NEW MICROBIOLOGICA, **33**, 359-372, 2010

Phylogenetic and evolutionary analysis of *Vibrio parahaemolyticus* and *Vibrio alginolyticus* isolates based on toxR gene sequence

Stefania Montieri[1]*, Elisabetta Suffredini[2]*, Massimo Ciccozzi[1], Luciana Croci[2]

[1]Istituto Superiore di Sanità, Dipartimento di Malattie Infettive, Parassitarie e Immunomediate, Reparto Epidemiologia, Epidemiologia molecolare e Filogenesi microbica, Roma, Italy
[2]Istituto Superiore di Sanità, Dipartimento di Sanità Pubblica Veterinaria e Sicurezza Alimentare, Reparto Adempimenti Comunitari e Sanità Pubblica, Roma, Italy

**These authors contributed equally to the work*

SUMMARY

The *Vibrio* genus contains a large number of closely related bacterial species differing, in some cases, less than 1% in 16S rRNA gene sequence. The present study evaluated the usefulness of *toxR* gene for phylogenetic and evolution analysis on *Vibrio* isolates of environmental or clinical origin belonging to the two closely related species *V. parahaemolyticus* and *V. alginolyticus*. The phylogenetic analysis based on *toxR* gene, contrary to 16S rRNA gene, allowed a clear differentiation of the isolates belonging to the two species and showed the presence of two separate, statistically supported clusters in *V. alginolyticus* (subgroup A and B). Such division, partially reflected in the biochemical features of the isolates, could not be explained by spatial and/or temporal distance in the isolation, leading to the hypothesis of two distinct, co-existing clusters in the *V. alginolyticus* isolates analysed. The evolutionary analysis on the *toxR* sequence showed that while the substitutions inferred from the alignment of *V. parahaemolyticus* are best explained by the negative/neutral selection model, in *V. alginolyticus* – and particularly in subgroup B – is acting a positive evolutionary pressure. The site detected as under diversifying selection (P164L) could be related to conformational changes of ToxR protein.

KEY WORDS: V. parahaemolyticus, *V. alginolyticus*, 16S rRNA gene, *toxR* gene, phylogenetic analysis

Received March 02, 2010 *Accepted* June 06, 2010

INTRODUCTION

The genus *Vibrio*, belonging to the *Gammaproteobacteria* on the basis of 16S rRNA gene sequence analysis (Bergey *et al.*, 1984), contains a large number of closely related bacterial species differing in nucleotide sequence from less than 1% up to 6% (Dorsch *et al.*, 1992; Ruimy *et al.*, 1994). Several novel *Vibrio* species, isolated mainly from the aquatic environment and from marine organisms, have been described in the last few years (Thompson *et al.*, 2005) and, in spite of the rapid advancement in the molecular taxonomy of *Vibrio* genus, the level of resolution between closely related species and between individual strains is still limited.

The use of 16S rRNA gene sequencing does not appear to be particularly suitable for the discrimination of closely related species, therefore the application of other genomic analyses, including DNA-DNA hybridisation, repetitive extragenic palindromic PCR, amplified fragment length polymorphism (AFLP) and multilocus sequence analysis has been considered (Thompson *et al.*, 2005). However, as these techniques are applied only by some specialized laboratories, the use of different genetic loci (*rpoA*, *recA*, *pyrH* and *atpA* genes) has been also proposed for phyloge-

Corresponding author
Dott.ssa Luciana Croci
Dipartimento di Sanità Pubblica Veterinaria
e Sicurezza Alimentare
Istituto Superiore di Sanità
Viale Regina Elena, 299 - 00161 Rome, Italy
E-mail: luciana.croci@iss.it

netic purposes (Thompson *et al.*, 2007; Thompson *et al.*, 2005) and, even if these genes have been recognized useful for discrimination of *Vibrio* species, Thompson *et al.* (2007) encouraged to analyse other genetic loci that present more discriminatory power to differentiate closely related species (Thompson *et al.*, 2007).

Among *Vibrio* species *V. parahaemolyticus* and *V. alginolyticus* are two important closely related species (99.8% of identical nucleotides) (Gomez-Gil *et al.*, 2004; Osorio and Klose, 2000), with similar phenotypic features that in some cases make difficult the biochemical differentiation of the isolates (Croci *et al.*, 2007).

V. parahaemolyticus, which includes toxigenic strains, is a recognized cause of seafood-related gastroenteritis and traveler's diarrhea in humans (Fuenzalida *et al.*, 2006; Martinez-Urtaza *et al.*, 2004) while *V. alginolyticus* is one of the most important pathogens in aquaculture, causing serious damage in shellfish and crustaceans (Hormansdorfer *et al.*, 2000) and has been sporadically associated with episodes of human infection (wound infection, seawater-related otitis, etc.) (Ho *et al.*, 1998; Matsiota-Bernard and Nauciel, 1993; Mukherji *et al.*, 2000).

Some authors have considered the virulence regulatory gene *toxR*, that appears to be well conserved among *Vibrio* species, a useful tool for the distinction of various species (Kim *et al.*, 1999; Osorio and Klose, 2000) particularly in relation to the significant gap between its maximun intraspecific and minimum interspecific distances (Pascual *et al.*, 2010). The regulatory protein ToxR (first described in *Vibrio cholerae* and then reported in other bacteria) contains a cytoplasmic DNA-binding-transcriptional activation domain, a transmembrane domain and a periplasmic domain (Lin *et al.*, 1993; Reich and Schoolnik, 1994). The alignment of ToxR amino acid sequences revealed a region between transcriptional activation and transmembrane domains ("tether" region) that displays wide divergence among *Vibrio* species. *V. parahaemolyticus* and *V. alginolyticus*, have only 61.7% identical nucleotides within the partial *toxR* sequence (Osorio and Klose, 2000).

The aim of the present study was to verify the usefulness of the *toxR* gene as a molecular marker in intraspecific phylogenetic analysis of *V. parahaemolyticus* and *V. alginolyticus*. The molecular evolution and diversity in the sequences of the two species was also examined to assess if a selective pressure is acting on *toxR* gene.

MATERIALS AND METHODS

Bacterial isolates: source and maintenance

A total of 48 isolates of environmental and clinical source, as well as two *Vibrio* reference strains (American Type Culture Collection - ATCC, Manassas, VA, USA), used for the present study are listed in Table 1.

The strains, selected from the Istituto Superiore di Sanità's culture collection, had been isolated during routine surveillances on national and imported fisheries, shellfish and seawater or had been provided from other national and international Laboratories. The isolates collected from Adriatic Sea in 2003 (Table 1) were obtained as a part of a single monitoring program carried out from June to December on shellfish production areas of the delta of Po river and samples collection was accompanied by recording of the environmental conditions (temperature and salinity ranging, respectively, between 12 to 29°C and from 22 to 34 ppm).

Long-term storage of isolates was performed at –80°C in commercially available cryogenic vials (Microbank™, Prolab Diagnostic, Neston, UK). All strains were subjected to biochemical and molecular identification before being included in the study.

Identification of bacterial isolates

Biochemical tests

The isolates were screened on the basis of colony morphology on TCBS, sucrose utilization, oxidase test, salt requirement (growth on 0, 3, 6, 8 and 10% NaCl medium), growth at different temperatures (37, 41.5 and 44°C), resistance to vibriostatic O129 (10 and 150 µg) and reactions on triple sugar iron (TSI) agar slant (Oliver and Kaper, 1997; Ripabelli *et al.*, 1997). Isolates with reactions characteristic for *V. parahaemolyticus* and *V. alginolyticus* were subjected to identification by means of two different miniaturized biochemical tests (API 20E and API 20NE, bioMerieux, Marcy l'Etoile, France) according to the manufacturer's instructions. Identification was obtained through the APILAB PLUS 3.3.3

software (bioMerieux) and isolates with identification probability higher than 95% in both tests were subjected to confirmation by molecular identification (PCR).

Molecular tests

Bacteria, cultured on 2.5% NaCl tryptone soya broth (TSB-S, Oxoid) at 30±1°C for 24 h were extracted by boiling. *Vibrio parahaemolyticus* iden-

TABLE 1 - *Vibrio strains used in the study.*

*Species and reference number**	*Source*	*Origin*	*Year*
Vibrio alginolyticus: 33			
Va020	fishery products	Cuba	2001
Va125, Va126	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va129	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va138	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va150	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va187	shellfish (*Mytilus galloprovicialis*)	Adriatic Sea (Italy)	2003
Va226, Va227	shellfish (*Mytilus galloprovicialis*)	Adriatic Sea (Italy)	2003
Va236	shellfish (*Mytilus galloprovicialis*)	Adriatic Sea (Italy)	2003
Va239	shellfish (*Mytilus galloprovicialis*)	Adriatic Sea (Italy)	2003
Va253, Va257	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va271, Va272	shellfish (*Mytilus galloprovicialis*)	Adriatic Sea (Italy)	2003
Va289	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va292, Va293	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va295	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va336	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va339	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va347	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va350	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va356, Va357	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va371, Va372	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va373, Va375, Va376	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va407	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va412	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Va487	fishery products	Adriatic Sea (Italy)	2002
Vibrio parahaemolyticus: 15			
Vp006	seawater	Adriatic Sea (Italy)	1994
Vp011	fishery products	Cuba	2001
Vp103	seawater	Adriatic Sea (Italy)	1994
Vp115	clinical sample	Mozambique	2002
Vp116	clinical sample	Mozambique	2002
Vp136	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Vp195	shellfish (*Mytilus galloprovicialis*)	Adriatic Sea (Italy)	2003
Vp250, Vp252	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Vp294	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Vp299	shellfish (*Tapes philippinarum*)	Adriatic Sea (Italy)	2003
Vp413	seawater	Adriatic Sea (Italy)	2002
Vp432	frozen cuttlefish	Senegal	2003
Vp481	shellfish	Adriatic Sea (Italy)	2002
Vp483	frozen fish	Japan	2003
Reference strains: 2			
Vibrio alginolyticus ATCC 17749			
Vibrio parahaemolyticus ATCC 43996			

*Isolates reported on the same line were obtained from the same shellfish sample.

TABLE 2 - *Regions amplified and primers used in the study.*

Species	*Region amplified*	*Primers*	*Expected PCR amplicon size*	*Position and GenBank reference*			*Reference*
V. parahaemolyticus	*toxR*	forw rev	5'- GTCTTCTGACGCAATCGTTG -3' 5'- ATACGAGTGGTTGCTGTCATG -3'	366 bp	600-619 946-966	L11929	(Kim et al., 1999)
V. alginolyticus	*collagenase*	forw rev	5'- CGAGTACAGTCACTTGAAAGCC -3' 5'- CACAACAGAACTCGCGTTACC -3'	737 bp	1526-1547 2242-2263	E03106	(Di Pinto et al., 2005)
V. cholerae	*ompW*	forw rev	5'- CACCAAGAAGGTGACTTTATTGTG -3' 5'- GAACTTATAACCACCCGCG -3'	587 bp	221-244 790-808	X51948	(Nandi et al., 2000)
V. vulnificus	*vvh*	forw rev	5'- CGCCGCTCACTGGGGCAGTGGCTG -3' 5'- CCAGCCGTTAACCGAACCACCCGC -3'	387 bp	1470-1493 1834-1857	M34670	(Brauns et al., 1991)
All species	*16S rRNA*	63f 763r	5'- CAGGCCTAACACATGCAAGTC -3' 5'- GCATCTGAGTGTCAGTATCTGTCC -3'	700 bp	43-63 740-763	J01859	(Marchesi et al., 1998) this study
V. parahaemolyticus *V. alginolyticus*	*toxR*	VaVpToxRf VaVpToxRr	5'- GAGCARGGKTTYGAGGTGG -3' 5'- CRCCWGTGGCRATYACYTC -3'	570 bp	343-362 895-913	L11929	this study

tification was confirmed through *toxR* gene detection as elsewhere described (Kim *et al.*, 1999), while *Vibrio alginolyticus* was tested for collagenase gene according to Di Pinto *et al.* (Di Pinto *et al.*, 2005). Primer sequences, region amplified and publication references are summarized in Table 2. All reactions were performed in a volume of 25 µl on a 9600 Applied Biosystems thermalcycler. *Vibrio parahaemolyticus* ATCC 43996 and *Vibrio alginolyticus* ATCC 17749 were used as positive controls respectively for *toxR* gene and collagenase gene detection; a negative control (sterile distilled water) was processed every ten strains. PCR products were visualized by electrophoresis on 1.5% agarose gel (Kodak, New Haven, CT, USA) (run at 90V for 50 min) and photographed using a Bio-Rad Gel Doc 2000 (Bio-Rad Laboratories, Hercules, CA, USA).

Evaluation of possible misidentifications of the isolates

In order to exclude misidentification of other species like *V. parahaemolyticus* and *V. alginolyticus*, the results of all screening and biochemical tests were compared with those available in literature for the species with closest phylogenetic relations according to previously published papers (Thompson *et al.*, 2005) (i.e. *V. diabolicus*, *V. mytili*, *V. proteolyticus*, *V. natriegens*, *V. harveyi*, *V. rotiferianus*, *V. campbellii*, *V. navarrensis*, *V. vulnificus*, *V. xuii*, *V. nereis*, *V. neptunis*, *V. corallilyticus*, *V. tubiashi*, *V. aestuarianus*, *V. anguillarum*, *V. ordalii*, *V. brasiliensis*).

Misidentification with certain species as *V. cholerae*, *V. vulnificus* and *V. harveyi* was also excluded by subjecting the isolates to the PCR assays for the identification of these species (Table 2).

Analysis of phenotypic features of V. alginolyticus

V. alginolyticus isolates were divided according to the results of phylogenetic analysis and the biochemical test results of API 20E and API 20NE were subjected to chi-square test with Yates' correction for continuity. Comparison of the other phenotypic features as growth at different temperatures (41.5 and 44°C) and salinity concentrations (8% and 10%), swarming on Marine Agar 2216 (BD Difco, Franklin Lakes, NJ, USA) at 25°C was performed with the same division criteria.

Amplification and sequencing of 16S rRNA and *toxR* gene

Nucleic acid from the isolates and from the two reference strains (Table 1) was extracted by a phenol-chloroform-CTAB procedure (Ausubel, 1995) and subjected to amplification and sequencing of 16S rRNA gene and *toxR* gene.

For 16S rRNA gene amplification a previously published primer, 63f (Marchesi *et al.*, 1998) and a newly designed primer, 763r (Table 2) were used; amplification was performed using the cycling conditions reported in Marchesi *et al.*

(Marchesi *et al.*, 1998) and the PCR mix described in Oakey *et al.* (Oakey *et al.*, 2003). Amplification of *toxR* gene was performed using two specifically designed primers, VaVpToxRf and VaVpToxRr (Table 2) and the above mentioned PCR conditions, with the exception of annealing temperature that was 56°C. Primer design was performed using the Oligo Perfect™ Designer with the default parameters. PCR products (700 bp for 16S rRNA gene and 570 bp for *toxR* gene), visualized on a 1.5% agarose gel (run at 90V for 50 min), were purified from the gel using the NucleoSpin Extract kit (Macherey-Nagel, Düren, Germany). Direct sequencing of purified PCR products was performed on both strands by MWG Biotech (Ebersberg, Germany). Sequences were deposited at GenBank with the accession numbers from EU155487 to EU155599.

Phylogenetic analysis

The genetic variation both at 16S rRNA locus and at the *toxR* gene were investigated to determine the evolutionary genetic relationships among our *V. parahaemolyticus* and *V. alginolyticus* isolates. All the sequences in the dataset were aligned using CLUSTAL X software (Thompson *et al.*, 1997) and then were manually edited with the BioEdit software (Hall, 1999). A 84 bp fragment of 16S rRNA gene (corresponding to nucleotides 412 to 496 of reference sequence accession number X56576), including the most variable region of the sequence, and a 459 bp fragment of *toxR* gene (amino acids 80 to 232 of reference sequences accession number L11929) were used. Sequences uploaded from NCBI GenBank (7 for 16S rRNA and 11 for *toxR* gene) were included in both cases.

The phylogenetic trees constructed by alignment of the partial sequences of *toxR* gene and 16S rRNA were estimated using the PAUP* package (Swofford, 2002). The models of nucleotide substitution used were HKY (Hasegawa, Kishino and Yano, 1985) (Hasegawa *et al.*, 1985) for *toxR* gene and K80 (Kimura, 1980) (Kimura, 1980) for 16S rRNA locus, as they consistently gave the highest likelihood value; both models incorporated Maximum Likelihood (ML) estimates of base composition and the shape parameter (α) of a gamma distribution (Γ) model of among-site rate variation with eight rate categories. The two models assume a time-reversible process, respectively a non-uniform and a uniform distribution of nucleotides and different rates for transitions and transversions (Yang, 1994). Maximum Likelihood trees were estimated under these models using tree bisection-reconnection (TBR) branch swapping. The statistical robustness and reliability of the branching order within each phylogenetic trees were confirmed through a bootstrap analysis using 1000 replicates for the Neighbor-Joining (NJ) tree and through the Zero Branch Length Test for the ML tree (Swofford, 2002).

Evolutionary analysis

Since *toxR* gene encodes for a protein product, an evolutionary analysis on the codon sites was performed to define the kind of evolutionary forces acting on this gene and to establish a more robust inference of the evolutionary history between *V. parahaemolyticus* and *V. alginolyticus*. Positively selected sites were identified using a series of models of codon substitution (Yang *et al.*, 2000). All models required a phylogenetic tree as input tree, generated excluding all reference sequences. The nonsynonymous/synonymous rate ratio (ω = dN/dS), is an important indicator of selective pressure at the protein level, the probabilistic models either estimated or fixed the ω parameters.

Six evolutionary models were used to detect positive selection. The simplest model, M0, calculated a single ω parameter for all sites, whereas M1a model accounted for nearly neutral evolution, where $0<\omega_0<1$ was estimated from the data, avoiding false positives at sites under weak selection, that would be absorbed in this neutral class rather than being claimed to be under positive selection. The M2a model accounted for a positive selection, adding, to the classes of sites under negative or purifying (p_0) and neutral (p_1) selection, an extra class (p_2) under positive or diversifying selection, for which ω_2 was estimated from the data. The M3 model provided a test for positive selection with a discrete distribution to estimate the ω ratio for three classes (p_0, p_1 and p_2). In particular, models M7 and M8, developed by Yang *et al.* (Yang, 1997) and implemented in the CODEML program of the PAML package (Yang, 1994), were also used: M7 assumed that codons are divided into ten classes, beta distributed, whose values lie between 0 and 1 (neutral- and negative-selection models), whereas M8 included the ten classes of M7 plus an additional codon class where ω could be >1 allowing

for positively selected sites. Nested models were compared directly using a Likelihood Ratio Test (LRT), namely M0 vs M3, M1a vs M2a and M7 vs M8 (Nielsen and Yang, 1998). The LRT indicated whether the substitutions inferred from an alignment was best explained by either the negative/neutral or the positive selection model. The LRT was performed by taking twice the difference in log likelihood between nested models and testing for significance using the χ^2 distribution with the degrees of freedom (df) equivalent to the difference in the number of parameters between models. If the LRT was significant, positive selection was inferred. The Bayes empirical Bayes (BEB) approach implemented in M2a and M8 was used to determine the positive selected sites by calculating the posterior probabilities (p) of ω classes for each sites (Yang *et al.*, 2005). The sites with high posterior probabilities (p>0.95) coming from the class with ω>1 were believed to be under positive selection. It is worth noting that PAML LRTs are reported to be conservative for short sequences (e.g. positive selection could be underestimated), although the Bayesian prediction of sites under positive selection is largely unaffected by sequence length (Anisimova *et al.*, 2001; Anisimova *et al.*, 2002).

The evolutionary models were chosen as best fitting models for the ML tree used as input tree in the program PAML, to investigate the presence of codons under positive selection in the bacterial gene.

A Maximum Likelihood approach was used to examine selection pressure on *toxR* gene fragment separately in the isolates of the two species, *V. parahaemolyticus* and *V. alginolyticus*.

GenBank accession numbers

The accession number of the references sequences used to perform the phylogenetic analysis were: *V. cholerae* (Vch1) M21249, (Vch2) AJ554204; *V. campbellii* (Vca1) AY946037, (Vca2) AY946038, (Vca3) DQ980029, (Vca4) AY738129, (Vca5) AB195982; *V. harveyi* (Vha1) AY247418, (Vha2) DQ403146, (Vha3) DQ503438, (Vha4) DQ517446, (Vha5) EF492029, (Vha6) AY750578; *V. vulnificus* (Vvu1) AF170883, (Vvu2) AB175476, (Vvu3) AB175477, (Vvu4) AB175478, (Vvu5) X76333; *V. parahaemolyticus* L11929; *V. alginolyticus* X56576.

RESULTS

Phylogenetic analysis

Within the 84 bp region of the 16S rRNA gene of the *V. alginolyticus* isolates examined there were 15 variable sites (containing at least two types of nucleotides) and 11 parsimony-informative sites (containing at least two types of nucleotides with two or more of them occurring with a minimum frequency of two). Among the *V. parahaemolyticus* isolates examined at the 16S rRNA locus, there were 11 variable sites and two parsimony-in-

TABLE 3 - *Genetic variation at the 16S rRNA locus and toxR gene among V. parahaemolyticus and V. alginolyticus isolates.*

Gene and group of isolates analysed	*Identity values median (min-max)*		*mean*	*SD*
16S rRNA gene				
V. parahaemolyticus	0,998	(0,981-1)	0,996	0,005
V. alginolyticus[1]	0,940	(0,892-1)	0,946	0,031
V. parahaemolyticus + *V. alginolyticus*	0,952	(0,833-1)	0,945	0,038
***toxR* gene**				
V. parahaemolyticus	0,984	(0,967-1)	0,984	0,006
V. alginolyticus[2]	0,967	(0,795-1)	0,909	0,082
V. alginolyticus subgroup A	0,986	(0,952-1)	0,984	0,012
V. alginolyticus subgroup B	0,984	(0,954-1)	0,982	0,010
V. parahaemolyticus + *V. alginolyticus*	0,816	(0,686-1)	0,825	0,123
V. parahaemolyticus + *V. alginolyticus* subgroup A	0,731	(0,694-1)	0,850	0,134
V. parahaemolyticus + *V. alginolyticus* subgroup B	0,958	(0,686-1)	0,842	0,141

[1]G+C content (%) = 46.9±2.3; [2]G+C content (%) = 45.5±0.7.

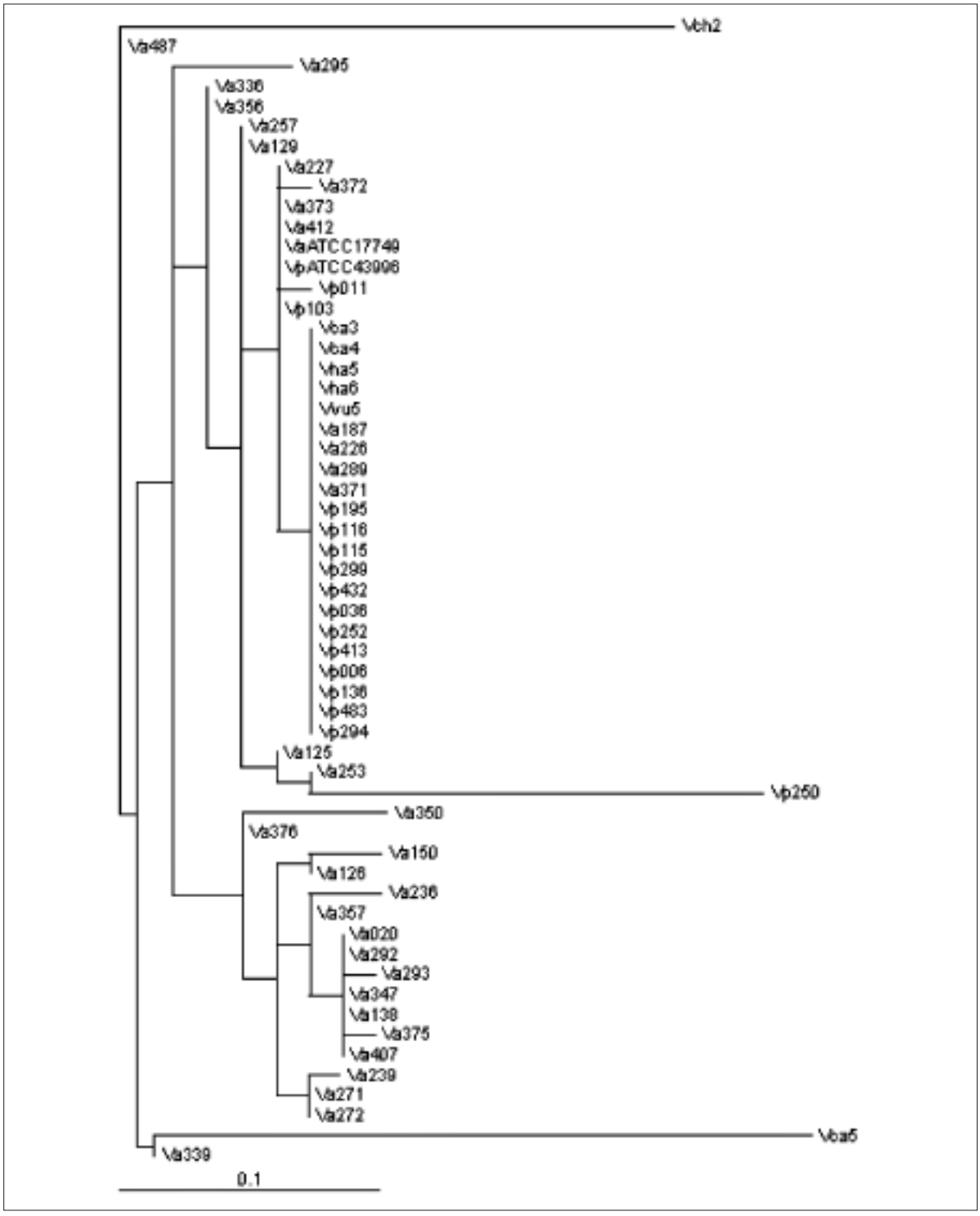

FIGURE 1 - *Phylogenetic relationship was defined using 34 sequences from V. alginolyticus (Va) and 16 sequences from V. parahaemolyticus (Vp) isolates; 7 sequences uploaded from NCBI GenBank were included in the analysis (Vca = V. campbellii, Vch = V. cholerae, Vha = V. harveyi, Vvu = V. vulnificus; see "GenBank accession numbers" in Materials and Methods). The scale bar indicates 10% nucleotide sequence divergence.*

formative sites. The similitude matrix for the nucleotide alignment (Table 3) gave a median value of 0.998 for *V. parahaemolyticus* isolates, 0.940 for *V. alginolyticus* isolates and 0.952 for the isolates of both species considered together (with a minimum value of 0.833). As expected, similitude among the sequences of the strains belonging to the two species was high. Clustering of the sequences in two clades (Fig. 1), a *V. parahaemolyticus* and a *V. alginolyticus* clade, confirmed the biochemical and molecular identifications performed. In both clades, however, some isolated 16S rRNA sequences appeared to intermingle with the other species.

Within the 459 bp region of *toxR* gene of the *V. alginolyticus* isolates under analysis there were 122 variable sites and 311 parsimony-informative sites, the number of synonymous substitutions was 72 and that of non-synonymous substitutions 64. Among the *V. parahaemolyticus* isolates, there were 21 variable sites and two parsimony-informative sites; the number of synonymous substitutions was 16 while non-synonymous substitutions were 10. The similitude matrix for the nucleotide alignment (Table 3) produced a median value of 0.984 for *V. parahaemolyticus* isolates (minimum 0.967), 0.967 for *V. alginolyticus* isolates (minimum 0.795) and 0.816 for the isolates of both species considered together (with a minimum value of 0.686).

The phylogenetic tree constructed using the *toxR* gene (Fig. 2) displayed greater divergence, based on nucleotide differences, than the phylogenetic tree constructed with the 16S rRNA gene from the same isolates (Fig. 1), with the *V. parahaemolyticus* and *V. alginolyticus* isolates forming separate divergent branches, without intermixing of sequences.

The Pearson's correlation coefficient of the similarity matrices of 16S rRNA gene and *toxR* gene was 0.019. Besides the different resolution of the two species, the *toxR* maximum likelihood tree showed that while the *V. parahaemolyticus* group clustered in a unique clade (100% of bootstrap value and a branch supported by a significant value of the Zero Branch Length Test), the *V. alginolyticus* clade split into two different subgroups (A and B), with the node of each subgroup supported by statistical robustness and reliability of the branching order (respectively a bootstrap value of 100% and 97% for the node of subgroup A and B and a p value <0.001 of the Zero Branch Length Test for both nodes). The analysis of the identity matrix of nucleotide alignment of *toxR* gene (Table 3) showed that while the median value of the similitude matrix for all *V. alginolyticus* isolates was 0.967 and the minimum value was 0.795, for subgroup A and B the median value was, respectively, 0.986 and 0.984 and the minimum 0.952 and 0.954.

An analysis of the biochemical features of the two subgroups of *V. alginolyticus* showed a statistically significant difference ($P<0.05$) in the frequency of positive results to the ornithine decarboxylase (ODC) and gelatine hydrolysis (GEL) tests in API 20E (for subgroup A and B respectively 18% and 64% in ODC test, 36% and 77% in GEL test) and PNPG (β-galactosidase) and arabinose assimilation (ARAa) in API 20NE (0% in subgroup A vs 80% in subgroup B for PNPG test and 0% vs 50% for ARAa). No significant differences were detected in other phenotypic features as growth at different temperatures and salinity concentrations or swarming.

The multiple alignment of *toxR* sequences (*V. alginolyticus* subgroup A and B and *V. parahaemolyticus*) showed that the membrane "tether" region is flanked by relatively conserved sequences corresponding to the transcription activation, transmembrane and periplasmic domains of the ToxR protein as described in a previous study by Osorio and Klose (Osorio and Klose, 2000). Although the "tether" region is considered the interspecific hypervariable domain, at intraspecific and intragroup level, the region contains conserved sequences distinctive of each species and group.

Focusing the attention on *V. alginolyticus* subgroup A, subgroup B and *V. parahaemolyticus*, there are six sites, T120A/S, P138H/S, V140A/N, L142I/V, A158T/V, T167P/E (aa of *V. alginolyticus* subgroup A _ site number _ aa of *V. alginolyticus* subgroup B / aa of *V. parahaemolyticus*; numbering based on *V. parahaemolyticus* ATCC 43996), absolutely conserved in each group, but different among them. In some cases, the biochemical properties of amino acid at this site change, for example in site 167, where the threonine (polar and neutral) present in *V. alginolyticus* subgroup A is substituted by a proline (nonpolar and neutral) in *V. alginolyticus* subgroup B and by a glutamic acid (polar and acid) in *V. parahaemolyti-*

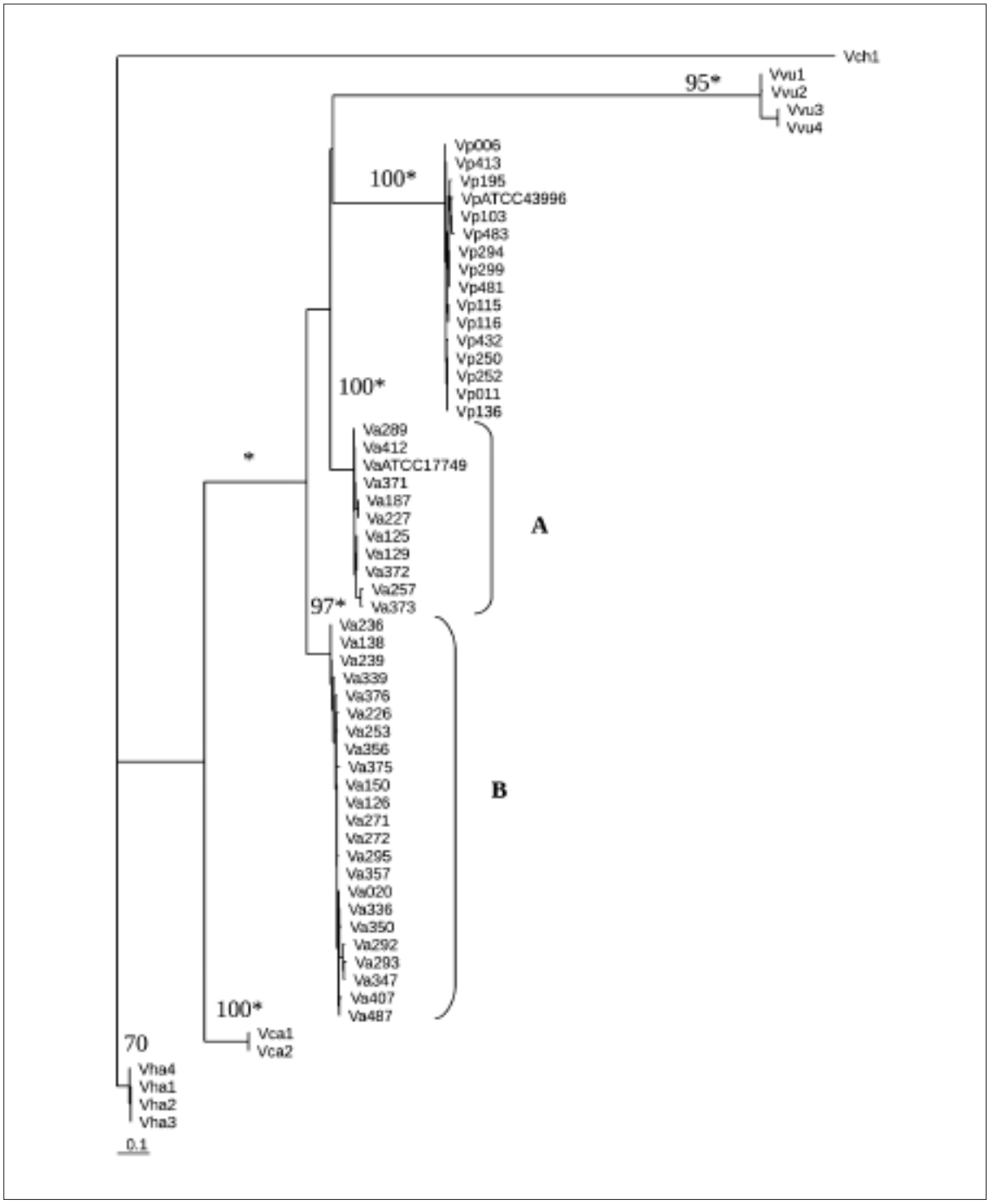

FIGURE 2 - *Phylogenetic relationship was defined using 34 sequences from V. alginolyticus (Va) and 16 sequences from V. parahaemolyticus (Vp) isolates; 11 sequences uploaded from NCBI GenBank were included in the analysis (Vca = V. campbellii, Vch = V. cholerae, Vha = V. harveyi, Vvu = V. vulnificus; see "GenBank accession numbers" in Materials and Methods). Bootstrap values below 70% were not shown. The scale bar indicates 10% nucleotide sequence divergence; * indicates a p value <0.001 in the Zero Branch Length test.*

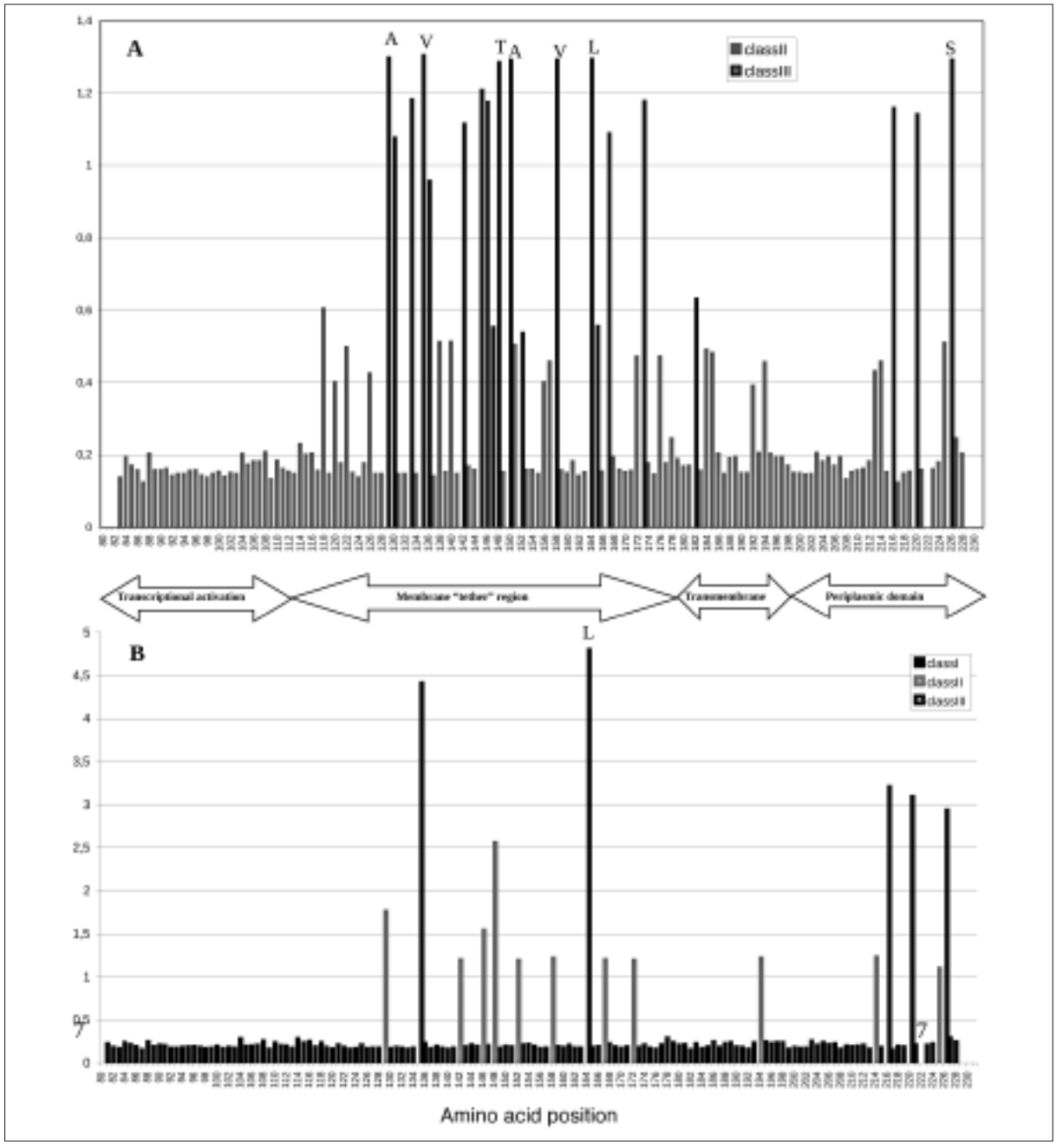

FIGURE 3 - *Panel A: mean of the posterior distribution of (postmean) for the classes of sites along the toxR gene in V. alginolyticus. The evolutionary model used was M3 (discrete), which assumes three classes of sites in the gene: sites under negative or purifying selection (posterior probability p_0), sites under neutral selection (p_1) and sites under positive or diversifying selection (p_2). The postmean ω is calculated as $p_0*\omega_0 + p_1*\omega_1 + p_2*\omega_2$. The posterior probabilities at site 135 (V), for example, are $p_0 = 0.00047$, $p_1 = 0.00051$, $p_2 = 0.99902$, and the site is almost certainly under diversifying selection, with postmean ω of 1.305 and P = 0.999 (ω>1). The letters on the top of columns indicates the amino acid sites under diversifying selection with probability >95%. Panel B: mean of the posterior distribution of (postmean ω) for the classes of sites along the toxR gene in V. alginolyticus subgroup B. Model M3 (discrete) was used. The estimated frequencies (p) and ratios for the three classes were: $p_0 = 0.69063$, $\omega_0 = 0.00000$ for conserved sites (negative selection), $p_1 = 0.26683$, $\omega_1 = 0.74534$ for neutral sites and $p_2 = 0.04254$, $\omega_2 = 4.86094$ for sites under diversifying selection.*

cus. In the transmembrane domain there is only one site (amino acid 184 I/V/L) conserved in each group but different between the groups, but the properties of the three amino acids are similar (nonpolar and neutral).

Evolutionary analysis

The evolutionary analysis, performed using the alignment of *toxR* fragment of *V. parahaemolyticus* isolates and the comparison of nested models using the LRT showed no significant p value for any comparison, indicating that the substitutions inferred from the *V. parahaemolyticus* isolates alignments are best explained by the negative/neutral selection model.

On the other hand, when the whole fragment of *toxR* locus of *V. alginolyticus* alignment was submitted to evolutionary analysis, the LRT of M3 vs M0 showed a strongly significant p value ($1.7 * 10^{-04}$). The estimated frequencies (p) and ratios for the three class of sites considered in model M3 were respectively $p_0 = 0.387$ and $\omega_0 = 0.104$ for conserved sites, $p_1 = 0.420$ and $\omega_1 = 0.104$ for neutral sites, $p_2 = 0.193$ and $\omega_2 = 1.306$ for sites under diversifying selection. This means that 19% of sites are under diversifying selection with $\omega_2 = 1.306$ and seven amino acid sites (129A, 135V, 148T, 150A, 158V, 164L, 226S; numbering based on *V. alginolyticus* ATCC 17749) were identified under positive selection at the 95% cut-off. Six of the sites considered under positive or diversifying selection belong to "tether" domain and one to the periplasmic domain (Fig. 3, panel A).

The evolutionary analysis performed on the whole fragment of *toxR* locus separately for the subgroup A and subgroup B alignments showed that only for subgroup B an evolutionary force is acting, as the p value of LTR M0 vs M3 was $4.25* 10^{-4}$. In this analysis only one site under diversifying selection, P164L, was identified with the posterior probability greater than 95% (Fig. 3, panel B); the site, belonging to the membrane "tether" region, was one of those identified under positive selection also in the evolutionary analysis performed on the whole *V. alginolyticus* group. An analysis of the four different protein domains (first 99 bp, second 198 bp, third 60 bp and fourth 102 bp) performed separately to detect which fragment produced a high value of significance of the LRT, did not present significant p values in any of the domains.

DISCUSSION

The results showed the usefulness of *toxR* gene for phylogenetic and evolution analysis on *Vibrio* isolates belonging to the species *parahaemolyticus* and *alginolyticus*. The results of the analysis on the 16S rRNA gene of the isolates used in this study confirmed that the this gene, which is useful to allocate species to different branches of the family Vibrionaceae (Thompson *et al.*, 2002), is largely inadequate for the discrimination of closely related species as *V. parahaemolyticus* and *V. alginolyticus*, that share nearly identical sequences in 16S rRNA gene (99.8% identity) (Marchesi *et al.*, 1998; Nagpal *et al.*, 1998). The phylogenetic tree (Fig.1) showed the approximate division of the *V. parahaemolyticus* and *V alginolyticus* isolates in two clades, in both of which, however, some isolates' 16S rRNA sequences appeared to intermingle with the other species. On the contrary, the use of *toxR* gene, which appears to be well conserved in *Vibrio* genus but shows higher divergence than 16S rRNA gene, provided a phylogenetic tree with separate divergent branches related to the *V. parahaemolyticus* and *V. alginolyticus* isolates and no occurrence of intermixed sequences. The difference in the results provided by the two analyses was indeed substantiated by the Pearson's correlation coefficient of the similarity matrices of 16S rRNA and *toxR* gene (0.019) showing that the two genes do not resolve the phylogenetic relationship in the same way. In fact, besides a different discrimination of the isolates belonging to the two species, the phylogenetic analysis performed using *toxR* gene divided the *V. alginolyticus* isolates into two separate clades (subgroup A and subgroup B), statistically supported by high bootstrap values and significant values in the Zero Branch Length tests. This result is particularly interesting in relation to the findings of Pascual *et al.* (2010) as it confirms the need to include in phylogenetic studies a reasonable number of strains for each species to account for intraspecies variability and, more significantly, highlights the relevance of *toxR* gene in genetic analysis of *Vibrio* species, due to the discriminatory potential of this sequence both at the interspecies (Le Roux *et al.*, 2005; Pascual *et al.*, 2010) and intraspecies level.

The two *V. alginolyticus* clades detected by *toxR* sequence analysis are at least partially reflected in

the phenotypic features, as some of the biochemical tests (ODC, GEL, PNPG and ARAa) performed on the isolates showed statistically significant differences in the response of the two subgroups. Such genetic and phenotypic divergences, however, do not seem to be related to the geographic and/or temporary origin of the isolates. While, in fact, the *V. parahaemolyticus* isolates were all included in a single cluster despite their level of variety – including both clinical and environmental sources as well as strains of Italian, African and Japanese origin collected over a ten year period – the clustering in two groups was detected among the *V. alginolyticus* isolates, whose majority was collected in the same geographic area (northern Adriatic Sea, Italy) in a short time span (from June to December 2003) and in environmental conditions (temperature, pH and salinity) varying in the range expected for different seasons. In more than one case, isolates attributed by phylogenetic analysis either to subgroup A or subgroup B (e.g. Va 125 and Va126, or Va226 and Va227) were collected from the very same food sample. The two subgroups could therefore represent separate bacterial populations, co-existing in the analysed samples.

The analysis of the protein codified by the *toxR* gene showed that the so-called "tether" region (corresponding to the cytoplasmic domain of ToxR and hypervariable at the interspecific level) is essentially conserved at the intraspecific level in *V. parahaemolyticus* isolates and at the intragroup level in *V. alginolyticus* subgroups A and B, with at least six amino acid sites (120, 138, 140, 142, 158 and 167) conserved in each cluster. According to the evolutionary analysis, in the "tether" region of the ToxR protein of *V. alginolyticus* six sites under diversifying selection were found. The evolutionary model for positive selection (M3) was identified by the LRT as the best fitting model with 19% of the sites under diversifying selection. In particular, although the genetic variability observed in *V. alginolyticus* (both subgroup A and B) was low, probably because they apparently cohabitate the same ecological niche, the evolutionary analysis showed subgroup B under positive evolutionary pressure. Evidence of diversifying selection was previously reported for *tcpA* gene of *V. cholerae* (Boyd and Waldor, 2002) and *pilA* of *V. vulnificus* (Chattopadhyay *et al.*, 2009), however this is, to our knowledge, the first study performed on *Vibrio* species showing evidence of diversifying selection in a gene not codifying for proteins subjected to response by host immune system or to phage attack (as the monomers forming the pili) but in a gene involved in a regulatory cascade. Moreover this study detected the presence of such positive evolutionary pressure only in one subgroup of the *V. alginolyticus* strains studied, highlighting a diversifying selection acting at sub-species level.

Separate analysis of *V. alginolyticus* subgroup B, identified position 164 as the only site under diversifying selection with a posterior probability greater than 95%. This site is characterized by the recurrence of two amino acids, leucine and proline. The whole "tether" region of ToxR protein is rich in prolines and shows conserved proline residue sites at both the intraspecies and intragroup level. Since the proline residues, due to their hydrophobicity and steric conformation, interfere with the secondary and tertiary structure of proteins (Freund *et al.*, 2008; Mazna *et al.*, 2008), this change in position 164, could interfere with the protein activity and studies on functional changes of ToxR protein should be performed to investigate this aspect.

In conclusion, the phylogenetic and evolutionary analysis on *toxR* gene sequences suggests the existence in the pool of our *V. alginolyticus* isolates of two genetically related but distinct clusters, one of which is subjected to a positive evolutionary pressure. Besides the ability to discriminate highly genetically related species, such locus showed an excellent discriminatory performance also with regard to different subgroups of the same species and could be useful for further studies on *Vibrio* genus.

Acknowledgements

This work was performed within the Integrated Project (IP) SEAFOODplus, partially granted by the EU Commission under contract No FOOD-CT-2004-506359.

REFERENCES

Anisimova M., Bielawski J.P., Yang Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**, 1585-1592.

ANISIMOVA M., BIELAWSKI J.P.,YANG Z. (2002). Accuracy and power of bayes prediction of amino acid sites under positive selection. *Mol. Biol. Evol.* **19**, 950-958.

AUSUBEL F.M. (1995). Short protocols in molecular biology: a compendium of methods from Current protocols in molecular biology. Brooklyn, NY: Greene Pub. Associates; New York, NY: Wiley, c1992.

BERGEY D.M., HOLT J.G.,KRIEG N.R. (1984). Bergey's Manual of systematic bacteriology. Baltimore, MD: Williams & Wilkins c1992.

BOYD E.F.,WALDOR M.K. (2002). Evolutionary and functional analyses of variants of the toxin-coregulated pilus protein TcpA from toxigenic Vibrio cholerae non-O1/non-O139 serogroup isolates. *Microbiol.* **148**, 1655-1666.

BRAUNS L.A., HUDSON M.C.,OLIVER J.D. (1991). Use of the polymerase chain reaction in detection of culturable and nonculturable Vibrio vulnificus cells. *Appl. Environ. Microbiol.* **57**, 2651-2655.

CHATTOPADHYAY S., PARANJPYE R.N., DYKHUIZEN D.E., SOKURENKO E.V., STROM M.S. (2009). Comparative evolutionary analysis of the major structural subunit of Vibrio vulnificus type IV pili. *Mol. Biol. Evol.* **26**, 2185-2196.

CROCI L., SUFFREDINI E., COZZI L., TOTI L., OTTAVIANI D., PRUZZO C., SERRATORE P., FISCHETTI R., GOFFREDO E., LOFFREDO G., MIONI R. (2007). Comparison of different biochemical and molecular methods for the identification of Vibrio parahaemolyticus. *J. Appl. Microbiol.* **102**, 229-237.

DI PINTO A., CICCARESE G., TANTILLO G., CATALANO D., FORTE V.T. (2005). A collagenase-targeted multiplex PCR assay for identification of Vibrio alginolyticus, Vibrio cholerae, and Vibrio parahaemolyticus. *J. Food Prot.* **68**, 150-153.

DORSCH M., LANE D., STACKEBRANDT E. (1992). Towards a phylogeny of the genus Vibrio based on 16S rRNA sequences. *Int. J. Syst. Bacteriol.* **42**, 58-63.

FREUND C., SCHMALZ H.G., STICHT J., KUHNE R. (2008). Proline-rich sequence recognition domains (PRD): ligands, function and inhibition. *Handb. Exp. Pharmacol.* 407-429.

FUENZALIDA L., HERNANDEZ C., TORO J., RIOSECO M.L., ROMERO J., ESPEJO R.T. (2006). Vibrio parahaemolyticus in shellfish and clinical samples during two large epidemics of diarrhoea in southern Chile. *Environ. Microbiol.* **8**, 675-683.

GOMEZ-GIL B., SOTO-RODRIGUEZ S., GARCIA-GASCA A., ROQUE A., VAZQUEZ-JUAREZ R., THOMPSON F.L., SWINGS J. (2004). Molecular identification of *Vibrio harveyi*-related isolates associated with diseased aquatic organisms. *Microbiol.* **150**, 1769-1777.

HALL T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series.* **41**, 95-98.

HASEGAWA M., KISHINO H., YANO T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160-174.

HO P.L., TANG W.M., LO K.S., YUEN K.Y. (1998). Necrotizing fasciitis due to Vibrio alginolyticus following an injury inflicted by a stingray. *Scand. J. Infect. Dis.* **30**, 192-193.

HORMANSDORFER S., WENTGES H., NEUGEBAUR-BUCHLER K., BAUER J. (2000). Isolation of Vibrio alginolyticus from seawater aquaria. *Int. J. Hyg. Environ. Health.* **203**, 169-175.

KIM Y.B., OKUDA J., MATSUMOTO C., TAKAHASHI N., HASHIMOTO S.,NISHIBUCHI M. (1999). Identification of *Vibrio parahaemolyticus* strains at the species level by PCR targeted to the *toxR* gene. *J. Clin. Microbiol.* **37**, 1173-1177.

KIMURA M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111-120.

LE ROUX F., GOUBET A., THOMPSON F.L., FAURY N., GAY M., SWINGS J., SAULNIER D. (2005). Vibrio gigantis sp. nov., isolated from the haemolymph of cultured oysters (Crassostrea gigas). *Int. J. Syst. Evol. Microbiol.* **55**, 2251-2255.

LIN Z., KUMAGAI K., BABA K., MEKALANOS J.J., NISHIBUCHI M. (1993). Vibrio parahaemolyticus has a homolog of the *Vibrio cholerae* toxRS operon that mediates environmentally induced regulation of the thermostable direct hemolysin gene. *J. Bacteriol.* **175**, 3844-3855.

MARCHESI J.R., SATO T., WEIGHTMAN A.J., MARTIN T.A., FRY J.C., HIOM S.J., DYMOCK D., WADE W.G. (1998). Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl. Environ. Microbiol.* **64**, 795-799.

MARTINEZ-URTAZA J., LOZANO-LEON A., DEPAOLA A., ISHIBASHI M., SHIMADA K., NISHIBUCHI M., LIEBANA E. (2004). Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. *J. Clin. Microbiol.* **42**, 4672-4678.

MATSIOTA-BERNARD P., NAUCIEL C. (1993). Vibrio alginolyticus wound infection after exposure to sea water in an air crash. *Eur. J. Clin. Microbiol. Infect. Dis.* **12**, 474-475.

MAZNA P., GRYCOVA L., BALIK A., ZEMKOVA H., FRIEDLOVA E., OBSILOVA V., OBSIL T., TEISINGER J. (2008). The role of proline residues in the structure and function of human MT2 melatonin receptor. *J. Pineal. Res.* **45**, 361-372.

MUKHERJI A., SCHROEDER S., DEYLING C., PROCOP G.W. (2000). An unusual source of Vibrio alginolyticus-associated otitis: prolonged colonization or freshwater exposure? *Arch. Otolaryngol. Head Neck Surg.* **126**, 790-791.

NAGPAL M.L., FOX K.F., FOX A. (1998). Utility of 16S-

23S rRNA spacer region methodology: how similar are interspace regions within a genome and between strains for closely related organisms? *J. Microbiol. Meth.* **33**, 211-219.

NANDI B., NANDY R.K., MUKHOPADHYAY S., NAIR G.B., SHIMADA T., GHOSE A.C. (2000). Rapid method for species-specific identification of Vibrio cholerae using primers targeted to the gene of outer membrane protein OmpW. *J. Clin. Microbiol.* **38**, 4145-4151.

NIELSEN R.,YANG Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*. **148**, 929-936.

OAKEY H.J., LEVY N., BOURNE D.G., CULLEN B., THOMAS A. (2003). The use of PCR to aid in the rapid identification of Vibrio harveyi isolates. *J. Appl. Microbiol.* **95**, 1293-1303.

OLIVER J.D., KAPER J.B. (1997). Vibrio Species. In: M.P. Doyle, L.R. Beuchat, T.J. Montville, eds. Food microbiology: fundamentals and frontiers. ASM Press, Washington, D.C. 228-264.

OSORIO C.R., KLOSE K.E. (2000). A region of the transmembrane regulatory protein ToxR that tethers the transcriptional activation domain to the cytoplasmic membrane displays wide divergence among Vibrio species. *J. Bacteriol.* **182**, 526-528.

PASCUAL J., MACIAN M.C., ARAHAL D.R., GARAY E., PUJALTE M.J. (2010). Multilocus sequence analysis of the central clade of the genus Vibrio by using the 16S rRNA, recA, pyrH, rpoD, gyrB, rctB and toxR genes. *Int. J. Syst. Evol. Microbiol.* **60**, 154-165.

REICH K.A., SCHOOLNIK G.K. (1994). The light organ symbiont Vibrio fischeri possesses a homolog of the Vibrio cholerae transmembrane transcriptional activator ToxR. *J. Bacteriol.* **176**, 3085-3088.

RIPABELLI G., GRASSO G.M., SAMMARCO M.L., LUZZI I. (1997). Procedure di isolamento e caratterizzazione di Vibrio spp. di importanza clinica. *Rapporti ISTISAN*. **31**, 1-55.

RUIMY R., BREITTMAYER V., ELBAZE P., LAFAY B., BOUSSEMART O., GAUTHIER M., CHRISTEN R. (1994). Phylogenetic analysis and assessment of the genera Vibrio, Photobacterium, Aeromonas, and Plesiomonas deduced from small-subunit rRNA sequences. *Int. J. Syst. Bacteriol.* **44**, 416-426.

SWOFFORD D.L. (2002). PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods) 4.0 Beta, Sinauer Associates, Sunderland, Massachussets.

THOMPSON C.C., THOMPSON F.L., VICENTE A.C., SWINGS J. (2007). Phylogenetic analysis of vibrios and related species by means of atpA gene sequences. *Int. J. Syst. Evol. Microbiol.* **57**, 2480-2484.

THOMPSON F.L., GEVERS D., THOMPSON C.C., DAWYNDT P., NASER S., HOSTE B., MUNN C.B., SWINGS J. (2005). Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl. Environ. Microbiol.* **71**, 5107-5115.

THOMPSON F.L., HOSTE B., THOMPSON C.C., GORIS J., GOMEZ-GIL B., HUYS L., DE VOS P., SWINGS J. (2002). Enterovibrio norvegicus gen. nov., sp. nov., isolated from the gut of turbot (Scophthalmus maximus) larvae: a new member of the family Vibrionaceae. *Int. J. Syst. Evol. Microbiol.* **52**, 2015-2022.

THOMPSON J.D., GIBSON T.J., PLEWNIAK F., JEANMOUGIN F., HIGGINS D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

YANG Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306-314.

YANG Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555-556.

YANG Z., NIELSEN R., GOLDMAN N., PEDERSEN A.M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics.* **155**, 431-449.

YANG Z., WONG W.S., NIELSEN R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107-1118.