MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3924–3937 doi: 10.1111/j.1365-294X.2010.04782.x

Origins and colonization history of pandemic *Vibrio parahaemolyticus* in South America

JUAN ANSEDE-BERMEJO,* RONNIE G. GAVILAN,* JOAQUÍN TRIÑANES,† ROMILIO T. ESPEJO‡ and JAIME MARTINEZ-URTAZA*

*Instituto de Acuicultura and †Instituto de Investigaciones Tecnológicas, Universidad de Santiago de Compostela, Campus Universitario Sur, 15782 Santiago de Compostela, Spain, ‡Laboratorio de Biotecnologia, Instituto de Nutricion y Tecnologıa de los Alimentos, Universidad de Chile, Santiago, Chile

Abstract

The dynamics of dissemination of the environmental human pathogen *Vibrio parahaemolyticus* are uncertain. The O3:K6 clone was restricted to Asia until its detection along the Peruvian coasts and in northern Chile in 1997 in phase with the arrival of El Niño waters. A subsequent emergence of O3:K6 strains was detected in austral Chile in 2004. The origin of these 1997 and 2004 population radiations has not yet been conclusively determined. Multiple loci VNTR analysis using seven polymorphic loci was carried out with a number of representative strains from Asia, Peru and Chile to determine their genetic characteristics and population structure. Asian and Chilean subpopulations were the most genetically distant groups with an intermediate subpopulation in Peru. Population structure inferred from a minimum-spanning tree and Bayesian analysis divided the populations into two genetically distinct groups, consistent with the epidemic dynamics of the O3:K6 clone in South America. One group comprised strains from the original Asiatic population and strains arriving in Peru and Chile in 1997. The second group included the remaining Peruvian Strains and Chilean strains obtained from Puerto Montt in 2004. The analysis of the arrival of the O3:K6 clone at the Pacific coasts of South America has provided novel insights linking the origin of the invasion in 1997 to Asian populations and describing the successful establishment of the O3:K6 populations, first in Peru and subsequently in the South of Chile owing to a possible radiation of Peruvian populations.

Keywords: climate change, invasive species, MLVA, Population Ecology, Population Genetics—empirical, waterborne diseases

Received 1 March 2010; revision revised 30 June 2010; accepted 6 July 2010

Introduction

The genus *Vibrio* comprises more than 50 recognized species found in aquatic habitats of a wide range of salinity. The particular biological features of *Vibrio* populations have triggered a growing interest over the last decades, converting this group into the best studied of all aquatic bacteria. *Vibrio* species have been used as models for the study of population structure and evolution (Bisharat *et al.* 2007), genomic plasticity (Han *et al.* 2008), niche adaptation and biogeography (Hunt *et al.* 2008). Such studies have acquired heightened interest due to human diseases associated with pathogenic Vibrios, which have been involved in a global expansion in recent years (Nair *et al.* 2007).

Vibrio parahaemolyticus is a marine bacterium and a natural inhabitant of estuarine areas worldwide. A marginal fraction of the populations naturally occurring in marine ecosystems have virulence traits and the capability, under specific environmental conditions, to cause infections in humans who consume raw or undercooked seafood. Until 1996, *V. parahaemolyticus* infections had shown a local distribution, emerging in

Correspondence: Jaime Martinez-Urtaza, Fax: +34 981 547165; E-mail: jaime.martinez.urtaza@usc.es

different areas of the world during the warmer months of the year. The epidemiology of *V. parahaemolyticus* changed drastically in 1996, when an atypical increment in *V. parahaemolyticus* infections arose in India linked to strains belonging to a clonal group of the serotype O3:K6 (Okuda *et al.* 1997; Chowdhury *et al.* 2000). This clone rapidly spread throughout the majority of southeast Asian countries within a single year (Okuda *et al.* 1997). The epidemic dissemination of the O3:K6 clone was restricted to Asia until 1997 when strains belonging to this group were detected in Peru in June of 1997 (Martinez-Urtaza *et al.* 2008) and subsequently in Chile by the end of the same year (Gonzalez-Escalona *et al.* 2005), initiating the pandemic expansion of the O3:K6 clone. From this point on, the pandemic clone began a global dissemination and was detected in the USA (Daniels *et al.* 2000), Europe (Martinez-Urtaza *et al.* 2005; Quilici *et al.* 2005) and Africa (Ansaruzzaman *et al.* 2008).

The routes and mechanisms of dissemination of the pandemic clone have been controversial from their origin. One of the most recurrent explanations has been based on the discharge of ballast waters from ships travelling from areas of *V. parahaemolyticus* endemicity. Ballast water discharges have been recognized as one of the major vehicles for the worldwide dissemination of marine species and biological invasions (Niimi 2004) and have been identified as a reliable mechanism for the propagation of pathogenic Vibrios (DePaola *et al.* 1992; McCarthy & Khambaty 1994; Ruiz *et al.* 2000).

Bacterial dispersal through ballast water, however, fails to provide a consistent and comprehensive explanation for the emergence of some epidemic episodes of *V. parahaemolyticus*. This is especially true where infections have emerged and spread rapidly over hundreds of kilometres of coastline, such as in the case of the arrival of the O3:K6 clone in Peru in 1997 (Martinez-Urtaza *et al.* 2008). The infections first surfaced in the north of the country and then spread southwards along more than 1500 km in just 4 months until they reached the Chilean city of Antofagasta (Gonzalez-Escalona *et al.* 2005). The origins and routes of dissemination of pandemic *V. parahaemolyticus* from its arrival in South America remained unknown until a recent revision of the oceanographic conditions existing during this period revealed that the emergence and dissemination of the pandemic clone in Peru correlated with the dynamics of progression and receding of the 1997 El Niño waters (Martinez-Urtaza *et al.* 2008). According to this study, the 1997 El Niño episode may have provided an extraordinary corridor for the displacement of Asian *Vibrio* populations to America. In 2004, new epidemic outbreaks of pandemic *V. parahaemolyticus* were detected emerging suddenly in the austral regions of Chile in areas located at more than 2000 km from Antofagasta. The origins of this second radiation of pandemic *V. parahaemolyticus* to Southern Chile also remain unknown.

Even if nonnative pathogenic populations of *V. parahaemolyticus* were introduced by either human-mediated dispersal (ballast water) or were naturally spread by the movement of waters, the epidemic dispersion of pandemic *V. parahaemolyticus* in South America should be considered as a biological invasion. However, the spread of bacterial pathogens through natural environments and the emergence of epidemics are rarely analysed from the perspective of population genetics or species invasion. In this study, employing this alternative perspective, we developed a novel genotyping scheme based on polymorphic regions of the genome of *V. parahaemolyticus* to analyse the clonal structure of the O3:K6 strains and discriminate among different geographical groups. Using multi-locus variable number tandem repeat analysis (MLVA), we investigated the genetic characteristics and population structure of a representative number of strains originating from Asia, Peru and Chile to infer the origin and routes of dissemination of the pandemic specimens of *V. parahaemolyticus* along the Pacific coasts of South America.

Material and methods

Strains and DNA preparation

In total, 69 strains of *Vibrio parahaemolyticus* belonging to the pandemic clonal complex were included in the present study as representatives of three geographical areas: southeastern Asia (Asia), Peru and Chile. Asian strains ($n = 13$) were isolated between 1996 and 1999 from multiple clinical cases reported throughout seven southeastern Asian countries: Bangladesh, India, Japan, Korea, Laos, Taiwan and Thailand (Table 1). Among the Asian strains, we have included the strain RIMD2210633, which represents the only *V. parahaemolyticus* strain with a completely sequenced genome (Makino *et al.* 2003). Peruvian strains ($n = 35$) were obtained between 1997 and 2003 from clinical sources in hospitals and public health laboratories located throughout the country (Martinez-Urtaza *et al.* 2008). Chilean strains ($n = 21$) included isolates obtained from the first epidemic outbreak of *V. parahaemolyticus* detected in this country in 1998 in the northern city of Antofagasta and from the later epidemic radiation of this pathogen in the austral region of Puerto Montt in 2004 (Table 1).

Strains were cultured at 37 °C in Luria–Bertani broth with 2% NaCl. One millilitre of an overnight culture was centrifuged at 9000 *g* for 7 min. The resulting pellet

Table 1 Characteristics of the *Vibrio parahaemolyticus* strains included in this study

Strain	Country of isolation	Year	Place	Serotype	MLVA genotype (locus TR-VP)							References
					5	10	16	18	19	21	25	
VP 81	India	1996	Calcutta	O3:K6	8	29	32	10	23	8	7	Matsumoto *et al.* (2000)
RIMD 2210633	IT (Japan)	1996	Osaka	O3:K6	7	28	35	11	22	9	8	Makino *et al.* (2003)
DOH 958 15	IT (Taiwan)	1997	ND	O3:K6	8	27	35	2	21	8	7	Chowdhury *et al.* (2000)
97LPV2	IT (Laos)	1997	ND	O3:K6	8	19	31	2	20	8	6	Matsumoto *et al.* (2000)
VP47	IT (Thailand)	1998	ND	O3:K6	8	29	18	10	23	9	7	Matsumoto *et al.* (2000)
AN 5034	IT (Bangladesh)	1998	ND	O4:K68	9	20	29	10	21	8	7	Matsumoto *et al.* (2000)
AV16000	IT (Bangladesh)	1998	ND	O1:Kut	8	30	22	9	22	8	7	Matsumoto *et al.* (2000)
VP2	IT (Korea)	1998	ND	O3:K6	8	30	22	9	22	8	7	Matsumoto *et al.* (2000)
JKY Vp6	IT (Japan)	1998	ND	O3:K6	8	32	29	2	24	8	7	Matsumoto *et al.* (2000)
AN8373	IT (Bangladesh)	1998	ND	O3:K6	10	32	35	10	23	8	7	Matsumoto *et al.* (2000)
VPHY67	IT (Thailand)	1999	ND	O3:K6	8	30	36	10	23	8	7	Laohaprertthisan *et al.* (2003)
VPHY145	IT (Thailand)	1999	ND	O4:K68	8	30	21	11	22	7	7	Laohaprertthisan *et al.* (2003)
VPHY191	IT (Thailand)	1999	ND	O1:K25	8	39	18	10	22	9	7	Laohaprertthisan *et al.* (2003)
763-97	Peru	1997	Lima	O3:K6	9	30	36	9	23	9	7	Martinez-Urtaza *et al.* (2008)
790-97	Peru	1997	Cajamarca	O3:K6	8	34	35	9	20	8	7	Martinez-Urtaza *et al.* (2008)
875-97	Peru	1997	Lambayeque	O3:K6	8	27	40	9	19	8	7	Martinez-Urtaza *et al.* (2008)
948-97	Peru	1997	Lambayeque	O3:K6	9	30	36	9	23	10	7	Martinez-Urtaza *et al.* (2008)
1171-97	Peru	1997	Monqueagua	O3:K6	8	31	36	9	23	9	7	Martinez-Urtaza *et al.* (2008)
780-98	Peru	1998	Lima	O3:K6	8	34	35	9	21	9	7	Martinez-Urtaza *et al.* (2008)
784-98	Peru	1998	Lima	O3:K6	9	31	37	9	23	10	7	Martinez-Urtaza *et al.* (2008)
971-98	Peru	1998	Lima	O3:K6	8	29	29	11	24	8	7	Martinez-Urtaza *et al.* (2008)
974-98	Peru	1998	Lima	O3:K6	9	29	36	9	23	10	7	Martinez-Urtaza *et al.* (2008)
3435-98	Peru	1998	Lima	O3:Kut	8	30	43	9	23	9	7	Martinez-Urtaza *et al.* (2008)
275-99	Peru	1999	Lima	O3:Kut	8	33	44	9	21	9	7	Martinez-Urtaza *et al.* (2008)
276-99	Peru	1999	Lima	O3:K6	8	29	39	10	25	9	7	Martinez-Urtaza *et al.* (2008)
278-99	Peru	1999	Lima	O3:K6	8	29	38	10	24	9	7	Martinez-Urtaza *et al.* (2008)
279-99	Peru	1999	Lima	O3:K6	8	33	44	9	21	9	7	Martinez-Urtaza *et al.* (2008)
698-99	Peru	1999	Lambayeque	O3:K6	8	29	27	11	24	8	7	Martinez-Urtaza *et al.* (2008)
330-00	Peru	2000	Lima	O3:K6	8	32	41	9	20	9	7	Martinez-Urtaza *et al.* (2008)
405-00	Peru	2000	Lambayeque	O3:K6	9	30	32	9	23	8	7	Martinez-Urtaza *et al.* (2008)
430-00	Peru	2000	Lima	O3:K6	8	32	37	9	20	9	7	Martinez-Urtaza *et al.* (2008)
462-00	Peru	2000	Lima	O3:K6	9	32	39	9	19	9	7	Martinez-Urtaza *et al.* (2008)
511-00	Peru	2000	Lima	O3:K6	10	27	42	10	22	9	7	Martinez-Urtaza *et al.* (2008)
706-00	Peru	2000	Lima	O3:K6	9	30	32	9	23	8	7	Martinez-Urtaza *et al.* (2008)
056-01	Peru	2001	Lambayeque	O3:K6	8	29	34	9	20	8	7	Martinez-Urtaza *et al.* (2008)
498-01	Peru	2001	Lima	O3:K6	7	33	31	10	22	8	7	Martinez-Urtaza *et al.* (2008)
565-01	Peru	2001	Lambayeque	O3:K6	9	33	39	9	21	9	7	Martinez-Urtaza *et al.* (2008)
568-01	Peru	2001	Iquitos	O3:K6	7	30	21	11	22	9	7	Martinez-Urtaza *et al.* (2008)
572-01	Peru	2001	Iquitos	O3:K6	9	23	39	9	18	9	7	Martinez-Urtaza *et al.* (2008)
003-02	Peru	2002	Lima	O3:K6	9	34	43	9	21	9	7	Martinez-Urtaza *et al.* (2008)
004-02	Peru	2002	Lima	O3:K6	9	34	45	9	21	9	7	Martinez-Urtaza *et al.* (2008)
020-02	Peru	2002	Lima	O3:K6	8	34	36	9	20	8	7	Martinez-Urtaza *et al.* (2008)
240-02	Peru	2002	Lima	O3:K6	8	37	35	9	19	7	7	Martinez-Urtaza *et al.* (2008)
651-02	Peru	2002	Lima	O3:K6	8	35	40	10	19	9	7	Martinez-Urtaza *et al.* (2008)
038-03	Peru	2003	Cajamarca	O3:K6	9	35	36	9	20	9	7	Martinez-Urtaza *et al.* (2008)
039-03	Peru	2003	Cajamarca	O3:K6	8	35	40	9	19	8	7	Martinez-Urtaza *et al.* (2008)
131-03	Peru	2003	Lima	O3:K6	8	33	36	9	20	8	7	Martinez-Urtaza *et al.* (2008)
302-03	Peru	2003	Cajamarca	O3:K6	8	36	37	9	19	8	7	Martinez-Urtaza *et al.* (2008)
ATC 210	Chile	1998	Antofagasta	O3:K6	8	31	36	9	23	7	6	Gonzalez-Escalona *et al.* (2005)
ATC 220	Chile	1998	Antofagasta	O3:K6	8	32	34	9	23	7	6	Gonzalez-Escalona *et al.* (2005)
ATC221	Chile	1998	Antofagasta	O3:K6	9	32	37	9	24	9	7	Gonzalez-Escalona *et al.* (2005)
PMC 37.5	Chile	2004	Puerto Montt	O3:K6	8	39	38	9	19	8	7	Gonzalez-Escalona *et al.* (2005)
PMC 39.5	Chile	2004	Puerto Montt	O3:K6	8	36	41	9	19	8	7	Gonzalez-Escalona *et al.* (2005)

Table 1 (Continued)

Strain	Country of isolation	Year	Place	Serotype	MLVA genotype (locus TR-VP)							References
					5	10	16	18	19	21	25	
PMC 42.5	Chile	2004	Puerto Montt	O3:K6	8	35	41	9	20	8	7	Gonzalez-Escalona *et al.* (2005)
PMC 48.5	Chile	2004	Puerto Montt	O3:K6	8	36	39	9	20	8	7	Gonzalez-Escalona *et al.* (2005)
PMC 49.5	Chile	2004	Puerto Montt	O3:K58	8	36	43	9	20	8	7	Gonzalez-Escalona *et al.* (2005)
PMC 57.5	Chile	2004	Puerto Montt	O3:K6	9	36	41	9	20	9	7	Gonzalez-Escalona *et al.* (2005)
PMC 61.5	Chile	2004	Puerto Montt	O3:K6	9	37	40	9	19	9	7	Gonzalez-Escalona *et al.* (2005)
PMC 65.5	Chile	2004	Puerto Montt	O3:K6	8	36	39	9	19	9	7	Gonzalez-Escalona *et al.* (2005)
PMC 14.7	Chile	2007	Puerto Montt	O3:K59	7	35	40	9	19	9	7	Harth-Chu *et al.* (2009)
PMC 16.7	Chile	2007	Puerto Montt	O3:K6	7	36	40	9	19	9	7	Harth-Chu *et al.* (2009)
PMC 28.7	Chile	2007	Puerto Montt	O3:K59	8	14	39	9	13	9	7	Harth-Chu *et al.* (2009)
PMC 29.7	Chile	2007	Puerto Montt	O3:K6,59	8	14	39	5	20	9	7	Harth-Chu *et al.* (2009)
PMC 41.7	Chile	2007	Puerto Montt	O3:K6	8	36	42	9	21	9	7	Harth-Chu *et al.* (2009)
PMC 44.7	Chile	2007	Puerto Montt	O3:K59	9	36	40	9	20	9	7	Harth-Chu *et al.* (2009)
PMC 50.7	Chile	2007	Puerto Montt	O3:K6	8	34	39	9	21	9	7	Harth-Chu *et al.* (2009)
PMC 55.7	Chile	2007	Puerto Montt	O3:K59	8	36	40	9	20	9	7	Harth-Chu *et al.* (2009)
PMC 70.7	Chile	2007	Puerto Montt	O3:K6,59	9	35	38	9	19	9	7	Harth-Chu *et al.* (2009)
PMC 73.7	Chile	2007	Puerto Montt	O3:K59	8	31	39	9	19	9	7	Harth-Chu *et al.* (2009)

IT, international travellers, origin of travel is indicated in parentheses; ND, not determined; MLVA, Multiple loci VNTR analysis.

was resuspended in 300 μL of 1 × TE buffer, and cell suspensions were lysed by boiling for 10 min. The lysate was then centrifuged at 10 000 *g* at 4 °C for 5 min, and the supernatant was transferred to a new tube and stored at −20 °C until it was used.

Identification and selection of tandem repeat loci

The identification and selection of the potentially polymorphic tandem repeat loci of interest for the study was carried out with a preliminary *in silico* search throughout the complete genome of the strain RIMD 2210633 of *V. parahaemolyticus* (GenBank accession numbers BA000031 and BA000032). Identification of loci was performed with the use of the Microorganisms Tandem Repeats online database (http://minisatellites.u-psud.fr) and the Tandem Repeat Sequence Finder program (Benson 1999). Tandem repeat loci were selected for further analysis according to the following criteria: (i) minimum repeat size of 6 bp, (ii) at least three copies of the repetitive motif and (iii) conservation between the tandem repeats >89%. In total, 25 VNTRs were identified according to the established criteria and designated according to the nomenclature proposed by Le Fleche *et al.* (2002).

Primer sets were designed in the flanking regions of the tandem repeats to produce fragment sizes no larger than 500 bp (Table 2) utilizing Vector NTI software (Invitrogen, Carlsbad, CA, USA). Predicted PCR product sizes in the RIMD2210633 strain were obtained using the 'Blast in the Tandem Repeats Database' (http://minisatellites.u-psud.fr/ASPSamp/base_ms/blast).

PCR amplification of microsatellites and genotyping

A PCR master mix (50 μL) containing 1 ng of DNA per μL, 1× *Taq* Reaction Buffer (Roche, Mannheim, Germany), 1 U of *Taq* DNA polymerase (Roche), 750 μM of each deoxynucleoside triphosphate and 0.2 μM of each flanking primer (for primer sequence, see Table 2). PCRs were performed in an MJ Research PTC200 thermocycler under the following conditions: initial denaturation at 94 °C for 5 min followed by 30 cycles of denaturation at 94 °C for 30 s, annealing between 60 °C and 64 °C (see Table 2) for 20 s, elongation at 72 °C for 30 s and final extension for 7 min at 72 °C. The lengths of PCR products were analysed by capillary electrophoresis on a microchip device using the 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) for accurate sizing. From these lengths, the number of repetitive units was deduced and converted into copy numbers for each locus, taking into account the genome of the strain RIMD221633. Additionally, PCR products for each selected locus from a subset of 17 strains were sequenced to evaluate the correct assignment of the copies of the repetitive motifs. Both strands of the PCR products were sequenced by Macrogen sequencing service on the 3730XL DNA Analyzer capillary electrophoresis system (AB, USA) using the BigDye version 3.1

Table 2 VNTR loci selected in this study and their corresponding primers and amplicon sizes

Locus[a]	Primer	Annealing temperature (°C)	Product size (bp)
TR-VP5_3160502 (I)_6 bp_7.3 U	F: 5′-ATCCGATGTTTGCTTTTGTACGGC-3′ R: 5′-CTAGCAGTGGAAGCGCCTTCTGAA-3	64	257
TR-VP10_3077908 (I)_6 bp_28.2 U	F: 5′-ATCGACCCACATCAATATTCATCTCG-3′ R: 5′-CTAACGTACGAAGGAGCTAACAACGCA-3′	60	316
TR-VP16_1548365 (II)_6 bp_35.2 U	F: 5′-ATCGCTGCTTGAAGAAAATCCTGAT-3′ R: 5′-CTAATTTTTCTGGTTGGGCTTGCG-3′	60	461
TR-VP18_2669177 (I)_7 bp_11.1 U	F: 5′-ATCGATGAAGAAAATGCCATTGCTG-3′ R: 5′-CTATGGAGAAGCCTTCAGGGAAAGTTTT-3′	62	231
TR-VP19_2339376 (I)_6 bp_21.5 U	F: 5′-ATCCATTGGTCGATTCGAGTTGACT-3′ R: 5′-CTATAACAGCGTACTTTCCAGCGAGC-3′	62	356
TR-VP21_447465 (I)_6 bp_9.5 U	F: 5′-ATCGGTTCGACCAAGCTTAAGAGTGA-3′ R: 5′-CTAATGTAACCGCTCAGCAAGGTGA-3′	62	330
TR-VP25_744172 (II)_6 bp_7.8 U	F: 5′-ATCCATGTCTTCTATTATCCATGCTTG-3′ R: 5′-CTACAACAGTACATTTTCCTTTGTGTGG-3′	62	264

[a]TR-locus reference_position (chromosome)_motive size (bp)_repeats (units) in RIMD2210633 genome.

terminator technology (AB, USA). The nucleotide sequences were deposited in GenBank under accession numbers HM154991 to HM155109.

The allele number for each locus was simply the number of copies of the repetitive motif observed in the strain studied. The resulting list of allelic values for each strain was the MLVA genotype (Table 1). Any two MLVA profiles differing by one or more loci were considered distinct genotypes.

Data analysis

Nei's diversity index (D) was used as a measure of the average genetic diversity for each marker for the total strain set and was calculated as $1 - \sum(\text{allele frequency})^2$ (Nei 1987).

For all the global populations and for each subpopulation, the number of alleles was determined. Genetic diversity was estimated at each locus and averaged across loci within each subpopulation by Nei's gene diversity (H; Nei 1987) and corrected for sample size using FSTAT version 2.9.3 (Goudet 2002). Additionally, the gene diversity was temporally and geographically examined in each subpopulation. Significant differences in the values obtained for the different years and between Peruvian and Chilean subpopulations were evaluated by the analysis of variance (ANOVA) performed with the SPSS software version 17.0 (SPSS Inc., Chicago, IL, USA). Associations among loci were tested using the index of multilocus linkage disequilibrium (LD) r_D (Agapow & Burt 2001) independently calculated from the overall population and within each subpopulation using MULTILOCUS version 1.3 software (Agapow & Burt 2001). The significance of r_D was tested by comparing the observed variance with the distribution of the variance expected under the null hypothesis of panmixia, as determined from 1000 randomizations of the genotype data. When the null hypothesis is rejected, a clonal population structure is suggested. The possible loci pair association was examined in each subpopulation using the FSTAT program. Significance levels were determined according to Bonferroni corrections based on the indicative adjusted *P*-value.

The validity of the mutation model was first evaluated through the construction of a minimum-spanning tree (MST) using Bionumerics version 5.10 (Applied Maths, Sint-Martens-Latern, Belgium). With the help of the MST, we estimated the proportion of strains that differed from their closest relatives by one or by multiple repeat changes to evaluate the validity of the Stepwise Mutation Model (SMM) for our data set. The SMM was further evaluated for each locus by calculating the allelic changes according to repeat-number differences using the MultiLocus Analyzer software (S. Brisse, unpublished).

The resolving power of the VNTR markers under the SMM may be affected by the presence of alleles identical in length but not identical by evolutionary descent, which may arise from convergent mutational events (homoplastic noise). Consequently, homoplasy was evaluated along the branches of the MST and estimated using the MultiLocus Analyzer (S. Brisse, unpublished). Homoplasy was defined as $1 - (K - 1)/M$, where K is the number of alleles, and M the number of changes along the MST. If each allele had been generated by a single evolutionary change, the number of changes since the ancestral state would be $K - 1$, and the homoplasy index (P) would be equal to zero. By contrast, if

alleles had been changed by convergent evolution or reversion to the ancestral state, *M* would become much greater than *K*, and the homoplasy index would increase towards 1.

The corresponding data files were analysed to examine genetic heterogeneity among the subpopulations studied based on the SMM model. To assess the relatedness among the subpopulations, the genetic distance, delta-mu squared $(\delta\mu)^2$, was computed as defined by Goldstein *et al.* (1995) using R_{ST} Calc version 2.2 software (Goodman 1997). Subsequently, the degree of genetic differentiation among the different subpopulations was evaluated with the use of the statistic R_{ST} (Slatkin 1995). Pairwise R_{ST} values were calculated using R_{ST} Calc version 2.2 with 1000 randomizations. The null hypothesis was no population differentiation. Gene flow was determined by measuring the number of migrants per generation (*M*) from the values of R_{ST} using the R_{ST} Calc version 2.2 package.

Population structure was inferred by combining an MST analysis of VNTR data and Bayesian analysis (Pritchard *et al.* 2000). The MST was built in Bionumerics version 5.10 using the Prim's algorithm and selecting the summed absolute distance as the coefficient for computing differences between strains. Priority rules were adopted from the BURST algorithm (Feil *et al.* 2004), with the highest priority given to profiles with the largest numbers of single locus variants, or double-locus variants in case of a match. Repeat-type (RT) complexes were defined as groups with a maximum neighbour distance of one change and a minimum size of two strains.

The Bayesian clustering algorithm implemented in the program Structure version 2.2 (Pritchard *et al.* 2000) was applied as an exploratory analysis to determine whether the different subpopulations of *V. parahaemolyticus* assigned according to their geographical origin could be subdivided into *K* genetically distinct groups based on allele frequencies. Genetic clusters were then constructed from a collection of individual genotypes, estimating the fraction of the genotype belonging to each cluster for each strain. The following parameters were used: admixture model, burn-in period of 10^5 iterations, and probability estimates based on 10 000 Markov Chain Monte Carlo iterations. To determine the most likely number of clusters (*K*) underlying our population samples, we conditioned our data on various values of *K* ranging from 1 to 6. Ten runs for each number of populations (*K*) were performed to quantify the variation of the likelihood for each *K*. The most likely value of *K* to indicate underlying genetic structure was determined using the statistic ΔK, which is based on the rate of change in the probability of an existing clustering between two successive *K* values (Evanno *et al.* 2005). CLUMPP version 1.1.2 (Jakobsson & Rosenberg 2007) was used to average multiple Structure runs. CLUMMP output was processed with Distruct version 1.1 (Rosenberg 2002), which allows for drawing barplots of CLUMPP.

Results

Microsatellite data analyses

A preliminary set of 25 potentially polymorphic VNTR loci were analysed across the collection of 69 clinical strains of *Vibrio parahaemolyticus* belonging to the pandemic complex originating from different Southeast Asian and Latin American countries. Of the 25 VNTR loci, 11 loci were monomorphic, and another seven additional loci showed null alleles and were consequently discarded. Finally, a set of seven loci were shown to be polymorphic and were selected for further analysis.

In total, 61 different alleles were identified among the seven microsatellites examined in the 69 *V. parahaemolyticus* strains. The loci studied showed tandem repeat units that varied in size between 6 to 7 bp for the different markers. According to the number of alleles observed, the loci studied varied in their degree of polymorphism ranging from 3 to 20 alleles per locus for TR-VP25 and TR-VP16, respectively (Table 3). MLVA showed 67 different genotypes among the 69 strains. The discriminatory capacity of each marker, measured through Nei's diversity index (D) for each individual locus, ranged from 0.11 for TR-VP25 to 0.92 for TR-VP16 (Table 3). Different allele sizes at the different VNTR loci resulted from the loss or gain of repeat units, which were verified by direct sequencing of the PCR products using the VNTR primers.

Genetic diversity

Strains were initially grouped for analysis into three groups according to their geographical origin: southeast Asian countries (Asia), Peru and Chile. Values of genetic diversity per locus presented highly variable results, with the lowest diversity found in loci TR-VP18 and TR-VP25 for the three subpopulations, except for TR-VP18 in Asian subpopulations, which in contrast showed a value of 0.679 (Table 4). The remaining loci showed values of diversity ranging from 0.5 to 0.91, with a homogenous diversity within the three subpopulations. The average genetic diversity across loci within each subpopulation showed similar values for the Asian and Peruvian subpopulations of 0.661 and 0.610, respectively, whereas Chile showed an overall diversity of 0.531. The average gene diversity for the total population investigated was 0.601 over the seven loci (Table 4).

Table 3 Characteristics of the VNTR loci studied into *Vibrio parahaemolyticus* genome

Locus	Identification[a]	Chr[b]	Motive length	Copy number	Total length	Match	Motive sequence	ORF	Alleles observed	D[c]	P[d]
TR-VP5	3160502–3160545	I	6	7.3	43	100%	CTTCTG	VP2956	4	0.519	0.800
TR-VP10	3077908–3078076	I	6	28.2	168	100%	ATAGAG	VP2892	16	0.908	0.594
TR-VP16	1548365–1548572	II	6	35.2	212	94%	AGCAAC	VPA1455	20	0.920	0.525
TR-VP18	2669177–2669254	I	7	11.1	77	100%	AGGTTCT	OUT	5	0.478	0.555
TR-VP19	2339376–2339507	I	6	21.5	132	89%	TCTGGC	VP2226	9	0.830	0.578
TR-VP21	447465–447520	I	6	9.5	55	94%	TGTGTC	VP0446	4	0.589	0.700
TR-VP25	744172–744218	II	6	7.8	46	93%	GTATCT	VPA0714	3	0.110	0.333

[a]Position into RIMD2210633 genome.
[b]Chromosome.
[c]Nei's diversity index.
[d]Homoplasy index.

Table 4 Gene diversity (H) estimates per locus and subpopulation

	TR-VP5	TR-VP 10	TR-VP16	TR-VP18	TR-VP19	TR-VP21	TR-VP25	Total
Asia	0.423	0.923	0.936	0.679	0.782	0.590	0.295	0.661 ± 0.241
Peru	0.568	0.903	0.934	0.388	0.857	0.597	0.029	0.610 ± 0.326
Chile	0.514	0.800	0.862	0.095	0.757	0.514	0.181	0.531 ± 0.301
Total	0.501	0.875	0.910	0.387	0.798	0.567	0.168	0.601 ± 0.290

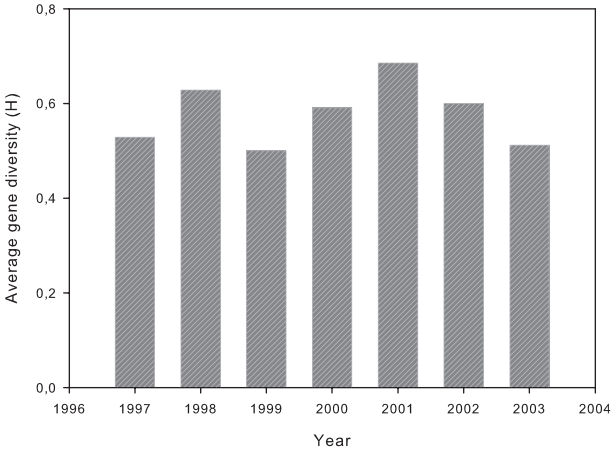

Fig. 1 Average gene diversity (H) over all loci within the Peruvian subpopulation throughout the period of study.

A temporal analysis of gene diversity within the Peruvian subpopulation revealed the highest values of H in 1998 and 2001, although no significant differences were detected in other years (Fig. 1).

Linkage disequilibrium

Analysis of the multilocus LD showed a significant ($P < 0.01$) positive weak LD for the overall population with a value of r_D of 0.038, suggesting a clonal structure within this group of strains. When analysing allelic profiles within each subpopulation, significant values were obtained only within the Peruvian subpopulation, which showed a low value of r_D of 0.061 ($P < 0.01$). Chilean and Asian subpopulations also showed low values of r_D although none of these subgroups yielded significant values. Throughout FSTAT program analysis, we examined locus pair associations in each subpopulation after Bonferroni correction ($N = 69$, adjusted alpha = 0.000794). Only the Peruvian set showed a statistical association ($P < 0.001$) between the locus TR-VP19 and the loci TR-VP10 and TR-VP18.

Genetic heterogeneity between populations

The SMM model was validated for the total set of the VNTR loci with the construction of an MST of all the *V. parahaemolyticus* strains. The number of differences in the VNTRs between closest relatives was evaluated and showed that at least 71% of the allelic changes were related to differences in one repeat, which fits the SMM. The evaluation of the allelic differences per locus showed a clear pattern of stepwise mutation for at least the loci TR-VP5, TR-VP19 and TR-VP21. The pattern was less clear for the two most polymorphic loci (TR-VP10 and TR-VP16), whereas the other loci (TR-VP18 and TR-VP25) did not have sufficient changes.

Genetic heterogeneity among the different subpopulations was investigated by computing the genetic distance, delta-mu squared $(\delta\mu)^2$, based on the SMM model. The highest genetic heterogeneity was observed between Asian and Chilean specimens (0.925), whereas a lower genetic distance was found within the South American strains suggesting a close proximity of the Peruvian and Chilean subpopulations (0.163). A closer genetic relatedness was detected between Peruvian and Asian strains (0.573) than between Asian and Chilean strains.

The R_{ST} index showed genetic relatedness values identical to those resulting from the genetic heterogeneity analysis. A significant genetic differentiation ($P < 0.05$) was detected between Asian and Latin American subpopulations, although the pairwise Asia-Peru subpopulations showed lower values of R_{ST} (0.182) than those obtained by comparing Asian and Chilean subpopulations ($R_{ST} = 0.243$). By contrast, a significant genetic differentiation was not observed between the Peruvian and Chilean subgroups ($R_{ST} = 0.082$), indicating a genetic association between these two populations. A significant ($P < 0.05$) genetic differentiation within each subpopulation was only observed among strains included in the Chilean group. Strains isolated during the first epidemic outbreak reported in 1997–1998 in Antofagasta showed a significant genetic differentiation from the strains obtained in the epidemic radiation of *V. parahaemolyticus* in Puerto Montt, whereas strains isolated from Peru were genetically homogeneous.

In accord with the results obtained from the genetic differentiation analyses, estimates of the gene flow (M) obtained from R_{ST} showed a close genetic relationship of all the South American strains, with a high rate of gene flow between the subpopulations from Peru and Chile (2.789). Differences in gene flow were detected between Asian subpopulations and the South American subgroups, with a higher genetic relatedness between Peruvian and Asian subpopulations (1.117) than between Asian and Chilean subpopulations (0.774).

Population structure and phylogenetic inferences

Results of Bayesian clustering analysis identified genetically distinct populations by allele frequencies and estimates of the fraction of the genotype that belonged to each cluster for each strain. Calculation of ΔK determined the most likely value of K, predicting the presence of two clusters (Fig. 2). The first cluster (cluster A) included the Asian strains and a group of closely related strains from South America isolated during the first epidemic expansion of pandemic *V. parahaemolyticus* in Peru and northern Chile (Antofagasta). The second cluster (cluster B) was composed of a second group of strains present in Peru in 1997 and dominating the

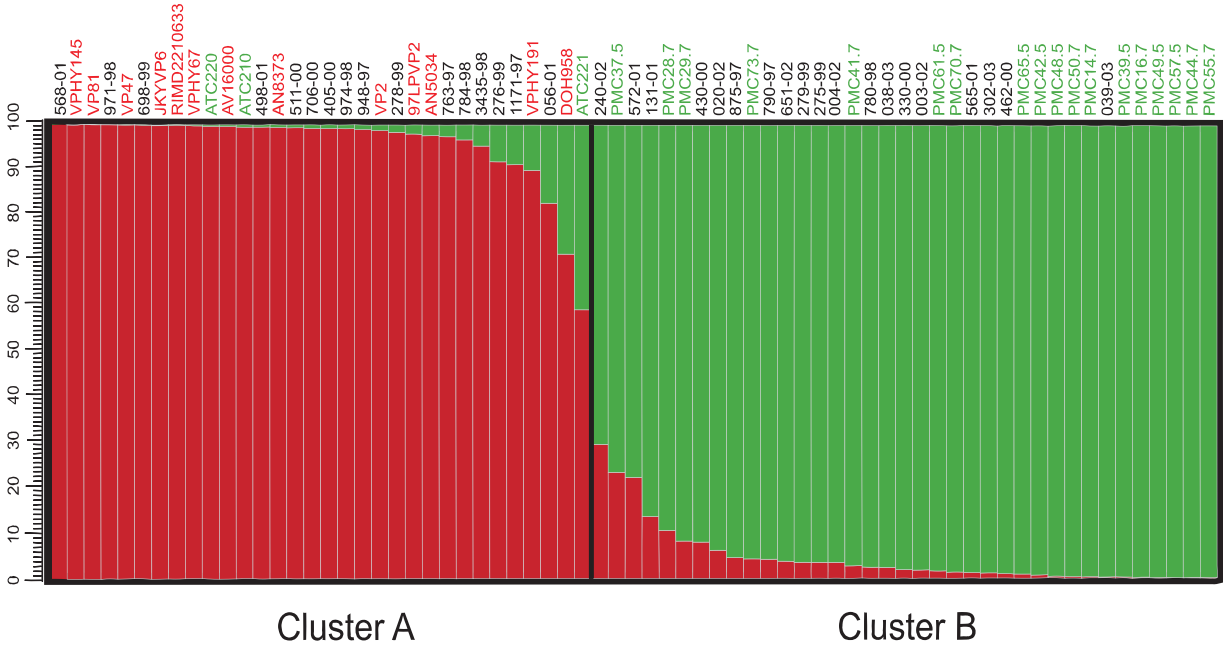

Fig. 2 Results of Bayesian clustering analysis inferred by the Structure for all the strains included in the study obtained from Asia (red codes), Peru (black codes) and Chile (green codes). Each individual is represented by a horizontal bar, the colours denote the different clusters, and the length of the coloured segment is the fraction of the genotype belonging to each cluster. Population structure was obtained using an admixture model where $K = 2$.

Peruvian subpopulations after 2000, as well as strains from the second epidemic radiation of pandemic *V. parahaemolyticus* in Puerto Montt. Clustering results showed a close relationship between pandemic strains arriving in Peru in 1997, the strains isolated in Antofagasta towards the end of 1997, and the theoretical population source originating in Asia, with these South American strains sharing more than 90% of their genotype with the Asian strains (Fig. 2). The second inferred cluster included strains that shared <10% of their genotype with the Asian strains.

MST of the 67 distinct MLVA genotypes identified among the 69 *V. parahaemolyticus* strains showed highly diverse genotypes with a T-shaped topology that distributed the different RT connected along two distinct branches distinguishing two groups (Fig. 3). Group A included all the strains from Asia and strains from Peru, primarily those isolated during the first years of the arrival of the pandemic clone in Peru, as well as the strains isolated in Antofagasta. A second group was composed of strains from Peru predominantly isolated before 2000 and strains from Chile exclusively detected in Puerto Montt. The topology of the MST correlated strongly with the population structure provided by Structure, with two exceptions. First, the Peruvian strains 790-97, 430-00, 020-02 and 131-03, which had been included within the cluster B in the Bayesian analysis, were located in the MST constituting a group in an intermediary position linking the two clusters identified by Structure. Second, two Peruvian strains, 3435-98 and 511-00, included in the Asian cluster by Structure, were included in the group B delineated by the MST. Most of the links in the MST corresponded to multiple repeat changes, although most of the Chilean and Peru-

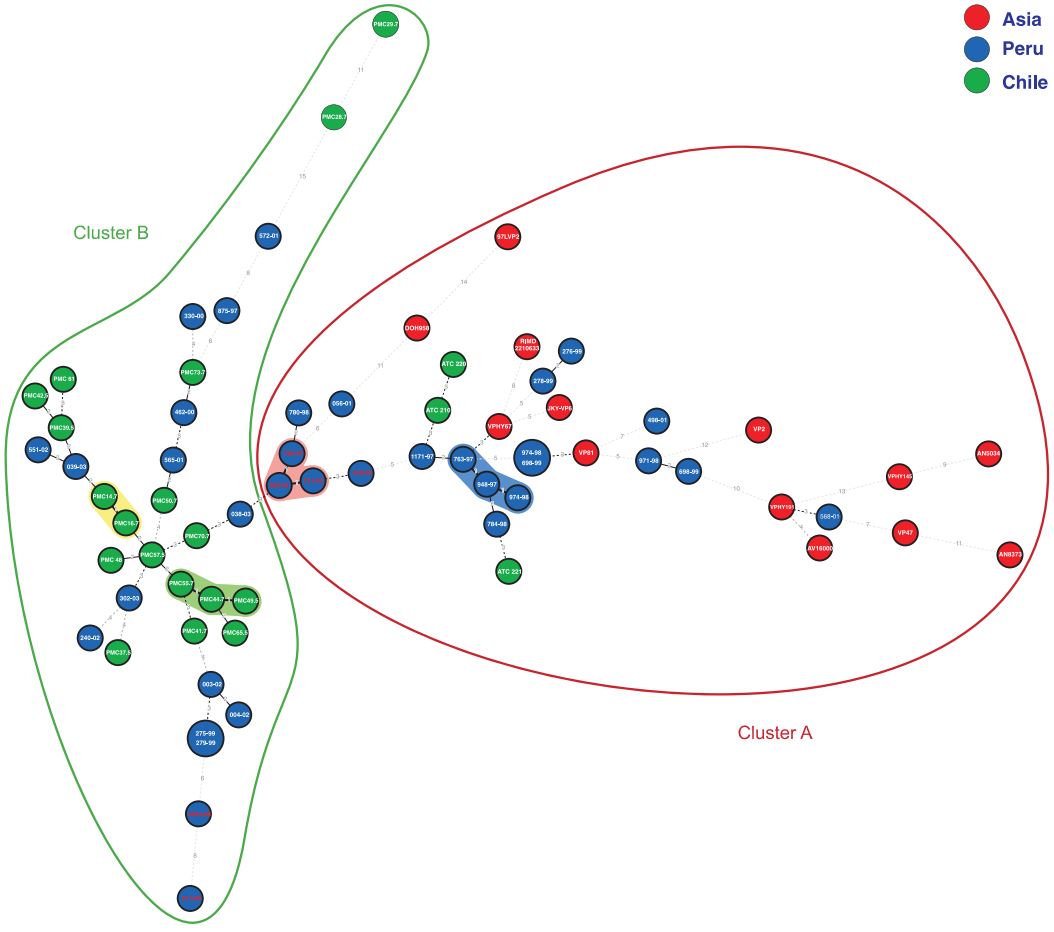

Fig. 3 Distribution of the different multiple loci VNTR analysis (MLVA) types in a minimum-spanning tree (MST). Each MLVA type is indicated by one node displayed as a circle that is connected by branches of MST. The lengths of the branches are proportional to the genetic distances (the sum of the absolute differences between the values of two MLVA profiles) between neighbouring profiles. The code within each circle identifies each strain, and circle colours indicate the origin of the strains. Shaded areas indicate repeat-type complexes defined as groups with a maximum neighbour distance of one change and a minimum size of two strains. Clusters A and B identified by Structure are delineated by red and green lines, respectively, whereas strain codes in red denote the discrepancies in the clustering results obtained by structure and MST.

vian strains showed a proximity to the four RT complexes identified in the MST with differences involving small changes among RTs. The overall homoplasy index calculated along the MST was 0.58, suggesting that the inferred links between genotypes would not be seriously affected by evolutionary convergence. The homoplasy was highly variable at the locus level (Table 3) with values ranging from 0.3 to 0.8 for TR-VP25 and TR-VP5, respectively, and intermediate indexes for the remaining loci.

To explore the robustness of the population structure suggested by Bayesian analysis and subsequently supported by the MST topology, Nei's gene diversity, associations among loci, genetic differentiation, and gene flow were estimated for the resulting clusters obtained from Structure. The cluster B showed a reduction in gene diversity (H = 0.558) in contrast to cluster A (H = 0.671). Estimated values of r_D across the inferred clusters were low and ranged from −0.020 to 0.042, showing only a significant association ($P < 0.01$) in cluster A between the loci TR-VP19 and TR-VP10. Results from R_{ST} indicated a significant population differentiation between the two clusters inferred by Structure ($R_{ST} = 0.260$; $P < 0.05$). No genetic differentiation was observed between strains isolated from Peru and Chile belonging to cluster A ($R_{ST} = 0.344$; $P > 0.05$), showing a close genetic relatedness between strains from Peru and Antofagasta. Additionally, no significant differences were detected comparing Peruvian and Chilean strains included in cluster B ($R_{ST} = 0.031$; $P > 0.05$), and between Asian and Chilean strains from Antofagasta ($R_{ST} = 0.213$; $P > 0.05$). Conversely, the largest genetic differences were detected comparing the Chilean populations belonging to clusters A (Antofagasta) and B (Puerto Montt) ($R_{ST} = 0.541$; $P < 0.05$). Estimates of the gene flow (M) showed results congruent with the data of genetic differentiation. The highest values of gene flow were obtained from the comparisons Peru B-Chile B (7.565), Asia-Peru A (1.142) and Asia-Chile A (0.921). Finally, the pairwise Chile A-Chile B showed the lowest number of migrants per generation (0.211), stressing the genetic differentiation of the subpopulations involved in the two epidemic outbreaks in Chile.

Remote sensing data were used to explore the possibility of an oceanic connection that allowed for the polewards shift of pandemic populations from Peru to southern Chile concurrently with the emergence of the infections in Chile in 2004 (Fig. 4). An extraordinary rise in seawater temperatures was observed towards the end of 2003 along the coast of Chile, as induced by the southwards movement of warm waters from southern Peru that reached the coasts of Puerto Montt in the final days of the year, just before the sudden emergence of infections in the area.

Discussion

One of the most extraordinary examples of the long distance spread of pathogenic *Vibrio parahaemolyticus* has probably been the arrival, dissemination and establishment of the O3:K6 clone populations along the Pacific Coast of South America since 1997. DNA restriction patterns of O3:K6 strains isolated in Peru and Chile have shown profiles closely related to strains from Asian countries obtained from the first epidemic radiation of the group in 1996–1997 (Gonzalez-Escalona *et al.* 2005; Martinez-Urtaza *et al.* 2008). Similar relationships were obtained through MLST using seven housekeeping genes (Gonzalez-Escalona *et al.* 2008). However, the high uniformity of restriction patterns and sequence types presented serious limitations in inferring the population sources and the origins of the invasive strains.

The use of the number of tandem repeats of microsatellite motifs has provided higher resolution for detecting genetic diversity among the uniform populations of O3:K6 strains (Kimura *et al.* 2008). MLVA analysis applied in this study showed a highly variable number of VNTR profiles with only two groups, each one comprised of two strains with identical allele profiles. Strains from Peru and Chile showed proximity to the different complexes inferred by the MST with differences involving a reduced number of repeat changes, which suggests a short evolutionary history and a recent differentiation. The compact clustering of the South American populations contrasts with the dispersion of the strains from Asia, which showed high diverse RT, in some cases, genetically distant from Peruvian and Chilean strains. The dispersion and dominance of multiple changes affecting the Asian subpopulations should be strongly influenced by the limited number of representative strains from this area, which should result in the absence of intermediary genotypes.

Evolutionary dynamics and mutations of microsatellite loci have had a substantial impact on estimations of population structure and levels of gene diversity. Although the patterns of microsatellite mutations are extremely complex (Primmer & Ellegren 1998; Brohede & Ellegren 1999), SMM seems to account for most of the observed patterns of microsatellite evolution (Estoup *et al.* 1995; Balloux & Lugon-Moulin 2002; Dettman & Taylor 2004), especially when short DNA repeat motifs are selected (Estoup *et al.* 1995). The evaluation of the SMM in the allelic changes observed in selected VNTR of the pandemic *V. parahaemolyticus* identified 71% of the allelic changes related through differences in one repeat. This value is close to the 64%, 75% and 81% observed in *Mycobacterium tuberculosis* (Wirth *et al.* 2008), *Escherichia coli* (Vogler *et al.* 2006) and yeast (Wierdl *et al.* 1997) VNTRs, respectively, and supports the

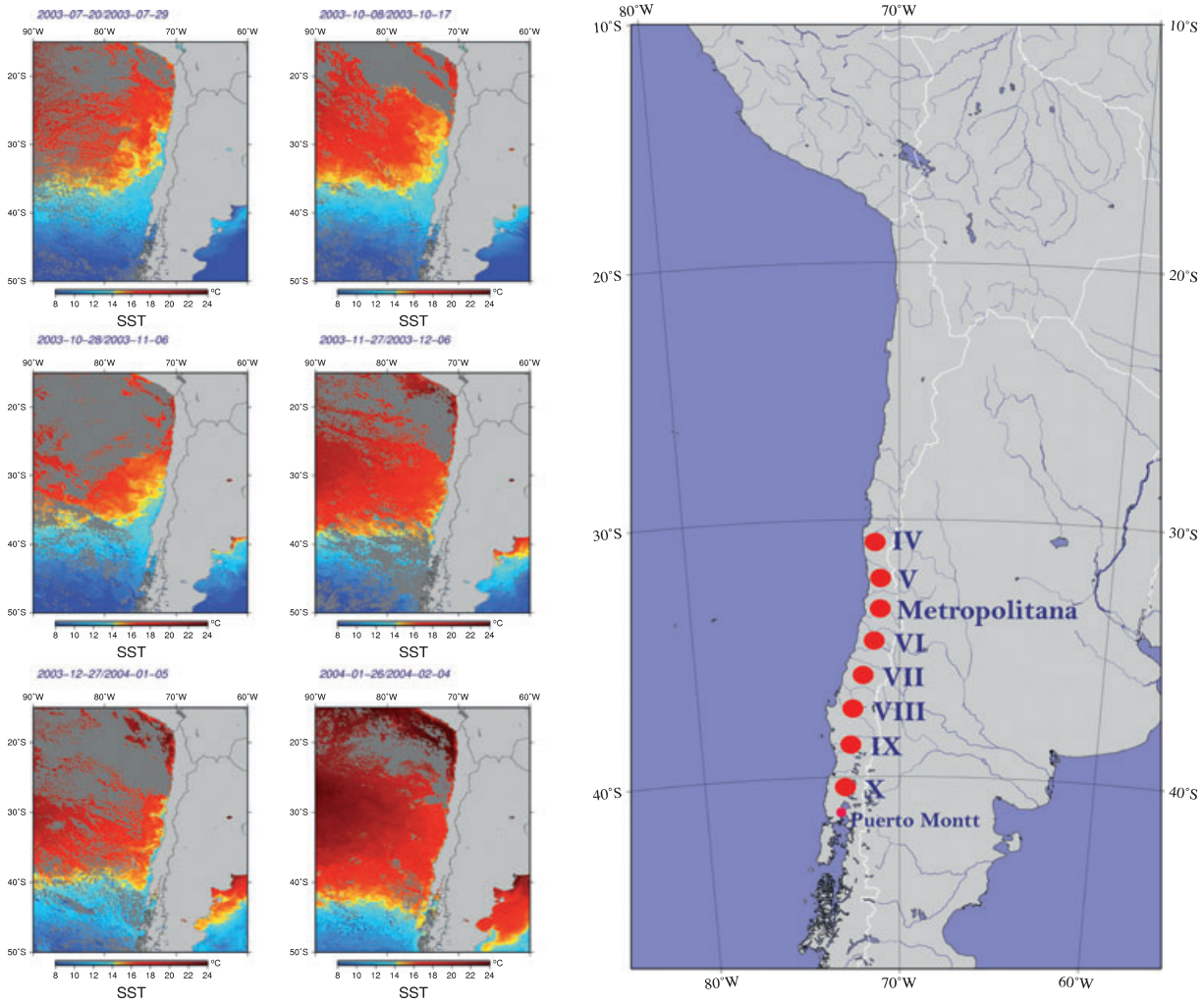

Fig. 4 Maps showing the polewards movement of warm waters from Peru to the south of Chile over the final months of 2003 and the beginning of 2004 and the parallel emergence of infection in different regions of Chile reported by the Ministry of Health on 15 January, 2004. Remote sensing imagery was constructed using 4 –km, 10-day composite SST fields obtained from night-time MODIS satellite passes. Infrared SST retrievals require cloud-free pixels but greatly improve coverage in coastal zones.

dominance of the SMM in the generation of variations in VNTR loci.

The resolution of the population structure is substantially affected by evolutionary convergence generated by homoplastic events, which may result in a misinterpretation of the genetic relationships among the *V. parahaemolyticus* strains resulting from the population structure inference. Homoplasy is inherent in microsatellite markers (Dettman & Taylor 2004) and is expected under SMM (Estoup *et al.* 1995); however, the values of homoplasy estimated for the populations of pandemic *V. parahaemolyticus* suggest that the inferred links between genotypes should not be seriously affected by homoplastic noise. A low value of homoplasy should also be consistent with the existence of a small effective population (Estoup *et al.* 1995; Dettman & Taylor 2004), as well as by the relatively short evolutionary history of the pandemic clone, leaving a reduced time for the evolutionary reversals or convergent mutations resulting in homoplasy (Pearson *et al.* 2009).

The population partitioning obtained by the MST and Bayesian analysis has been consistent with the epidemic spreading of the O3:K6 clone in South America, clearly reflecting two epidemic radiations involving specimens with a marked difference in genetic diversity and allele dominance. The first radiation has been linked to the arrival of genetically diverse specimens to Peru and to northern Chile with genotypes closely related to Asian strains. The second, less diverse group, cluster B, comprised strains exclusively from South America, with no genetic links to Asian strains. Both clusters were con-

nected in the MST by an intermediary group composed of four strains, which were included in cluster B by Structure. These strains showed a hybrid VNTR profile with four alleles distinctive to group B and three VNTR types characteristic of group A. However, no hybrid populations were detected among the population partitioning inferred by Structure, which suggests the simultaneous arrival of two independent groups in South America in 1997. From the time of their arrival, both clusters showed a differentiated evolutionary history, with the group A dominating at the onset until its decline in 2001, when group B initiated its principle radiation in Peru and its subsequent propagation to Chile in 2004.

Biological invasions of *Vibrio* populations can be mediated by human activities such as ballast water discharges, or induced by natural events such as the movement of oceanic waters. In either case, the propagule pressure, or number of dispersing specimens, and the genetic characteristics of the invading populations should be substantially different. Discharge of ballast waters should be related to a low propagule pressure or introductory effort: a unique introduction event of reduced genetic variation with low chances of success of establishment in the invading area (Lockwood *et al.* 2005). Ballast water-mediated invasions have been proposed for the entrance of O3:K6 strains into Texas in 1998 (Daniels *et al.* 2000) and A Coruña, Spain, in 2004 (Martinez-Urtaza *et al.* 2005). In both episodes, infections caused a unique epidemic outbreak in areas close to ports, although strains belonging to the O3:K6 clone were never detected in clinical cases or in the environment after the outbreaks (DePaola *et al.* 2000; Rodriguez-Castro *et al.* 2010). These observations support the theory that ballast-associated pathogen transport may usher transient populations of *V. parahaemolyticus* into an area, rather than establish long-term populations of pathogenic strains.

By contrast, the O3:K6 subpopulation arriving in Peru in 1997 showed an extraordinary genetic diversity, including specimens from multiple genetically distinct sources. This admixture represents clear support for an invasion mediated by an event of even higher magnitude than a single ballast water discharge. The El Niño phenomenon is characterized by the arrival of equatorial warm waters in a sequence of invasive waves that, in 1997, affected the coast for more than 6 months. The repetitive invasion of tropical masses of water may have resulted in a recurrent source of *V. parahaemolyticus* populations that may produce an invasion process characterized by a high propagule size in a single event or during a number of discrete invasion events. Additionally, the chance of success of a biological invasion positively correlates with the level of ecological disturbance of the invaded environment (Crawley 1986; D'Antonio *et al.* 2001; Lockwood *et al.* 2005). The arrival of El Niño waters in South America causes a general disruption of the environmental conditions of coastal areas, displacing the native species southwards. Under these exceptional conditions, nonnative populations of *V. parahaemolyticus* arriving in Peru may have found suitable conditions for success in their establishment. The restoration of cold waters in the area may have begun the retreat of O3:K6 populations towards tropical areas located in northern Peru, where they remain endemic to date (Martinez-Urtaza, personal communication).

The interpretation of these results provides novel information about the uncertain origin of the epidemic radiation of the O3:K6 clone in southern Chile. Strains from the epidemic radiation O3:K6 in the austral regions of Chile showed genotypes unrelated to Chilean O3:K6 strains from Antofagasta. According to these results, strains from Puerto Montt may have originated from an invasion of specimens belonging to cluster B from areas of the natural occurrence of O3:K6 strains in Peru. The Chilean coast is influenced by the highly productive Humboldt Current, which flows northwards along the west coast of South America ushering in low saline cold waters (Montecino & Lange 2009) and limiting the hypothetical biological transport from Peru to Chile. However, the Peru-Chile Countercurrent has been described in the area as a conduit for transporting equatorial warm waters in the surface layers from Northern Peru (8°S) to 30°S in Chile and occasionally to 40°S (Strub *et al.* 1995). The analysis of the surface seawater temperature data obtained by satellite in the days leading up to the emergence of infections in Chile in 2004 showed a polewards progression of warm waters from the south of Peru towards Chile that crossed the southern boundary of 40°S and reached Puerto Montt some days before the onset of infections. This specific pattern of spreading suggests a coastal dispersion of the aetiological agent of the infections from north to south in phase with the displacement of the warm waters. The Peru-Chile Countercurrent has been identified as a reliable vehicle for the transport of subtropical zooplankton species to Chilean areas located to the south of Puerto Montt (Hirakawa 1989).

The study of the arrival of the O3:K6 clone on the Pacific coasts of South America from a perspective of biological invasion has provided new information linking the origin of the invasion to Asian populations and has described the successful establishment of the O3:K6 subpopulations, first in Peru and subsequently in Chile, through population radiation of the Peruvian subpopulation. The results obtained in this study stress the importance of the analysis of population genetics for inferring the dynamics of disease and the propagation potential of environmental pathogens.

Acknowledgements

We thank Veronica Blanco Abad, Alba Rodriguez Castro, Isabel Mayán Barreiro and Silvia Carlés Gonzalez for technical assistance and Mitsuaki Nishibuchi (Kyoto University, Japan) for providing the Asian strains included in this study. We acknowledge Sylvain Brisse of Genotyping of Pathogens and Public Health at Institut Pasteur for his analyses with the MultiLocus Analyzer Software.

This study was funded by project '2007/CP381' from the Dirección Xeral de Cooperación Exterior, Xunta de Galicia.

References

Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102; Online: http://www.bio.ic.ac.uk/evolve/ software/multilocus.

Ansaruzzaman M, Chowdhury A, Bhuiyan NA *et al.* (2008) Characteristics of a pandemic clone of O3:K6 and O4:K68 *Vibrio parahaemolyticus* isolated in Beira, Mozambique. *Journal of Medical Microbiology*, **57**, 1502–1507.

Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**, 573–580.

Bisharat N, Cohen DI, Maiden MC *et al.* (2007) The evolution of genetic structure in the marine pathogen, *Vibrio vulnificus*. *Infection, Genetics and Evolution*, **7**, 685–693.

Brohede J, Ellegren H (1999) Microsatellite evolution: polarity of substitutions within repeats and neutrality of flanking sequences. *Proceedings Biological Sciencies*, **266**, 825–833.

Chowdhury NR, Chakraborty S, Eampokalap B *et al.* (2000) Clonal dissemination of *Vibrio parahaemolyticus* displaying similar DNA fingerprint but belonging to two different serovars (O3:K6 and O4:K68) in Thailand and India. *Epidemiology and Infection*, **125**, 17–25.

Crawley MJ (1986) The population biology of invaders. *Philosophical Transactions of the Royal Society B*, **314**, 711–731.

Daniels NA, MacKinnon L, Bishop R *et al.* (2000) *Vibrio parahaemolyticus* infections in the United States, 1973–1998. *Journal of Infectious Diseases*, **181**, 1661–1666.

D'Antonio CM, Levine J, Thomsen V *et al.* (2001) Ecosystem resistance to invasion and the role of propagule supply: a California perspective. *Journal of Mediterranean Ecology*, **2**, 233–245.

DePaola A, Capers GM, Motes M *et al.* (1992) Isolation of Latin American epidemic strain of *Vibrio cholerae* O1 from US Gulf Coast. *Lancet*, **339**, 624.

DePaola A, Kaysner CA, Bowers J, Cook DW (2000) Environmental investigations of *Vibrio parahaemolyticus* in oysters after outbreaks in Washington, Texas, and New York (1997 and 1998). *Applied and Environmental Microbiology*, **66**, 4649–4654.

Dettman JR, Taylor JW (2004) Mutation and evolution of microsatellite loci in Neurospora. *Genetics*, **168**, 1231–1248.

Estoup A, Garnery L, Solignac M, Cornuet JM (1995) Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. *Genetics*, **140**, 679–695.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Feil EJ, Li BC, Aanensen DM *et al.* (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *Journal of Bacteriology*, **186**, 1518–1530.

Goldstein DB, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 6723–6727.

Gonzalez-Escalona N, Cachicas V, Acevedo C *et al.* (2005) *Vibrio parahaemolyticus* diarrhea, Chile, 1998 and 2004. *Emerging Infectious Diseases*, **11**, 129–131.

Gonzalez-Escalona N, Martinez-Urtaza J, Romero J *et al.* (2008) Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. *Journal of Bacteriology*, **190**, 2831–2840.

Goodman SJ (1997) RST Calc: A collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Molecular Ecology*, **6**, 881–885.

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Online: http://www2.unil.ch/popgen/ softwares/fstat.htm.

Han H, Wong HC, Kan B *et al.* (2008) Genome plasticity of *Vibrio parahaemolyticus*: microevolution of the 'pandemic group'. *BMC Genomics*, **9**, 570.

Harth-Chu E, Espejo RT, Christen R, Guzman CA, Hofle MG (2009) Multiple-locus variable-number tandem-repeat analysis for clonal identification of *Vibrio parahaemolyticus* isolates by using capillary electrophoresis. *Applied and Environmental Microbiology.*, **75**, 4079–4088.

Hirakawa K (1989) Planktonic copepods from Aysen Fjord and adjacent waters, southern Chile. *Proceedings of the NIPR Symposium of Polar Biology*, **2**, 46–50.

Hunt DE, David LA, Gevers D *et al.* (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science*, **23**, 1081–1085.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Kimura B, Sekine Y, Takahashi H, *et al.* (2008) Multiple-locus variable-number of tandem-repeats analysis distinguishes Vibrio parahaemolyticus pandemic 03:K6 strains. *Journal of Microbiological Methods*, **72**, 313–320.

Laohaprertthisan V, Chowdhury A, Kongmuang U, *et al.* (2003) Prevalence and serodiversity of the pandemic clone among the clinical strains of *Vibrio parahaemolyticus* isolated in southern Thailand. *Epidemiology and Infection.*, **130**, 395–406.

Le Fleche P, Fabre M, Denoeud F *et al.* (2002) High resolution, on-line identification of strains from the *Mycobacterium tuberculosis* complex based on tandem repeat typing. *BMC Microbiology*, **2**, 37.

Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology and Evolution*, **20**, 223–228.

Makino K, Oshima K, Kurokawa K *et al.* (2003) Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. *Lancet*, **361**, 743–749.

Martinez-Urtaza J, Simental L, Velasco D *et al.* (2005) Pandemic *Vibrio* parahaemolyticus O3:K6, Europe. *Emerging Infectious Diseases*, **11**, 1319–1320.

Martinez-Urtaza J, Huapaya B, Gavilan RG *et al.* (2008) Emergence of Asiatic Vibrio diseases in South America in phase with El Nino. *Epidemiology*, **19**, 829–837.

Matsumoto C, Okuda J, Ishibashi M, *et al.* (2000) Pandemic spread of an 03:K6 clone of Vibrio parahaemolyticus and emergence of related strains evidenced by arbitrarily primed PCR and toxRS sequence analyses. *Journal of Clinical Microbiology*, **38**, 578–585.

McCarthy SA, Khambaty FM (1994) International dissemination of epidemic *Vibrio cholerae* by cargo ship ballast and other nonpotable waters. *Applied and Environmental Microbiology*, **60**, 2597–2601.

Montecino V, Lange CB (2009) The Humboldt Current System: ecosystem components and processes, fisheries, and sediment studies. *Progress in Oceanography*, **83**, 65–79.

Nair GB, Ramamurthy T, Bhattacharya SK *et al.* (2007) Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. *Clinical Microbiology Reviews*, **20**, 39–48.

Nei M (1987) *Molecular Genetics*. Columbia University Press, New York.

Niimi AJ (2004) Role of container vessels in the introduction of exotic species. *Marine Pollution Bulletin*, **49**, 778–782.

Okuda J, Ishibashi M, Hayakawa E *et al.* (1997) Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of strains from the same clonal group from Southeast Asian travelers arriving in Japan. *Journal of Clinical Microbiology*, **35**, 3150–3155.

Pearson T, Okinaka RT, Foster JT, Keim P (2009) Phylogenetic understanding of clonal populations in an era of whole genome sequencing. *Infection, Genetics and Evolution*, **9**, 1010–1019.

Primmer CR, Ellegren H (1998) Patterns of molecular evolution in avian microsatellites. *Molecular Biology and Evolution*, **15**, 997–1008.

Pritchard JK, Stephens M, Donelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Quilici ML, Robert-Pillot A, Picart J, Fournier JM (2005) Pandemic *Vibrio parahaemolyticus* O3:K6 spread, France. *Emerging Infectious Diseases*, **11**, 1148–1149.

Rodriguez-Castro A, Ansede-Bermejo J, Blanco-Abad V *et al.* (2010) Prevalence and genetic diversity of pathogenic populations of *Vibrio parahaemolyticus* in coastal waters of Galicia, Spain. *Environmental Micobiology Reports*, **2**, 58–66.

Rosenberg NA (2002) Distruct: a program for the graphical display of structure. *Molecular Ecology Notes*, **4**, 137–138.

Ruiz GM, Rawlings TK, Dobbs FC *et al.* (2000) Global spread of microorganisms by ships. *Nature*, **408**, 49–50.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Strub PT, Mesias J, James C (1995) Satellite observations of the Peru-Chile countercurrent. *Geophysical Research Letters*, **22**, 211–214.

Vogler AJ, Keys C, Nemoto Y *et al.* (2006) Effect of repeat copy number on variable-number tandem repeat mutations in *Escherichia coli* O157:H7. *Journal of Bacteriology*, **188**, 4253–4263.

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*, **146**, 769–779.

Wirth T, Hildebrand F, Allix-Beguec C *et al.* (2008) Origin, spread and demography of the *Mycobacterium tuberculosis* complex. *PLoS Pathogens*, **4**, e1000160.