños, B., Núñez, S. et al., 2011. Pathogenic potential of *Edwardsiella*
463 *tarda* strains isolated from turbot. Fish. Pathol. 46, 27-30.
464
465 Castro, N., Toranzo, A.E., Barja, J.L. et al., 2006. Characterization of *Edwardsiella*
466 *tarda* strains isolated from turbot, *Psetta maxima* (L.). J. Fish Dis. 29, 541-547.
467
468 Castro, N., Toranzo, A.E., Bastardo, A. et al., 2011b. Intraspecific genetic variability of
469 *Edwardsiella tarda* strains from cultured turbot. Dis. Aquatic. Org. 95, 253-258.
470
471 Castro, N., Toranzo, A.E., Devesa, S. et al., 2012. First description of *Edwardsiella*
472 *tarda* in Senegalese sole, *Solea senegalensis* (Kaup). J. Fish Dis. 35, 79-82.
473
474 De las Rivas, B., Marcobal, A., Muñoz, R., 2006. Development of a multilocus
475 sequence typing method for analysis of *Lactobacillus plantarum* strains. Microbiology.
476 152, 85-93.
477
478 Diaz, J.H., 2014. Emerging causes of superficial and invasive infections following
479 marine injuries and exposures. J. La State Med. Soc. 166, 103-108.
480
481 Didelot, X., Falush, D., 2007. Inference of bacterial microevolution using multilocus
482 sequence data. Genetics. 175, 1251-1266.
483
484 Didelot, X., Maiden, M.C., 2010. Impact of recombination on bacterial evolution.
485 Trends Microbiol. 18, 315-322.
486
487 Du, M., Chen, J., Zhang, X. et al., 2007. Retention of virulence in a viable but non-
488 culturable *Edwardsiella tarda* isolate. Appl. Environ. Microbiol. 73, 1349-1354.
489
490 Ewing, W.H., Mcwhorter, A.C., Escobar, M.R., et al., 1965. *Edwardsiella*, a new genus
491 of Enterobacteriaceae based on a new species, *E. tarda*. Intern. Bull. Bacteriol. Nomen.
492 Taxon. 15, 33-38.
493
494 Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to
495 perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour.
496 10, 564-567.
497
498 Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using
499 Multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 164,
500 1567-1587.

501
502 Falush, D., Stephens, M., Pritchard, J.K., 2007. Inference of population structure using
503 Multilocus genotype data, dominant markers and null alleles. Mol. Ecol. Notes. 7, 574-
504 578.
505
506 Feil, E.J., Holmes, E.C., Bessen, D.E., et al., 2001. Recombination within natural
507 populations of pathogenic bacteria, short-term empirical estimates and long-term
508 phylogenetic consequences. Proc. Natl. Acad. Sci. USA. 98, 182-187.
509
510 Fogelson, S.B., Petty, B.D., Reichley, S.R., et al., 2016. Histologic and molecular
511 characterization of *Edwardsiella piscicida* infection in large-mouth bass (*Micropterus*
512 *salmoides*). J. Vet. Diagn. Invest. 28, 338-344.
513
514 Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth,
515 hitchhiking and background selection. Genetics. 147, 915-925.
516
517 Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics. 133, 693-
518 709.
519
520 Griffin, M.J., Quiniou, S.M., Cody T., et al., 2013. Comparative analysis of
521 *Edwardsiella* isolates from fish in the eastern United States identifies two distinct
522 genetic taxa amongst organisms phenotypically classified as *E. tarda*. Vet. Microbiol.
523 165, 358-372.
524
525 Grimont, P.A.D., Grimont, F., Richard C., et al., 1980. *Edwardsiella hoshinae*, a new
526 species of *Enterobacteriaceae*. Curr. Microbiol. 4, 347-351.
527
528 Guttman, D., Dykhuizen, D., 1994. Clonal divergence in *Escherichia coli* as a result of
529 recombination, not mutation. Science. 266, 1380-1383.
530
531 Hawke, J.P., Mcwhorter, A., Steigerwalt, A.G., Brenner, D.O.N.J., 1981. *Edwardsiella*
532 *ictaluri* sp. nov., the causative agent of enteric septicemia of catfish. Int. J. Syst. Evol.
533 Microbiol. 31, 396-400.
534
535 Hubisz, M., Falush, D., Stephens M., et al., 2009. Inferring weak population structure
536 with the assistance of sample group information. Mol. Ecol. Resour. 9, 1322-1332.
537
538 Huson, D.H., Bryant D., 2006. Application of phylogenetic networks in evolutionary
539 studies. Mol. Biol. Evol. 23, 254-267.
540
541 Jensen, J.L., Bohonak, A.J., Kelley, S.T., 2005. Isolation by distance, web service.
542 BMC Genet. DOI: 10.1186/1471-2156-6-13.
543
544 Jolley, K.A., Maiden, M.C.J., 2010. BIGSdb: scalable analysis of bacterial genome
545 variation at the population level. BMC Bioinformatics. DOI: 10.1186/1471-2105-11-
546 595.
547
548 Kimura, M., 1983. The neutral theory of molecular evolution. Cambridge University
549 Press, Cambridge.
550

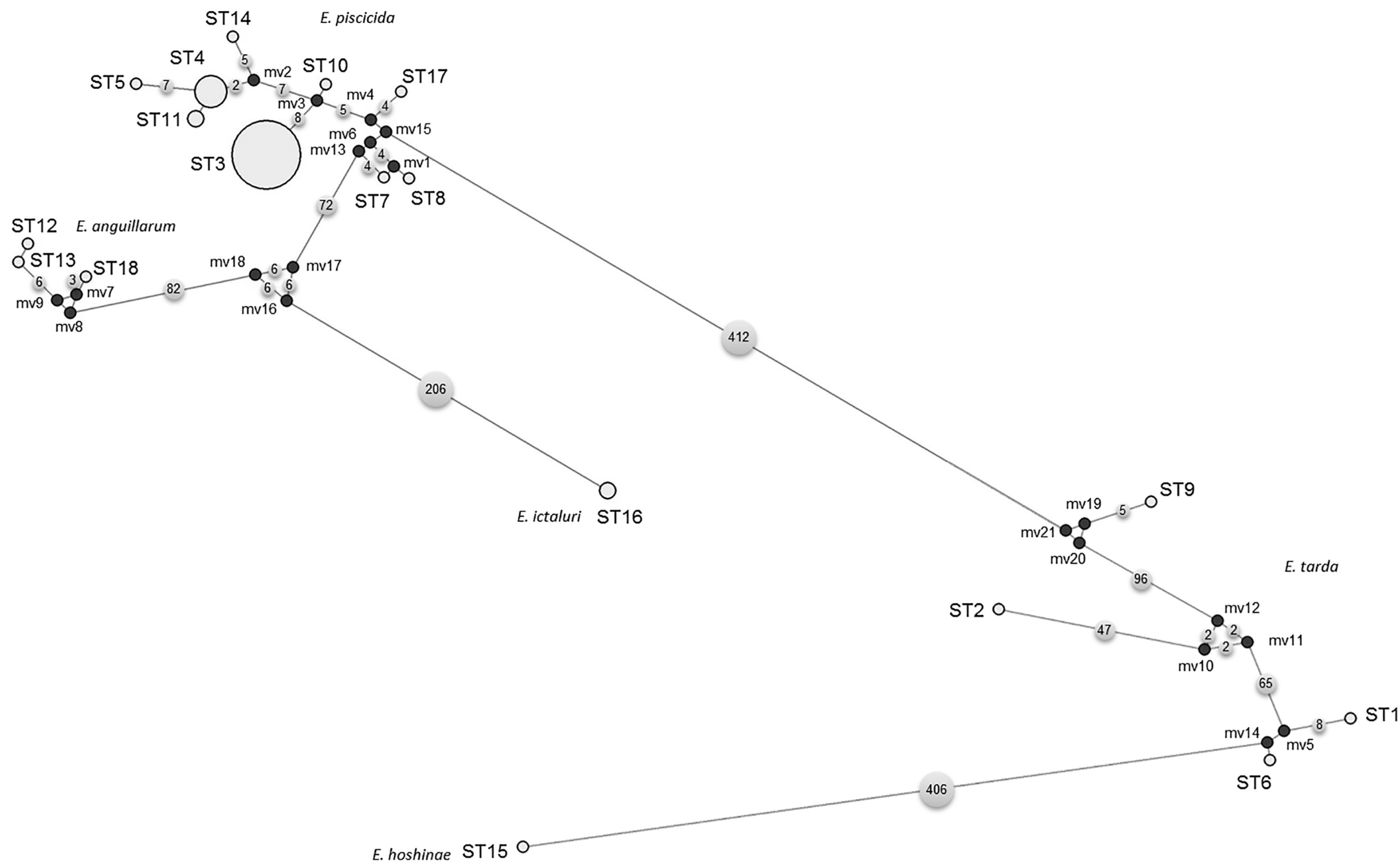
551 Kumar, S, Stecher, G, Tamura, K., 2016. MEGA7: Molecular evolutionary genetics
552 analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870-1874.
553
554 Kotetishvili, M., Kreger ,A., Wauters, G., et al., 2005. Multilocus Sequence typing for
555 studying genetic relationships among *Yersinia* species. J. Clin. Microbiol. 43, 2674-
556 2684.
557
558 Leung, K.Y., Siame, B.A., Tenkink, B.J., et al., 2012. *Edwardsiella tarda* - virulence
559 mechanisms of an emerging gastroenteritis pathogen. Microbes Infect. 14, 26-34.
560
561 Librado, P., Rozas, J., 2009. DnaSP v5, A software for comprehensive analysis of
562 DNA polymorphism data. Bioinformatics. 25, 1451-1452.
563
564 Maiden, M.C., 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol.
565 60, 561-588.
566
567 Maiden, M.C.J., Bygraves, J.A., Feil, E., et al., 1998. Multilocus sequence typing, a
568 portable approach to the identification of clones within populations of pathogenic
569 microorganisms. Proc. Natl. Acad. Sci. USA. 95, 3140-3145.
570
571 Maiti, N.K., Mandal, A., Mohanty, S., et al., 2009. Phenotypic and genetic
572 characterization of *Edwardsiella tarda* isolated from pond sediments. Comp. Immunol.
573 Microbiol. Infect. Dis. 32, 1-8.
574
575 Mantel, N., 1967. The detection of disease clustering and a generalized regression
576 approach. Cancer Res. 27, 209-220.
577
578 Martino, M.E., Fasolato, L., Montemurro, F., et al., 2011. Determination of microbial
579 diversity of *Aeromonas* strains on the basis of multilocus sequence typing, phenotype,
580 and presence of putative virulence genes. Appl. Environ. Microbiol. 77, 4986-5000.
581
582 Maynard-Smith, J., Smith, N.H., et al., 1993. How clonal are bacteria? Proc. Natl. Acad.
583 Sci. USA. 90, 4384-4388.
584
585 Milkman, R., Bridges, M.M., 1990. Molecular evolution of the *Escherichia coli*
586 chromosome. III. clonal frames. Genetics. 126, 505-517.
587
588 Mohanty, B.R., Sahoo, P.K., 2007. Edwardsiellosis in fish: a brief review. J. Biosci. 32,
589 1331-1344.
590
591 Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small
592 number of individuals. Genetics. 89, 583-590.
593
594 Paradis, S., Boissinot, M., Paquette, N., et al., 2005. Phylogeny of the
595 Enterobacteriaceae based on genes encoding elongation factor Tu and F-ATPase beta-
596 subunit. Int. J. Syst. Evol. Microbiol. 55, 2013-2025.
597
598 Pérez-Losada, M., Browne, E.B., Madsen, A., et al., 2006. Population genetics of
599 microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect.
600 Genet. Evol. 6, 97-112.

601
602 Pham, H.N., Ohkusu, K., Mishima, N., et al., 2007. Phylogeny and species
603 identification of the family *Enterobacteriaceae* based on *dna*J sequences. Diagn.
604 Microbiol. Infect. Dis. 58, 153-161.
605
606 Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of Population Structure
607 Using Multilocus Genotype Data. Genetics. 155, 945-995.
608
609 Ramo, A., Monteagudo, L.V., Del Cacho, E., Sánchez-Acedo, C., Quílez, J., 2016.
610 Intra-Species Genetic Diversity and Clonal Structure of *Cryptosporidium parvum* in
611 Sheep Farms in a Confined Geographical Area in Northeastern Spain. PLoS One.
612 DOI:10.1371/journal.pone.0155336.
613
614 Ramos-Onsis, S.E., Rozas, J., 2002. Statistical properties of new neutrality tests against
615 population growth. Mol. Biol. Evol. 19, 2092-2100.
616
617 Reichley, S.R., Waldbieser, G.C., Tekedar, H.C., et al., 2016. Complete genome
618 sequence of *Edwardsiella piscicida* isolate S11-285 recovered from channel catfish
619 (*Ictalurus punctatus*) in Mississippi, USA. Genome Announc.
620 DOI:10.1128/genomeA.01259-16.
621
622 Sakai, T., Yuasa, K., Ozaki, A., et al., 2009. Genotyping of *Edwardsiella ictaluri*
623 isolates in Japan using amplified-fragment length polymorphism analysis. Lett. Appl.
624 Microbiol. DOI:10.1111/j.1472-765X.2009.02686.x.
625
626 Salerno, A., Delétoile, A., Lefevre, M., et al., 2007. Recombining population structure
627 of *Plesiomonas shigelloides* (*Enterobacteriaceae*) revealed by Multilocus Sequence
628 Typing. J. Bacteriol. 189, 7808-7818.
629
630 Sawabe, T., Kita-Tsukamoto, K., Thompson, F.L., 2007. Inferring the evolutionary
631 history of vibrios by means of Multilocus Sequence Analysis. J. Bacteriol. 21, 7932-
632 7936.
633
634 Shafiei, S., Viljamaa-Dirks, S., Sundell, K., et al., 2016. Recovery of *Edwardsiella*
635 *piscicida* from farmed white- fish, *Coregonus lavaretus* (L.), in Finland. Aquaculture.
636 454, 19-26.
637
638 Shao, S., Lai, Q., Liu, Q., et al., 2015. Phylogenomics characterization of a highly
639 virulent *Edwardsiella* strain ET080813(T) encoding two distinct T3SS and three T6SS
640 gene clusters, Propose a novel species as *Edwardsiella anguillarum* sp. nov. Syst.
641 Appl. Microbiol. 38, 36-47.
642
643 Smith, J.M., Smith, N.H., O'Rourke, M., et al., 1993. How clonal are bacteria? Proc.
644 Natl. Acad. Sci. USA. 90, 4384-4388.
645
646 Strobeck, C., 1987. Average number of nucleotide differences in a sample from a single
647 subpopulation, a test for population subdivision. Genetics. 117, 149-153.
648
649 Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA
650 polymorphism. Genetics. 123, 585-595.

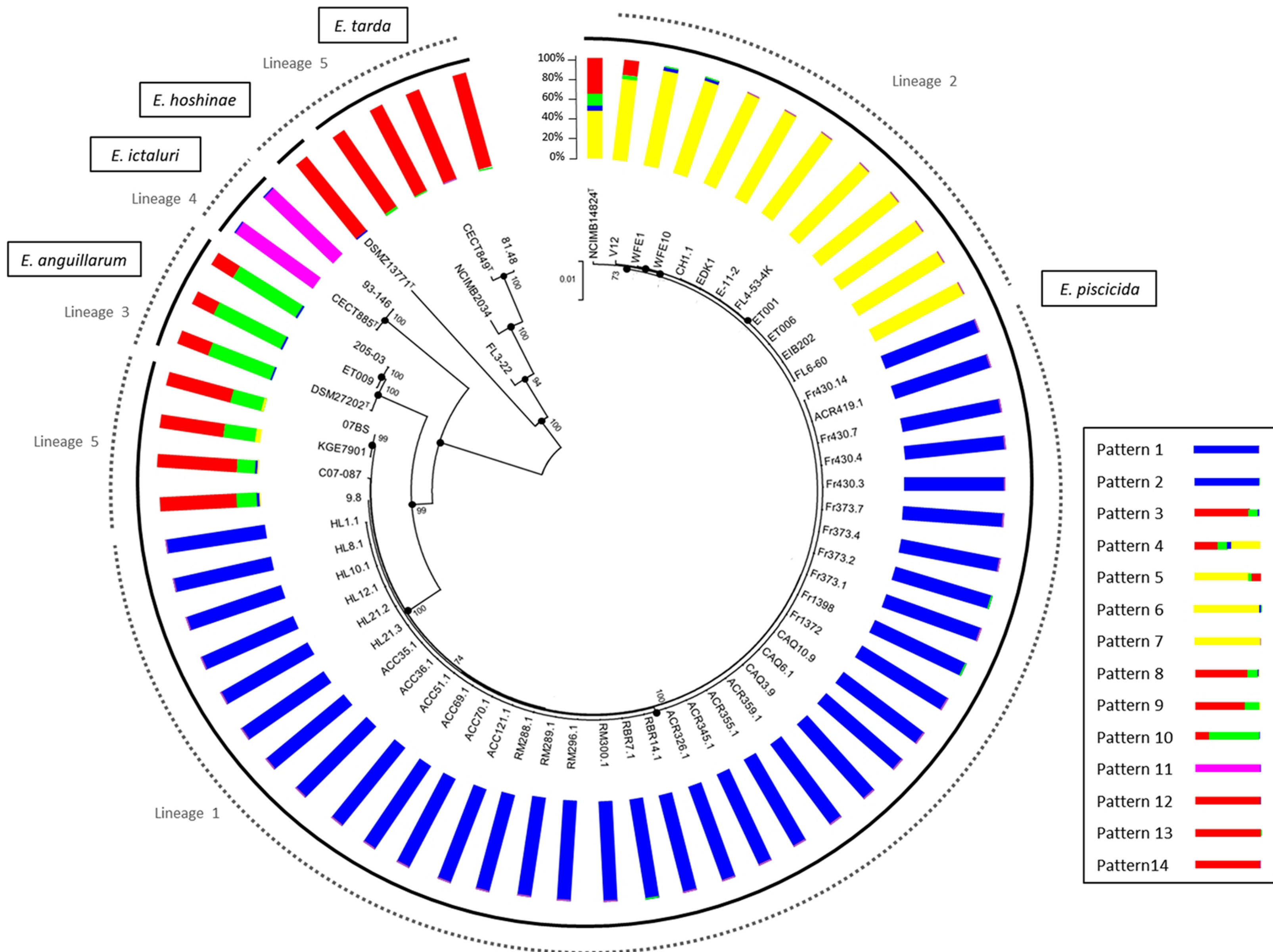
651
652 Thompson, F.L., Gevers, D., Thompson, C.C., et al., 2005. Phylogeny and molecular
653 identification of vibrios on the basis of multilocus sequence analysis. Appl. Environ.
654 Microbiol. 71, 5107-5115.
655
656 Urwin, R., Maiden, M.C., 2003. Multilocus sequence typing: a tool for global
657 epidemiology. Trends Microbiol. 11, 479-487.
658
659 Wright, S., 1951. The genetical structure of populations. Ann. Eugen. 15, 323-354.
660
661 Yang, M., Shao, S., Xiao J., et al., 2013. Phylogenetic investigation of *Edwardsiella*
662 *tarda* with multilocus sequence typing (MLST) and pulsed field gel electrophoresis
663 (PFGE) typing methods. Aquaculture. 410, 79-85.



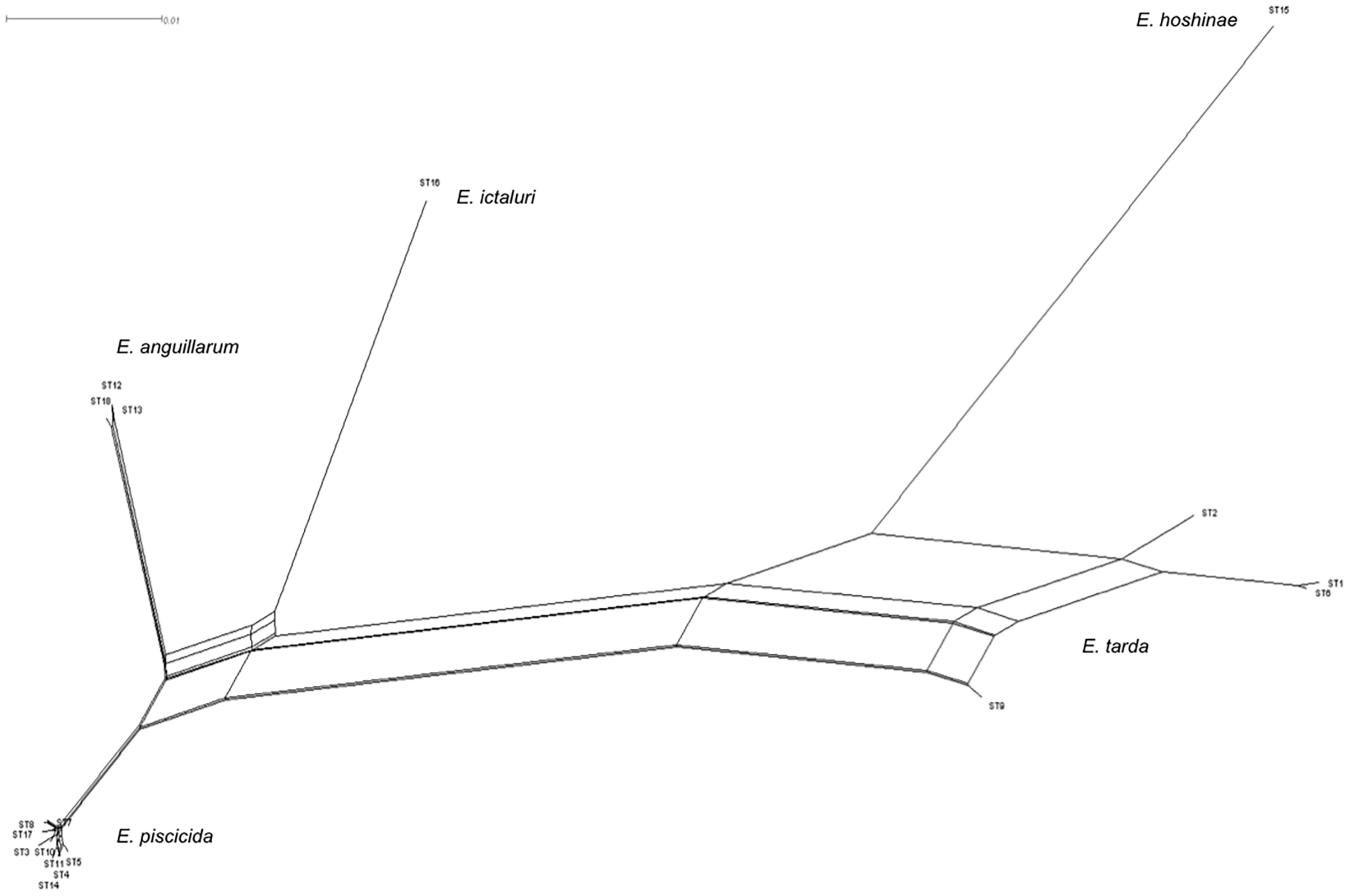
E. piscicida
14
17
CC2
7
8
3
5
10
CC1
4
11
15
E. hoshinae
18
13
12
E. anguillarum
16
E. ictaluri
6
2
9
1
E. tarda



ST14
E. piscicida
ST4
ST10
ST17
ST5
mv2
mv3
mv4
ST11
mv6
mv15
mv13
mv1
ST3
ST7
ST8
ST12
E. anguillarum
ST13
ST18
mv18
mv17
mv7
mv9
mv16
mv8
E. ictaluri
ST16
ST9
mv19
mv21
mv20
E. tarda
ST2
mv12
mv11
mv10
ST1
mv14
mv5
ST6
E. hoshinae
ST15



E. tarda
E. hoshinae
E. ictaluri
E. anguillarum
E. piscicida
Lineage 1
Lineage 2
Lineage 3
Lineage 4
Lineage 5
100%
80%
60%
40%
20%
0%
0.01
Pattern 1
Pattern 2
Pattern 3
Pattern 4
Pattern 5
Pattern 6
Pattern 7
Pattern 8
Pattern 9
Pattern 10
Pattern 11
Pattern 12
Pattern 13
Pattern14


0.01
E. hoshinae
ST15
ST16
E. ictaluri
E. anguillarum
ST12
ST18
ST13
ST2
ST1
ST6
E. tarda
ST9
ST8
ST7
ST17
ST3
ST10
ST11
ST5
ST4
ST14
E. piscicida

1 **Figure legends**
2
3 Fig. 1. Population snapshot of all sequences types (ST) using eBurst program. SLVs
4 (Single Locus Variable) are connected by black lines and DLVs (Doble Locus Variable)
5 by grey lines. Clonal complexes (CC) are labeled with a discontinuous circle. The ST
6 size of circle is proportional to the number of strains within each one.
7
8 Fig. 2. Median joining network of *Edwardsiella* population based on the concatenated
9 sequence (*adk*, *atp*A, *dna*J, *gap*A, *gln*A, Y-*hsp*60, *pho*R, *pyr*G, *rpo*A, *tuf*). The size of
10 the circles is related with the number of strains within each ST. Open circles represent
11 putative intermediate sequences. Number in grey circles indicate the nucleotide
12 substitutions. Lines without number correspond whit one nucleotide substitution. mv,
13 medium vector.
14
15 Fig. 3. Phylogenetic reconstruction based on the concatenated sequences (*adk*, *atp*A,
16 *dna*J, *gap*A, *gln*A, Y-*hsp*60, *pho*R, *pyr*G, *rpo*A, *tuf*) using 62 isolates of genus
17 *Edwardsiella* by Neighbor-Joining algorithm. Only bootstrap values above 70% are
18 shown. Bold circles indicating the coincident nodes in the tree generated by Maximum
19 Likelihood algorithm (TN93+G). Coloured bar represent the five genetic ancestors
20 identified with Structure analysis. The length of the coloured bar indicates the
21 proportion of the genetic ancestor from each five lineages.
22
23 Fig. 4. Neighbor-Net graph based on the concatenated sequences of the housekeeping
24 genes of the 18 STs. Formation of an interconnecting network structure was suggestive
25 of recombination. Scale represents genetic distance.

1 **Table 1.** Origin and typing data of *Edwardsiella* strains employed in this study.

Id PubMLST	Strain	Species	ST (allelic profile)	Host	Country	Year
1	CECT849[T]	*E. tarda*	1 (1,1,1,1,1,1,1,1,1,1)	Human feces	Southeast USA	1959
2	NCIMB2034	*E. tarda*	2 (2,2,2,2,2,2,2,2,2,2)	Unknow fish	Northeast USA	1977
3	HL1.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*Scophthalmus maximus*	West Netherlands	2004
4	HL8.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	West Netherlands	2005
5	HL10.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	West Netherlands	2005
6	HL12.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	West Netherlands	2005
7	HL21.2	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	West Netherlands	2006
8	HL21.3	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	West Netherlands	2006
9	ACC35.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	North Portugal	2005
10	ACC36.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	North Portugal	2005
11	ACC51.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	North Portugal	2006
12	ACC69.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	North Portugal	2006
13	ACC70.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	North Portugal	2006
14	ACC121.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	North Portugal	2009
15	RM288.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2006
16	RM289.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2006
17	RM296.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2006
18	RM300.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2006
19	RBR7.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	North Portugal	2008
20	RBR14.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	North Portugal	2008
21	ACR326.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2009
22	ACR345.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2009
23	ACR355.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2009
24	ACR359.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2009
25	CAQ3.9	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2009
26	CAQ6.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2009
27	CAQ9.10	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest Spain	2009
28	Fr1372	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
29	Fr1398	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
30	Fr373.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
31	Fr373.2	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
32	Fr373.4	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
33	Fr373.7	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
34	Fr430.3	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
35	Fr430.4	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012

36	Fr430.7	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
37	Fr430.14	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*S. maximus*	Northwest France	2012
38	ACR419.1	*E. piscicida*	3 (2,3,3,3,3,3,3,3,3,3)	*Solea senegalensis*	Northwest Spain	2010
39	CH1.1	*E. piscicida*	4 (3,4,4,3,4,4,4,4,3,4)	*S. maximus*	Northeast China	2011
40	ET001	*E. piscicida*	4 (3,4,4,3,4,4,4,4,3,4)	*Paguellus bogaraveo*	East Japan	2002
41	ET006	*E. piscicida*	4 (3,4,4,3,4,4,4,4,3,4)	*P. bogaraveo*	East Japan	2002
42	E-11-2	*E. piscicida*	4 (3,4,4,3,4,4,4,4,3,4)	*A. japonica*	East Japan	1981
43	EDK1	*E. piscicida*	4 (3,4,4,3,4,4,4,4,3,4)	*Anguilla japonica*	East Japan	1981
44	FL4-53-4K	*E. piscicida*	4 (3,4,4,3,4,4,4,4,3,4)	*Morone saxatilis*	Southeast USA	1997
45	V12	*E. piscicida*	5 (3,4,4,3,4,4,5,5,3,4)	*A. anguilla*	Southeast Spain	2003
46	81.48	*E. tarda*	6 (1,1,1,4,1,5,6,6,1,2)	*Ictalurus punctatus*	Southeast USA	1979
47	KGE7901	*E. piscicida*	7 (3,4,3,5,5,6,4,7,3,3)	*Tilape nilotica*	East Japan	1979
48	07BS	*E. piscicida*	8 (3,5,3,5,5,6,4,7,3,3)	*T. nilotica*	East Japan	1979
49	FL3-22	*E. tarda*	9 (2,1,3,6,1,5,7,8,1,1)	*Pterophyllum scalare*	Southeast USA	1994
50	9.8	*E. piscicida*	10 (4,4,3,3,5,7,8,9,3,3)	*Morone saxatilis*	Northeast USA	1986
51	WFE1	*E. piscicida*	11 (3,3,4,3,4,4,4,4,3,4)	*Paralichthys olivaceus*	East Japan	2002
52	WFE10	*E. piscicida*	11 (3,3,4,3,4,4,4,4,3,4)	*P. olivaceus*	East Japan	2002
53	ET009	*E. anguillarum*	12 (5,6,5,7,6,8,9,10,3,5)	*P. bogaraveo*	East Japan	2002
54	EIB202	*E. piscicida*	4 (3,4,4,3,4,4,4,4,3,4)	*Anguilla anguilla*	Northeast China	2006
55	205/03	*E. anguillarum*	13 (2,6,5,7,6,8,10,10,3,5)	*Sparus aurata*	South Spain	2003
56	NCIMB14824^T	*E. piscicida*	14 (2,5,4,3,4,9,4,4,3,3)	*A. anguilla*	Southeast Norway	1989
57	DSMZ13771^T	*E. hoshinae*	15 (6,7,6,8,7,10,11,11,4,6)	*Fratercula arctica*	Northwest France	1980
58	CECT885^T	*E. ictaluri*	16 (7,8,7,9,8,11,12,12,5,7)	*I. punctatus*	Southeast USA	1976
59	93-146	*E. ictaluri*	16 (7,8,7,9,8,11,12,12,5,7)	*I. punctatus*	Southeast USA	1993
60	FL6-60	*E. piscicida*	4 (3,4,4,3,4,4,4,4,3,4)	*M. saxatilis*	Southeast USA	1994
61	C07-087	*E. piscicida*	17 (3,9,8,3,5,12,8,13,3,3)	*I. punctatus*	Southeast USA	2007
62	DSM27202^T	*E. anguillarum*	18 (8,6,9,7,6,8,13,14,3,8)	*Anguilla marmorata*	Southeast China	2008

2

Table 2. Primers used for amplification and PCR condition.

Locus	Primer name	Primer sequence (5'---3')	Amplicon size (bp)	Size used in MLST (bp)	Annealing temp (ºC)	Reference
adk	*adk*F	ATTCCGCAGATCTCCAC	508	470	55	Buján *et al*. 2017
	*adk*R	TTCACATAGCGAGTATTGC				
*atp*D	A1	RTIATIGGIGCIGTIRTIGAYGT	884	790	45	Paradis *et al.* 2005
	A2	TCRTCIGCIGGIACRTAIAYIGC				
*dna*J	1F	GATYTRCGHTAYAACATGGA	758	729	52	Pham *et al.* 2007
	2R	TTCACRCCRTYDAAGAARC				
*gap*A	150F	AACTCACGGTCGTTTCAAC	750	719	52	Sawabe *et al.* 2007
	899R	CGTTGTCGTACCAAGATAC				
*gln*A	*gln*A-1	CGATTGGTGGCTGGAAAGG	530	505	52	Kotetishvili *et al.* 2005
	*gln*A-2	TTGGTCATRGTRTTGAAGCG				
Y-*hsp*60	Y-*hsp*60-1	GACGTNGTAGAAGGTATGYA	565	534	52	Kotetishvili *et al.* 2005
	Y-*hsp*60-2	CGCCGCCAGCCAGTTTAGC				
*pho*R	*pho*R-F	GACGGAGTTCTGACTGT	659	600	52	Abayneh *et al.* 2012
	*pho*R-R	ATGGCTGACGTTGGCAA				
*pyr*G	*pyr*G-F	TGAACTCCGTTGAGTTG	722	684	52	Abayneh *et al.* 2012
	*pyr*G-R	TCGTGAACACACCATGT				
*rpo*A	1F	ATGCAGGGTTCTGTDACAG	951	795	52	Thompson *et al.* 2005
	3R	GHGGCCARTTTTCHARRCGC				
tuf	T1	AAYATGATIACIGGIGCIGCICA	884	801	60	Paradis *et al.* 2005
	T2	CCIACIGTICKICCRCCYTCRCG				

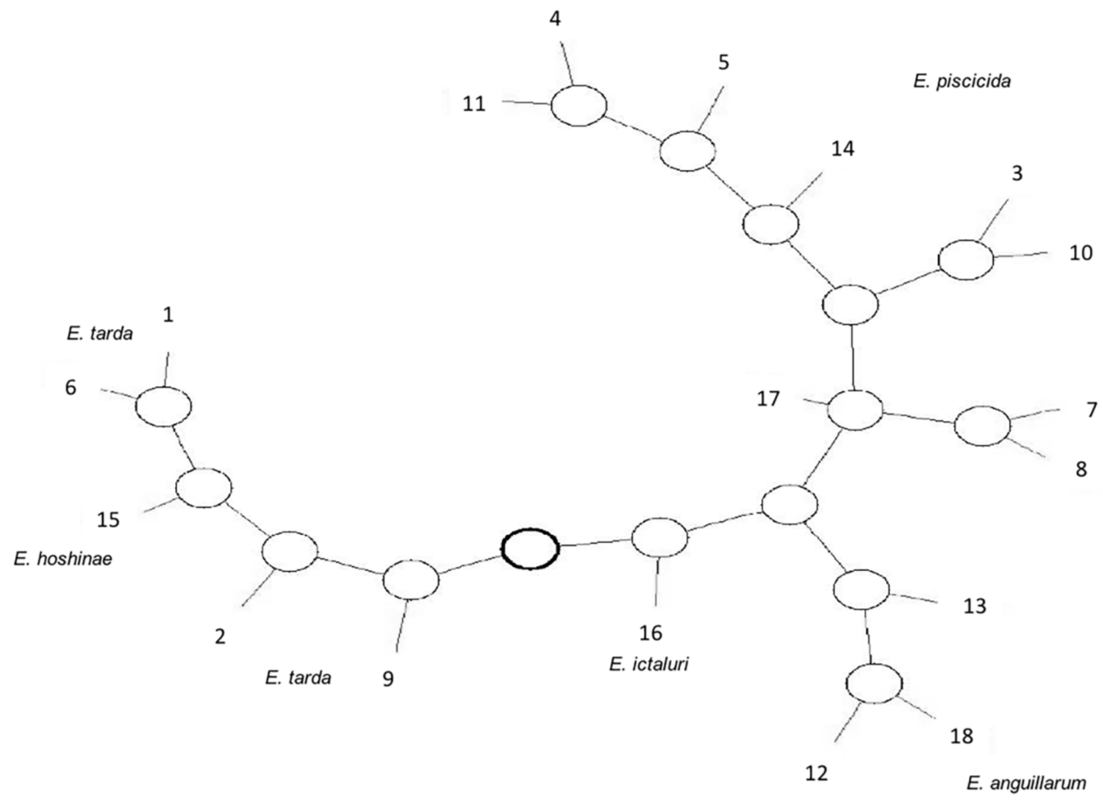
Table 3. Genetic characteristics, evolutionary variation and demographic history statistics of *Edwardsiella* population.

Locus	Number of alleles	Number of polymorphic site (PIM)	R_{min}	π	θ	d_N/d_S	D	$D*$	$F*$	F_S	S	R_2
adk	8	79 (48)	3	0.065	0.071	0.020	-1,9915*	0,2117	-0,77	8,193	0,001	0,24193
*atp*D	9	121(88)	21	0.067	0.066	0.006	-1,4896	0,5322	-0,31	17,76	0	0,10026
*dna*J	9	158 (113)	25	0.095	0.095	0.008	-1,8399*	-0,4131	-1,17	20,03	0	0,10015
*gap*A	9	68 (32)	10	0.036	0.040	0.017	-1,8292*	-0,2706	-1,04	6,756	0,004	0,10075
*gln*A	8	111 (82)	19	0.104	0.100	0.011	-1,4181	0,2032	-0,52	18,76	0	0,10047
Y-*hsp*60	12	74 (34)	6	0.042	0.050	0.029	-1,9449*	-2,0576	-2,41*	3,589	0,057	0,1009
*pho*R	13	154 (111)	24	0.100	0.088	0.058	-1,5395	0,0855	-0,67	13,66	0	0,10055
*pyr*G	14	119 (85)	18	0.067	0.060	0.008	-1,4244	0,2814	-0,46	9,374	0	0,10032
*rpo*A	5	31 (16)	4	0.018	0.017	0.010	-1,7082	-0,0203	-0,77	4,914	0,025	0,10068
tuf	8	48 (30)	5	0.027	0.027	0.017	-1,5648	0,4434	-0,38	5,71	0,01	0,10044
MLST	95	956 (689)	139	0.046	0.047	NA	-1,6978	-0,0789	-0,88	49,14	0	0,1003

PIM, parsimony informative sites; R_{min}, minimal number of recombination events; π, average number of nucleotides differences per site; θ, number of segregating sites; d_N/d_S,ratio of mean non-synonymous substitutions per non-synonymous site/mean synonymous substitution per synonymous site; D, Tajima´s D statistic; $D*$ and $F*$, Fu and Li statistics; F_S, Fu´s statistic; S, Strobecks´s statistic; R_2, Onsis R_2 statistic; *significate values; NA, not available

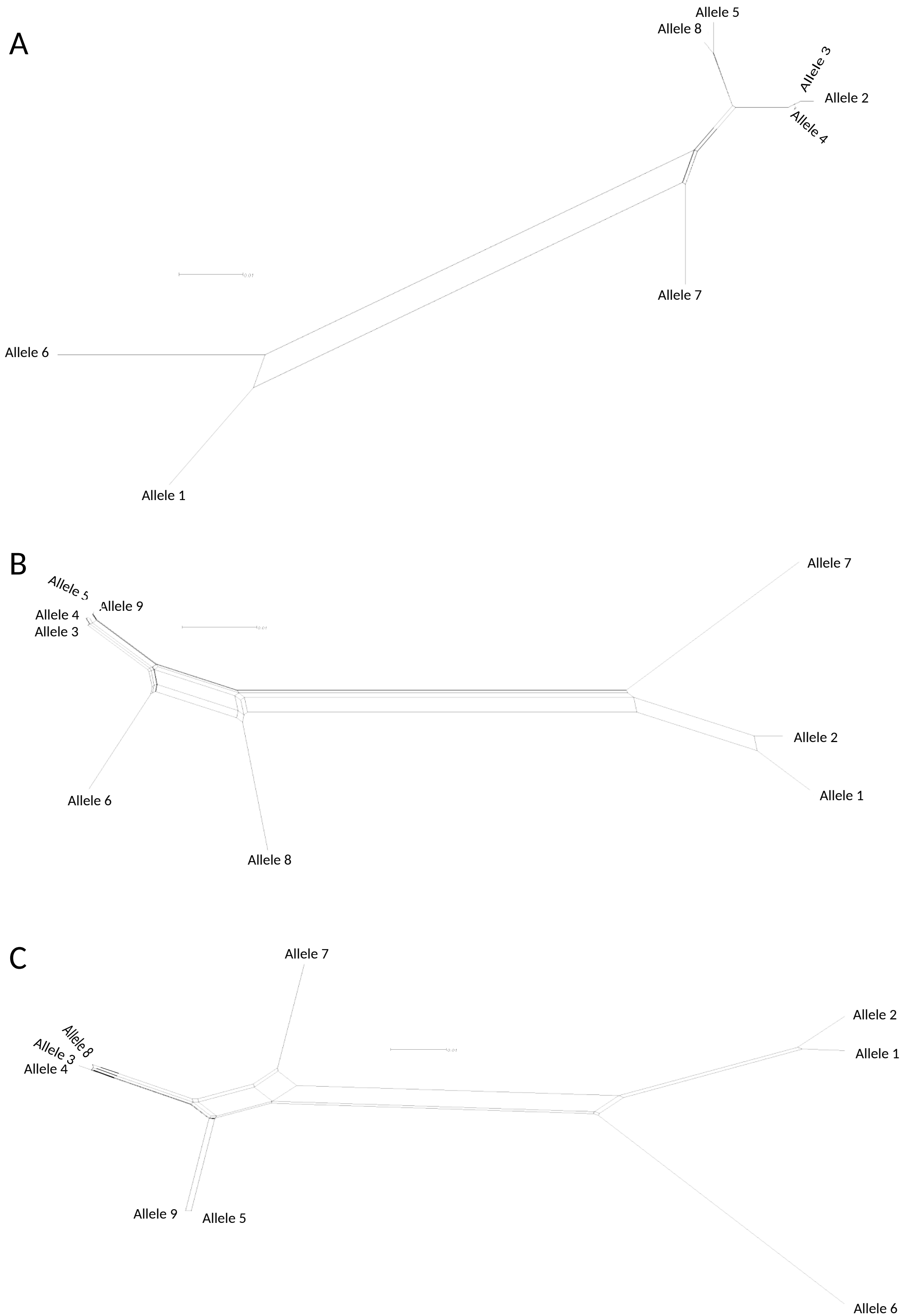
1
2
3 # Population Genetic and Evolution Analysis of Controversial
4 # Genus *Edwardsiella* by Multilocus Sequence Typing
5
6
7
8
9
10
11
12 Noemí Buján, Sabela Balboa, Jesús L. Romalde, Alicia E. Toranzo, Beatriz Magariños
13
14
15 Departamento de Microbioloxía y Parasitoloxía, CIBUS-Facultade de Bioloxía and
16 Instituto de Acuicultura. Universidade de Santiago de Compostela, Santiago de
17 Compostela, Spain.
18
19
20
21
22
23
24
25 # Supplementary material
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 Fig. S1. Dot graph based on majority rule consensus tree. Ancestral node is represented
54 in black. Grey lines indicate the STs. The empty circles indicate putative ancestral
55 nodes.
56
57



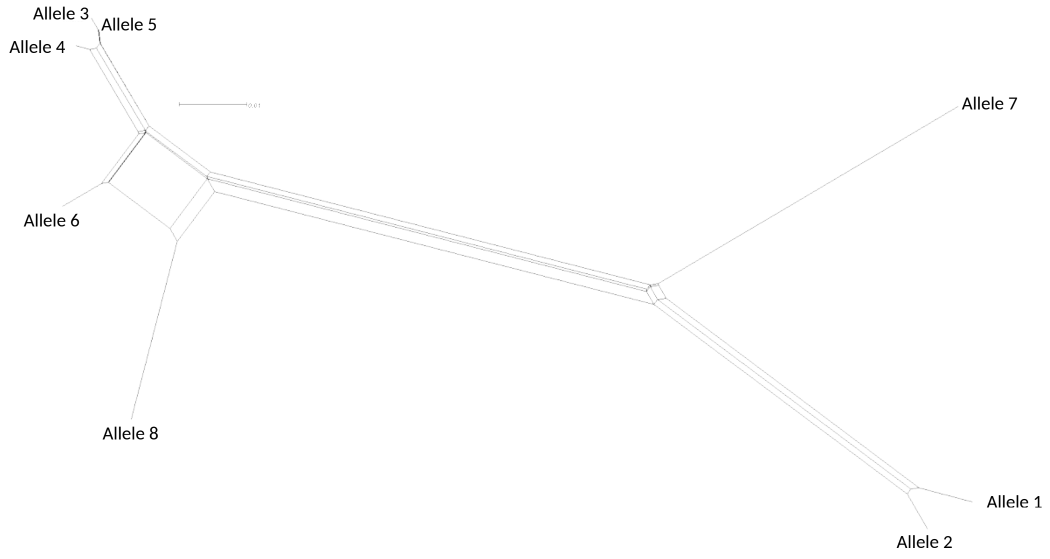
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82

83 Fig. S2. Neighbor-Net graph based on the individual housekeeping genes of the
84 different alleles. A, *adk*; B, *atp*A ; C, *dna*K; D, Y-*hsp*60; E, *gap*A; F, *gln*A; G, *pho*R; H,
85 *pyr*G; I, *rpo*A; J, *tuf*. Bar, 0.01 substitutions per nucleotide position.
86



87
88
89
90
91
92
93
94

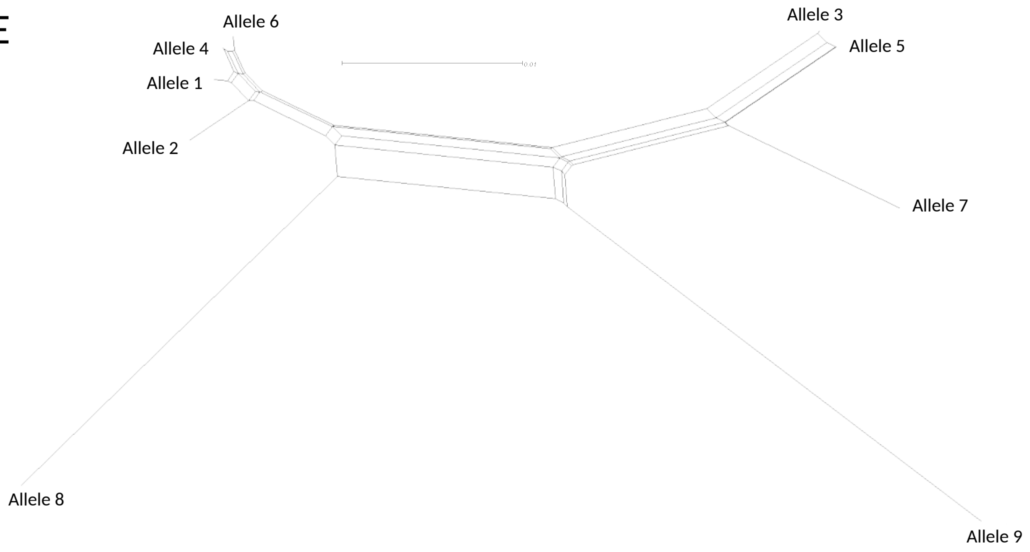
D



95

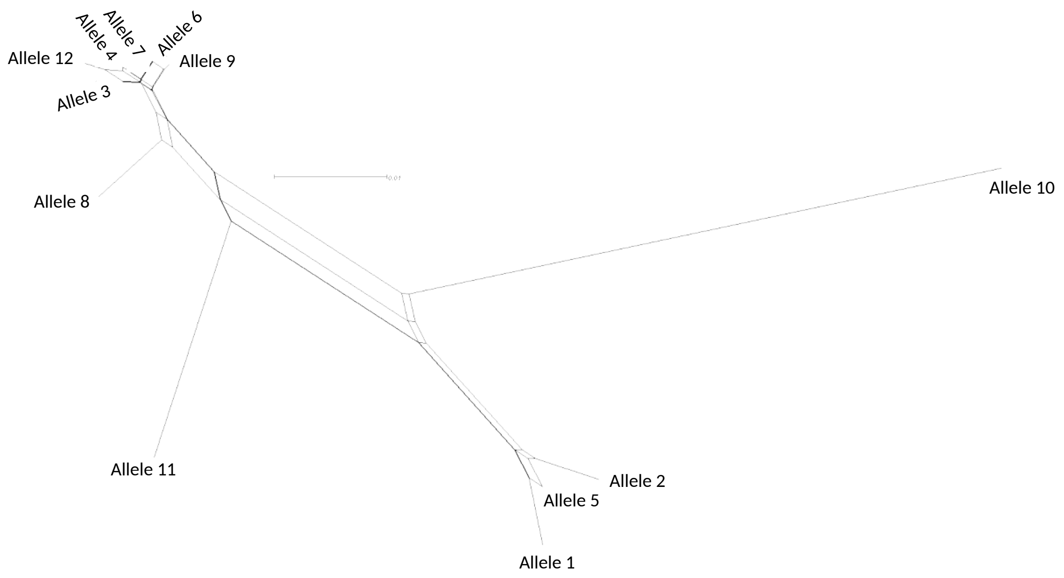
96

E

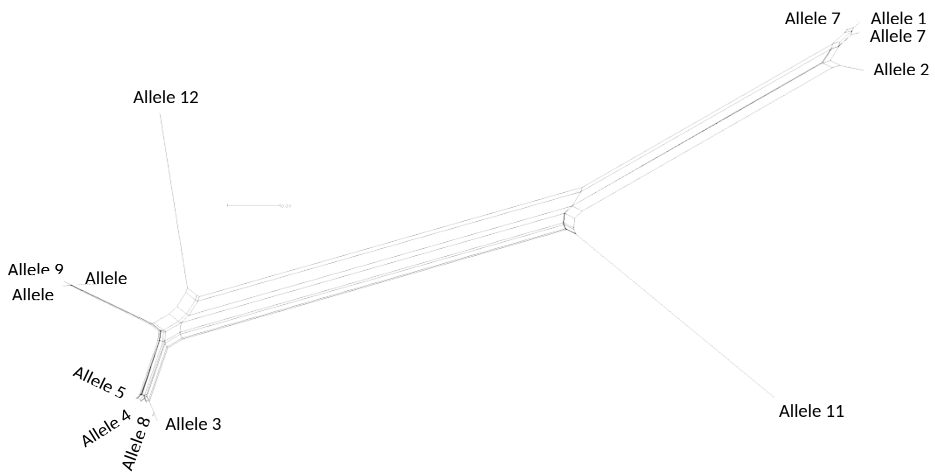


97

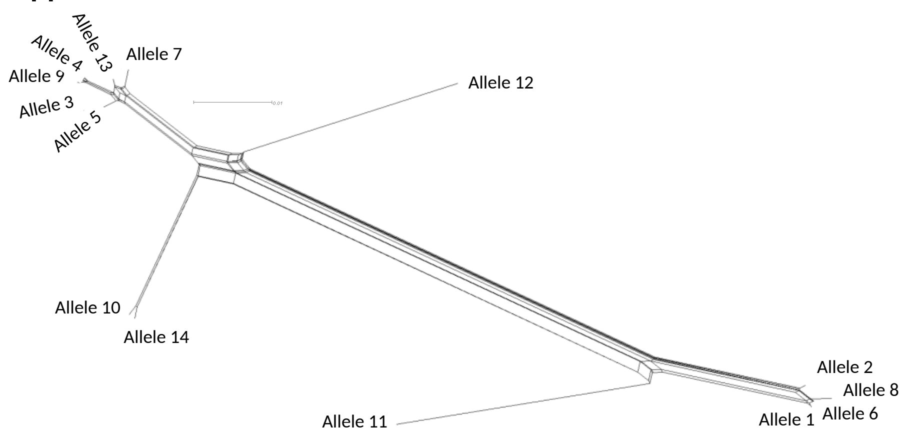
98



99

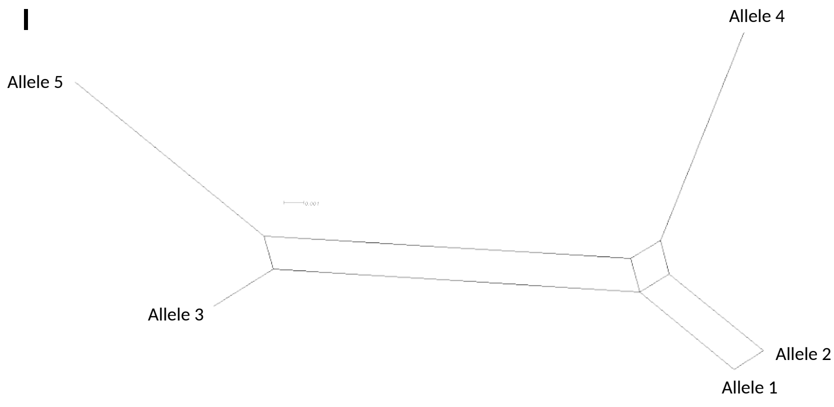


100



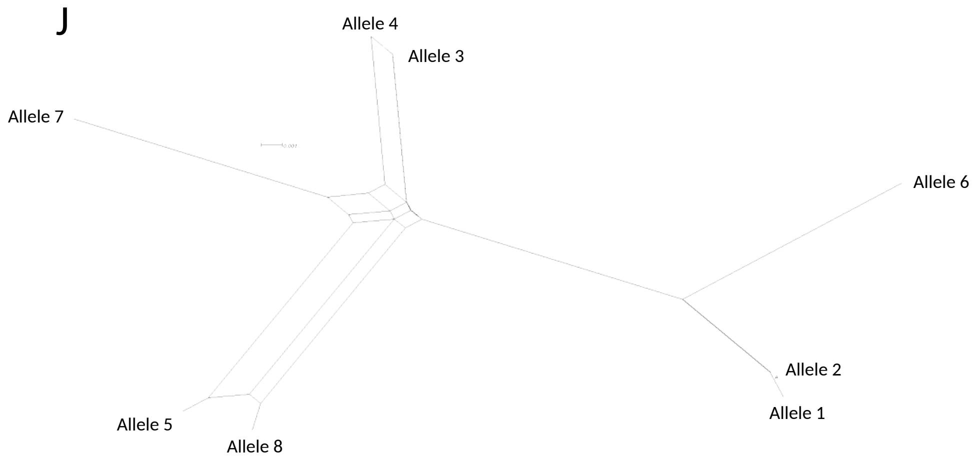
101

I



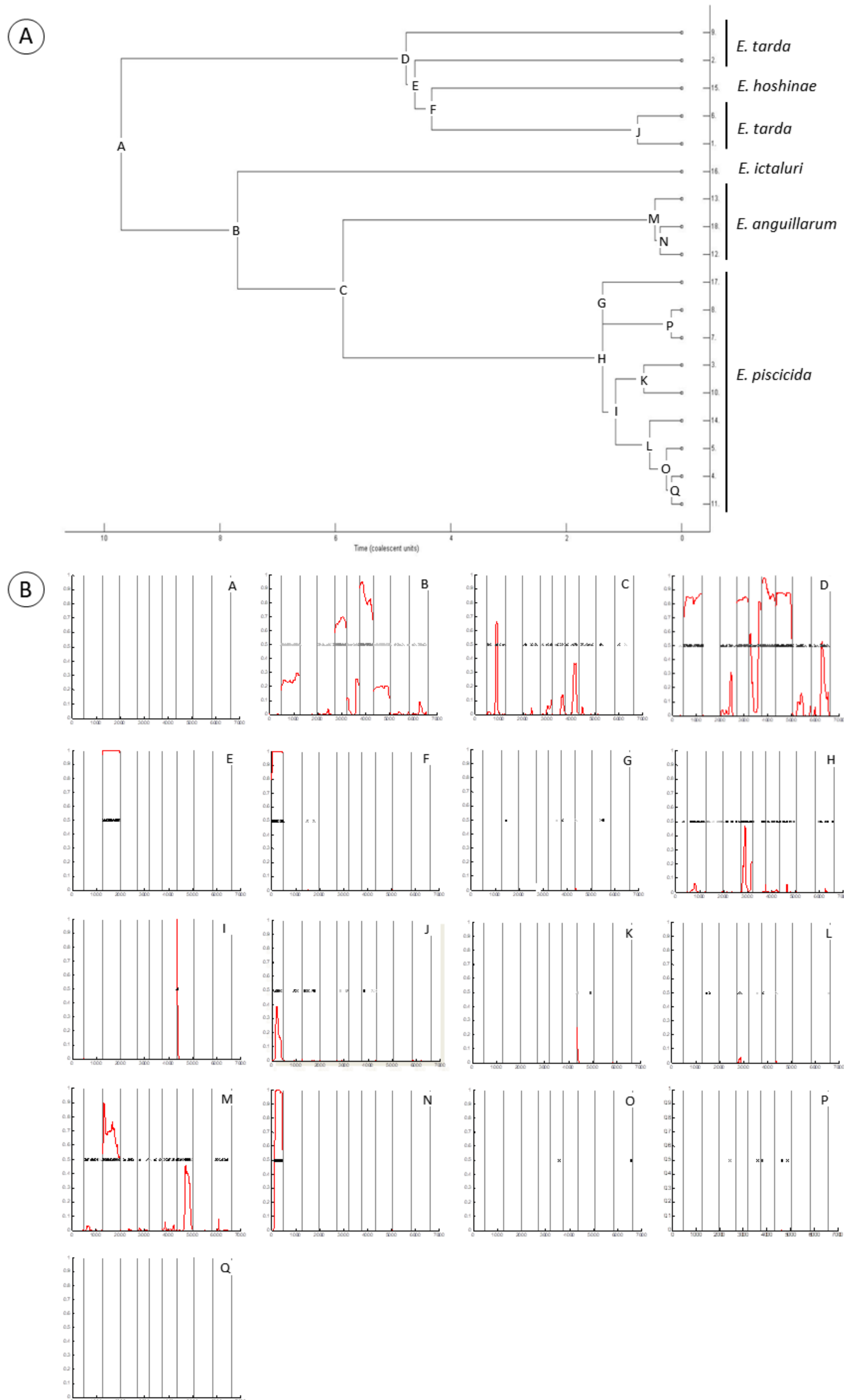
102

J



103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128

129 Fig. S3. Clonal genealogy (A) and evolutionary events of each node (B) reconstructed
130 following the majority rule from the concatenated sequence. In figure S3 (B), the x-axis
131 is the length of concatenate sequences, the red line indicates the probability of
132 recombination on a scale from 0 to 1 (y-axis) and black crosses represents each
133 nucleotide substitution.

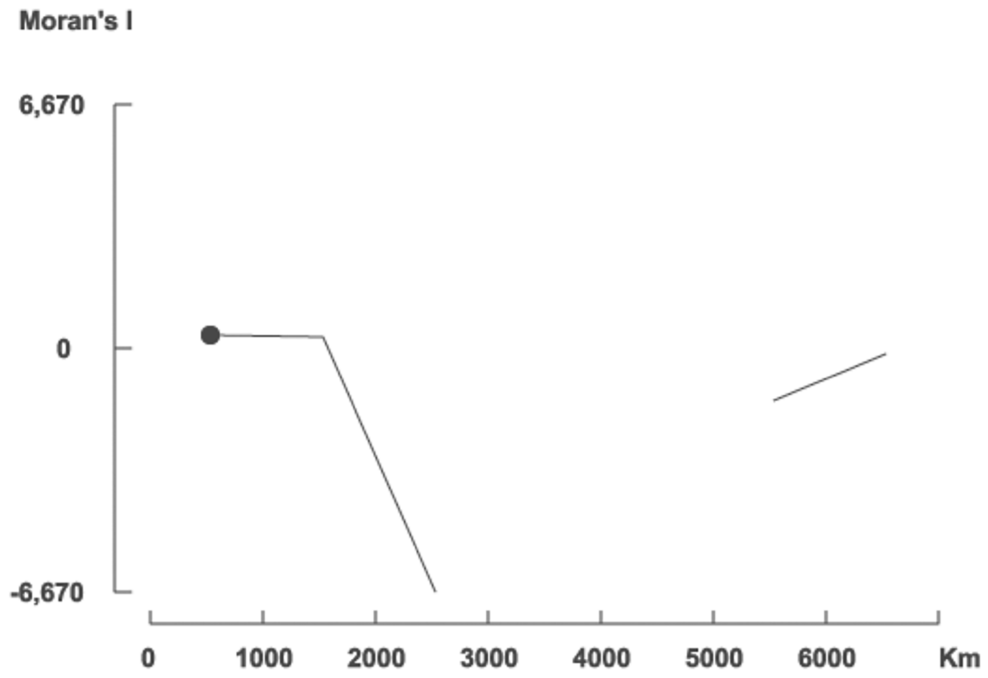


134

135

136 Fig. S4. Moran’s correlogram of individual allele frequencies. Significant value (black
137 dot) of Moran´s *I* indicate positive spatial dependence at *P*<0.05.

138



139