Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South

Sophie Plouviez, T. M. Shank, Jean-Baptiste Faure, Claire Daguin-Thiébaut, Frédérique Viard, François Lallier, Didier Jollivet

Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South

S. PLOUVIEZ[1, 2, *], T.M. SHANK[3], B. FAURE[4], C. DAGUIN-THIEBAUT[1, 2], F. VIARD[1, 2], F.H. LALLIER[1, 2] and D. JOLLIVET[1, 2]

[1] Université Pierre et Marie Curie-Paris 6, Laboratoire Adaptation et Diversité en Milieu Marin, Roscoff, France

[2] CNRS, UMR7144 Station Biologique de Roscoff, France

[3] Biology Department MS#33, Woods Hole Oceanographic Institution, Woods Hole, MA 02543

[4] Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

Keywords: cytochrome oxidase I, deep-sea, Approximate Bayesian Computation, genetic isolation, allopatry, past demography

*Corresponding author: Sophie Plouviez, Génétique de l'Adaptation en Milieux Extrêmes, Station Biologique de Roscoff, Place Georges Teissier, BP 74, 29682 Roscoff cedex, France, fax number, E-mail: plouviez@sb-roscoff.fr

Running Title: Species vicariance on the East Pacific Rise

Abstract

The use of sequence polymorphism from individual mitochondrial genes to infer past demography has recently proved controversial because of the recurrence of selective sweeps acting over genes and the need for unlinked multi-locus datasets. However, comparative analyses using several species for one gene and/or multiple genes for one species can serve as a test for potential selective effects and clarify our understanding of historical demographic effects. This study compares nucleotide polymorphisms in mitochondrial Cytochrome Oxidase I across seven deep-sea hydrothermal vent species that live along the volcanically-active East Pacific Rise. Approximate Bayesian Computation (ABC) method, developed to trace back shared vicariant events across species pairs, indicates the occurrence of two across-species divergence times, and suggests that the present geographic patterns of genetic differentiation may be explained by two periods of significant population isolation. The oldest period dates back 11.6 Mya, and is associated with the vent limpet *Lepetodrilus elevatus,* while the most recent period of isolation is 1.3 Mya, apparently affected all other species examined and coincides with a transition zone across the equator. Moreover, significant negative Tajima's D and star-like networks were observed for all southern lineages, suggesting that these lineages experienced a concomitant demographic and geographic expansion about 100,000 to 300,000 generations ago. This expansion may have initiated from a wave of range expansions during the secondary colonization of new sites along the Southern East Pacific Rise (founder effects below the equator) or recurrent bottleneck events due to the increase of eruptive phases associated with the higher spreading rates of the ridge in this region.

Introduction

Patterns of genetic variation are powerful tools for elucidating population history over time and space. The extent to which current genetic signals reveal past demography mostly depends on the mutational model of the molecular marker used (e.g., microsatellites, introns or exons) and the analytical methods, either allele frequency (allele size, RFLP or SNPs) or sequence-based approaches (e.g., Sunnucks 2000). Among them, coalescence-based approaches using nucleotidic polymorphisms from one or more species/lineages have arisen as a powerful tool for tracing historical demographic events, genetic exchange, and population isolation (e.g., Hey & Nielsen 2004). Particularly powerful are coalescent analyses in a geographic context (i.e., phylogeography) with neutral markers lacking recombination. In these cases, gene genealogies and phylogenetic networks (Posada & Crandall 2001) together with DNA sequence diversity indexes can enable the reconstruction of population histories and identification of the possible reduction and/or expansion of populations (e.g., Emerson *et al.* 2001). Comparative approaches using multiple co-distributed species sharing congruent patterns of genetic structure between reciprocally monophyletic populations can reveal major isolating mechanisms or barriers to dispersal (Bermingham & Avise 1986).

Mitochondrial genes are the most widely used genes for phylogeographic analyses in animals (Avise 1998). The absence of recombination due to the maternal inheritance of the mitochondria (but see Skibinski *et al.* 1994), the supposed neutrality, relatively rapid mutation rates makes it sensitive to population subdivision (Avise 1998) and ease of data generation of this haploid sequence (Folmer *et al.* 1994) provide major advantages to using this gene. The use of these genes has been recently found to be controversial for the inference of population size and past demographic events because of the recurrence of possible

selective sweeps, which mimic the reduction of the effective population size following a bottleneck or a founder event (Bazin *et al.* 2006) and the need for unlinked multi-loci datasets. However, even if rarely done in the same study, comparison across multiple co-living species could help in disentangle demographic events from selective sweeps. For example, Lessa *et al.* (2003) showed concordant evidence for shared demographic expansion in North American mammals by comparing different species using the mitochondrial Cytochrome b gene only and Smith & Farrell (2005) revealed concomitant range expansions in *Moneilema gigas* and *M. armatum* beetles using the Cytochrome Oxidase I gene (mtCOI) gene.

Deep-sea hydrothermal vents on intermediate to fast-spreading mid-ocean ridges provide opportunities for testing shared vicariant events and concomitant demographic changes as populations form metapopulations subjected to frequent local extinctions and bottlenecks due to the recurrence of eruptive activity (Haymon *et al.* 1991; Jollivet *et al.* 1999; Tunnicliffe *et al.* 1997). Vent habitats are distributed along oceanic ridges where sulfide-rich fluid emissions are ephemeral both in time (years to a few decades; Shank *et al.* 1998) and space (Jollivet *et al.* 1999). Distances between sites vary from tens to hundreds of kilometres and can play an important role in population differentiation and allopatric speciation. Previous studies have focussed on the role of plate tectonics in favoring allopatry and subsequent secondary contacts across topographic (transform faults) and/or oceanographic (gyres) barriers (O'Mullan *et al.* 2001; Hurtado *et al.* 2004; Young *et al.* 2008; Faure *et al.* accepted). Moreover, at the scale of geological times, vent field (cluster of sites) displacements imposed by the dynamics of the underlying magmatic chamber are likely to favor a succession of isolation phases and secondary contacts (Jollivet *et al.* 1999). Plate

tectonics is thus one of the major forces acting on faunal composition, distinguishing seven hydrothermal vent biogeographical provinces (Tunnicliffe 1991; Van Dover *et al.* 2002; Bachraty *et al.* 2009). Differences in the faunistic composition of vent communities at a global scale are however difficult to reconcile with the origin of speciation processes, which in turn need to be assessed at a more restricted spatial scale (i.e., the ridge scale). At this level of spatial organization, well-dated topographical discontinuities that offset the ridge (like transform faults, overlapping spreading centers or microplates) are largely responsible for bottom current disruption, favoring breaks in gene flow that may contribute to vicariance affecting nearly the entire regional species pool. In this geological context, the concomitant isolation of taxa may be followed or not by secondary contacts and population admixtures, the extent of a contact zone and its age mostly depending on both the life-history traits of each species and the level of habitat connectivity. Comparative phylogeographic analyses are thus helpful to identify potential sources of vicariance shared between species. Sequence-based methods based on coalescence theory are then necessary to detect the major ecological/tectonic events that might have influenced most co-evolving species population dynamics and subsequently blurred the phylogenetic information (i.e., divergence time).

Based on faunal composition, the East Pacific Rise (EPR) has been recently sub-divided into two different biogeographic provinces one from each side of the equator (Bachraty *et al.* 2009), suggesting the presence of a potential barrier to gene flow at this latitude. The equatorial barrier was however established on faunal differences between the two communities based on the presence/absence of species. The present study aims at examining whether this hypothesized biogeographic break gave rise to distinct genetic divisions among vent species that co-occur in both Northern and Southern EPR. Indeed, Burton (1998) showed that the abrupt change in the faunal composition of nearshore marine

communities at Point Conception (California) did not always correlate with the genetic isolation of its species components. If the biogeographic split coincides with a species genetic break, two additional questions must be addressed: 1) is the position of the break at the same geographic location for all species, and 2) is the timing of separation consistent among taxa and thus represent a 'true' vicariant event between communities. Differences in species' dispersal capabilities may indeed allow them to track spreading vent fields differentially after the time of isolation. Finally, the study also aims to examine hypotheses regarding the possible stepwise population expansion in a given direction along the EPR. Gene flow asymmetries across the barrier would provide useful information about the history of colonization along ridges, particularly if taxa display similar patterns of directional gene flow. Comparative phylogeographic analysis of mtCOI sequences from seven broadly-distributed species from three different taxonomic groups (three gastropod limpets, three polychaetes and one bivalve species) was therefore conducted to test for shared vicariant events and demographic histories among the vent fauna.

Materials and methods

Biological specimens and molecular methods

Specimens from seven morphologically well-described species (*Bathymodiolus thermophilus, Alvinella pompejana, Hesiolyra bergi, Branchipolynoe symmytilida, Eulepetopsis vitrea Lepetodrilus ovalis, Lepetodrilus elevatus*) were collected along the East Pacific Rise during either the *Nautile* or DSV *Alvin* submersible expeditions. Nine hydrothermal vent localities were sampled in 1999, 2002, 2003 and 2004 (Table 1). The primary biological and habitat characteristics of the seven sampled species are summarized in Table 2. Animals were collected from sulfide chimneys, mussel beds, and vestimentiferan assemblages by grabbing the fauna with the submersible manipulator arm and storing them in an insulated box until the sub was recovered on board the support ship. Animals were individually preserved in absolute alcohol.

Genomic DNA of the three species of gastropods (*L. elevatus, L. ovalis* and *E. vitrea*) was extracted using NucleoSpin Tissue-Kit (Macherey-Nagel) following the manufacturer's instructions. Genomic DNA of polychaetes (*A. pompejana, B. symmytilida,* and *H. bergi)* and the bivalve *B. thermophilus* was extracted using a CTAB extraction procedure (Doyle & Dickson 1987). Tissues were digested in 600 µl of a 2% CTAB buffer solution (1.4 M NaCl, 0.2% 2-mercaptoethanol, 20 mM EDTA, 100 mM Tris-HCl pH 8, 0.1 $mg.ml^{-1}$ proteinase K) for two hours at 60°C. DNA was then purified by adding Chloroform-Isoamyl alcohol (24:1) and precipitated together with 1 ml of 100% isopropanol at -20°C for two hours. Finally, DNA pellets were washed with 70% ethanol and re-suspended in 50 µl of sterile H_2O.

Species-specific primers (Table 3) were developed using a first set of sequences obtained from an initial set of amplification and sequencing using 'universally' applicable

mitochondrial Cytochrome Oxidase I primers (LCOI, HCO1, Folmer *et al.* 1994). This procedure strengthened the amplification of DNA for *B. thermophilus*, *A. pompejana*, *B. symmytilida*, *H. bergi* and *L. elevatus*. Polymerase chain reaction amplifications were performed in a 25 µl reaction volume containing 1x reaction buffer (supplied by manufacturer), 2 mM $MgCl_2$, 0.12 mM each dNTPs, 0.5 µM each primers, 20 $\mu g.ml^{-1}$ Bovine Serum Albumin, 0.75 U Thermoprime plus DNA polymerase (Thermo Scientific), 2 µl of template DNA and sterile H_20. Thermal cycling parameters used an initial denaturation at 94°C for 2 min, followed by 40 cycles at 94°C for 35 s, appropriate annealing temperature for 35 s (Table 3), and 72°C for 1min20s, before a final 10-min extension at 72°C.

PCR products were purified with Millipore Montage^{TM} µPCR_{96} Cleanup kit and sequenced on an ABI 3130XL DNA analyser using BigDye® Terminator v3.1 (Applied Biosystems) sequencing chemistry following the manufacturer's protocol. Sequences were proofread in Chromas version 2.23 (http://www.technelysium.com.au/chromas.html) and aligned in BioEdit version 6.0.6 (Hall 1999).

Phylogeographic structure and divergence

For each species, number of haplotypes (*h*) and haplotype diversities (*Hd*) within each hydrothermal vent site were determined using DNAsp 4.10.3 (Rozas *et al.* 2003). The genealogical relationships among haplotypes were estimated using the median joining algorithm (Bandelt *et al.* 1999) of Network software (version 4.5.0.0; www.fluxus-engineering.com), allowing for the definition of clades (based on divergence up to 0.5%). The haplotype diversity (*Hd*), nucleotide diversity (π_n) and Watterson's theta (θ_w) were estimated within each divergent clade using DNAsp 4.10.3.

Geographic distribution of the clades was then assessed by plotting the dominant (ancestral) haplotype-frequency distributions for each locality. In order to detect putative geographic barriers to gene flow, F-statistics (ϕ_{st}; Hudson *et al.* 1992) were computed via DNAsp 4.10.3 with a sliding window of 3 populations from 21°S to 21°N in order to avoid sampling size effect. For example, for *E. vitrea*, the ϕ_{st} calculation grouped populations from 21°S-18°S-17°34'S, the second one from 18°S-17°34'S-17°25'S and so on until reaching the last grouping 9°N-13°N-21°N. Groups were not always the same among species depending on the locality sampling schedule (sites without sample). For each sliding window, departure of the F-statistic value from zero was tested using a permutation test. Clade admixtures (i.e. presence of divergent clades in the population) were examined for each population by looking at the proportion of synthetic haplotypes typifying each mitochondrial lineage in order to reveal potential geographic clines.

Substitution rates consistent with a molecular clock were tested for each species via the BEAUti/BEAST 1.4.8 MCMC package (Drummond & Rambaut 2007) using a subset of 15-20 informative sequences. A GTR + G + I substitution model was chosen to run an uncorrelated lognormal relaxed molecular clock model using a constant population size coalescent with 10,000,000 steps, a sampling every 500 steps and a burnin of 100,000 steps to reach convergence. The goodness of fit to a strict clock model was performed by examining of the posterior distribution of the standard deviation of the uncorrelated lognormal relaxed clock (Std ulrc). Divergence times (T) were then estimated between clades for each species when possible using the formula T = D/(2r), where D is the average divergence between clades and r the evolutionary rate per site per million years (Kumar *et al.* 1996). Three independent clock calibrations were previously performed on mtCOI, using known historical vicariance events of the ridge system that may have affected vent population

demography: (1) the Farallon ridge subduction under the American plate, 28.5 Mya (Chevaldonné *et al.* 2002), the formation of the Cascadia depression in the 450 km-long Blanco transform fault, 5 Mya (Johnson *et al.* 2006) and the formation of the Easter microplate, 5.9 Mya (Faure *et al.* accepted). Calibrations led to the estimation of a mutation rate of 2.2%, 2.8% and 3.8% for several vent annelids, *Lepetodrilus* limpets and *Bathymodiolus* bivalves, respectively, indicating a close evolution rate across vent taxa for the mtCOI gene. Consequently, estimations of divergence times between each taxon-pair were performed using a consensual (2.8%) mutation rate for the seven species.

To test for simultaneous divergence across clade pairs and common barriers to gene flow among the seven species, an integrative approach using the MsBayes Approximate Bayesian Computational (ABC) software (Hickerson *et al.* 2006b) was performed. The ABC method simulates sequence datasets for a series of taxon pairs that fit a divergence population model in which one ancestral population of size (N_A) splits into two daughter populations a and b (of size N_a and N_b) that independently endure a bottleneck of varying length τ' (and leading to actual sizes of N_a' and N_b'). Runs consist of three steps, allowing for the estimation of inter-species parameters (i.e., hyperparameters such as the number of possible divergence times (Ψ) across taxon pairs, and the corresponding mean ($E(\tau)$) and variance (Ω) of the divergence time τ). First, a vector of observed summary statistics is obtained from the observed dataset of each taxon pair. Second, through a series of hundreds of thousands of replicates, the hyperparameters and sub-parameters that typify the coalescent model are randomly drawn from hyper-prior and a sub-prior distributions in order to simulate a corresponding finite sequence dataset for the Y taxon pairs. These simulated datasets are then used to calculate a series of pseudo-summary statistics. These pseudo-summary statistics are compared to the observed ones to produce an approximate sample of simulations (i.e.,

simulations that yield the nearest summary statistics values to the observed ones), that is then used to draw the posterior distributions of the hyper- and sub-parameters of the taxon pairs following an arbitrary rejection/acceptance ratio of 0.1-0.2 and a local weighted linear regression.

Twenty 329 bp- long sequences per clade constituted the input dataset. Because of a lack of transversion in the between-clade divergence of some species, the transition-transversion ratio was fixed to two (Jukes & Cantor 1969). Joint prior distributions were used to perform 500,000 simulated draws. Out of these simulated draws, the closest to the observed dataset (1,000) were used to define the joint posterior distribution. Upper bounds for the prior population mutation parameter for the ancestral population size (θ_A) were chosen to be equal to 5 given the observed pairwise differences within clades (Hickerson *et al.* 2006b). These parameters were unchanged across runs allowing the cross-taxa comparison of results. ABC analysis was run on the seven species without any constraint on the number of divergence times (Ψ) to estimate the hyperparameters. ABC analysis was also performed fixing $\Psi = 2$ because obtained Ψ values were greater than 1. A final ABC analysis was computed on species after suppressing the pair-clade *L. elevatus* (for which divergence between clades was considered extremely high), without any constraint on Ψ.

Demographic history

Significantly non-null Tajima's *D* and its derived statistics (e.g., the Fu & Li's F) have been widely used to trace back demographic events (Glinka *et al.* 2003; Akey *et al.* 2004) providing that the studied gene region evolved under a lack of selective pressures. In order to test the hypothesis that vent fauna have undergone demographic changes in response to shared vicariant events, Tajima's (1989) statistic *D* was estimated (using DNAsp 4.10.3)

within each species for both the southern and northern clades located from each side of the equatorial biogeographic break following results of Bachraty *et al.* (2009). Tajima's D departure from zero was tested by a two tailed test assuming that *D* follows the beta distribution (Tajima 1989).

To further examine whether or not species endured the same demographic change, effective population size (N_e) and the exponential population growth parameter *g* were estimated from clades with significantly non-null Tajima's *D* values using Fluctuate (version 1.4, Kuhner *et al.* 1998). This allowed for the timing of expansion to be generated, following the formulation $\theta(t) = \theta_0 e^{-g\mu t}$ in which $\theta = N_e\mu$ with μ the mutation rate, *t* the time elapsed (in generations) and θ_0 the present state of θ for the clade under scrutiny. For each species, Fluctuate analyses were run with different values of short (i.e., 10, 50, 100, 150) and long (i.e., 2, 10, 20, 30) chains to evaluate convergence between runs. Sampling increment and the number of steps were respectively three and fifteen times greater than the number of sequences found within a clade for both short and long chains. To test if expansion signature was due to a wave of colonization (stepwise foundations) or from the recovering of a recent bottleneck, asymmetrical gene flow was checked across the barrier. For each species, the sliding window of Φ_{st} was used to locate the barrier and populations were then separated into two geographic groups from each part of the barrier, respectively. Watterson's theta (θ_w) was calculated for each group using DNAsp 4.10.3 (Rozas *et al.* 2003) and used as a starting parameter in Migrate 3.0.3 (Beerli & Felsenstein 1999; 2001) to estimate the gene flow parameters (θ*M) with MCMC runs of 10 short and 3 long chains.

Results

Phylogeographic structure and divergence across taxon pairs

For each species, the number of haplotypes (*h*) and haplotype diversities (*Hd*) within each population were assigned (Table 1). Median-joining networks exhibited similar topological patterns across all species, but differed in their structure (Fig. 1). Two divergent clades were detected in all networks except for: 1) *H. bergi,* in which 3 were divergent but markedly rare haplotypes were found in two of the most southern populations; and 2) for the scaleworm *B. symmytilida*, in which connections are more complex with at least 9 nearly equally-frequent haplotypes. Clades were geographically structured (with the exception of *B. symmytilida*), in particular with regard to the geographic location of the equator (Fig. 1). In each network, at least the southern clade (B) displayed a 'star'-like topology, with one central and frequent (ancestral) haplotype surrounded by a crown of derived singletons or multiple singletons (see however *B. thermophilus* for the exception of two closely-related equally-frequent haplotypes in the southern lineage). Derived haplotypes were often "unique" within a given locality. The northern clade was more diversified with more distant ('older') COI lineages. Divergence between the northern and the southern clades ranged from 0.9% for *B. thermophilus* to 6.5% for *L. elevatus*. By comparison, divergence for the two most frequent haplotypes of the scaleworm *B. symmytilida* was only 0.4%.

The geographic distribution of haplotypes (Fig. 1) together with Φ_{st} differentiation tests (Fig. 2) allowed us to discriminate clear patterns of geographic isolation for all species. When grouping localities from south to north, the sliding-window Φ_{st} became significantly different from zero between 17°S and the equator in all species except *H. bergi*. Occurrence for possible admixtures between divergent haplotypes typifying each clade along the EPR

was detected (Table 1). Except for *B. thermophilus* and *L. ovalis*, admixture is often caused by less than 10% of individuals inside the investigated populations.

The vent polychaete *A. pompejana* (*Ap*) and the gastropod limpet *E. vitrea* (*Ev*) both displayed exactly the same network architecture (Fig. 1). Coalescence trees are separated into two distinct clades (A and B) across the equator with 1 % (*Ap*) and 0.9 % (*Ev*) mutations accumulated into the divergence, respectively. Just one sequence from the southern clade (B) of *A. pompejana* was sampled in the northern EPR.

The vent limpet *L. elevatus* (*Le*) also displayed a pronounced geographic structure with two highly-divergent clades (divergence = 6.5 %), however clades overlap on the northern part of the EPR at 9°N (Fig. 1). In the northern clade, the population from 21°N displayed only one haplotype (i.e., Hap 1, the most frequent one in other populations) leading to a significant pairwise Φ_{st} between these two disjunct vent fields with and without considering the presence of clade B at 13°N. In the southern clade, the population from 9°N sharply differed from the SEPR ones by the presence of a unique haplotype (Hap 4) at a high frequency. This haplotype diverged from the most frequent (ancestral) one (Hap 3) by 3 mutations and possessed a crown of derived haplotypes, indicating that it had time to diversify since the southern *versus* northern isolation of lineages.

The bivalve *B. thermophilus* (*Bt*) and the gastropod *L. ovalis* (*Lo*) showed a nearly similar but reverse situation in which haplotypes of one clade were distributed along the entire ridge system whereas haplotypes from the other clade A are restricted to the northern (*Bt*) or the southern (*Lo*) parts of the EPR.

The two other vent species, *H. bergi* (*Hb*), which lives in sympatry with *A. pompejana*, and *B. symmytilida (Bs)*, which lives commensally with the bivalve *B. thermophilus*, were characterized by a lack of obvious differences across localities throughout

the EPR (Fig. 1). However, three individuals of *H. bergi* located in the south part of the EPR exhibited haplotypes with a clear divergence of nearly 1 % that may represent 'old' surviving mitochondrial lineages whereas an isolation-by-distance structure was observed in *B. symmytilida*.

Using Migrate 3.0.3, *A. pompejana* and *E. vitrea* displayed a nearly complete absence of mtCOI gene flow across the equatorial region, whereas other species showed high asymmetric gene flow (Table 4) from North to South for *H. bergi* and South to North for the remaining species.

For each species, standard deviation of the uncorrelated lognormal relaxed clock (Std ulrc) was closed to zero (mean ranged from 0.416 to 0.608 with maximum probabilities close to zero and upper HPDs under 1), indicating no mutation rate heterogeneity among clades. Thus, these datasets can be utilized for divergence time estimates that assume a constant rate. Clade splitting dates calculated with 0.28% per My and the T = D/(2r) formula leads to the following divergence times: T_{Le} = 11.6 My, T_{Lo} = 3.6 My, T_{Hb} = 1.9 My, T_{Ap} = 1.8 My, T_{Ev} = 1.6 My, T_{Bt} = 1.6 My, and T_{Bs} = 0.7 My. Estimates of Ω (= Variance(τ)/E(τ)) and the number of shared vicariant events Ψ with MsBayes using our 7 taxon pairs did not support a history of simultaneous divergence (Fig. 3 A, C). Indeed, the average Ψ value (1.858) is close to 2, indicating the occurrence for two possible isolation times. Moreover, the amount of variance (Ω = 0.384) is high, possibly indicating a biased estimation of the mean divergence time E(τ) (= 1.521, corresponding to 1.65 My using the T = (100 E(τ)) / (r * length of gene) equation (Hickerson *et al.* 2006a). Simulating vicariance events across the seven taxon pairs by fixing Ψ=2 indicated that six species have a simultaneous divergence time (with $E(\tau_1)$ = 1.088 corresponding to 1.2 My) and one species had a greater divergence time (with $E(\tau_2)$ = 2.639 corresponding to 2.9 My). The most divergent taxon-pair was expected to correspond to *L.*

344 *elevatus* because this species shown the greatest single T value across species. Simulation
345 excluding *L. elevatus* pairs (Fig. 3 B, D) showed a clear-cut estimate of Ψ nearly equal to one
346 (1.316) together with a markedly small variance ($\Omega = 0.047 < 0.1$). This latter estimate seems
347 to be robust, if the ancestral coalescent variance has a larger effect on the total genetic
348 divergence when divergence times are recent: a plausible explanation for EPR vent fauna in
349 light of the allopatric distribution across the equator. Here, the high variance of coalescent
350 estimates across taxon pairs was greatly compensated by the choice of the prior for θ_A (0.5-
351 5), highly stringent boundaries for this parameter being the consequence of the exceptionally
352 small ranges of observed pairwise differences within populations within each species pair.
353 Simultaneous divergence time E(τ) was estimated to 1.187, corresponding to a vicariant event
354 1.3 Mya. Divergence time (1.1 Mya using the T = D/2r equation) between Hap 3 and Hap 4
355 for *L. elevatus* was congruent with this simultaneous divergence time estimated for the six
356 other species, suggesting that all species were subjected to this vicariant event 1.3 Mya.
357 Results using BEAST were consistent with a common and recent vicariant event for all taxa
358 (including the southern lineage of *L. elevatus*) with average times of the most recent common
359 ancestor (T_{MRCA}) ranging from $9.2\ 10^{-4}$ to $2.3\ 10^{-3}$ with a strong overlapping of HPDs.
360
361 *Shared demographic histories in the South*
362 To examine how clades evolved after being putatively isolated, Tajima's D was calculated for
363 each clade separately over the seven species pairs. Results (Table 5) indicated a difference
364 between the southern and northern clades. Tajima's D was always significantly negative in
365 the southern clade regardless of the species analyzed. A significant excess of rare variants
366 was also detected in the polychaetes *H. bergi* and *B. symmytilida* for which only one clade
367 was detected over the entire range. Moreover, when only considering populations without
368 admixture (Table 1), 25%, 7% and 42% of populations had pairwise distributions differing

369 significantly from the neutral hypothesis in 21°N-13°N, 9°N-14°S and 17°25'S-21°S regions,
370 respectively. Fluctuate analyses of the southern lineages (Fig. 4; using 100 short and 20 long
371 chains allowing convergence between runs for all species) estimated the beginning of a
372 global population expansion between 100,000 to 300,000 generations for all species except *B.*
373 *symmytilida* (~1 million generations), with exponential population growth parameters *g*
374 ranging from 3,640 (*L. elevatus*) to 10,000 (*H. bergi*) when discarding *B. symmytilida* (g =
375 2,160).

Discussion

The seven deep-sea hydrothermal-vent species studied here possess different life-history traits (Table 2), which are likely to influence their ability to expand their range and colonize new localities. Despite biological disparities across species, mtCOI network topologies yielded similar topologies – two divergent clades located in Northern and Southern EPR, respectively, with only one exception, the commensal scaleworm *B. symmytilida*. Similar topologies raise the question of whether the equatorial barrier previously described by Bachraty *et al.* (2009) may have played a role in promoting a genetic break(s) across vent species and subsequent gene flow limitations after the isolation of populations. In order to test the 'vicariance' hypothesis, two main questions need to be answered: (1) did ancestral species split at the same time and (2) was the barrier impermeable enough to impose the same exact geographic patterns across the northern and southern vent lineages.

Number of splitting events in vent populations along the East Pacific Rise

In order to date vicariant events, the most currently used approaches consist of: 1) estimating divergence times for each taxon-pairs; 2) comparing these dates across taxa accounting for stochastic processes associated with the coalescence process; and 3) estimating the demographic evolution of populations. This technique has been widely used to propose vicariant events when speciation has occurred in the distant past (several millions years) or when the target taxa represent conspecific species displaying the same life-history traits across the same habitat. For example, simultaneous vicariant events have been documented for coastal invertebrates inhabiting sediments along the European coasts of the North Atlantic (Jolly *et al.* 2006) or the intertidal limpets separated across the Greater Cook Strait in New

Zealand (Goldstien *et al.* 2006). In our case, the vicariance date of 11.6 millions years estimated for *L. elevatus* coincides with two cryptic, possibly hybridizing, species detected by Matabos *et al.* (2008) at 9°50N/EPR using allozymes and mtCOI sequences. However, divergence times of the other remaining species are more recent and markedly close to each other (from 3.6 Mya for *L. ovalis* to 1.6 Mya for *B. thermophilus*, 0.7 Mya for *B. symmytilida*), suggesting that they may be the result of a same vicariant event.

However, testing whether co-distributed taxa share a common history of simultaneous vicariance just by superimposing divergence times may lead to erroneous conclusions if the vent community mixes species groups of different origins possibly via arriving into the ridge system at different times. According to Hickerson *et al.* (2006b), variation between estimated divergence times is largely explained by the mutational and coalescent variance. A recent approach (Hickerson *et al.* 2006b) allows testing simultaneous divergence using Approximate Bayesian Computation (ABC) by incorporating differences in the demographic history of each sister population during the isolation process. Simulations using our seven taxon-pairs indicated two possible splitting events (simulation without fixing Ψ), with the most recent event affecting six species and the second one with a much older divergence time (simulation with Ψ=2), affecting only one species, *L. elevatus* according to the previous species-by-species estimation of divergence times. Although the most ancient divergence time estimated for one species (*L. elevatus*) seems to be underestimated (2.9 My), the simultaneous divergence time (1.2 My) estimated with Ψ=2 is consistent with the 1.3 My detected for the most recent event when using Ψ=1 and the six remaining taxon-pairs. Moreover, *L. elevatus* divergence between Hap 3 and 4 (1.1 Mya) inside the clade B are congruent with this simultaneous divergent time, indicating that *L. elevatus* was submitted to this more recent isolation event as well. Consequently, results are congruent with the hypothesis of two

424 independent isolation events: 11.6 Mya for *L. elevatus* and 1.3 Mya for the other taxon-pairs.
425 Because allopatric speciation is usually caused by gene flow disruption due to
426 physical/tectonic barriers that modify hydrothermalism along ridge axes (Jollivet 1996), the
427 two possible splitting dates sub-dividing these vent species into two distinct phylogeographic
428 clades are likely explained by the formation of geological discontinuities that progressively
429 offset the ridge crest.
430
431 *Geological barriers to gene flow along the East Pacific Rise*
432 Plate tectonics most likely influence genetic structure and speciation of the hydrothermal vent
433 fauna, but the exact events are tough to pin point. Modeling the formation of the East Pacific
434 ridge system based on fossil records and magnetic/gravimetric anomalies suggested that the
435 subduction of the Farallon plate under the American plate provoked multiple reorganizations
436 of the now-extinct Pacific/Farallon ridges while forming the present East Pacific Rise
437 (Mammerickx *et al.* 1980). Two major co-occurring tectonic events could coincide with the
438 divergence date causing the *L. elevatus* split 11.6 Mya: (1) the Bauer microplate rotation and
439 (2) the Mathematician Ridge reorientation. The formation of the Bauer Overlapping
440 Spreading Centre initiated about 17 Mya (at latitudes located between 10 to 15°S), evolved
441 into a microplate between 15-11 Mya (Eakins & Lonsdale 2003). Such reorganization
442 provoked ridge offset and the formation of two parallel active ridges that could have been
443 responsible for modifying the bottom current patterns in this region and subsequently the *L.*
444 *elevatus* divergence. During the same period (12.5-11 Mya, Mammerickx & Klitgord 1982),
445 the Mathematician Ridge situated further North (12°N to 17°N) was subjected to abrupt
446 changes in magnetic and bathymetric orientations, provoking the fossilization of transform
447 faults and the formation of a parallel ridge system, the Moctezuma trough in this northern

region (Mammerickx & Klitgord 1982). Because the Mathematician ridge reorganization event appears to be much closer to latitudes at which the two present divergent lineages overlap (between 9°N and 13°N), this latter scenario seems to be more consistent with the *L. elevatus* split.

The seven-studied vent species seem to share a simultaneous vicariant event ~ 1.3 Mya. Moreover, this vicariant event coincides with a significant north/south differentiation of vent populations for all species with the exception of the polychaete *H. bergi.* However, this polychaete species displays three divergent haplotypes sampled at the most southern sites, suggesting that either the second clade was under-sampled or went nearly extinct. Even if *B. symmytilida* and *L. elevatus* seem to display a more complex population history, both species conform to the hypothesis of a recent North/South isolation. This is particularly clear when considering the genetic differentiation observed between populations located at 9°N and 14°S for *B. symmytilida* and the southern lineage (clade B) of *L. elevatus* at the same geographical sites. By superimposing geographic distribution of haplotypes and the position at which genetic differentiation increases significantly across species, this barrier is likely to be positioned between the Equator and 17°S, and defines a clear transition zone between the Northern and Southern EPR. Most transform faults located between 9°N and 17°S (i.e., Quebrada/Discovery/Gofar fracture zone system, or Wilkes and Garrett transform faults) deeply offset the ridge axis into separate segments around 1 to 2 millions years ago (Kureth & Rea 1981; Naar & Hey 1989; Francheteau *et al.* 1990), a scenario consistent with the simultaneous 1.3 Mya divergent time. As an example, the 450-km-long Blanco Transform fault is known to be responsible for the speciation of *Lepetodrilus fucensis* and *L. gordensis* (Johnson *et al.* 2006) with a divergence of 7.3% on the mtCOI gene and allele frequency inversion at the phosphoglucomutase gene. However, the impact of this fault as a barrier to

gene flow was less intense for the tubeworm *Ridgeia piscesae,* for which isolation did not lead to reciprocal monophyletic clades but only abrupt haplotype frequency differentiation in populations located at each part of the ridge offset (Young *et al*. 2008).

The slow and progressive formation of a tectonic barrier (rate at which a ridge offset is typically 5-10 cm/year), could explain why slight discrepancies in population differentiation still hold across species, with some species able to cross the barrier when others are not (Knowlton & Weigt 1998). The slight differences observed in the barrier positioning therefore may be attributable to differing dispersal capabilities. Indeed, egg size, egg vitellogenin content, and larval developmental mode (see Table 2) are good apparent indicators of dispersal capabilities and vary greatly across species. These characteristics affect the buoyancy of propagules and are likely to change the vertical dispersal of a larva in the water column (e.g., Mullineaux *et al.* 2005). Larvae could be thus subjected to different water currents and subsequent divergent trajectories. This may be particularly the case of *Bathymodiolus* bivalves which have planktotrophic larvae able to reach the upper water column (Arellano & Young 2009) as opposed to *Lepetodrilus* gastropods whose larvae are mainly found beneath hydrothermal plume layers, less than 200 meters above the seafloor (Mullineaux *et al.* 2005). The extremely great size of mature oocytes (400 μm) of the scaleworm *Branchipolynoe* spp. may also explained why this species is so weakly affected by the equatorial barrier as large yolky eggs may delay larval metamorphosis for months (Jollivet *et al.* 2000). The frequency of available vent habitat may also affect the ability of vent species to cross the barrier, as diffuse venting systems are more prevalent along the ridge compared to vent chimneys. However, the habitat distribution does not seem to play an important role to this extent as lineages from the two chimney-living species (*Ap* and *Hb*) have a markedly contrasted geographic range.

Regardless of the species, the precise spatial positioning of the barrier remains difficult to establish, firstly because some of these barriers may have disappeared (i.e., the Bauer microplate) and, secondly because migration events and subsequent secondary contacts likely mask the exact position of the barriers. Secondary contact zones detected in this study often display clines of clade-specific haplotype frequencies (responsible for north/south EPR differentiation). Such clines have been detected particularly in *L. ovalis* and *B. thermophilus* and might be attributed to preferential migration routes along the EPR, as suggested by the strong South to North asymmetric gene flow revealed by Migrate results. However, differential lineage extinctions between northern and southern EPR (regional extinctions of *L. ovalis* clade B in northern EPR and/or *B. thermophilus* clade A in the southern EPR) could also explain such haplotype distributions.

Evidence for simultaneous population expansions in the south

Comparing haplotype coalescence trees and overall gene diversities between reciprocal monophyletic clades of the examined vent taxa (with exception of *H. bergi* and *B. symmytilida* for which a single clade is preponderant) indicated large discrepancies between the northern and southern lineages. Star-like topologies together with reduced gene diversities were indeed observed for southern lineages (generally not for the northern ones) and are strongly suggestive of non-equilibrium dynamics. These results were supported by significant negative Tajima's D for the southern clades. Furthermore, a greater proportion of populations showed departure to equilibrium between 17°25'S and 21°S when compared to the most northern populations. These departures to equilibrium may be explained by either a recent demographic expansion along the Southern EPR (*sensu* Harpending *et al.* 1998) or alternatively recurrent selective sweeps at the mitochondrial locus (Bazin *et al.* 2006).

However, the positive selection hypothesis here imposes that a series of nearly-simultaneous fixations of advantageous alleles co-occurred in different species, and only for the most southern cryptic lineage: a situation in which mitochondria would have been sensitive to positive fixation and able to sweep at an extremely high rate in the south but not in the north. In addition, Fluctuate analyses performed on our seven vent taxa revealed a nearly simultaneous expansion of southern populations for all species except the commensal scaleworm *B. symmytilida*. Altogether the selective hypothesis is highly unlikely because simultaneous selective sweeps across different vent taxa is highly improbable.

If we accept the idea that populations expanded on one side of the barrier after having been separated about 1.3 Mya, the nearly simultaneous expansion would date back to 100,000 to 300,000 generations for all species except *B. symmytilida* (around 800,000 generations). Making the assumption that one to two generations occur per year for nearly all vent species, the date of expansion would fall in the last 500,000 years and coincide well with an expansion subsequent to the splitting of the two vicariant lineages. Two hypotheses are proposed: (1) this demographic expansion coincides with a geographic expansion of most lineages to the south by subsequent founder events until discouraged by the Eastern Microplate around 23°S; and/or (2) the southern EPR endured a large catastrophic/eruptive event that destroyed most of the vent fauna throughout the region, thus causing a large bottleneck, about 0.5 Mya. The second hypothesis appears to be more probable, as eruptive events can be frequent occurrences along the EPR, with the observation of two eruptions in the last twenty years near 9°50N (Haymon *et al.* 1993; Cowen *et al.* 2007). Such an assumption is also in agreement with the lack of concordant asymmetric flow across species from north to south. Indeed, vent species did not display similar patterns of orientated gene flow across the barrier, whereby gene flow was mainly orientated in the opposite direction

from south to north. Such eruptive events may also to be responsible for a possible bottleneck at 21°N (only one haplotype has been sampled for *L. elevatus*). Multiple catastrophic events along the southern EPR could therefore explain this cross-species demographical pattern.

Recent theoretical studies argue that mitochondrial genes may not be appropriate to perform phylogeographic analyses despite some advantages such as the non-recombination of the gene (Bazin *et al.* 2006). The present study illustrates the usefulness of mitochondrial markers when used across a set of species sharing the same environment, regional distribution, and more or less the same history. Comparing multiple phylogeographic patterns can help in discriminating demographic *versus* selective effects, and thus yielding a better understanding on the micro-evolutionary processes that shapie the geographic structure of populations.

Acknowledgments

We thank the chief scientists and 'Nautile' and 'Alvin' crews for their technical support and efforts during our oceanographic expeditions: HOPE99, PHARE2002 (F. Gaill and N. Lebris), Extreme 2003 (AT11-4 cruise) and BIOSPEEDO2004. We are very grateful to Stéphane Hourdez and Eric Thiébaut for collecting and sorting polychaetes and gastropods and to Marjolaine Matabos for her help in the diagnosis of Lepetodrilid gastropods. We are also truly indebted to the sequencing genomic plateform (GENOMER, Station Biologique de Roscoff, France) for DNA direct sequencing. This work was supported by the GDR Ecchis, ANR-06-BDV-005 (Deep Oases) and ANR-05-BLAN-0407 (Alvi_Stress_Adapt). S. Plouviez was supported by a PhD grant from the Université Pierre et Marie Curie. T.M. Shank was supported by the National Science Foundation (OCE-01-29394 and OCE-03-16348) and a Fellowship from the Deep-Ocean Exploration Institute, Woods Hole

568 Oceanographic Institution.

569 **Table 1** Location, size of faunistic samples, number of haplotypes and admixture between clades at population level along the East Pacific
570 Rise for the seven species.
571 *N*, number of mtCOI sequences used in the study per sample and species (Abs stands for the absence of the species at the studied site. 0
572 means that no sample could be collected although the species was observed at the site) including those coming from GenBank (in
573 brackets); *h*, number of haplotypes; *Hd*, haplotype diversity; *Adm*, number of sequences from clade A / clade B; *Bt*, *B. thermophilus*; *Ap*, *A.*
574 *pompejana*; *Hb*, *H. bergi*; *Bs*, *B. symmytilida*; *Ev*, *E. vitrea*; *Lo*, *L. ovalis*; *Le*, *L. elevatus*
575

Vent field	Geographical position	Depth (m)		Species						
				Bt	*Ap*	*Hb*	*Bs*	*Ev*	*Lo*	*Le*
21°N	20°49'N	2606	*N*	Abs	7 (7)	0	Abs	1	0	37
	109°06'W		*h* (*Hd*)		6 (0.952)			1		1 (0.000)
			Adm		7/0			1/0		37/0
13°N	12°43-50'N	2560-2700	*N*	12 (12)	72 (5)	153	0	7	0	150
	103°53-57'W		*h* (*Hd*)	4 (0.561)	26 (0.911)	23 (0.720)		3 (0.667)		13 (0.523)
			Adm	8/4	72/0	153/0		7/0		147/3
9°N	9°31-51'N	2500-2585	*N*	59 (12)	31 (21)	10	39(8)	10	15	63
	104°15-18'W		*h* (*Hd*)	16 (0.729)	16 (0.929)	4 (0.533)	27 (0.968)	4 (0.533)	12 (0.962)	25 (0.828)
			Adm	32/25	30/0	10/0	9/30	10/0	15/0	2/61

7°S	7°25'S	2735-2752	*N*	58 (12)	14	0	20	27	6	1
	107°47-49W		*h* (*Hd*)	21 (0.837)	7 (0.758)		15 (0.953)	9 (0.687)	2 (0.333)	1
			Adm	22/38	0/14		3/17	0/27	6/0	0/1
14°S	13°59'S	2623-2632	*N*	30	23 (3)	8	17	48	22	44
	112°29'W		*h* (*Hd*)	14 (0.825)	8 (0.581)	3 (0.464)	13 (0.971)	11 (0.647)	7 (0.714)	8 (0.368)
			Adm	0/30	0/23	8/0	6/11	0/48	11/11	0/44
17°25'S	17°25'S	2578-2590	*N*	33 (12)	51 (21)	0	29	72	52	49
	113°12'W		*h* (*Hd*)	8 (0.585)	19 (0.610)		22 (0.982)	17 (0.725)	18 (0.749)	10 (0.489)
			Adm	0/33	0/51		4/25	0/72	7/45	0/49
17°35'S	17°35-36'S	2591-2597	*N*	60	36	34	49	51	46	64
	113°15'W		*h* (*Hd*)	24 (0.817)	10 (0.483)	7 (0.373)	30 (0.957)	15 (0.503)	13 (0.650)	17 (0.588)
			Adm	2/58	0/36	33/1	7/42	0/51	5/41	0/64
18°S	18°26-37'S	2636-2680	*N*	30	29	11	29	47	24	0
	113°23-24'W		*h* (*Hd*)	14 (0.825)	11 (0.724)	3 (0.345)	20 (0.961)	13 (0.677)	8 (0.757)	
			Adm	0/30	0/29	11/0	6/23	0/47	5/19	
21°S	21°25-33'S	2804-2840	*N*	27	55	19	24	34	22	31
	114°16-18'W		*h* (*Hd*)	12 (0.818)	14 (0.613)	3 (0.292)	19 (0.982)	15 (0.779)	9 (0.745)	6 (0.353)
			Adm	0/27	0/55	17/2	2/22	0/34	3/19	0/31

576

577 **Table 2** Biological and ecological characteristic of the seven studied vent invertebrates

Species	Fertilization	Fecundity (oocytes female^{-1})	Egg size (µm)	Dispersal mode	Reproduction	Habitat	Reference
Bt	External	1,000000	50	Plantoctrophic larva	Discontinuous	Diffuse venting	Le Pennec *et al.* 1984; Tunnicliffe 1991; Jollivet 1996
Ap	Internal	200,000	180	Lecithotrophic larva	Nearly-continuous	Vent chimney	Jollivet 1996; Chevaldonné *et al.* 1997; Faure *et al.* 2007
Hb						Vent chimney	Jollivet 1996
Bs	Internal	1,000	400	Lecithotrophic larva	Continuous or frequently intermittent	Palleal cavity of *mussels*	Van Dover *et al.* 1999; Jollivet *et al.* 2000; Plouviez *et al.* 2008
Ev	Internal	200	230	Lecithotrophic larva	Continuous	Basaltic rocks and mussel shells	Tyler *et al.* 2008
Lo	Internal	1,000	90	Lecithotrophic larva	Continuous	Mussel shells	Fretter 1988; Mullineaux *et al.* 1995; Sadosky *et al.* 2002; Tyler *et al.* 2008
Le	Internal	1,000	90	Lecithotrophic larva	Continuous	Vestimentiferan tubes and mussel shells	Fretter 1988; Mullineaux *et al.* 1995; Sadosky *et al.* 2002; Tyler *et al.* 2008

578 **Table 3** Species-specific primers

Species	Primer sequences (5'-3')	T_a (°C)
B. thermophilus	**F:** TGTGGTCTGGAATAATTGGAAC	50
	R: ATAAAAAGATGTATTRAARTGACG	
A. pompejana	**F:** TATTTGGTATTTGGGCAGGTC	57
	R: GATGGGTCGAAGAATGATGTG	
B. symmytilida	**F:** CCCTTTACTTTCTATTTGGC	51
	R: ATTTCGATCTGTTAGGAGTATG	
H. bergi	**F:** CATACAAATAGTGGTACTCGTTC	51
	R: TTCCTTTTCGGACTATGAG	
L. elevatus	**F:** TGARCTYGGACAACCRGGAG	56
	R: RGGGTCAAAGAARGARGTGTT	

579 F, forward primer; R, reverse primer; T_a, annealing temperature

580 **Table 4** Theta (θ) and migration (θ*M) parameters in each species. *Bt*, *B. thermophilus*; *Ap*,
581 *A. pompejana*; *Hb*, *H. bergi*; *Bs*, *B. symmytilida*; *Ev*, *E. vitrea*; *Lo*, *L. ovalis*; *Le*, *L. elevatus*

Parameters	*Bt*	*Ap*	*Hb*	*Bs*	*Ev*	*Lo*	*Le*
θ North population	0.014	0.016	0.024	0.050	0.007	0.018	0.040
θ South population	0.034	0.024	0.009	0.065	0.030	0.018	0.013
θ*M North to South	0.000	0.000	2.338	1.142	0.280	0.377	0.000
θ*M South to North	21.676	0.572	0.000	90.882	0.000	15.292	28.077

582 **Table 5** Haplotype diversity (*Hd*), nucleotide diversity (π_n), Watterson's theta per site from
583 number of segregating sites (θ_w), and Tajima's *D* of clades from studied species and overall
584 Φ_{st} values for each species

Species/clade (Φ_{st})	*n*	*Hd* (SD)	π_n (SD)	θ_w (SD)	*D*
B. thermophilus (0.255***)					
Northern clade	64	0.208 (0.068)	0.000 (0.000)	0.002 (0.001)	-2.011*
Southern clade	245	0.812 (0.019)	0.003 (0.000)	0.016 (0.004)	-2.452***
A. pompejana (0.533***)					
Northern clade	108	0.911 (0.014)	0.009 (0.000)	0.014 (0.004)	-0.983^{NS}
Southern clade	210	0.618 (0.040)	0.002 (0.000)	0.015 (0.004)	-2.564***
H. bergi (0.018**)					
Major clade	232	0.614 (0.036)	0.002 (0.000)	0.015 (0.004)	-2.379**
Southern minor clade	3	-	-	-	-
B. symmytilida (0.031***)	208	0.970 (0.005)	0.007 (0.000)	0.022 (0.005)	-2.137**
E. vitrea (0.578***)					
Northern clade	18	0.562 (0.134)	0.002 (0.001)	0.005 (0.003)	-1.849*
Southern clade	279	0.646 (0.032)	0.003 (0.000)	0.016 (0.004)	-2.369**
L. ovalis (0.484***)					
Northern clade	51	0.722 (0.047)	0.003 (0.000)	0.004 (0.002)	-1.143^{NS}
Southern clade	136	0.628 (0.048)	0.002 (0.000)	0.012 (0.003)	-2.477**
L. elevatus (0.900***)					
Northern species	186	0.432 (0.036)	0.001 (0.000)	0.005 (0.002)	-1.599^{NS}
Southern species	253	0.590 (0.037)	0.003 (0.000)	0.016 (0.004)	-2.326**

585 *n*, number of sequences; SD, standard deviation; -, too small sample size to estimate indices;
586 ^{NS} $P > 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

Fig. 1 Median joining networks and haplotype frequency distributions of all sampled populations for the seven species. Size of haplotype circles and connections are proportional to number of sequences and mutation step, respectively. A and B represent the two divergent clades for each species. On the haplotype-frequency distributions, shared haplotypes with greater than 2% frequency within its corresponding clade were coloured. Private haplotypes and shared haplotypes with lower than 2% frequency within its clade were put in white. *Bt*, *B. thermophilus*; *Ap*, *A. pompejana*; *Hb*, *H. bergi*; *Bs*, *B. symmytilida*; *Ev*, *E. vitrea*; *Lo*, *L. ovalis*; *Le*, *L. elevatus*

595 **Fig. 2** Distribution of Φ_{st} values calculating for groups of three populations using a sliding
596 windows as a function of distance to 21°S. Each point represents Φ_{st} values relative to a
597 barycentric position of the three vent fields latitudes used in the sliding window. * $P < 0.05$.
598 *Bt*, *B. thermophilus*; *Ap*, *A. pompejana*; *Hb*, *H. bergi*; *Bs*, *B. symmytilida*; *Ev*, *E. vitrea*; *Lo*, *L.*
599 *ovalis*; *Le*, *L. elevatus*

Fig. 3 (A, B) Three-dimensional joint posterior probability densities for $E(\tau)$ and Ω. (C, D) Posterior probability densities for Ψ, the number of divergence times given Y clade pairs. Estimates in (A) and (C) are based on data from all seven species clade pairs, whereas estimates in panels (B) and (D) are based on a dataset in which *Lepetodrilus elevatus* was excluded. These estimates use the same uniform prior for θ_A bounded by 0.5 and 5.0 and are based on 500,000 simulated draws from the joint hyperprior and 1,000 draws from the joint posterior using MsBayes Approximate Bayesian Computational software (Hickerson *et al.* 2006b). In panels (C) and (D), the dotted line is the prior for Ψ and the solid line is the posterior for Ψ.

609 **Fig. 4** Graph of theta parameter (where $\theta = 2N_e\mu$, N_e = effective population size and μ = the
610 mutation rate) over number of generations based on Metropolis-Hastings Monte Carlo
611 coalescent analysis using Fluctuate version 1.4. Pattern of growth are based on estimates of g
612 (the exponential growth rate of the population) generated jointly with θ.
613 *Bt*, *B. thermophilus*; *Ap*, *A. pompejana*; *Hb*, *H. bergi*; *Bs*, *B. symmytilida*; *Ev*, *E. vitrea*; *Lo*, *L.*
614 *ovalis*; *Le*, *L. elevatus*
615

References

Akey JM, Eberle MA, Rieder MJ, *et al.* (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *Plos Biology* **2**, 1591-1599.

Arellano SM, Young CM (2009) Spawning, development and the duration of larval life in a deep-sea cold-seep mussel. *Biological Bulletin* **216**, 149-162.

Avise JC (1998) The history and purview of phylogeography: a personal reflexion. *Molecular Ecology* **7**, 371-379.

Bachraty C, Legendre P, Desbruyères D (2009) Biogeographic relationships among deep-sea hydrothermal vent faunas at global scale. *Deep Sea Research Part I: Oceanographic Research Papers* **doi:10.1016/j.dsr.2009.01.009**.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Bazin E, Glemin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570-572.

Beerli P, Felsenstein J (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* **152**, 763-773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4563-4568.

Bermingham E, Avise JC (1986) Molecular Zoogeography of Fresh-Water Fishes in the Southeastern United-States. *Genetics* **113**, 939-965.

Chevaldonné P, Jollivet D, Desbruyeres D, Lutz RA, Vrijenhoek RC (2002) Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial COI clock calibration. *Cahiers De Biologie Marine* **43**, 367-370.

Chevaldonné P, Jollivet D, Vangriesheim A, Desbruyeres D (1997) Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific .1. Influence of vent site distribution, bottom currents, and biological patterns. *Limnology and Oceanography* **42**, 67-80.

Cowen JP, Fornari DJ, Shank TM, *et al.* (2007) Volcanic Eruptions at East Pacific Rise near 9°50'N. *EOS, Transactions of the American Geophysical Union* **88**, 81-83.

Doyle JJ, Dickson E (1987) Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon* **36,** 715–722.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7 **doi:10.1186/1471-2148-7-214**.

Eakins B, Lonsdale P (2003) Structural patterns and tectonic history of the Bauer microplate, Eastern Tropical Pacific. *Marine Geophysical Researches* **24**, 171-205.

Emerson BC, Paradis E, Thebaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology & Evolution* **16**, 707-716.

Faure B, Chevaldonné P, Pradillon F, Thiébaut E, Jollivet D (2007) Spatial and temporal dynamics of reproduction and settlement in the Pompeii worm *Alvinella pompejana* (Polychaeta: Alvinellidae). *Marine Ecology Progress Series* **348**, 197-211.

Faure B, Jollivet D, Tanguy A, Bonhomme F, Bierne N (accepted) Speciation in the deep-sea: multilocus analysis of divergence and gene flow between two hybridizing species of hydrothermal vent mussels. *PlosOne*.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294-299.

Francheteau J, Armijo R, Cheminee JL, *et al.* (1990) 1 Ma East Pacific Rise Oceanic-Crust and Uppermost Mantle Exposed by Rifting in Hess Deep (Equatorial Pacific-Ocean). *Earth and Planetary Science Letters* **101**, 281-295.

Fretter V (1988) New archaeogastropod limpets from hydrothermal vents; superfamily *Lepetodrilacea*. II. Anatomy. *Philosophical Transactions of the Royal Society of London Series B* **318**, 33-82.

Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D (2003) Demography and natural selection have shaped genetic variation in Drosophila melanogaster: A multi-locus approach. *Genetics* **165**, 1269-1278.

Goldstien SJ, Schiel DR, Gemmell NJ (2006) Comparative phylogeography of coastal limpets across a marine disjunction in New Zealand. *Molecular Ecology* **15**, 3259-3268.

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

Harpending HC, Batzer MA, Gurven M, *et al.* (1998) Genetic traces of ancient demography. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1961-1967.

Haymon RM, Fornari DJ, Edwards MH, *et al.* (1991) Hydrothermal Vent Distribution Along the East Pacific Rise Crest (9-Degrees-09'-54'n) and Its Relationship to Magmatic and Tectonic Processes on Fast-Spreading Midocean Ridges. *Earth and Planetary Science Letters* **104**, 513-534.

Haymon RM, Fornari DJ, Vondamm KL, *et al.* (1993) Volcanic-Eruption of the Midocean Ridge Along the East Pacific Rise Crest at 9-Degrees-45-52'n - Direct Submersible Observations of Sea-Floor Phenomena Associated with an Eruption Event in April, 1991. *Earth and Planetary Science Letters* **119**, 85-101.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167**, 747-760.

Hickerson MJ, Dolman G, Moritz C (2006a) Comparative phylogeographic summary statistics for testing simultaneous vicariance. *Molecular Ecology* **15**, 209-223.

Hickerson MJ, Stahl EA, Lessios HA (2006b) Test for simultaneous divergence using approximate Bayesian computation. *Evolution* **60**, 2435-2453.

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of Levels of Gene Flow from DNA-Sequence Data. *Genetics* **132**, 583-589.

Hurtado L, Lutz R, Vrijenhoek R (2004) Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. *Molecular Ecology* **13**, 2603-2615.

Johnson SB, Young CR, Jones WJ, Waren A, Vrijenhoek RC (2006) Migration, isolation, and speciation of hydrothermal vent limpets (Gastropoda; Lepetodrilidae) across the Blanco transform fault *Biological Bulletin* **210**, 140-157.

Jollivet D (1996) Specific and genetic diversity at deep-sea hydrothermal vents: An overview. *Biodiversity and Conservation* **5**, 1619-1653.

Jollivet D, Chevaldonne P, Planque B (1999) Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution* **53**, 1128-1142.

Jollivet D, Empis A, Baker MC, *et al.* (2000) Reproductive biology, sexual dimorphism, and population structure of the deep sea hydrothermal vent scale-worm, *Branchipolynoe seepensis* (Polychaeta: Polynoidae). *Journal of Marine Biological Association of the United Kingdom* **80**, 55-68.

Jolly MT, Viard F, Gentil F, Thiebaut E, Jollivet D (2006) Comparative phylogeography of two coastal polychaete tubeworms in the Northeast Atlantic supports shared history and vicariant events. *Molecular Ecology* **15**, 1841-1855.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro HN), pp 21-123. Academic Press, New York.

Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 2257-2263.

Kuhner MK, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**, 429-434.

Kumar S, Balczarek KA, Lai Z-C (1996) Evolution of the *hedgehog* gene family. *Genetics* **142**, 965-972.

Kureth CL, Rea DK (1981) Large-Scale Oblique Features in an Active Transform-Fault, the Wilkes Fracture-Zone near 9-Degrees-S on the East Pacific Rise. *Marine Geophysical Researches* **5**, 119-137.

Le Pennec M, Hily A, Lucas A (1984) Structures gonadiques particulières d'un Mytilidae profound des sources hydrothermales du Pacifique oriental. *Comptes Rendus de l'Académie des Sciences de Paris, Série III* **299**, 725-730.

Lessa EP, Cook JA, Patton JL (2003) Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10331-10334.

Mammerickx J, Herron E, Dorman L (1980) Evidence for two fossil spreading ridges in the southeast Pacific. *Geological Society of America Bulletin* **91**, 263-271.

Mammerickx J, Klitgord KD (1982) Northern East Pacific Rise - Evolution from 25 My Bp to the Present. *Journal of Geophysical Research* **87**, 6751-6759.

Matabos M, Thiebaut E, Le Guen D*, et al.* (2008) Geographic clines and stepping-stone patterns detected along the EPR in the vetigastropod *Lepetodrilus elevatus* reflects species crypticism. *Marine Biology* **153**, 545-563.

Mullineaux LS, Mills SW, Sweetman AK, Beaudreau AH, Metaxas A, Hunt HL (2005) Vertical, lateral and temporal structure in larval distributions at hydrothermal vents. *Marine Ecology Progress Series* **293**, 1-16.

Mullineaux LS, Wiebe PH, Baker ET (1995) Larvae of Benthic Invertebrates in Hydrothermal Vent Plumes over Juan-De-Fuca Ridge. *Marine Biology* **122**, 585-596.

Naar DF, Hey RN (1989) Speed Limit for Oceanic Transform Faults. *Geology* **17**, 420-422.

O'Mullan GD, Maas PAY, Lutz RA, Vrijenhoek RC (2001) A hybrid zone between hydrothermal vent mussels (Bivalvia : Mytilidae) from the Mid-Atlantic Ridge. *Molecular Ecology* **10**, 2819-2831.

Plouviez S, Daguin C, Hourdez S, Jollivet D (2008) Juvenile and adult scale-worms, *Branchipolynoe seepensis*, in Lucky Strike hydrothermal vent mussels are genetically unrelated. *Aquatic Biology* **3**, 79-87.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution* **16**, 37-45.

Rogers AR, Harpending H (1992) Population-Growth Makes Waves in the Distribution of Pairwise Genetic-Differences. *Molecular Biology and Evolution* **9**, 552-569.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

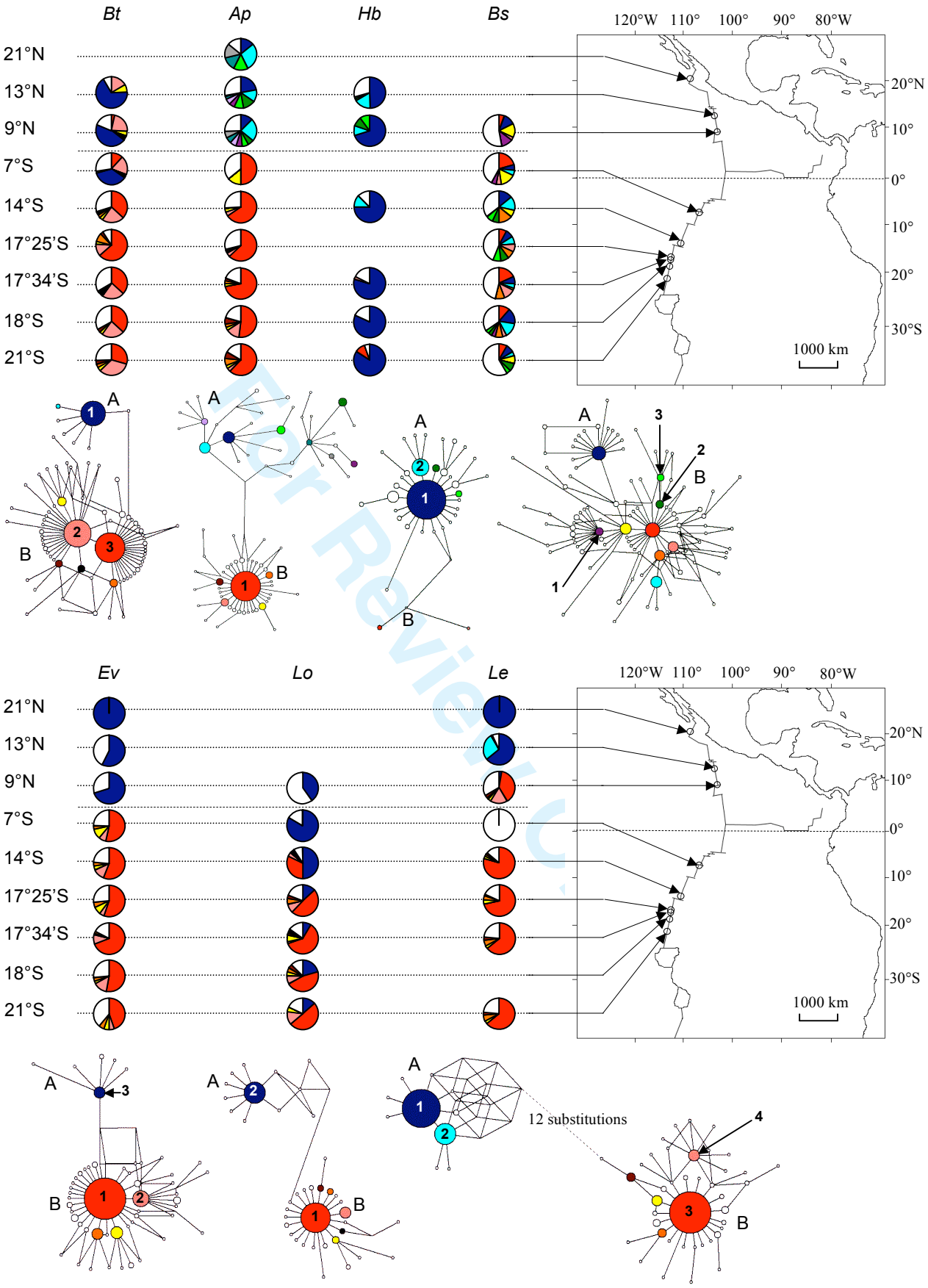
Sadosky F, Thiebaut E, Jollivet D, Shillito B (2002) Recruitment and population structure of the vetigastropod Lepetodrilus elevatus at 13 degrees N hydrothermal vent sites on East Pacific Rise. *Cahiers de Biologie Marine* **43**, 399-402.

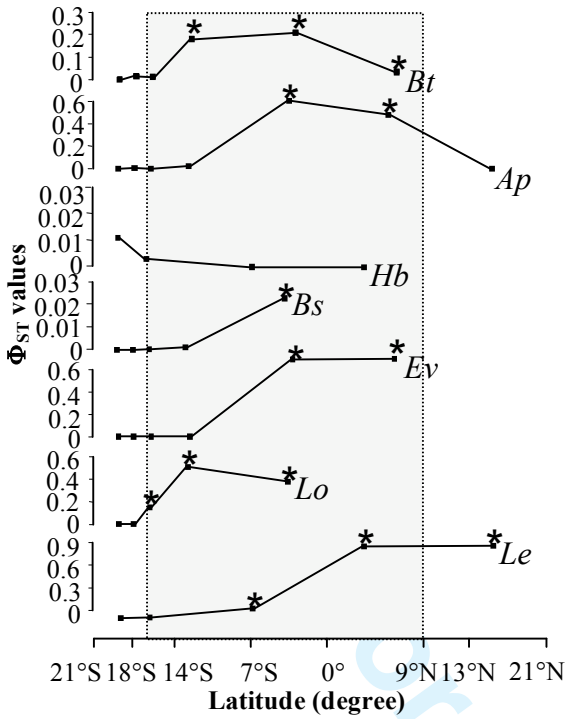
Shank TM, Fornari DJ, Von Damm KL *et al.* (1998) Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50'N, East Pacific Rise). *Deep Sea Research Part II: Topical Studies in Oceanography* **45**, 465-515.

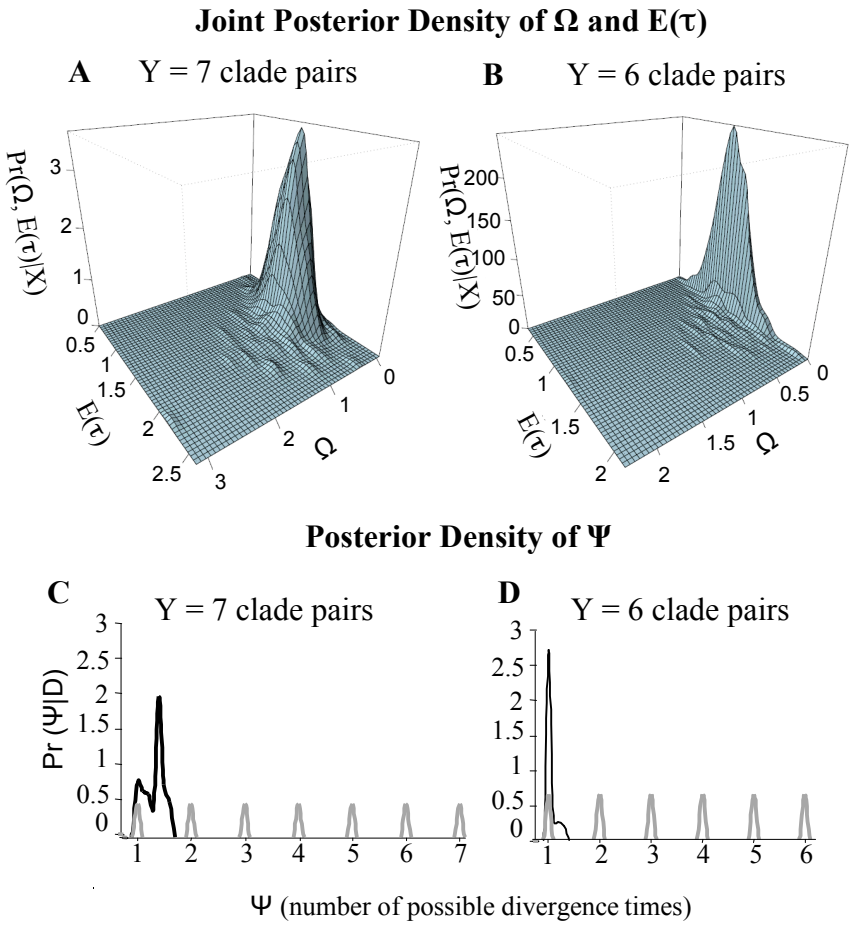
766 Skibinski DOF, Gallagher C, Beynon CM (1994) Mitochondrial-DNA Inheritance. *Nature*
767 **368**, 817-818.
768 Smith CI, Farrell BD (2005) Range expansions in the flightless longhorn cactus beetles,
769 Moneilema gigas and Moneilema armatum, in response to Pleistocene climate
770 changes. *Molecular Ecology* **14**, 1025-1044.
771 Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology &*
772 *Evolution* **15**, 199-203.
773 Tajima F (1989) Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA
774 Polymorphism. *Genetics* **123**, 585-595.
775 Tunnicliffe V (1991) The biology of hydrothermal vents: ecology and evolution.
776 *Oceanography and Marine Biology. An annual review* **29**, 319-407.
777 Tunnicliffe V, Embley RW, Holden JF*, et al.* (1997) Biological colonization of new
778 hydrothermal vents following an eruption on Juan de Fuca Ridge. *Deep-Sea Research*
779 *Part I-Oceanographic Research Papers* **44**, 1627-&.
780 Tyler PA, Pendlebury S, Mills SW*, et al.* (2008) Reproduction of gastropods from vents on
781 the East Pacific Rise and the Mid-Atlantic Ridge. *Journal of Shellfish Research* **27**,
782 107-118.
783 Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC (2002) Evolution and
784 biogeography of deep-sea vent and seep invertebrates. *Science* **295**, 1253-1257.
785 Van Dover CL, Trask JL, Gross J, Knowlton A (1999) Reproductive biology of free-living
786 and commensal polynoid polychaetes at the Lucky Strike hydrothermal vent field
787 (Mid-Atlantic Ridge). *Marine Ecology-Progress Series* **181**, 201-214.
788 Young CR, Fujio S, Vrijenhoek RC (2008) Directional dispersal between mid-ocean ridges:
789 deep-ocean circulation and gene flow in *Ridgeia piscesae*. *Molecular Ecology* **17**,
790 1718-1731.

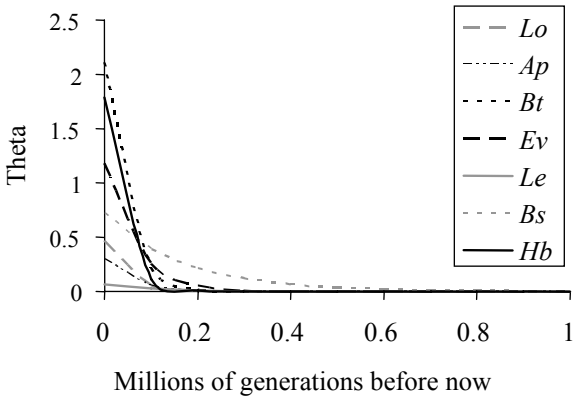
Author information box

This study is a component of S. Plouviez's Ph.D. project, which investigates comparative phylogeography of deep-sea hydrothermal vent species. Her research uses multiple species and markers to examine demographic processes and species dispersion along the East Pacific Rise. The Ph.D. is co-supervised by D. Jollivet and F.H. Lallier. T.M. Shank, B. Faure, C. Daguin and F. Viard made significant contributions to this study through the design of sampling strategies, field collections, and editing that improved the manuscript, initially written by the first author.






Joint Posterior Density of Ω and E(τ)
A
Y = 7 clade pairs
B
Y = 6 clade pairs
Pr(Ω, E(τ)|X)
E(τ)
Ω
Posterior Density of Ψ
C
Y = 7 clade pairs
D
Y = 6 clade pairs
Pr (Ψ|D)
Ψ (number of possible divergence times)


2.5
2
1.5
1
0.5
0
Theta
0
0.2
0.4
0.6
0.8
1
Millions of generations before now
Lo
Ap
Bt
Ev
Le
Bs
Hb