We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK
CITATION
INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

From the Gene Sequence to the Phylogeography through the Population Structure: The Cases of *Yersinia ruckeri* and *Vibrio tapetis*

Asmine Bastardo, Sabela Balboa and
Jesús L. Romalde

Additional information is available at the end of the chapter

Abstract

Multilocus sequence analysis (MLSA) and multilocus sequence typing (MLST) are nowadays considered as gold standards in the study of microbial systematic, being both techniques based on the interpretation of the sequences of several housekeeping genes. In this context, the sequences can be analyzed from different points of view. On the one hand, the phylogeny of the bacterial species can be estimated using the MLSA approach and on the other hand, the structure of the population can be inferred by means of MLST. Moreover, most species display some degree of population structure that can be interpreted in geographic and chronological contexts, that is, phylogeographic studies. In this review, the phylogeny and population structure of two important fish and shellfish pathogens, *Yersinia ruckeri* and *Vibrio tapetis*, exhibiting very different evolutive patterns will be analyzed. In both cases, the species form robust and monophyletic groups from a phylogenetic point of view. Regarding to the population structure, very different results were found. While *Y. ruckeri* follows an epidemic model of clonal expansion with well-adapted clones that explode to be widely distributed, *V. tapetis* appears to have a mixed structure in where the paradox of clonality and high level of variability coexist. Furthermore, phylogeographical studies provided the evolutionary and geographical context for the species, allowing the determination of historical and spatial influences on the diversification of both species.

Keywords: *Yersinia ruckeri*, *Vibrio tapetis*, MLST, population diversity, population dynamics, mutation, recombination.

1. Introduction

Phylogenetic analysis has long played a central role in basic microbiology. Sequence data offer direct genealogical information that can be efficiently used to estimate phylogenetic relationships and parameters associated with population dynamics. Furthermore, sequencing methods provide standardized and unambiguous data that are portable through online databases with direct access to the information needed to identify and monitor emerging pathogenic agents [1, 2]. Reconstructing the patterns of descent for a group of organisms can yield important awareness into why and how members of that group have specific characteristics and how those organisms are distributed across the environment. Integrating population patterns with phylogeny knowledge provides insights into epidemiological tracking of an organism at different evolutionary scales, from a single host to across the globe [3, 4]. On the other hand, more recently emerging fields of microbiology, including comparative genomics and phylogenomics, require substantial expertise in phylogenetic analysis and computational skills to handle the large-scale data involved [5]. Understanding the ways in which current and emerging technologies can be used to maximize phylogenetic knowledge is advantageous only with a complete proficiency of the strengths and weaknesses of these methods.

Since the conception of phylogenetic trees [6], morphological comparisons have been utilized to determine patterns of descent. Historically, numerous DNA-based approaches have been used to discriminate, subtype and build phylogenies for groups of organisms. Multilocus sequence analysis (MLSA) represents the standard in microbial molecular systematics. In this context, MLSA is implemented in a relatively straightforward way, consisting essentially in the concatenation of several gene fragments for the same set of organisms, resulting in one matrix which is used to infer a phylogeny by means of purely algorithmic methods [7–9].

For microbial pathogens, phylogenetic analyses are often conducted in order to determinate whether one particular outbreak may be related to another during times of an epidemic. While the clonal nature of an outbreak could be readily measured and predicted, Maynard-Smith et al. [10] pointed out the potential importance of homologous recombination as a determinant in the overall population structure of many bacterial species. These notions are now supported by several typing methods including multilocus sequence typing (MLST). Unambiguous genotyping systems are a key to describing epidemiological and ecological patterns and highlighting the evolutionary processes that shape microbial populations. Levels of genetic diversity are sufficiently high in most of microbial taxa that the sequences of several housekeeping gene fragments can provide a medium-resolution overview of their population genetic structure [11]. For the pathogenic bacteria whose members exhibit varying degrees of virulence, the integration of population genetic, evolutionary and epidemiological studies can provide important insights into the origins and spread of bacterial disease.

MLSA and MLST are based on housekeeping genes, which are subject to purifying selection and slow evolution and the variation within these genes is nearly neutral [12]. Although there are normally fewer polymorphic sites in individual housekeeping genes compared with hypervariable genes, the use of the combined sequences of multiple housekeeping genes has been shown to provide high discriminatory power while retaining signatures of longer-term

evolutionary relationships or clonal stability. Furthermore, analysis of multiple loci can buffer against potentially skewed evolutionary pictures obtained by single-locus analysis [13, 14].

Most species display some degree of population structure that can be interpreted in geographical and chronological context [15]. Phylogeography uses genetic information to study the geographical distribution of genealogical lineages, especially those found within species [16, 17]. Because the discipline has deep roots in historical biogeography and population genetics, phylogeography was heralded as a bridge linking the study of micro- and macroevolutionary processes providing the empirical and conceptual link between systematics and population genetics. Based on appropriated sampling of individuals and genes, this approach allows the assessment of the biogeographic hypothesis, the description of the evolution of isolated reproductive population units and the inference of processes underlying the origin, distribution and maintenance of diversity [18]. Detecting concordance of geographical variation in genotypes, or their genealogies and the environment is therefore at the core of phylogeographic studies.

The generation of large volumes of sequence data, combined with the development of novel analytical techniques and conceptual advances, promises a better understanding of the complexity of the evolution of bacterial populations. The application, advantages and constrains of the MLSA, MLST and phylogeographic analysis in taxonomic studies will be illustrated in this chapter with two examples: *Yersinia ruckeri* and *Vibrio tapetis*.

2. MLSA: inferring phylogeny of bacteria species

16S rRNA gene was the most common phylogenetic marker during 40 years. This molecule has a slow rate of evolution so very often it is difficult to establish phylogenetic relationships among taxa with recent divergence. MLSA represents nowadays the novel standard in microbial molecular systematics [19]. This is a rapid and robust classification method to study phylogenetic relationships of very diverse taxa of prokaryotes, including entire genera, by combining the information contained in the sequences of several specific genes [19]. This technique consists essentially in the concatenation of the sequence of several housekeeping genes (more than five), being the relationships among taxa established by phylogenetic inference [7, 8, 20, 21]. The use of such amount of data provides increased resolution power than the use of a unique gene as in the case of 16S rRNA gene, although this marker is considered still useful at taxonomic levels above the species.

MLSA can be used for bacterial identification and classification as well as for inferring evaluative relationships and variability among different groups of bacteria. At identification level, there are several studies demonstrating that MLSA using the concatenation of eight housekeeping genes provides a robust phylogenetic resolution for microorganisms sharing 70–95% of average nucleotide identity (ANI) and therefore, it could distinguish species of the same genus [22]. It has been even proposed to replace DNA-DNA hybridization (DDH), although concerns have arisen about this replacement [23–26]. At systematic level, MLSA is considered an intermediate resolution technique between the 16S rRNA gene and

the whole-genome-based approaches [19]. At evolutionary level, MLSA is a useful tool for studying the variability of different evolutionary identities, from families to species, as long as the selected genes for the analysis reflect properly the similarity of the complete genome among the studied group and the evolutionary ratios of the genes represent the evolution of the species.

The critical point for the MLSA is the suitability of the genes chosen for the analysis. In fact, genes that are perfectly informative within a given species, genus, or family may not be useful or even present in other taxa [19, 27]. Ideally, the best strategy to get a reasonable estimation of the species tree is to consider multiple genealogies inferred from unlinked loci and to use multiple individuals per species [28–30]. To date, there is not a general criterion for determining which genes are more useful for taxonomic purposes, but some attributes have been described for the genes to be used in the analysis [9, 26, 31, 32]. Genes should contain enough genetic information and although there is not a specification regarding to length, they should be small enough to be easily sequenced. Very often the fragments used in phylogenetic reconstructions are the same of those employed for MLST, resulting in too short fragments of the studied genes. The genes should also reflect the evolutionary history of the studied taxon [33]. Therefore, conserved genes must be selected for higher taxa and more evolved genes for species or subspecies levels. In that concern, the so-called core genes, the orthologous genes, should be used preferably than the accessory genes [34].

The availability of a universal set of conserved orthologous loci on a given taxon and, therefore, a set of primers that could amplified them across the studied group often precludes the comparative analysis of evolutionary process and patterns among closely related species and genera [26]. The strongest conflicting signals are usually derived from the existence of horizontal gene transfer (HGT) events in the dataset [35–37]. The resulting phylogenetic hypothesis may be distorted since standard treeing methods assume a single underlying evolutionary history [20, 38, 39].

There are no official recommendations about the inclusion of amino acid-base sequence analysis in MLSA studies although it is recommended because the study of the nucleotide sequences by themselves can lead to an "overinterpretation" of phylogenetic differentiation in closely related taxa [32]. Usually, the exchange in a base on the third position of a given codon has no influence in the resulting protein sequence and therefore in the structure and/or function of the protein, but also it can have the opposite effect. Because of that, nucleotide alignments should be done regarding their amino acid sequence. It must be taking into account that a bacterium is not only a sequence of DNA and for taxonomic purposes, the living unit at all its levels should be considered.

3. MLST: establishing the bacterial population structure

Nucleotide sequence data from multiple housekeeping genes in an appropriately sampled population can be used in a variety of analyses to determine population structure. The simplest of these analyses is MLST, which establishes the allele present at each locus and use a

clustering algorithm to determine the relationships among strains from the matrix of pairwise differences between their allelic profiles [40]. The major advantage of MLST over others typing methods, such as multilocus enzyme electrophoresis (MLEE), is the unambiguous nature of the data obtained and the simple storage and electronically exchange, meaning that any isolate that is typed using the method can be rapidly compared with all previously typed strains.

The number of alleles obtained for each locus is much higher using MLST than MLEE and the information obtained by MLST is more precise. Publically available databases such as http://pubmlst.org and http://www.mlst.net/ provide examples where clinical subtyping has allowed epidemiological, geographical, and/or evolutionary hypotheses were for pathogens like *Neisseria meningitidis*, *N. gonorrhoeae*, *Streptococcus pneumoniae*, *Vibrio parahemolyticus* and *Staphylococcus aureus* [40–43]. National and international surveillance of bacterial clones can be performed using this resource.

Unweighted pair group method with arithmetic mean (UPGMA) dendrograms based on pairwise comparisons among allelic profiles can be structured on the website to detect relationships between query and/or isolates database. However, although clustering algorithms are useful for detecting the genetic relatedness of small number of isolates, they can result infeasible when visualizing larger sample sizes (e.g., >1000) in MLST database. As these methods are not based on an evolutionary model, they are often inaccurate in reconstructing evolutionary events [44]. The recent development of the algorithm eBURST [1] has addressed both issues. The model incorporated into eBURST assumes that, due to selection or genetic drift, some genotypes will occasionally increased the frequency in the population and then gradually diversify by the accumulation of mutation(s) and/or recombinational replacements, resulting in slight variants of the founding genotype. Using allelic profile data, one sequence type (ST) is assigned to each isolate. STs sharing high genetic similarity are grouped into clonal complexes (CCs). The founding genotype for each CC is then identified parsimoniously as the genotype differs from the highest number of the other genotypes in the CC at only one locus. Further diversification will produce variants of the founder allelic profile that differ at two or more locus. Thus, the simple principal underlying eBURST is that bacterial populations will consist of a series of clonal complexes (set of variants of a funding genotype) that can be recognized from the allelic profiles of the strains within a MLST database [1].

While MLST is very effective for establishing which isolates are identical or closely related, the approach will not provide major information about the relationships between more distantly related isolates, unless the population is strictly clonal. However, additional phylogenetic information can be gathered if the nucleotide sequences themselves are studied by analyzing the extent of linkage disequilibrium between alleles and looking for recombination by the congruence of gene trees, or the presence of mosaic structures [45, 46]. Knowledge of the recombination extent in bacterial pathogens is important since low levels of recombination result in a highly clonal population, where lineages persist with little variation over hundreds or thousands of years. At the other extreme, high rates of recombination lead to weakly and/or non-clonal populations in which lineages diversify so rapidly that the isolates recovered in one decade may be completely different from those recovered in the next [47].

For highly clonal species such as *Salmonella enterica* [48], most of substitutions in the genome have appeared by mutations. Alleles that arise independently multiple times in different branches are therefore incongruous with the tree. The phylogenetic relationships between isolates can be inferred from the dendrogram derived from the pairwise differences between STs and independently from a consensus tree constructed from the gene sequences. Characterization of weakly clonal pathogens (e.g., *N. meningitidis*, *S. pneumoniae*) is more problematic since clones diversify rapidly by the accumulation of recombinational exchanges. However MLST is very useful for the identification of the currently circulating hypervirulent lineages because these are recognized as clusters of isolates with identical, or very similar, multilocus sequence types.

4. Phylogeography: putting the geography into phylogeny

Phylogeography attempts to infer history from the geographical variation of genes and genetically controlled characters. In the phylogenetic/population genetic approach, graphical phylogenetics trees, networks, or clades are visualized from the observed variation data [49–51]. Thus, the usefulness of this approach is to integrate both phylogeny and geography within a quantitative analytical framework that encompasses the diverse aspects of phylogeography concordance [16, 52]. In this context, several classes of analytical techniques are used according to their function. The first class of techniques (i.e., AMOVA, Wombling, Monmonier's maximum difference algorithm, cline model by maximum likelihood) extracts spatial pattern from geographically distributed genetic data to identify either geographical partitions or clines (first-order pattern, in the terminology of spatial statistic), or alternatively, patterns of isolation-by-distance (second-order pattern) [53–55]. The second class (i.e., analysis of distance matrices, allelic aggregation index) attempts to infer historical scenarios directly from observed distributions of genes or taxa and one or more phylogenetic model [56]. A third class of techniques, such as Slatkin's distribution, provides statistical testing for the previously inferred scenario [57]. Phylogenetic trees and networks are often visualized over a cartographic background. Spatial interpolation algorithms [58] estimate parameter values at unsampled locations from a spatial distribution of observed points, providing a mean of interpreting and visualization the sampled data at different sets of locations [59, 60].

Many species show pronounced phylogeographic structure, or even regional or continental endemism, which counteracts the previously held paradigm of continuous and global panmixia. However, biogeographic and macro-ecological studies at the community level have shown that relatively few free-living microbial eukaryotes have cosmopolitan distribution [61, 62]. However, prokaryotes are generally smaller and have faster reproduction cycles than eukaryotic microorganisms that were the subject of these biogeographic studies [63]. Several studies have reported clear phylogeographic structuring in bacterial communities including marine, soil and soil-freshwater bacteria [64–66]. Conversely, the absence of spatial structuring in other prokaryotes has been corroborated by molecular data for bacteria from those same environments, including cyanobacteria [67–69]. For microorganisms occurring in

extreme environments, phylogeographic structure indicates the effects of strong geographic isolation and dispersal constraints, although not all show clear spatial structure [70, 71]. For the more widely distributed bacteria, biogeographic patterns may result from historical and/or contemporary environmental processes. The importance of these processes in structuring microbial systems is still poorly understood [72] and few studies have focused on the phylogeographic structure and dispersal limitation in bacteria on a truly global scale in discontinuous but globally common habitats.

5. Case studies

5.1. Yersinia ruckeri

Yersinia ruckeri is a Gram-negative bacterium and the causative agent of enteric redmouth (ERM) disease or yersiniosis in salmonid and non-salmonid fish reared in both fresh and marine waters. *Y. ruckeri*, initially isolated from rainbow trout (*Oncorhynchus mykiss*) in the Hagerman valley of Idaho (USA) in the 1950s [73], is now widely found in fish populations throughout North and South America, Australia, Africa and Europe [74]. The pathogen *Y. ruckeri* is a serologically variable, highly clonal species. It includes two biotypes: biotype 1 strains are positive for motility and lipase activity, whereas biotype 2 strains are negative for both tests [75]. The species has been grouped into 6 serovars [76], 5 O-serotypes [77], or 4 O-serotypes with different subgroups [78] by using different serotyping systems. In addition, *Y. ruckeri* strains can be grouped into clonal types on the basis of biotype, serotype and outer membrane protein (OMP) profiles [79]. Strains of serotypes O1a (classic serovar I) and O2b (classic serovar II) cause most epizootic outbreaks and serotype O1a is predominant in cultured salmonids [76, 78].

ERM has been successfully controlled for decades by vaccination with commercial monovalent killed whole-cell vaccines. Although formulations of most commercial vaccines are based only on serotype O1a (Hagerman strain), different degrees of cross-protection among serotypes have been described [76]. In recent years, reports of ERM vaccine breakdown have emerged in Europe and USA mostly attributed to biotype 2 strains [80–82]. Other epizooties have occurred in vaccinated Atlantic salmon (*Salmo salar*) from Chile, caused by serotype O1b/biotype 1 *Y. ruckeri* strains [83].

Molecular techniques have been used to study the intraspecific genetic variability showing a low genetic diversity. By using of MLEE was identified only four electropherotypes for 47 isolates of *Y. ruckeri* indicating that the genetic structure of *Y. ruckeri* is clonal, with one predominant clonal group [84]. The ribotypes, patterns of pulsed-field gel electrophoresis (PFGE) and interspersed repetitive sequences (IRS)-PCR of 30 *Y. ruckeri* O1a strains have been studied, reporting a high level of genetic homogeneity for all the isolates [85]. On the other hand, a total of 44 pulsotypes from 160 isolates identified by PFGE have provided better insights into the relationship between similar *Y. ruckeri* clones responsible for recent ERM outbreak among salmonid [86]. Heterogeneous assembly of phenotypes in serotype O1a *Y. ruckeri* strains with respect to pathogenicity and host has been reported [87]) and the need

of expansion of the clonal group theory in this species was suggested, highlighting therefore the existence of new clonal groups [88].

In the context of the genetic approach, none of the studies have focused on the sequencing and analysis of housekeeping genes to understand the *Y. ruckeri* population structure. The existing studies have been limited to MLSA analysis in which few isolates were included for the comparison and description of new species within the *Yersinia* genus [89–91].

Using a sequence-based approach, new studies were developed by our research group to reconstruct the phylogeny and to characterize the molecular epidemiology and population structure with a collection of 103 strains of *Y. ruckeri* (**Table 1**). Studies included the sequencing of six housekeeping genes *gln*A (glutamine synthetase), *gyr*B (DNA gyrase B subunit), *rec*A (recombinase A), Y-HSP60 (heat-shock protein 60 kDa), *dna*J (heat-shock protein 40 kDa) and *thr*A (aspartokinase-homoserine dehydrogenase), as described by [92].

Origin	Isolates (*n*)	Biotype	Serotype	ST[*]	Host/source
USA	19	1, 2	O1a, O1b, O2b, O3, O4	8	*O. mykiss, S. trutta, O. tshawytcha, Ictalurus punctatus*
UK	6	1, 2	O1a, O1b	4	*O. mykiss*
Portugal	21	1, 2	O1a, O3	5	*O. mykiss,* water, sediment
Chile	11	1	O1a, O1b, O2b	3	*S. salar*
Peru	28	1, 2	O1a	7	*O. mykiss*
Denmark	4	1	O1a, O2b,	4	*S. trutta, Anguilla anguilla, O. mykiss*
Spain	4	1	O1a, O2b	2	*O. mykiss*
Finland	3	1, 2	O1a,	2	*O. mykiss, S. salar*
Canada	4	1	O1a, O1b, O2a	3	*S. salar, Rutilus rutilus, O. tshawytscha, S. malma, Ondatra zibethicus*
Germany	1	1	O2b	1	*O. mykiss*
Scotland	1	1	O2b	1	*O. mykiss*
Norway	1	1	O1b	1	*S. salar*

[*]Established by Bastardo et al. [92].

Table 1. Summary of *Y. ruckeri* strains studied.

Similarity matrix of intraspecific sequence for the individual genes ranged 97.2–100% for *dna*J, 98.5–100% for *gyr*B, 98.8–100% for *gln*A, 99.4–100% for Y-HSP60, 98.1–100% for *rec*A and 99.0–100% for *thr*A. For the concatenated sequence (2786 bp) of encoding-protein genes, the similarity was determined between 98.9 and 100%. The best-fit model of nucleotide evolution was determined using Modeltest 3.7 and following the Akaike information criterion (AIC). The maximum-likelihood (ML) estimation was implemented in PHYML 3.0 without substitution rate heterogeneity correction or invariant estimation as recommended by Modeltest. Clade support was evaluated by analyzing 1000 bootstrap pseudo-replicates. To further probe the robustness of our MLSA-based phylogeny, the DNA sequence data were analyzed

by the neighbor joining (NJ), maximum parsimony (MP) and Bayesian inference methods. All three analyses yielded similar tree topologies, which were highly congruent with our previous findings based on maximum likelihood (**Figure 1**).
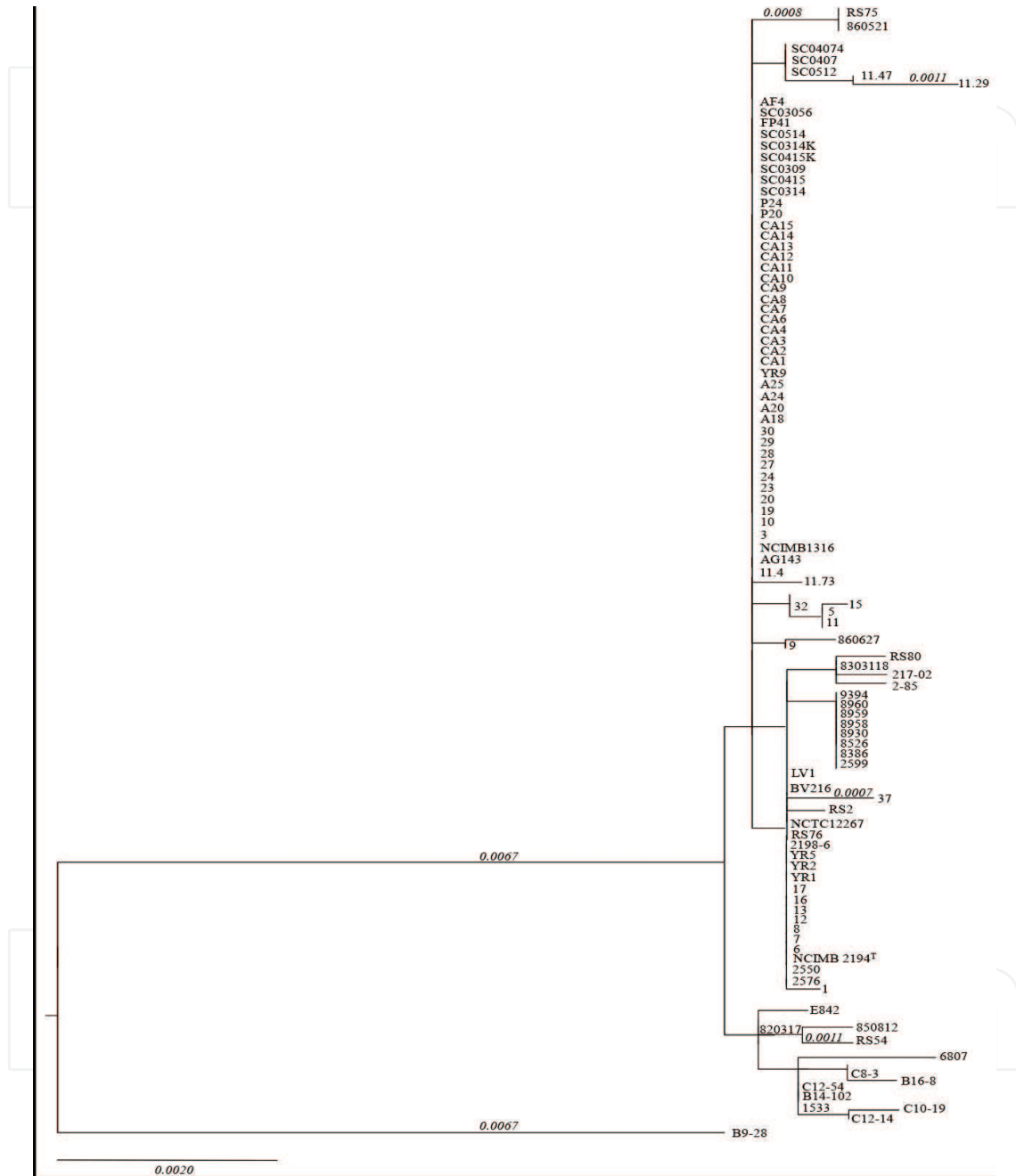
Figure 1. Phylogenetic tree based on concatenated sequences of six housekeeping genes of *Y. ruckeri* obtained by the ML method.

The results of the MLSA analysis confirm that there is significant diversity within *Yersinia ruckeri* isolates showing that they formed distinct clusters. Except one isolate, all *Y. ruckeri* strains joined in a major cluster with a complex topology that does not seem to reflect previous typing schemes of the species [84, 85]. Using MLSA, the ML tree topology suggests major

genetic diversity among the isolates of the serotype O1a. Thus, isolates belonging to serotype O1a appear spread among different branches together with other serotype/biotype representatives. Interestingly, only the groups of isolates associated with recent outbreaks in USA, Chile and Peru fall into well-defined branches.

Based on the sequences of the six housekeeping genes available in the public database htpp://publmst.org/yruckeri/, a MLST scheme for *Y. ruckeri* has been developed [92]. **Table 2** shows a descriptive analysis of nucleotide and allele diversity for each locus. Synonymous substitutions (d_S) occurred more frequently than non-synonymous substitutions in every gene (d_N). Furthermore demonstration that all loci are under purifying pressure ($d_N/d_S < 1$) was obtained, which means a strong selection such that most amino acid substitutions are deleterious, as being typically observed for housekeeping genes.

	Size of fragment (pb)	Alleles	Polymorphic sites	*Pi*	*H*	d_N	d_S	d_N/d_S
*gln*A	416	9	7	0.0064	0.2528	0.0042	0.0123	0.3427
*gyr*B	454	6	8	0.0075	0.3107	0.0043	0.0194	0.2238
Y-HSP60	509	4	4	0.0039	0.0577	0.0000	0.0159	0.0000
*rec*A	472	4	11	0.0125	0.5773	0.0047	0.0354	0.1339
*dna*J	632	7	16	0.0078	0.4717	0.0000	0.0340	0.0000
*thr*A	303	4	3	0.0055	0.4841	0.0074	0.0000	-

Table 2. Genetic characteristics and evolutionary variation among the six loci included in the MLST scheme for *Yersinia ruckeri*. *Pi*, average number of nucleotide differences per site; d_N/d_S, ratio of mean non-synonymous substitutions per non-synonymous site/mean synonymous substitution per synonymous site; *H*, Nei's index of diversity.

Among all isolates of *Y. ruckeri*, 30 different sequence types (ST) were established (**Table 3**), 21 of which were represented by a single isolate, evidencing high genetic diversity. From these MLST scheme, eBURST analysis identified two clonal complexes (CC) showing a common evolutionary origin for 94 isolates forming 21 STs into CC1 and for six isolates of six STs in the CC2. ST 14 and ST 21 were identified as founder sequences into CC1 and CC2, respectively (data not shown). Furthermore, the formation of three STs (singletons), no associated to any clonal complexes, suggests genetic diversification among *Y. ruckeri* strains into the population.

All alleles analyzed showed to be in nonrandom distribution or linkage disequilibrium (I_A value of 0.5563, $p = 0.000$) in the *Y. ruckeri* population, suggesting that mutation drives the diversification of this species and supporting a clonal structure for the population [10]. Furthermore, recombination events were not detected for *gln*A, *rec*A, Y-HSP60, *dna*J and *thr*A when DnaSP5 and RDP4 software was used. However, recombination was found among *gyr*B alleles ($R_{min} = 1$), indicating that the relative low genetic diversity present in all of alleles analyzed could have obliterated the chance of detecting recombination in the other five loci and suggesting that recombination could occur within different subpopulations.

ST	Profile						Frequency %	
	*gln*A	*gyr*B	Y-HPS60	*rec*A	*dna*J	*thr*A		
2	1	1	1	1	1	2	43	41.75
1	1	1	1	1	1	1	16	15.53
7	1	1	1	1	2	1	8	7.77
3	1	2	1	1	2	2	4	3.88
16	6	1	1	1	1	2	3	2.91
9	1	3	1	1	1	1	2	1.94
14	1	1	1	1	2	2	2	1.94
23	1	1	1	1	6	2	2	1.94
26	1	5	1	1	2	2	2	1.94
4	2	1	1	1	2	2	1	0.97
5	1	1	2	1	2	2	1	0.97
6	3	1	1	1	1	2	1	0.97
8	1	2	3	1	3	1	1	0.97
10	4	1	1	1	1	1	1	0.97
11	1	4	1	1	1	1	1	0.97
12	1	3	1	1	1	2	1	0.97
13	1	2	1	1	1	1	1	0.97
15	5	1	1	1	2	2	1	0.97
17	7	1	1	1	4	4	1	0.97
18	1	5	1	1	2	1	1	0.97
19	8	6	4	4	5	2	1	0.97
20	9	1	1	1	6	1	1	0.97
21	7	2	1	1	6	1	1	0.97
22	7	2	1	1	4	4	1	0.97
24	1	1	1	1	4	4	1	0.97
25	7	2	1	1	1	2	1	0.97
27	7	2	1	1	7	2	1	0.97
28	1	1	1	1	1	3	1	0.97
29	1	1	1	2	1	2	1	0.97
30	1	1	1	3	2	2	1	0.97

Table 3. Allelic profiles of MLST scheme established for *Y. ruckeri* (htpp://publmst.org/yruckeri/).

Based on the single-locus variables (SLVs) found between the two clonal complexes and the different subgroups identified by eBURST algorithm, the variant alleles can be used to determine the events responsible for the evolution into the population [93]. Thus, the per-allele and per-site recombination/mutation (r/m) parameter was calculated empirically from 25 SLVs identified within the two clonal complexes in *Y. ruckeri*. Twenty out of 25 SLVs arose from a recombination event, whereas only five arose by mutation. This resulted in a per-allele r/m parameter of 4:1. In the case of the per-site analysis, r/m parameter ratio was 7.5:1. These two parameters suggest that the initial steps of *Y. ruckeri* clonal diversification at allele or individual nucleotide sites are 4- and 7-fold more likely to occur by recombination than by point mutation. Recombination appears then to play a greater role than mutation for the generation and maintenance of the genetic diversity of *Y. ruckeri* (**Figure 2**).
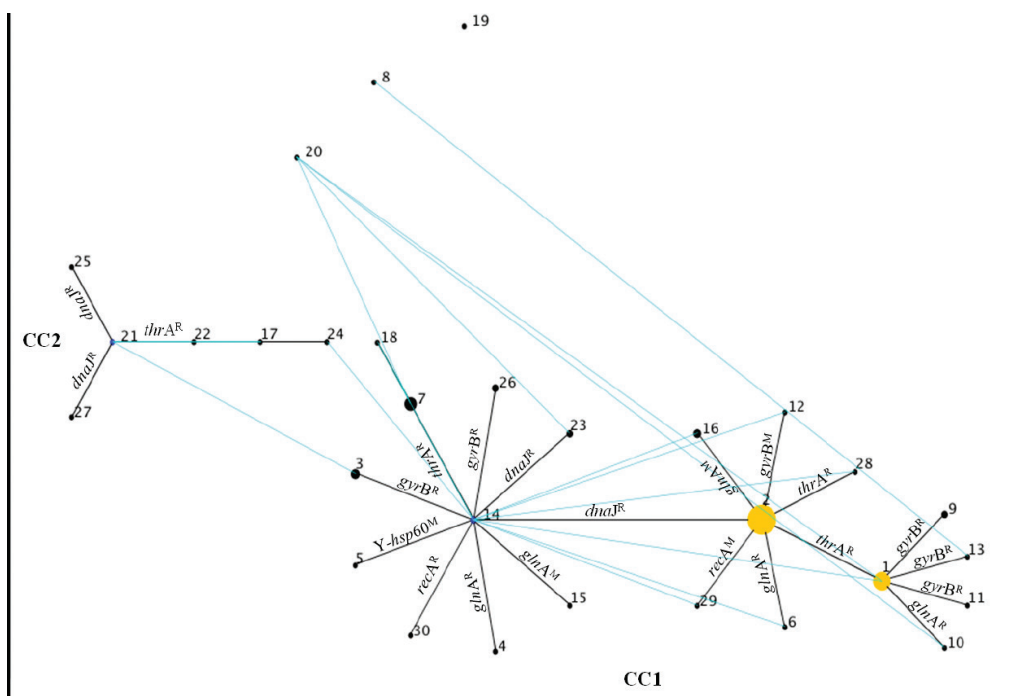
Figure 2. Population snapshot with all the sequence types (STs) included in the MLST database for *Y. ruckeri*. The snapshot was made using the eBURST program. STs that are SLVs of each other are shown connected by black lines. DLV STs are shown connected by aqua lines. Clonal complexes (CC) are marked. The sizes of the circle are related to the number of strains within each ST. The founder and cofounder genotypes are colored blue and yellow, respectively. A superscripted R or M indicates recombination or mutation, respectively (from [92]).

Epidemic model is also consistent with the epidemiology of *Y. ruckeri*, which suggest that ERM started as a geographically isolate disease that relatively quickly became widely disseminated [74, 92]. In an epidemiological approach, the strong association of sub-founders ST 1 and ST 2 to the majority of the ERM outbreaks in salmonid cultures also allows to link these STs to virulence of *Y. ruckeri* strains. On the other hand and although serotypes were not strictly associated with the STs in this MLST study, our results suggest that serotypes O1a and O1b are an example of recently emerged and disseminated variants. In addition, nonmotile *Y. ruckeri* strains (biotype 2), causing recent outbreaks in vaccinated fish, were included into

the sub-founders ST 1 and ST 2, indicating that biotype 2 phenotype may have evolved from related motile *Y. ruckeri* strains.

The phylogeographic analysis showed concordance with the eBURST diagram obtained previously for the *Y. ruckeri* population. Thus, it was possible to construct the complete evolutionary networks, showing the missed putative variants linking the established STs separated into two clonal complex and three singletons by using eBURST algorithm (**Figure 3**). The inclusion of the geographical data into this analysis indicates high genetic differentiation into *Y. ruckeri* caused by the fixation of different alleles into one geographical area (data not shown). These findings explain the high diversity of STs found in Europe and USA, including those observed in not salmonid fish species and supports the hypothesis that the majority of *Y. ruckeri* STs has evolved independently in specific areas. Furthermore, the presence in USA, UK and Peru of the different STs grouping nonmotile isolates provides strong evidence of the independent emergence and dissemination of biotype 2 *Y. ruckeri* strains in different geographical areas [92, 94, 95].
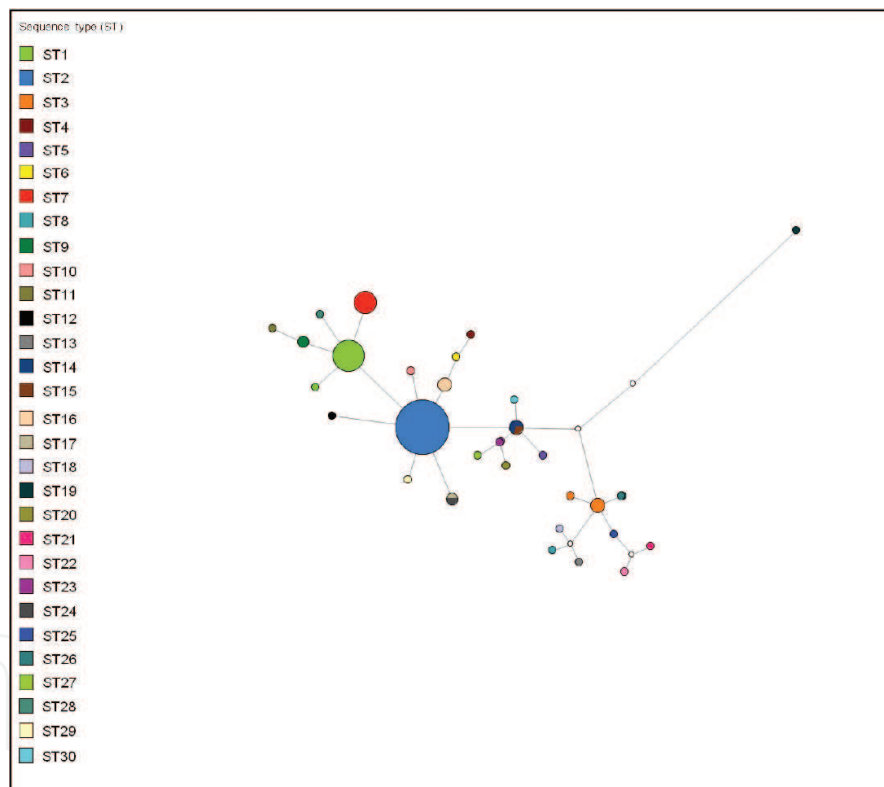
Figure 3. Parsimony network (95%) of *Y. ruckeri* based on the ST alignment. Colors indicate sequences correspondent to STs. A line between STs represents mutational step and open circles represent STs not present in the sample. Radius of the circle represents number of sequences.

The sequence dataset was divided into 29 predefined subpopulations consisting of sequences from STs that present in each geographical origin and the geographical distances between different populations were measured using geographical coordinates. The Mantel test ("isolation-by-distance" analysis) for the matrix of correlation between genetic and geographic

distance showed no significant correlation positive for the full dataset ($Z = 18 \times 10^{11}$, $r = 0.0139$ one side $p = 0.5959$), indicating a lack of overall genetic differentiation between the different geographical area. Furthermore, demographical analyses indicate a recent global expansion of *Y. ruckeri* revealed by both Tajima's test ($D = -1.669$, $p < 0.001$) and Fu's test ($Fs = -13.83056$, $p < 0.001$). This fact can explain the emergence of genetic variants that have caused the recent outbreaks in farmed salmonids in several areas, as occurred in Chile, Canada and Portugal where uncommon serotype and phenotype isolates were involved (**Figure 4**) [95].
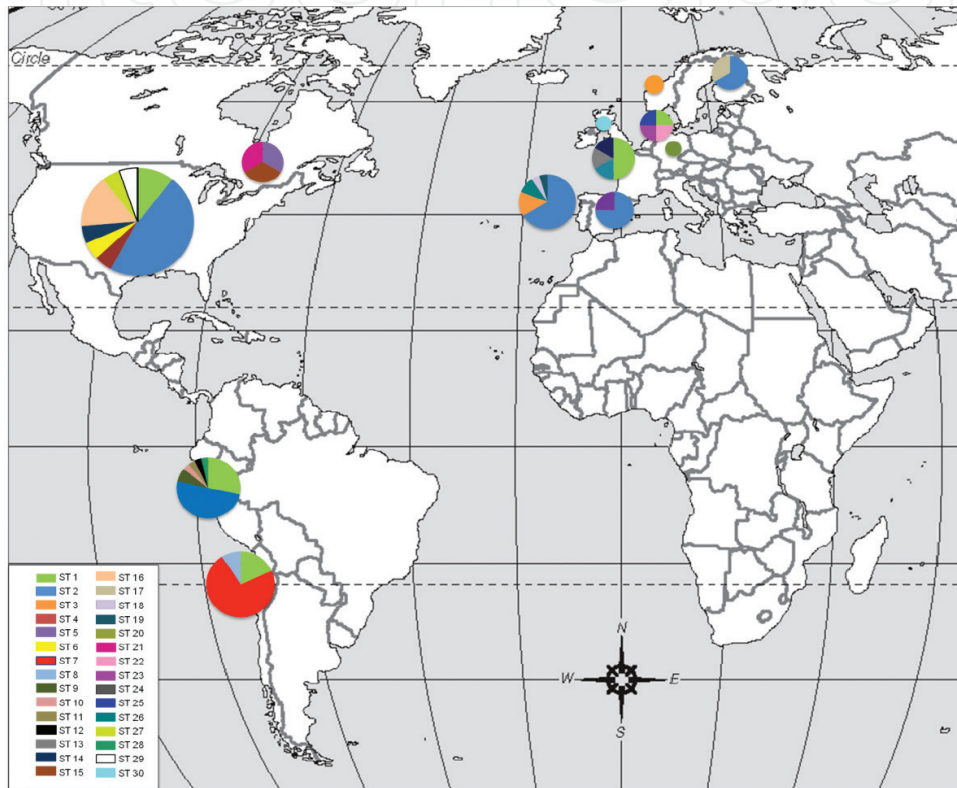
Figure 4. Global distribution of STs of *Y. ruckeri* population based on multilocus allelic profiles defined for the eBURST analysis.

5.2. Vibrio tapetis

Vibrio tapetis is the causative agent of brown ring disease (BRD), a major limiting factor for the culture of Manila clams (*Ruditapes philippinarum*) in Europe which has been associated with large economic losses in the sector [96]. It is considered the only disease with demonstrated bacterial etiology that affects adult clams. This disease was first described in Landeda (France) in 1987 associated with an episode of mass mortality of Manila clam. Since then, it has been detected throughout the European Atlantic coast and occasionally in the Mediterranean and the Adriatic coast as well as in Korea and Japan [97, 98]. The disease received its name because of the most visible symptom in affected animals, the presence of an abnormal deposit of brown organic (composed mostly of conchiolina) on the inner surface of the valves, usually

located between the pallial line and the edge of the shell and not subjected to calcification processes [97, 98]. *V. tapetis* was initially considered as a homogeneous taxon but further studies of new isolates with different geographic and host origin, including fish and shellfish species, demonstrated the existence of intraspecific variability, at phenotypic, serological and genetic levels. Serologically, at least three serogroups were detected using slide agglutination and dot blot [99, 100]. At genetic level, differences were first detected in the plasmid content and ribotypes of the different strains [101, 102]. More recently, based on ERIC-PCR (enterobacterial repetitive intergenic consensus), REP-PCR (repetitive extragenic palindromic) and RAPD (randomly amplified polymorphic DNA analysis), three major groups associated with the host and the serogroup were established [103]. The existence of these three groups was confirmed by preliminary studies of MLSA and protein expression by 2D-PAGE studies. This MLSA study was performed on the basis of five protein-coding housekeeping genes but only with three representative strains of the described groups [104].

Population structure and phylogenetic analysis (as well as its relationship with the geography) of *V. tapetis* was performed using thirty strains with different host and geographic origin: seven strains isolated from *R. philippinarum* in France between 1988 and 1994; eight strains isolated also from *R. philippinarum* in Spain in the years 1994, 2005 and 2007; one strain of French cockle isolated in France in 1990 as well as other isolate from the same location and year isolated from Venus clam (*Venerupis pullastra*); two Irish isolates obtained from *R. philippinarum* seed (2005); two isolates from Spanish carpet shell clam (*R. decussatus*) isolated in 1994; one isolate from halibut (*Hippoglossus hippoglossus*) cultured in UK isolated in 2001; three isolates from shi drum (*Umbrina cirrosa*); four isolates from wedge sole (*Dicologoglossa cuneata*) isolated in Spain in 2007 and 2005, respectively; and one isolate from corkwing wrasse (*Symphodus melops*) obtained in Norway in 1999.

The partial sequences of ten housekeeping genes were used: *atp*A (α subunit of ATPase), *fst*Z (cell division protein), *gap*A (glyceraldehyde-3-phosphodehydrogenase), Y-HSP60, (heat-shock protein 60 kD), *pyr*H (uridyl monophosphate kinase), *rct*B (replication origin-binding protein), *rec*A (recombinase A), *rpo*A (α subunit of RNA polymerase), *rpo*D (RNA polymerase sigma factor), *top*A (topoisomerase I). The selected genes were demonstrated to be phylogenetically resolutive in other *Vibrio* species [105] and the results were also compared with those obtained by DDH [106] as well as by other typing methods [101–103, 107]. The concatenation of these ten genes rendered a fragment of 5826 bp in length. The intraspecific sequence similarities ranged between 84.8 and 100% for individual genes and between 93.3 and 100% for the concatenated sequence.

The phylogenetic reconstruction for the concatenated gene sequences was done using three different methods, NJ, MP and ML, using in all cases 1000 bootstraps. Topology of all the trees was the same, showing only some differences at bootstrap values. Visual inspection of the *V. tapetis* concatenated alignment tree reveals the existence of two tight clusters, one formed by most of the isolates and a second one, smaller, composed for three isolates (**Figure 5**). These two clusters shared sequences similarities of 93.3–93.5%, while strains within each cluster showed less than 0.6% of gene sequence divergence variation in cluster 1 and less than 0.4% in cluster 2, being most of the substitutions located at third position of each codon.
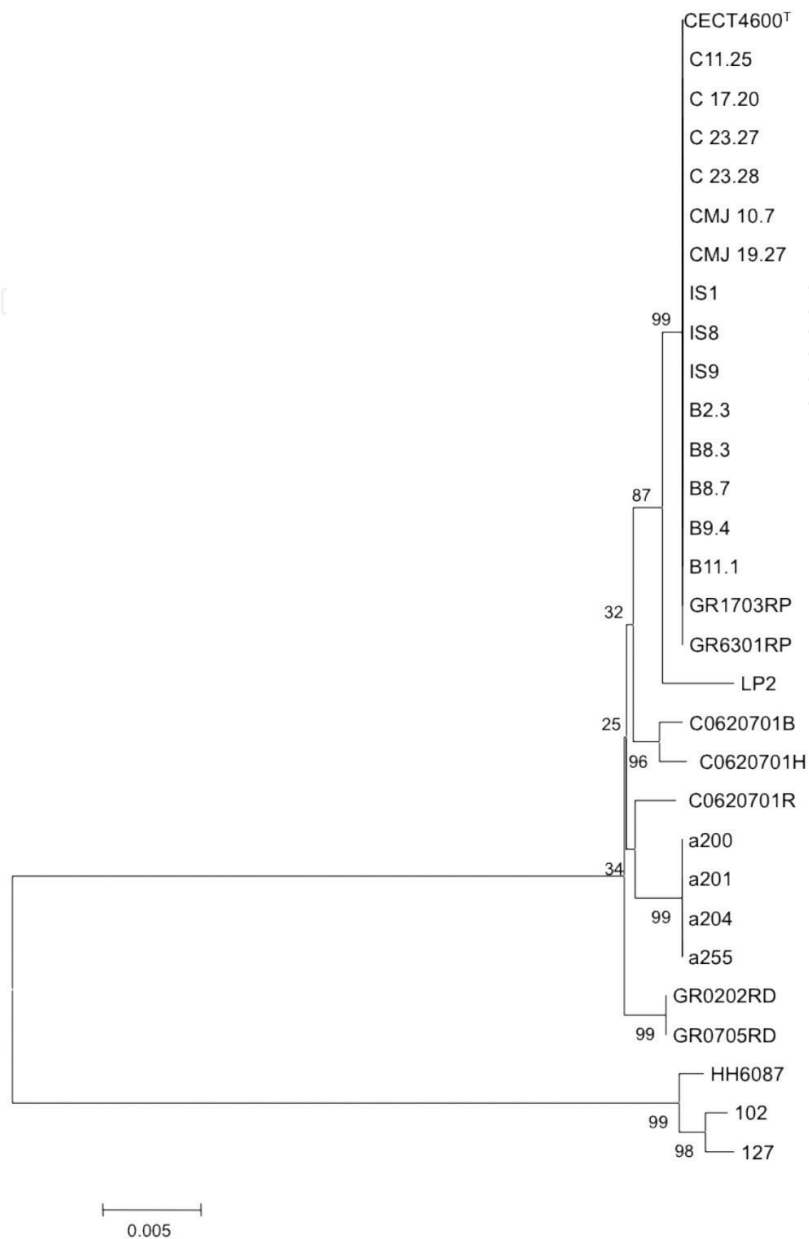
Figure 5. Phylogenetic reconstruction based on concatenated alignment of ten housekeeping genes of *V. tapetis* using the NJ method. Horizontal branch lengths are proportional to evolutionary divergence. Bootstrap from 1000 replicates appears next to the corresponding branch.

In the biggest cluster, high diversity is observed regarding to their host and geographical origin, containing the isolates classified as group one (represented by the type strain) and group two (represented by the isolate GR0202RD) by Rodríguez et al. [103]. As can be observed, different branches are formed, most of them related with host origin: the adult Manila clam isolates together with those from cockle and Venus clam cluster in the major branch and related to them appears the corkwing wrasse isolate. The carpet shell clam isolates fall into an individual branch as well as the wedge sole isolates, which form a cluster close to shi drum isolates. The second cluster, formed by isolates HH6087 (halibut), 102 and 127 (*R. philippinarum* seed), is a very robust branch supported by a 99% bootstrap. These isolates have in

common the geographic origin, being all of them isolated in British Isles (UK and Ireland) although from different hosts. The two-cluster topology is supported by the trees generated individually for each gene.

MLST analysis revealed the heterogeneity of the population of this clam pathogen. The high variability of the population is reflected in the number of identified alleles ranging from 3 to 9 depending on the gene analyzed. The allele combination leads to the description of 10 STs (**Table 4**), all of them constituting singletons. Even when the stringent SLV criterion was relaxed (from 9/10 to 1/10 shared alleles), none of the SLV or DLV was found (data not shown). This variability is also reflected in the 450 single-nucleotide polymorphisms (SNPs) detected across the 5826 bp surveyed. The majority of the SNPs were biallelic, being only 7 of them were triallelic. The nucleotidic substitutions found throughout the concatenated sequence showed, as usual for housekeeping genes, more frequency in synonymous substitutions (d_N) than synonymous one (d_S). The ratio d_N/d_S shows that all the genes except by *rpo*D are under positive selection (**Table 5**).

ST	*atp*A	*fst*Z	*gap*A	Y-HSP60	*pyr*H	*rct*B	*rec*A	*rpo*A	*rpo*D	*top*A	n
1	1	1	1	1	1	1	1	1	1	1	17
2	2	1	2	2	2	2	2	1	2	2	2
3	3	2	3	3	3	3	3	2	3	3	1
4	4	3	3	4	4	4	4	2	4	4	1
5	5	4	3	5	5	5	5	2	4	5	1
6	6	1	1	6	6	6	6	1	5	6	1
7	6	5	1	6	6	2	7	1	6	6	1
8	7	1	1	7	1	2	7	1	6	7	1
9	8	1	4	1	7	7	8	3	7	8	4
10	1	6	5	8	8	8	9	1	8	9	1

**Table 4.** Allelic profile of MLST scheme established for *V. tapetis* (htpp://publmst.org/vtapetis/).

The alleles showed to be in linkage disequilibrium (I_A value for the whole strain collection was 6.3008 ($p = 0.000$)), being therefore the mutation of the main cause of diversification. These data are in agreement with the fact that even when a number of approaches were used to achieve recombination events, only using R_{min} and Phi test some events were found in *atp*A, *pyr*H, *rct*B, *rec*A, *rpo*D and *top*A genes, but not utilizing RDP4 (**Table 5**). However, the SplitsTree generated for all the isolates included in this study shows some structures typical from high recombinatory populations (**Figure 6**). It has been described that the simplest method to detect recombination in aligned sequences is to look for mosaic structures by eye. Significant mosaic structure is indicative of recombinatorial exchange, usually among isolates of the same species [93].

The contradiction between the results inferred for I_A index and those from the SplitsTree can be explained analyzing the deduced genealogies after stripping sequences of recombinational

| Gene	Loci		Polymorphisms					Index of diversity			Recombination	
	Length (bp)	G+C content	Alleles	SNPs	SVS	PIS	S_{nonsyn}	Pi	d_N/d_S	H	R_{min}	Phi test
*atp*A	600	46.51	8	25(25)	2	23	1	0.01257	0.0087	0.9556	1	$p =0.081$
*fst*Z	609	46.38	6	49(51)	12	37	6	0.01291	0.0204	0.7778	0	$p =1.000$
*gap*A	588	45.23	5	8(10)	0	8	0	0.00292	0.0000	0.8000	0	$p =1.000$
*hsp*60	525	44.91	8	35(37)	5	30	6	0.01167	0.0270	0.9556	0	$p =1.000$
*pyr*H	516	48.42	8	38(38)	2	36	1	0.01332	0.0036	0.9556	1	$p =1.000$
*rct*B	600	43.14	8	95(96)	3	92	7	0.02975	0.0095	0.9333	1	$p =1.000$
*rec*A	588	45.86	9	64(66)	4	60	6	0.02117	0.0237	0.9778	4	$p =0.397$
*rpo*A	600	47.00	3	7(7)	0	7	1	0.00226	0.0528	0.6000	0	$p =1.000$
*rpo*D	600	43.52	8	47(47)	2	45	13	0.01598	0.1111	0.9556	5	$p =0.586$
*top*A	600	47.75	9	82(86)	5	77	10	0.02633	0.0188	0.9778	1	$p =0.650$

Table 5. Genetic characteristics, evolutionary variations and recombination among the ten loci included in the MLST scheme of *V. tapetis* population.
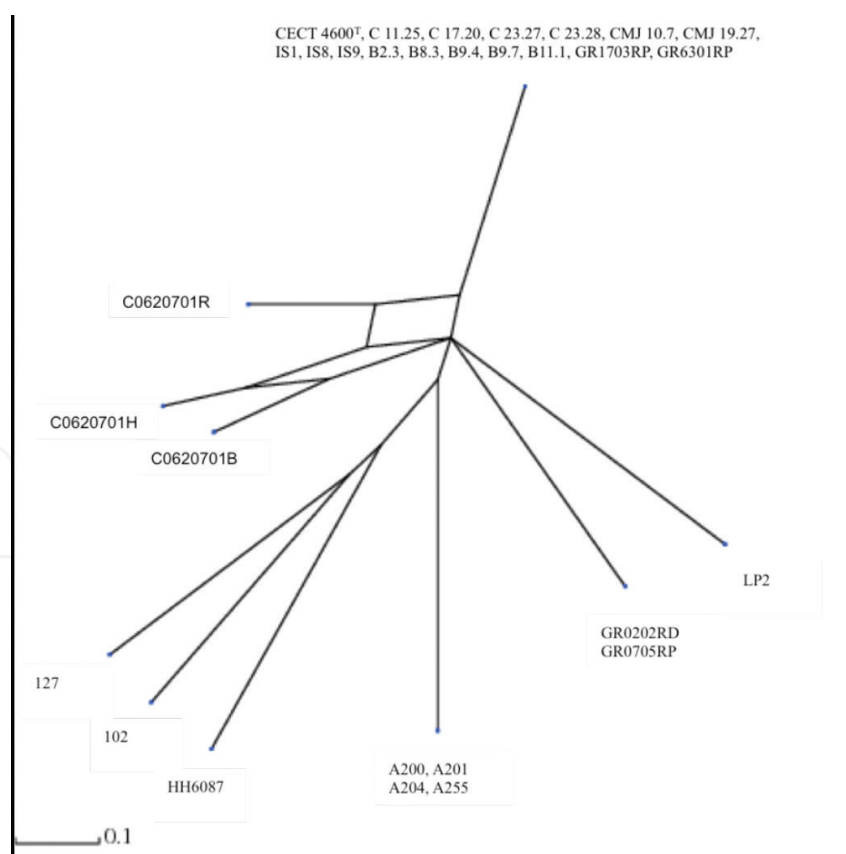
Figure 6. Split decomposition analysis of the concatenated sequences of the 30 *V. tapetis* strains was studied. The splitgraph was generated using SplitsTree v4.

events performed with the ClonalFrame [108]. **Figure 7** shows the clonal genealogy constructed for the 30 isolates (**Figure 7A**) and the evolutionary events for the three nodes tasked in the tree (**Figure 7B**) in where the height of the red line indicates the probability of recombination on a scale from 0 (row bottom) to 1 (row top) and each nucleotide substitution is represented by a black cross. Node A represents the divergence among the two clusters generated by MSLA analysis and it can be seen in the representation of evolutionary events are exclusively generated by mutation events. The opposite occurs at nodes B and C, which represent the diversification between cluster one and two, respectively and their evolution is produced mainly by recombination.

Since the three isolates of the cluster two of *V. tapetis* have in common the geographic origin (British Isles), a phylogeographic approach was used with the aim of correlate geographic
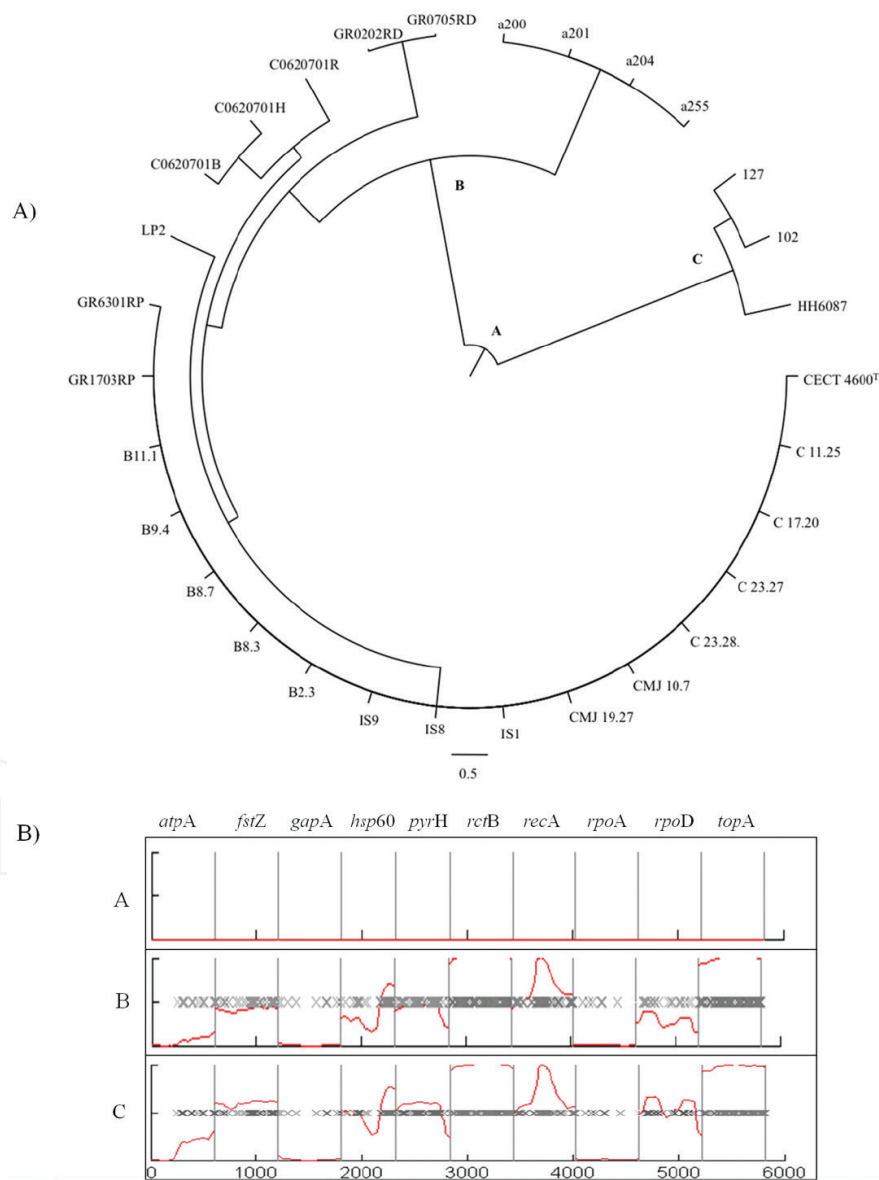
Figure 7. Clonal genealogy (A) and evolutionary events (B) reconstructed by ClonalFrame from concatenated sequences of *V. tapetis* population.

origin and genetic evolution of the isolates. The concatenated sequences were divided into four subpopulations consisting in sequences of STs that belong to each geographic origin. For this study, regions of the same country were not taking in account. Despite the two well-differentiated groups showed in the phylogeograpic network (**Figure 8**), significant correlation was not observed among the identified STs and geographic distance in the "isolation-by-distance" analysis (Mantel test: $Z = 25 \times 10^{11}$, $r = 0.1314$, $p = 0.9750$). Values obtained for Tajima's test ($D = -0.99118$, $p < 0.10$), negative and non-significative and Fu's test ($Fs = 29.227$, $p = 0.000$), positive and significative, are indicatives of population expansion. These indexes indicate that the population has suffered a recent expansion following a bottleneck or a selective sweep.
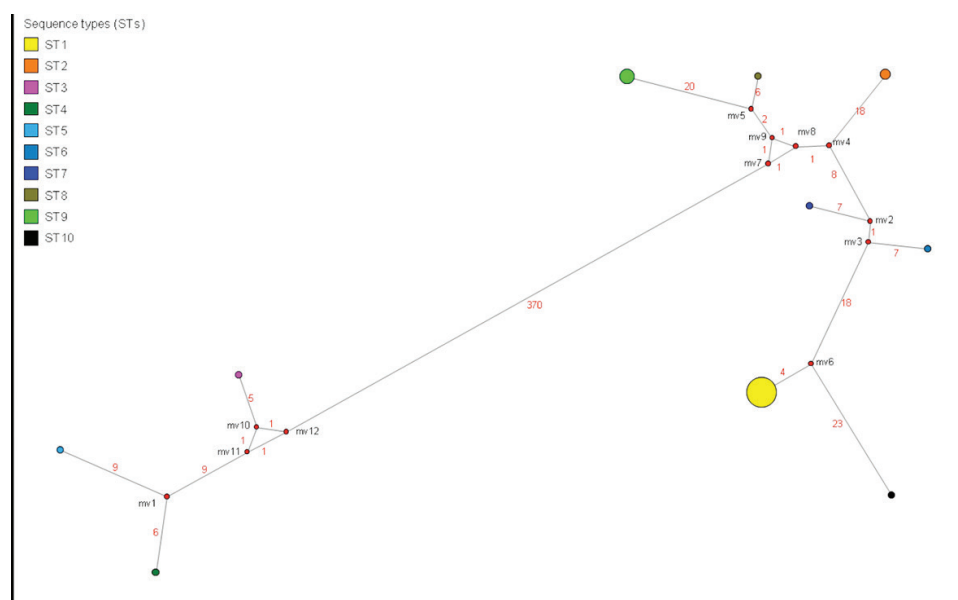
Figure 8. Representation of the evolutionary network using the combination of allelic variants (STs) of *V. tapetis* population. Red dots indicate possible intermediates in the evolution of the species.

The phylogeographic network is very useful to clarify all the previous data. First, the two groups of isolates that can be observed are the same two clusters of the phylogenetic study (MLSA) and in the inferred clonal genealogy. The node A in the clonal genealogy is represented in this phylograph by the long branch (note that this is not an evolutive method, so that length is not representative of distance) generated by 370 nucleotidic substitutions, which according to the reconstructed evolutionary study are likely mutations. At the ends of this branch are located nodes B and C on the clonal genealogy (and the two clusters in MLSA), which are generated essentially by recombination according to ClonalFrame. These recombinatory events can be seen in the topology of the graphic. On the other hand, the inconsistence between the results in I_A index (predominance of mutation) and the splitgraph (predominance of recombination) can be explain on the basis of the amount of nucleotide substitutions produced by mutation between the two clusters (370 substitutions between the two groups) that are probably masking recombination events in the evolution of each group. Moreover,

the lack of clonal groups in the population structure obtained by eBURST algorithm can be explain for the amount of intermediated evolutive isolates missed in the population, represented as red dots in the graphic.

To date, the groups defined for *V. tapetis* have been associated to their host origin. In this work, we described species diversification for the first time on the basis of geographic origin. Distribution of identified STs for *V. tapetis* population is shown in **Figure 9**. In the analysis performed, both at phylogenetic and population levels, three isolates appear in an independent group. These isolates show different host origin, halibut and seed of Manila clam, but it have a common geographical origin, British Isles. These groups present huge genetic distance between them, produced mostly by mutation, supporting this finding that the description of two subspecies for *V. tapetis: V. tapetis* subsp *tapetis and V. tapetis* subsp *britanniensis* [106]. The former comprises the majority of the isolates regardless their geographical origin, whereas the latter includes the British isolates. *V. tapetis* shows a non-clonal, panmictic population in where the two subspecies are
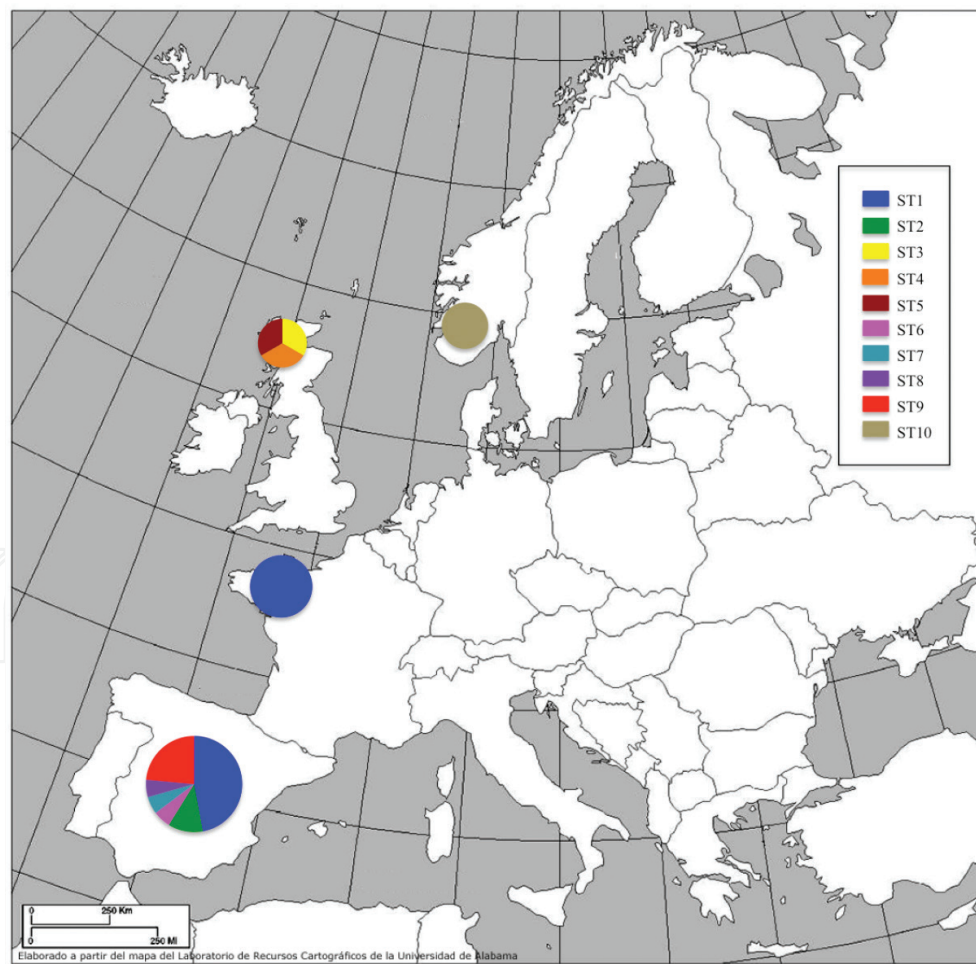
Figure 9. Geographic distribution of the ten STs identified for *V. tapetis*. Regions of the same country are considered as a unique localization.

generated by mutational events but the diversification within each of them is produced mostly by recombination [109].

6. Conclusions

In conclusion, MLSA, MLST and phylogeographic analysis are successful for (i) unambiguously genotyping both *Yersinia ruckeri* and *Vibrio tapetis* species, (ii) establishing evolutionary relationships among the bacterial populations at different levels and (iii) capturing geographical structure of these pathogens. The case studies reviewed here constitute good examples of the usefulness of these powerful tools for understanding the evolution, epidemiology and genetic population/landscape of bacterial pathogens.

The results obtained from our works suggest that the processes involved in the genetic variability and evolution in both species are different. Using the MLST approximation, two different expansion models of population were detected, a mutation-based epidemic model for *Y. ruckeri* and panmictic for *V. tapetis*, where recombination represented the genetic event contributing mostly to diversification.

The phylogeographic approach indicated that well-adapted clones of *Y. ruckeri* exploded to be widely distributed, while *V. tapetis* was divided into two defined groups being one of them associated to a specific geographical area.

It is noteworthy that the observed diversification, no matter the process suffered, could be related with host specificity to some extent, which may be indicating the existence of certain degree of function specialization. Further studies using "omics" techniques will allow to confirm such hypothesis.

Acknowledgements

Author details

Asmine Bastardo[1, 2], Sabela Balboa[1] and Jesús L. Romalde[2]*

*Address all correspondence to: jesus.romalde@usc.es

1 Universidade de Santiago de Compostela, Santiago de Compostela, Spain

2 Fundación La Salle de Ciencias Naturales, Ciudad Guayana, Venezuela

References

[1] Feil EJ, Li BC, Aanensen DM, Manage WP, Spratt EG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. Journal of Bacteriology. 2004;**186**:1518–1530. DOI: 10.1128/JB.186.5.1518-1530.2004.

[2] Spratt BG, Hanage WP, Li B, Aanensen, DM, Fell EJ. Displaying the relatedness among isolates of bacterial species—the eBURST approach. FEMS Microbiology Letters. 2004;**241**:129–134. DOI: 10.1016/j.femsle.2004.11.015.

[3] Stacy A, McNally L, Darch SE, Brown SP, Whiteley M. The biogeography of polymicrobial infection. Nature Reviews Microbiology. 2016;**14**:93–105. DOI: 10.1038/nrmicro.2015.8.

[4] Moodle Y, Linz B. *Helicobacter pylori* sequences reflect past human migrations. Genome Dynamics. 2009;**6**:62–74. DOI: 10.1159/000235763.

[5] Ekblom R, Wolf JBW. A field guide to whole-genome sequencing, assembly and annotation. Evolutionary Applications. 2014;7:1026–1042. DOI: 10.1111/eva.12178.

[6] Darwin C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 1st ed. John Murray. Bromley, Kent (UK), 1859.

[7] Devulder G, Perouse de Montclos M, Flandroies JP. A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. International Journal of Systematic and Evolutionary Microbiology. 2005;**55**:293–302. DOI: 10.1099/ijs.0.63222-0.

[8] Vinuesa P, Rojas-Jimenez K, Contrearas-Moreira B, Mahna SK, Prasad VN, Moe H, Selvaraju SB, Thierfelder H, Werner D. Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybeans on the Asiatic continent. Applied and Environmental Microbiology. 2008;**74**:6987–6996. DOI: 10.1128/AEM.00875-08.

[9] Glaeser SP, Kämpfer P. Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. Systematic and Applied Microbiology. 2015;**38**:237–245. DOI: 10.1016/j.syapm.2015.03.007

[10] Maynard-Smith SJ, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria? Proceedings of the Natural Academy of Sciences USA.1993;**90**:4384–4388. DOI: 10.1073/pnas.90.10.4384.

[11] Maiden MC. Multilocus sequence typing of bacteria. Annual Reviews of Microbiology. 2006;**60**:561–588. DOI: 10.1146/annurev.micro.59.030804.121325.

[12] Enright MC, Spratt BG. Multilocus sequence typing. Trends in Microbiology. 1999;**7**:482–487. DOI: 10.1016/S0966-842X(99)01609-1.

[13] Spratt BG. Multilocus sequence typing: Molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Current Opinion in Microbiology. 1999;**2**:312–316. DOI: 10.1016/S1369-5274(99)80054-X.

[14] Urwin R, Maiden MC. Multi-locus sequence typing: A tool for global epidemiology. Trends in Microbiology 2003;**11**:479–487. DOI: http://dx.doi.org/10.1016/j.tim.2003.08.006.

[15] Martyni et al., 2015

[16] Avise, JC (2000). Phylogeography: The history and formation of Species. Harvard University Press, Harvard. 464 p. ISBN: 9780674666382.

[17] Avise JC, Bowen BW, Ayala FJ. In the light of evolution X: Comparative phylogeography. Proceedings of the Natural Academy of Sciences USA. 2016;113:7957-7961. DOI: 10.1073/pnas.1604338113.

[18] Beheregaray LB, Caccone A. Cryptic biodiversity in a changing world. Journal of Biology. 2007;**6**:9. DOI: 10.1186/jbiol60.

[19] Gevers D, Cohan F, Lawrence J, Spratt B, Coenye T, Feil E, Stackebrandt E, Van de PY, Vandamme P, Thompson F, Swings J. Opinion: Reevaluating prokaryotic species. Nature Reviews Microbiology. 2005;**3**:733–739. DOI: 10.1038/nrmicro1236.

[20] Vinuesa P, Silva C, Werner D, Martínez-Romero E. Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in *Bradyrhizobium* species cohesion and delineation. Molecular Phylogenetics and Evolution. 2005;**34**:29–54. DOI: 10.1016/j.ympev.2004.08.020.

[21] Wertz JE, Goldstone C, Gordon DM, Riley MA. A molecular phylogeny of enteric bacteria and implications for a bacterial species concept, Journal of Evolutionary Biology. 2003;**16**:1236–1248. DOI: 10.1046/j.1420-9101.2003.00612.x.

[22] Konstantidinis KT, Ramett A, Tiedje. JM. The bacterial species definition in the genomic era", Philosophical Transactions of the Royal Society B. 2006;**361**:1929–1940. DOI: 10.1098/rstb.2006.1920.

[23] Lee I, Kim YO, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. International Journal of Systematic and Evolutionary Microbiology. 2015;**66**:1100–1103. DOI: 10.1099/ijsem.0.000760.

[24] Konstantidinis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Current Opinion in Microbiology. 2007;10:504–509. DOI: 10.1016/j.mib.2007.08.006.

[25] Richter M, Rosselló Mora R. Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the Natural Academy of Sciences USA. 2009;**106**:19126–19131. DOI: 10.1073pnas.0906412106.

[26] Vinuesa P. Multilocus sequence analysis and bacterial species phylogeny estimation. In Oren A, Papke RT. Molecular Phylogeny of Microorganisms. Caister Academic Press, Norfolk (UK), 2010. pp. 41–64. 2010 ISBN: 978-1-904455-67-7.

[27] Santos S, Ochman H. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. Environmental Microbiology. 2004;**6**:754–759. DOI: 10.1111/j.1462-2920.2004.00617.x.

[28] Degnan JH, Rosenber, NA. Discordance of species trees with their most likely gene tres. PLoS Genetics 2006;**2**:e68. DOI: 10.1371/journal.pgen.0020068.

[29] Maddison WP, Knowles LL. Inferring phylogeny despite incomplete lineage sorting. Systematic Biology. 2006;**55**:21–30. DOI: 10.1080/10635150500354928.

[30] Nichols, R. Gene trees and species trees are not the same. Trends in Ecology and Evolution. 2001;**16**:358–364. DOI: http://dx.doi.org/10.1016/S0169-5347(01)02203-0.

[31] Gevers D, Coenye T. (2007) Phylogenetic and genomic analysis. In Hurst CJ, Crawford RL,Garland JL, Lipson DA, Mills AL, Stetzenbach LD, editors. Manual of Environmental Microbiology 3rd ed. ASM Press, Washington DC. 2007. pp. 157–168. DOI: 10.1128/9781555815882.

[32] Kämpfer P, Glaeser SP. Prokaryote taxonomy in the sequencing era-the polyphasic approach revisited. Environmental Microbiology. 2012;**14**:291–317. DOI: 10.1111/j.1462-2920.2011.02615.x.

[33] Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K, Tobe T, Hattori M, Hayashi T. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. Proceedings of the Natural Academy of Sciences USA. 2009;**106**:17939–17944. DOI: 10.1073/pnas.0903585106.

[34] Young JPW, Crossman LC, B Johnston AWB, Thomson NRT, Ghazoui ZF, Hull KH, Wexler M, Curson ARJ, Todd JD, Poole,PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biology. 2006;**7**:R34. DOI: 10.1186/gb-2006-7-4-r34.

[35] Gogarte JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. Nature Reviews Microbiology. 2005;**3**:679–687. DOI: 10.1038/nrmicro1204.

[36] Ochman H, Lerat E, Daubin V. Examining bacterial species under the specter of genen transfer and exchange. Proceedings of the Natural Academy of Sciences USA. 2005;**102**:595–6599. DOI: 10.1073/pnas.0502035102.

[37] Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau ME, Nesbø CL, Case RJ, Doolittle, WF. Lateral gene transfer and the origins of prokaryotic groups. Annual Reviews of Genetics. 2003;**37**:283–328. DOI: 10.1146/annurev.genet.37.050503.084247

[38] Posada D, Crandall KA. The effect of recombination on the accuracy of phylogeny estimation. Journal of Molecular Evolution. 2002;**54**:396–402. DOI: 10.1007/s00239-001-0034-9.

[39] Schierup MH, Hein J. Consequences of recombination on traditional phylogenetic analysis. Genetics 2000;**156**:879–891.

[40] Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. Multilocus sequence typing:

A portable approach to the identification of clones within populations of pathogenic microorganisms. Proceedings of the Natural Academy of Sciences USA. 1998;**95**:3140–3145. DOI: 10.1073/pnas.95.6.3140.

[41] Brueggemann AB, Spratt, BG. Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. Journal of Clinical Microbiology. 2003;**41**:4966–4970. DOI: 10.1128/JCM.41.11.4966-4970.2003.

[42] Enright MC, Day NPJ, Davies CE, Peacock SJ, Spratt BG. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. Journal of Clinical Microbiology. 2000;**38**:1008–1015.

[43] Gonzalez-Escalona N, Martinez-Urtaza, J, Romero, J, Espejo RT, Jaykus LA, De Paola A. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing, Journal of Bacteriology. 2008;**90**:2831–2840. DOI: 10.1128/JB.01808-07.

[44] Feil EJ, Enright M. Analyses of clonality and the evolution of bacterial pathogens. Current Opinion in Microbiology. 2004;**7**:308–313. DOI: 10.1016/j.mib.2004.04.002.

[45] Boyd EF, Wang F-S, Whittam TS, Selander RK. Molecular genetic relationships of the *Salmonellae*. Applied and Environmental Microbiology. 1996;**62**:804–808.

[46] Maynard-Smith SJ. Analyzing the mosaic structure of genes. Journal of Molecular Evolution. 1992;**34**:126–129.

[47] Spratt BG, Maiden MC. Bacterial population genetics, evolution and epidemiology. Philosophical Transactions of the Royal Society B.1999;**354**:701–710. DOI: 10.1098/rstb.1999.0423.

[48] Navarro, F, Llovet, T, Echeita, MA, Coll, P, Aladuena, A, Usear, MA and Prats, G. Molecular typing of *Salmonella enterica* serovar *typhi.* Journal of Clinical Microbiology. 1996;**4**:2831–2834.

[49] Mane S, Schwartz MK, Luikart G, Taberlet P. Landscape genetics: combining landscape ecology and population genetics", Trends in Ecology and Evolution. 2003;**18**:189–197. DOI: 10.1016/S0169-5347(03)00008-9.

[50] Sites, JW, Jr and Marshall, JC. Delimiting species: a Renaissance issue in systematic biology. Trends in Ecology and Evolution. 2003;**18**:462–470. DOI: 0.1016/S0169-5347(03)00184-8.

[51] Thompson JN. The geographical mosaic of coevolution. The University of Chicago Press, Chicago, USA. 2005. 400 p. ISBN: 978-0226797618.

[52] Riddle BR, Hafner DJ. The past and future roles of phylogeography in historical biogeography. In Lomolino MV, Heaney LR, editors. Frontiers In Biogeography: New Directions in the Geography of Nature. M.V. Sinauer Associates, Inc., Sunderland, Mass, 2004. pp. 93-110. ISBN: 0878934790.

[53] Bridle JR, Baird SJE, Butlin, RK. Spatial structure and habitat variation in a grasshopper hybrid zone. Evolution. 2001;**55**:1832–1843. DOI: 10.1111/j.0014-3820.2001.tb00832.x.

[54] Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;**131**:479–491.

[55] Monmonier MS. Maximum-difference barriers: an alternative numerical regionalization method. Geographical Analysis. 1973;**5**:245–261. DOI: 10.1111/j.1538-4632.1973.tb01011.x.

[56] Miller MP. Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. Journal Heredity 2005;**96**:722–724. DOI: 10.1093/jhered/esi119.

[57] Slatkin M, Maddison WP. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics. 1989;**123**:603–613.

[58] Mitas L, Mitasova H. Spatial Interpolation. In Longley, P, Goodchild, MF, Maguire, DJ and Rhind, DW, editors. Geographical Information Systems: Principles, Techniques, Management And Applications. John Wiley and Sons, Hoboken, NJ. 1999. pp. 481-492. ISBN: 978-0-471-73545-8.

[59] Miller MP, Bellinger MR, Forsman ED, Haig SM. Effects of historical climate change, habitat connectivity and vicariance on genetic structure and diversity across the range of the red tree vole (*Phenacomys longicaudus*) in the Pacific Northwestern United States. Molecular Ecology. 2006;**15**:145–159. DOI: 10.1111/j.1365-294X.2005.02765.x.

[60] Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling, MA. Y-chromosomal diversity in Europe in clinal and influenced primarily by geography, rather than language. American Journal of Human Genetics. 2000;**67**:1526–1534. DOI: 10.1086/316890.

[61] Casteleyn G, Leliaert F, Backeljau T, Debeer AE, Kotaki Y, Rhodes L, Lundholm N, Sabbe K, Vyverman, W. Limits to gene flow in a cosmopolitan marine planktonic diatom. Proceedings of the Natural Academy of Sciences USA. 2010;**107**:12952-12957. DOI: 10.1073/pnas.1001380107.

[62] Verleyen E, Vyverman W, Sterken M, Hodgson DA, De Wever A, Juggins S, Van de Vijver B, Jones VJ, Vanormelingen P, Roberts D, Flower R, Kilroy C, Souffreau C, Sabbe K. The importance of dispersal related and local factors in shaping the taxonomic structure of diatom metacommunities. Oikos 2009;**118**:1239–1249. DOI: 10.1111/j.1600-0706.2009.17575.x.

[63] Hillebrand H, Azovsky AI. Body size determines the strength of the latitudinal diversity gradient. Ecography 2001;**24**:251–256. DOI: 10.1034/j.1600-0587.2001.240302.x.

[64] Cho JC, Tiedje JM. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil, Applied and Environmental Microbiology. 2000;**66**:5448–5456. DOI: 10.1128/AEM.66.12.5448-5456.2000.

[65] Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, Tuanyok A, Price EP, Glass MB, Leadem B, Beckstrom-Sternberg JS, Allan GJ, Foster JT, Wagner DM, Okinaka RT, Sim SH, Pearson O, Wu Z, Chang J, Kaul R, Hoffmaster AR, Brettin TS, Robison RA, Mayo M, Gee JE, Tan P, Currie BJ, Keim, P. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. BMC Biology. 2009;**7**:78. DOI: 10.1186/1741-7007-7-78.

[66] Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, Not F, Massana R, Ulloa O, Scanian DJ. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. Environmental Microbiology. 2008;**10**:147–161. DOI: 10.1111/j.1462-2920.2007.01440.x.

[67] Fierer N, Jackson, RB. The diversity and biogeography of soil bacterial communities. Proceedings of the Natural Academy of Sciences USA. 2006;**103**:626–631. DOI: 10.1073/pnas.0507535103.

[68] Jungblut AD, Lovejoy C,Vincent, WF. Global distribution of cyanobacterial ecotypes in the cold biosphere. The ISME Journal. 2010;**4**:191–202. DOI: 10.1038/ismej.2009.113.

[69] Van der Gucht K, Cottenie K, Muylaert K, Vloemans N, Cousin S, Jeppesen E, Conde-Porcuna J-M, Schwenk K. Zwart G, Degans H, Vyverman,W, de Meester L. The power of species sorting: Local factors drive bacterial communitycomposition over a wide range of spatial scales. Proceedings of the Natural Academy of Sciences USA. 2007;**104**:20404–20409. DOI: 10.1073/pnas.0707200104.

[70] Ionescu D, Hindiyeh M, Malkawi H, Oren, A. Biogeography of thermophilic cyanobacteria: insights from the ZerkaMa in hot springs (Jordan). FEMS Microbiology Ecology. 2010;**72**:103–113. DOI: 10.1111/j.1574-6941.2010.00835.x.

[71] Miller SR, Castenholz RW, Pedersen, D. Phylogeography of thethermophilic cyanobacterium *Mastigocladus laminosus*, Applied and Environmental Microbiology. 2007;**73**:4751-4759. DOI: 10.1128/AEM.02945-06.

[72] Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. Environmental Microbiology. 2009;**11**:823–832. DOI: 10.1111/j.1462-2920.2008.01803.x.

[73] Rucker R. Redmouth disease of rainbow trout (*Salmo gairdneri*). Bulletin Office International des Epizooties 1966;**65**:825–830.

[74] Austin B, Austin, D, editors. Bacterial fish pathogens. Diseases of farmed and wild fish. 4th edn. Praxis Publishing, Chichester. 2007. 552 p. DOI: 10.1007/978-1-4020-6069-4

[75] Davies RL, Frerichs, GN. Morphological and biochemical differences among isolates of *Yersinia ruckeri* obtained from wide geographical areas. Journal of Fish Diseases. 1989;**12**:357–365. DOI: 10.1111/j.1365-2761.1989.tb00324.x.

[76] Stevenson RMW, Airdrie DW. Serological variation among *Yersinia ruckeri* strains. Journal of Fish Diseases. 1984;7:247-254. DOI: 10.1111/j.1365-2761.1984.tb00930.x.

[77] Davies, RL. O-serotyping of *Yersinia ruckeri* with special emphasis on European isolates. Veterinary Microbiology 1990;**22**:299–307. DOI: 10.1016/0378-1135(90)90016-O.

[78] Romalde JL, Margariños B, Barja JL, Toranzo AE. Antigenic and molecular characterization of *Yersinia ruckeri*. Proposal for a new intraspecies classification. Systematic and Applied Microbiology. 1993;**16**:411–419. DOI: 10.1016/S0723-2020(11)80274-2.

[79] Davies RL. Clonal analysis of *Yersinia ruckeri* based on biotypes, serotypes and outer membrane protein types. Journal of Fish Diseases. 1991;**14**:221–228. DOI: 10.1111/j.1365-2761.1991.tb00591.x.

[80] Arias CR, Olivares-Fuster O, Hayden, K. First report of *Yersinia ruckeri* biotype 2 in the USA. Journal of Aquatic Animal Health. 2007;**19**:35–40. DOI: 10.1577/H06-011.1.

[81] Austin DA, Robertson PAW, Austin, B. Recovery of a new biogroup of *Yersinia ruckeri* from diseased rainbow trout (*Oncorhynchus mykiss*, Walbaum). Systematic and Applied Microbiology. 2003;**26**:127–131. DOI: 10.1078/072320203322337416.

[82] Fouz B, Zarza C, Amaro C. First description of non-motile *Yersinia ruckeri* serovar I strains causing disease in rainbow trout, *Oncorhynchus mykiss* (Walbaum), cultured in Spain. Journal of Fish Diseases. 2006;**29**:339–346. DOI: 10.1111/j.1365-2761.2006.00723.x.

[83] Bastardo A, Bohle H, Ravelo C, Toranzo AE, Romalde JL. Serological and molecular heterogeneity among *Yersinia ruckeri* strains isolated from farmed Atlantic salmon *Salmo salar* in Chile. Diseases of Aquatic Organisms. 2011;**93**:207–214. DOI: 10.3354/dao02296.

[84] Schill WB, Phelps SR, Pyle SW. Multilocus electrophoretic assessment of the genetic structure and diversity of *Yersinia ruckeri*. Applied and Environmental Microbiology. 1984;**48**:975–979.

[85] Lucangeli C, Morabito A, Caprioli A, Achene L, Busani L, Mazzolini E, Fabris A, Macri, A. Molecular fingerprinting of strains of *Yersinia ruckeri* serovar O1 and *Photobacterium damsela* subsp. *piscicida* isolated in Italy. Veterinary Microbiology. 2000;**76**:273–281. DOI: 10.1016/S0378-1135(00)00241-8.

[86] Wheeler RW, Davies RL, Dalsgaard I, Garcia J, Welch TJ, Wagley S, Bateman KS, Verner-Jeffreys DW. *Yersinia ruckeri* biotype 2 isolates from mainland Europe and the UK likely represent different clonal groups. Diseases of Aquatic Organisms. 2009;**84**:25–33. DOI: 10.3354/dao02039.

[87] Verner-Jeffreys DW, Haig S, Welch TJ, Pond MJ, Davies RL, Gradnier R. Characterisation of a serotype O1 *Yersinia ruckeri* isolates from the Isle of Man: further evidence that O antigen serotype is not a reliable indicator of virulence. Bulletin of the European Association of Fish Pathologists. 2011;**31**:86–91.

[88] Tinsley JW, Austin DA, Lyndon AR, Austin B. Novel non-motile phenotypes of *Yersinia ruckeri* suggest expansión of the current clonal complex theory. Journal of Fish Diseases. 2011;**34**:311–317. DOI: 10.1111/j.1365-2761.2011.01237.x.

[89] Merhej V, Adékambi T, Pagnier I, Raoult D and Drancourt M. *Yersinia massiliensis* sp. nov. isolated from fresh water. International Journal of Systematic and Evolutionary Microbiology. 2008;**58**:779–84. DOI: 10.1099/ijs.0.65219-0.

[90] Murros-Kontiainen A, Johansson P, Niskanen T, Fredriksson-Ahomaa M, Korkeala H, Björkroth J. *Yersinia pekkanenii* sp. nov. International Journal of Systematic and Evolutionary Microbiology. 2010;**61**:2363–2367. DOI: 10.1099/ijs.0.019984-0.

[91] Murros-Kontiainen A, Fredriksson-Ahomaa M, Korkeala H, Johansson P, Rahkila R, Björkroth J. *Yersinia nurmii* sp. nov. Int J Syst Evol Microbiol 2011;61:2368-2372. DOI: 10.1099/ijs.0.024836-0.

[92] Bastardo A, Ravelo C, Romalde, JL. Multilocus sequence typing (MLST) reveals high genetic diversity and epidemic population structure for the fish pathogen *Yersinia ruckeri*. Environmental Microbiology. 2012;**14**:1888–1897. DOI: 10.1111/j.1462-2920.2012.02735.x.

[93] Feil EJ, Enright MC, Spratt, BG. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. Research in Microbiology. 2000;**151**:465–469. DOI: 10.1016/S0923-2508(00)00168-6.

[94] Bastardo A, Ravelo C, Romalde JL. Phylogeography of *Yersinia ruckeri* reveals effects of past evolutionary events on the current strain distribution and explains variations in the global transmission of enteric redmouth (ERM) disease. Frontiers in Microbiology. 2015;**6**:1198. DOI: 10.3389/fmicb.2015.01198.

[95] Welch TJ, Verner-Jeffreys DW, Dalsgaard I, Wiklund T, Evenhuis JP, Cabrera JA, Hinshaw JM, Drennan JD, LaPatra SE. Independent emergence of *Yersinia ruckeri* biotype 2 in the United States and Europe. Applied and Environmental Microbiology. 2011;**77**:3493–3499. DOI: 10.1128/AEM.02997-10.

[96] Borrego JJ, Castro D, Luque A, Paillard C, Maes P, García MT, Ventosa A. *Vibrio tapetis* sp. nov., the causative agent of the brown ring disease affecting cultured clams. International Journal of Systematic Bacteriology. 1996;**46**:480–484. DOI: 10.1099/00207713-46-2-480.

[97] Borrego JJ, Romalde JL and Castro D. Brown ring disease. In: Figueras A, Novoa B, editors. Diseases of bivalve molluscs of interest in aquaculture. 1st ed. Fundación OESA, Madrid (Spain), 2011. pp. 149-179 (in Spanish). ISBN: 978-84-00-09288-7.

[98] Paillard C. A short-review of brown ring disease, a vibriosis affecting clams, *Ruditapes philippinarum* and *Ruditapes decussatus*. Aquatic Living Resources. 2004;17:467-475. DOI: 10.1051/alr:2004053.

[99] López JR, Balboa S, Núñez S, de la Roca E, de la Herran R, Navas JI, Toranzo AE, Romalde JL. Characterization of *Vibrio tapetis* strains isolated from diseased cultured Wedge sole (*Dicologoglossa cuneata* Moreau). Research in Veterinary Sciences. 2011;**90**:189–195. DOI: 10.1016/j.rvsc.2010.05.030.

[100] Novoa B, Luque A, Castro D, Borrego JJ, Figueras A. Characterization and infectivity of four bacterial strains isolated from brown ring disease-affected clams. Journal of Invertebrate Pathology. 1998;**1**:34–41. DOI: 10.1006/jipa.1997.4704.

[101] Castro D, Romalde JL, Vila J, Magariños B, Luque A, Borrego, JJ. Intraspecific characterization of *Vibrio tapetis* strains by use of pulsed-field gel electrophoresis, ribotyping and plasmid profiling. Applied and Environmental Microbiology. 1997;**63**: 1449–1452.

[102] Romalde JL, Castro D, Magariños B, López-Cortés L, Borrego JJ. Comparison of ribotyping, randomly amplified polymorphic DNA and pulsed-field gel electrophoresis for molecular typing of *Vibrio tapetis*. Systematic and Applied Microbiology. 2002;**25**:544–550. DOI: 10.1078/07232020260517689.

[103] Rodríguez JM, López-Romalde S, Beaz R, Alonso MC, Castro D, Romalde JL. Molecular fingerprinting of *Vibrio tapetis* strains using three PCR-based methods: ERIC-PCR, REP-PCR and RAPD. Diseases of Aquatic Organisms. 2006;**69**:175–183. DOI: 10.3354/dao069175.

[104] Balboa S, Bermúdez-Crespo J, Gianzo C, Romalde JL. Proteomics and multilocus sequence analysis confirm intraspecific variability of *Vibrio tapetis*. FEMS Microbiology Letters. 2011;**324**: 80–87. DOI: 10.1111/j.1574-6968.2011.02393.x.

[105] Pascual J, Macián MC, Arahal DR, Garay E, Pujalte MJ. Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, *rec*A, *pyr*H, *rpo*D, *gyr*B, *rct*B and *tox*R genes. International Journal of Systematic and Evolutionary Microbiology. 2010;**60**:154–65. DOI: 10.1099/ijs.0.010702-0.

[106] Balboa S, Romalde JL. Multilocus sequence analysis of *Vibrio tapetis*, the causative agent of Brown Ring Disease: description of *Vibrio tapetis* subsp. *britannicus* subsp. nov. Systematic and Applied Microbiology. 2013;**36**:183–7. DOI: 10.1016/j.syapm.2012.12.004. DOI: 10.1016/j.syapm.2012.12.004.

[107] Castro D, Santamaría JA, Luque A, Martínez-Manzanares E, Borrego, JJ. Antigenic characterization of the etiological agent of the brown ring disease affecting manila clams. Systematic and Applied Microbiology. 1996;**19**:231–239. DOI: 10.1016/S0723-2020(96)80049-X.

[108] Didelot X, Falush, D. Inference of bacterial microevolution using multilocus sequence data. Genetics 2007;**175**:1251–1266. DOI: 10.1534/genetics.106.063305.

[109] Balboa S, Bastardo A, Romalde JL. Disentangling the population structure and evolution of the clam pathogen *Vibrio tapetis*. Microbial Ecology. 2014;**67**:145–154. DOI: 10.1007/s00248-013-0318-9.