*Syst. Biol.* 61(1):127–137, 2012

DOI:10.1093/sysbio/syr088
Advance Access publication on August 18, 2011

# Biogeography Revisited with Network Theory: Retracing the History of Hydrothermal Vent Communities

YANN MOALIC[1], DANIEL DESBRUYÈRES[1], CARLOS M. DUARTE[2,3],
ALEJANDRO F. ROZENFELD[2], CHARLEYNE BACHRATY[4], AND SOPHIE ARNAUD-HAOND[1,*]

[1]*Département Etude des écosystèmes profonds (DEEP), IFREMER, Institut Français de Recherche pour l'Exploitation de la Mer, centre de Brest, BP70, 29280 Plouzané, France; E-mail: Yann.Moalic@ifremer.fr; Daniel.Desbruyeres@ifremer.fr;* [2]*Department of Global Change Research, IMEDEA, IMEDEA (CSIC-UIB), Instituto Mediterraneo de Estudios Avanzados, 07190 Esporles, Mallorca, Spain; E-mail: carlosduarte@ifisc.uib.es; alex@ifisc.uib.es;* [3]*The UWA Oceans Institute, University of Western Australia, 35 Stirling Highway, Crawley 6009, Western Australia, Australia; and* [4]*Département de sciences biologiques, Université de Montréal, C.P. 6128, succursale Centre Ville, Montreal, Québec, Canada H3C 3J7; E-mail : charleyne.bachraty@umontreal.ca;*
[*]*Correspondence to be sent to: Département Etude des écosystèmes profonds (DEEP), IFREMER, Institut Français de Recherche pour l'Exploitation de la Mer, centre de Brest, BP70,29280 Plouzané, France;*
*E-mail: Sophie.Arnaud@ifremer.fr.*

*Received 18 August 2010; reviews returned 8 October 2010; accepted 14 June 2011*
*Associate Editor: Adrian Paterson*

*Abstract*.—Defining biogeographic provinces to understand the history and evolution of communities associated with a given kind of ecosystem is challenging and usually requires a priori assumptions to be made. We applied network theory, a holistic and exploratory method, to the most complete database of faunal distribution available on oceanic hydrothermal vents, environments which support fragmented and unstable ecosystems, to infer the processes driving their worldwide biogeography. Besides the identification of robust provinces, the network topology allowed us to identify preferential pathways that had hitherto been overlooked. These pathways are consistent with the previously proposed hypothesis of a role of plate tectonics in the biogeographical history of hydrothermal vent communities. A possible ancestral position of the Western Pacific is also suggested for the first time. Finally, this work provides an innovative example of the potential of network tools to unravel the biogeographic history of faunal assemblages and to supply comprehensive information for the conservation and management of biodiversity. [Biogeography; deep sea ecology; hydrothermal vents; network analysis; systems biology.]

Biogeographic provinces are areas of animal and plant distribution that have similar or shared characteristics throughout. Distinct provinces support different biomes resulting from divergent ecological and/or evolutionary processes. The identification of biogeographic provinces and their delimitation are important goals within biogeography but such work is seldom straightforward. This is illustrated by the diversity of approaches that have been proposed for this task (Nelson and Platnick 1981; Hausdorf 2002; Hausdorf and Hennig 2003), which often involve a somewhat arbitrary delimitation of boundaries (Hausdorf 2002) such as the ones based on the geographical distribution of sites. Approaches are needed that limit the a priori hypotheses about the predominant factors that shape provinces and therefore deliver more objective delineations of biogeographic provinces and the pathways of connectivity between them. Network analysis appears to be a promising tool, as it can help us to understand geographical landscape connectivity (Urban and Keitt 2001; McRae and Beier 2007; Dos Santos et al. 2008) and genetic relationships between populations or individuals (Rozenfeld et al. 2007, 2008). Networks provide representations of complex data sets where nodes (e.g., communities) are interconnected by links that are scaled to the connectivity between them, as indicated by metrics of distances or the degree of interaction between nodes, all of which depend on the system under analysis. The topology of the ensuing network can then be analyzed with a range of tools and models developed in network theory, allowing inferences on the past or present dynamic properties of the system. Network analysis provides a holistic approach, free of a priori data set assumptions, which is better than classical pairwise interaction analyses for assessing interactions within complex data sets (Proulx et al. 2005). Therefore, network analysis has been shown to be a powerful tool for understanding the behavior of biological systems composed of interacting units, with nodes representing units from genes to communities, and links illustrating their interactions. Analysis of biological systems as complex networks has allowed, for example, an understanding of species specialization in mutualistic interactions (Bascompte and Jordano 2007), species interactions in a food web (Krause et al. 2003), and the identification of specific molecules with a crucial role in the stability of specific metabolic pathways (Albert et al. 2000; Strogatz 2001; Sales-Pardo et al. 2007). At the intraspecific scale, network analysis has allowed the elucidation of genetic relationships and structure among individuals or populations and provided important insights into dispersal migration (Rozenfeld et al. 2007; Fortuna et al. 2009) and source–sink dynamics (Rozenfeld et al. 2008). At the community level, network analysis has allowed the identification of differentiated communities and biogeographic entities along a continuously distributed space, based on patterns of sympatry (Dos Santos et al. 2008).

Here, we use network analysis to resolve the biogeography of hydrothermal vent sites discretely scattered along oceanic ridges, using the faunal composition of

63 such fields (Desbruyères, Segonzac, et al. 2006). Hydrothermal vents occur across all oceans due to volcanic activity at active deep-sea ridges, back-arc spreading centers, volcanic arcs and fore-arcs, and some intraplate seamounts. Widely spaced and often ephemeral, these ecosystems develop in the darkness of the oceans. They depend almost entirely on prokaryote chemosynthesis, which allows the proliferation of a specialized fauna exhibiting limited biodiversity but extraordinarily high biomass (Van Dover 2000). Subsequent to the initial discovery of vent ecosystems in 1977 (Corliss et al. 1979), research has strongly focused on the origins and processes driving the distribution patterns and dynamics of the impressive communities they support. The fauna present at hydrothermal sites appears to be strongly endemic, with only 8% of species previously found outside the vent habitat (McArthur and Tunnicliffe 1998). First discovered on the East Pacific Rise (EPR) (Hekinian et al. 1983; Spiess et al. 1980), hydrothermal vent fields have now been found over all of the world's ocean ridges. Other chemosynthetically based ecosystems, such as cold seeps and whale carcasses, have also been found to share common taxa with hydrothermal vents. Increased knowledge of hydrothermal vent distribution across the ocean floor has prompted interest about the origins, evolutionary history, and dynamics of these very unusual ecosystems, which rely on the chemosynthetic production of organic matter.

Due to high species endemism, it has been rather difficult to define comprehensive clusters of biogeographic provinces that could help to reconstruct the evolutionary history of hydrothermal vent ecosystems. Indeed, the number of distinct biogeographic provinces recognized ranges between four and seven, depending on the number of vent fields explored and the methods and the underlying hypotheses used to delineate them (Tunnicliffe 1997; Van Dover et al. 2002; Desbruyères, Hashimoto, et al. 2006). Most approaches relied on a priori hypotheses about faunal composition, geographical distance, ridge history, topography, and vicariant events, or on an arbitrary defined cutting distance in UPGMA (Unweighted Pair Group Method with Arithmetic Mean) trees, to cluster hydrothermal sites into provinces. For example, in different studies, the EPR has been considered to contain one (Van Dover et al. 2002), two (Bachraty et al. 2009), or three (Tunnicliffe 1997) biogeographic provinces, depending on the underlying hypothesis and information. The similarities in hydrothermal vent assemblages are particularly important because, in addition to helping delineate biogeographical provinces, they can also help elucidate the links between different biogeographical provinces, which may contribute to understanding their history and colonization pathways, and the dispersal patterns of the fauna they contain. The extreme level of endemism across vent sites can be attributed to 1) the partial colonization of limited ridge segments by some species and 2) the divergence of species over geological time. The occurrence of a large number of common genera, represented by different species across vent sites, supports the idea of vicariance as one of the main drivers of the spatial heterogeneity observed in community composition (Tunnicliffe et al. 1998), resulting from species divergence on both sides of geographical barriers such as major ridge discontinuities or transforming faults (Plouviez et al. 2009). High endemism, combined with the extremely fragmented distribution of vents along mid-oceanic ridges, means that the network analysis method proposed by (Dos Santos et al. 2008), is not possible based on species; nor was it necessary, as the fragmented distribution of sites offers a natural delineation of geographic units (i.e., hydrothermal fields).

An alternative approach to examine the relationships between the hydrothermal vent fauna can be taken, using network analysis to represent each of the 63 hydrothermal vent fields, investigated to date, as nodes can be connected among them with a distance reflecting the common genera among sites (links). These networks of hydrothermal vent communities can then be formally analyzed to explore their structures; identifying groups of closely related locations that are helpful in delineating biogeographic provinces and the hydrothermal vent regions connecting these provinces. Moreover, such networks can be constructed without entering any information concerning the geological or geographic properties of the sites, thereby allowing subsequent examination of the consistency between the network structure and the independent geological or geographic properties as a validation exercise. In particular, we aimed at testing 1) the various delineation of biogeographic provinces hitherto proposed by different authors and 2) the hypothesis of an ancient history of colonization and evolution of hydrothermal communities by comparing the pattern of clustering and connectivity revealed by network analysis with the spreading of oceanic ridges across geological time (Tunnicliffe 1997; Van Dover et al. 2002; Desbruyères, Hashimoto, et al. 2006).

## MATERIALS AND METHODS

### *Data Set Used in This Study*

Hydrothermal vent fields ($n = 63$) were selected across the world's oceans according to available information on their faunal composition. A taxonomic database from a recent compilation (Desbruyères, Segonzac, et al. 2006) was assembled from information from the literature, databases on the World Wide Web (ChEss: http://www.noc.soton.ac.uk/chess/database/, NOAA: http://www.pmel.noaa.gov/, NeMO: http://www.pmel.noaa.gov/vents/nemo/), and researchers we contacted directly. At each of the 63 hydrothermal vent fields, the presence or absence of taxa was noted among the total 591 species and 331 genera described. All the taxa recorded in the biogeographic provinces found by network analysis are described in Table S1 (available from http://www.sysbio.oxfordjournals.org).

### *Networks*

We began analysis with a fully connected network of 63 hydrothermal vent fields considered as nodes. Each link joining a pair of fields was assigned the corresponding Jaccard's index distance. This assemblage distance is given by:

$$\text{Jaccard's index distance} = 1 - \left(\frac{a}{a+b+c}\right),$$

where $a$ is the number of taxa (i.e., genera or species) common to two fields (e.g., field 1 and field 2), $b$ is the total number of taxa present in field 1 but absent from field 2, and $c$ is the total number of taxa present in field 2 but absent from field 1.

This coefficient ranges from 0 (both fields are identical) to 1 (the fields share no items) and gives similar weight to the presence or absence of a taxon. This property makes the Jaccard's index distance more appropriate for the present case because the available description of these remote communities is far from exhaustive, and some absences may also be due to the lack of observation of a given taxon, rather than to true absences from the community. Nevertheless, when tests with Sorensen index, which gives double weighting to double presence, were realized on the same data set, results yielded a similar network topology, with identical clusters. Consequently, we kept the simplest index and focused on analyses at different taxonomic levels. Therefore, we analyzed and compared networks of Jaccard's index at both the species and genus levels. Like previous authors (Tunnicliffe et al. 1996), we consider that genus-level analysis provides a more appropriate picture of the patterns of dispersal and/or vicariance between hydrothermal fields, based on the high levels of endemism observed at the species level. Jaccard distance interpretation in this context is that hydrothermal vent areas once connected by strongest dispersal paths would have exchanged more species, resulting in a lower Jaccard distance at both species and genus levels among contemporary communities.

The similarity matrix provides a fully connected network that does not allow recognition of biogeographic provinces. Various approaches have been developed to extract biogeographic provinces from networks, including those based on community detection (Krause et al. 2003; Newman 2004; Allesina and Pascual 2009) and percolation theory (Rozenfeld et al. 2008). A percolation approach was used to break down the fully connected network into discrete clusters of biogeographic provinces because these are better able to handle continuous link weights (Jaccard's index distance). Moreover, Rozenfeld et al. (2008) successfully applied the percolation distance approach to resolve gene flow within species, based on a fully connected network of populations. We transposed this method to the community level to resolve the biogeographic provinces emerging from historical disruption of gene flow through ancient vicariant events.

The percolation method is based on the analysis of a network built only with links illustrating the minimum Jaccard distance necessary to maintain the connexion across most components of the system. At this distance precisely, the network still depicts a single giant cluster composed of subclusters linked through primary connections. Below this distance, the network collapses into disconnected subclusters and the pattern of global connectivity is lost. This critical threshold distance is also named percolation distance (Dp) (Stauffer and Aharony 1994). For a finite system, this point is derived by calculating the average cluster size of all clusters excluding the largest,

$$\langle S\rangle * = \frac{1}{N}\sum_{s\langle S\max} s^2 n_s,$$

as a function of the last threshold distance value beyond which links were removed. $N$ is the total number of nodes not included in the largest cluster and $n_s$ is the number of clusters containing $s$ nodes. $S$max is the size of the largest cluster. The Dp is then identified heuristically in the transitional region characterized by a strong decrease in $<S>*$. Inside this transitional stage, the single giant cluster network is found at Dp $= 0.84$. The network topology and its characteristics were analyzed at this percolation distance, meaning that the links retained in the network analyzed are the ones corresponding to ecological distances lower or equal to Dp, whereas all links beyond that value were discarded. Additionally, the network was also explored at different distance thresholds around this percolation point, in order to assess the consistency of its topology and the inferred properties and interpretation. The network characteristics (see Appendix) were quite robust to variability with the applied threshold distance Dp: major clusters, nodes, and paths remained unchanged indicating in the robustness of these key properties to uncertainty in the percolation threshold and, therefore, supporting the reliability of the network for displaying the biogeographic structure of the communities.

In order to assess the respective influence of the three main phyla that encompass more than 80% of the genera (Annelida, Mollusca, and Arthropoda, which represent 20.5%, 25%, and 37%, respectively) on the global network topology, three subnetworks were then built based on genera within these three major phyla.

*Estimation of the global and local properties of the network.*— The *connectivity degree*, $k_i$ of a given node $i$, is the number of other nodes linked to it (i.e., the number of neighbor nodes). We named $E_i$ as the number of links existing among the neighbors of node $i$. This quantity takes values between 0 and $E_i^{(\max)} = k_i(k_i-1)/2$, which is the case in a fully connected neighborhood. This value is used to calculate the *clustering coefficient* $C_i$ of node $i$, defined as:

$$C_i = \frac{E_i}{E_i^{(\max)}} = \frac{2E_i}{k_i(k_i-1)},$$

$C_i$ quantifies how close the node $i$ and its neighbors are to being a clique (See Fig. S1a for an illustrative diagram).

The *clustering coefficient* (Watts and Strogatz 1998) of the whole network <CC> is defined as the average of all the individual clustering coefficients in the system. $C_i$ values vary between 0 and 1. The *clustering coefficient* helps us to understand how nodes are organized into clusters within the system as a whole. In order to test the significance of <CC>, its value was compared with the average value $<CC_o>$ of 10,000 random simulations of our data set. To perform this test, we randomly rewired our network with the same number of links present at a given threshold (here percolation), thereby yielding random networks with the same overall structure of taxa presence and absence.

The *betweenness centrality* (Freeman 1977) of node $i$, bc($i$), counts the fraction of shortest paths between pairs of nodes that pass through node $i$. Let $\sigma_{st}$ denote the number of shortest paths connecting nodes $s$ and $t$, and $\sigma_{st}(i)$ the number of those passing through the node $i$; then,

$$\mathrm{bc}(i)=\sum_{s\neq t\neq i}\frac{\sigma_{st}\,(i)}{\sigma_{st}}.$$

The *betweenness centrality* determines the relative importance of a node within the network as an intermediary in the flow of information and its vulnerability to fragmentation (See Fig. S1b for an illustrative schema).

Networks were visualized and analyzed using Pajek software (Batagelj and Mrvar 2002). Random simulations were compiled by C++ scripts.

*Comparative analysis.*—A set of four methods based on community detection were additionally tested on this data set to infer their potential for biogeographic analysis on our data set (Girvan and Newman 2002; Blondel et al. 2008; Dongen 2008; See Table S2 for details on these methods.

## RESULTS

### *Delineation of Biogeographic Provinces*

In order to reveal the clustering and interconnectivity of hydrothermal vents, we built networks with data from the 63 fields (nodes) referenced in the database (See Table S3), connected by links (edges) on the basis of their faunal distance in terms of genera, as assessed with Jaccard's index in an adjacency matrix (Legendre and Legendre 1998). Genera were chosen, rather than species, because of the extremely high level of endemism at the species level (95.3%, compared with 76.2% for genera; see Fig. S2), indicating that biogeographic history is ancient and likely to be better reflected by the distribution of genera (Tunnicliffe and Fowler 1996).

The Jaccard's distances were, as expected, lower for neighboring vent sites, but the resulting network had a percolation value at a Jaccard distance of Dp = 0.84, below which distinct groups of vent sites emerged suggesting distinct biogeographic regions (Fig. 1). In particular, the network topology clearly shows the existence of five strong clusters of hydrothermal vents (Figs. 1 and 2, see Table 1 for the network properties), even though the data used contained no geographic information and the process did not involve any subjective dissection like those needed for a "classic" dichotomic UPGMA tree (See Fig. S3). The ensuing clusters of hydrothermal vents correspond to five well-defined regional provinces, coherent both at the genus (Fig. 2) and species (Figs. S4 and S5) levels, in agreement with Tunnicliffe et al. (1998). At the species level, the network topology looses its integrity at a Dp value higher than the one of the genus level (0.95 vs. 0.84). This is due to the high species endemism; the much lower number of shared species than genera among fields results in a higher threshold below which the connectivity across the entire network breaks down (Fig. S5).

Analyzes performed at the genus level showed that this hierarchical structure is supported by a significantly higher average clustering coefficient, <CC> = 0.52, than expected by chance (i.e., by randomly rewiring the links, $<CC_o> = 0.17$ with $\sigma_o = 0.01$, after 10,000 random simulations, thus illustrating the modularity of the network composed of clusters of hydrothermal fields, with greater internal interconnection than would be expected by chance. Moreover, the recognition of the West Pacific (WP), Northeast Pacific (NP), EPR, and Mid-Atlantic Ridge (MAR) as provinces is confirmed by their individual <CC> values, which are higher than the average (Table 1). This highlights the strong endemism of hydrothermal vent fauna among the provinces defined by the network clusters (Fig. 2).

### *Connectivity among Provinces*

In addition to an objective identification of biogeographic provinces and subprovinces, network analysis offers another fundamental advantage: The study of biogeographic patterns in hydrothermal fields as the patterns of connectivity among fields and provinces can suggest past and present dynamics of information flow reflecting species divergence. Putative pathways can be assessed from the connections among biogeographic provinces in the network topology, identifying regions that possibly play or have played a role in the foundation and/or connectivity of the system. The first observation that emerges when these "intercluster" links are considered is the central position of the WP province. WP is intermediate among all other provinces, which themselves share no links (Fig. 2). An accurate diagnostic for this analysis is the *betweenness centrality* (see Materials and Methods section), which estimates the relative importance of a node in relaying the flow of information through the system, corresponding here to past gene flow, which will result in common genera across fields represented in the network (Fig. 2 and Table S4). Here, the two strongest *betweenness centrality* values indeed correspond to fields in the WP region: the Mariana Trough and the Pacmanus system, which are present in 28% and 26%, respectively, of all shortest paths among fields. As noted above, the MAR province

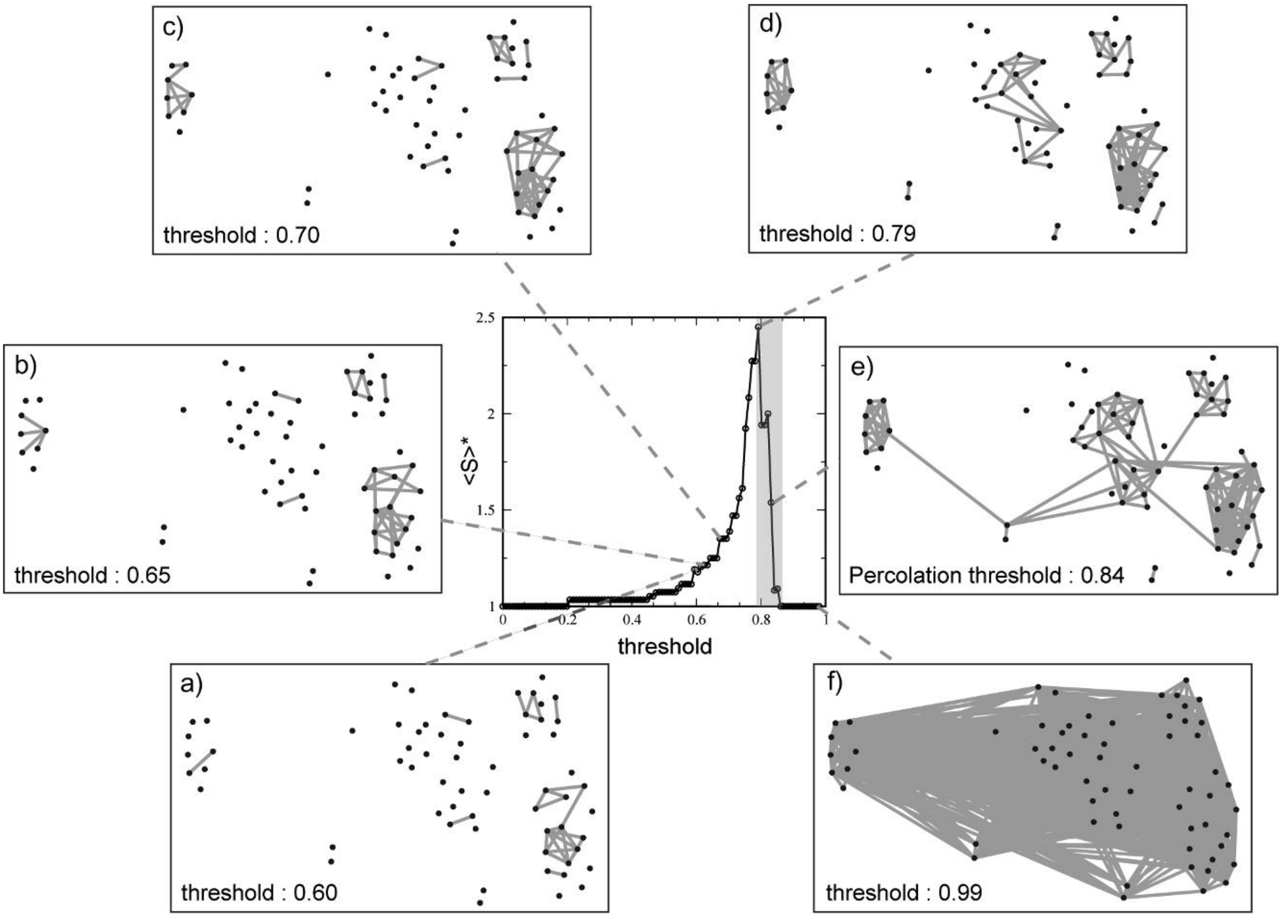

FIGURE 1. Evolution of network topology according to the average cluster size imposed by the Jaccard's distance at the genus level. Six threshold values were selected to visualize networks as a function of the average cluster size, excluding the largest one. The critical point is the Percolation Threshold. It is located in the transitional region (gray area). Percolation Threshold is reached just after all the clusters become connected. In our study, this critical value is 0.84. Each hydrothermal field is represented by a node, and the networks are distorted to accommodate the geographical distribution of hydrothermal sites. Link lengths are meaningless and only their presence/absence is finally taken into account in function of the threshold distance applied.

has no direct link with EPR, so that the Eastern Pacific and Atlantic Oceans are only connected in the network through the WP and Indian Ocean (IO).

The taxon-specific networks represented in Figure S6 show slightly lower percolation threshold than the global network (Dp = 0.84) revealing a higher overall proximity of communities of Annelida (Dp = 0.68), Arthropoda (Dp = 0.77), and Mollusca (Dp = 0.75) but generally support the number and identity of the biogeographic provinces revealed by the global network. These patterns are clearly similar for Annelida and Arthropoda (Fig. S6a,b) with the same five provinces distinguished. For Mollusca (Fig. S6c), WP, MAR, and EPR provinces still appear as a single cluster with NP divided into two clusters connected either to WP or to EPR and IO disconnected.

### *Comparative Analysis*

Among the methods tested based on community detection, only the method from Rosvall and Bergstrom (2008) showed resolution and cluster discrimination, whereas no coherent clusters emerged with the other methods (see Table S1). The analysis with the Rosvall and Bergstrom (2008) method shows a lower resolution than the percolation approach, although still supporting the discrimination of NP and EPR, and their sequential branching with and through the cluster including WP (merged with MAR and IO).

## DISCUSSION

The overall picture of hierarchical structure and connectivity among vent sites depicted by this network analysis allows the definition of five biogeographic provinces at the percolation threshold distance of the network, which is defined on the basis of the global analysis of taxonomic similarity among sites. Compared with previous studies based on other approaches that rely on a more "arbitrary" identification of clusters, some previously recognized distinct provinces appear to be compounded into a single one here, whereas others

TABLE 1. Network properties of identified provinces at the percolation threshold

| | MAR | IO | WP | NP | EPR |
|---|---|---|---|---|---|
| *n* | 8 | 2 | 26 | 10 | 17 |
| Cluster size | 7 | 2 | 17 | 8 | 16 |
| bc max/<bc> | 0.14/0.017* | 0.21/0.107*,[a] | 0.28/0.038*** | 0.16/0.023** | 0.13/0.021*** |
| <*k*> | 5.57 | 3.00 | 4.53 | 3.88 | 7.81 |
| <CC> | 0.92 | — | 0.68 | 0.88 | 0.81 |

Note: *n* is the number of fields localized inside the province, cluster size is the effective number of fields within the "province" cluster at the percolation threshold. For each province, <*k*> is the average connectivity degree, bc max is the maximum value of betweenness centrality, and <CC> is the average value of clustering coefficient.

[a] The mean value for IO is based on only two sampled fields and is therefore subjected to high variance.

*P* values: * <0.1, ** <0.05, *** <0.001.

are well differentiated. In fact, the splitting of EPR or MAR proposed in previous studies (Tunnicliffe 1997; Van Dover et al. 2002) is not supported here for MAR (Fig. 1), which corresponds, at best, to slightly differentiated subprovinces for EPR, disconnecting at a threshold far lower than that at which the other provinces are identified (just before the threshold of 0.60; Fig. 1a). The integrity of the IO province remains questionable due to its extremely weak representation in the data gathered so far, due to a lack of scientific cruises in this area. The two fields explored in IO do however remain strongly linked together, which, coupled with the robust intermediate position of IO between MAR and WP (Fig. 2), strongly supports the existence of at least one discrete province in this region.

The central position of WP in all connection paths between provinces was supported both by the global analysis at the genus and species (Fig. S4) levels, by the analysis specific to each of the three main phyla (Annelida, Mollusca, and Arthropoda). The network topology derived here calls into question the hypothesis of a major ancestral pathway between the EPR and MAR

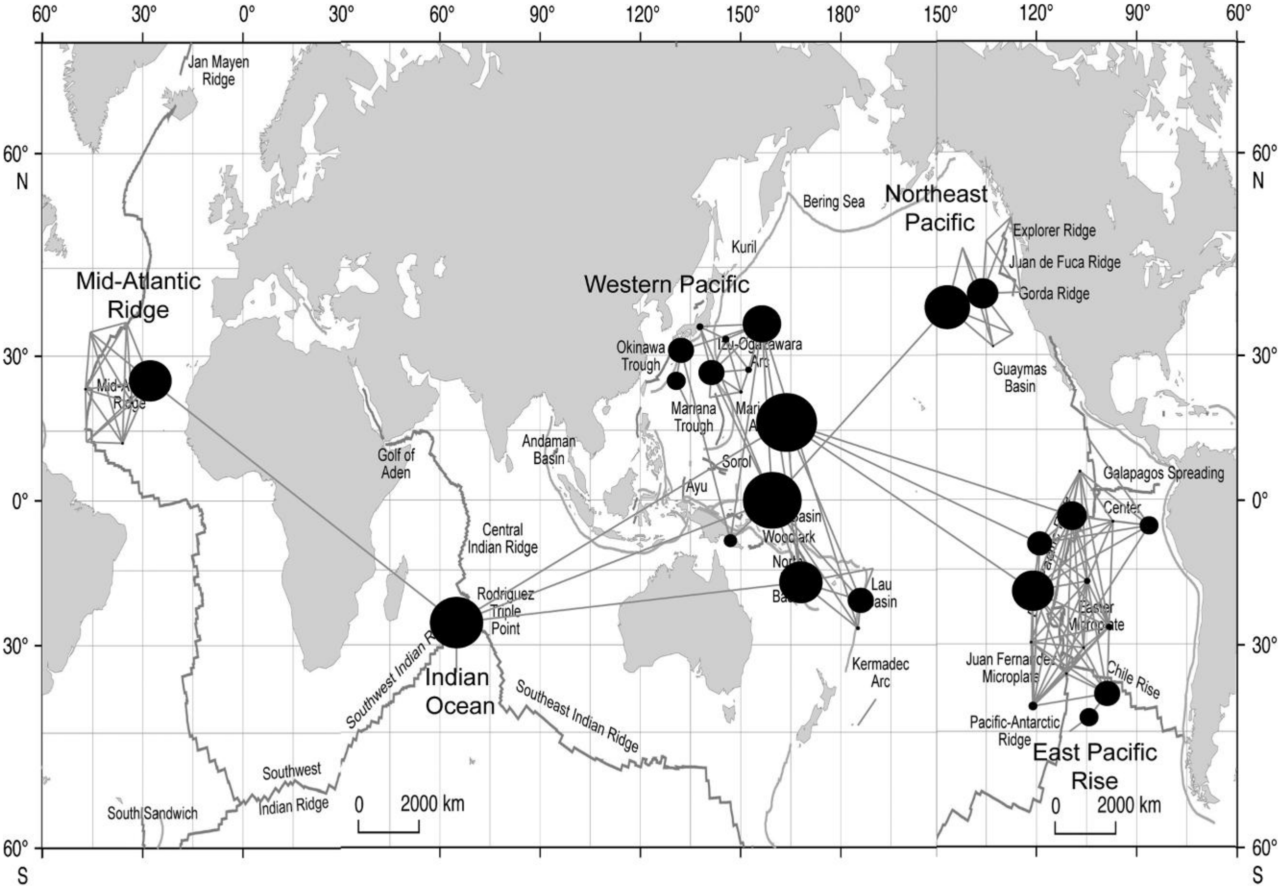

FIGURE 2. Global network of hydrothermal vent fauna diversity built on the basis of Jaccard's distance among fields and represented here at the percolation threshold (Dp = 0.84. See Materials and Methods section). Circle size represents the betweenness centrality values of the corresponding field. Five provinces are highlighted by this network analysis: MAR, IO, WP, NP, and EPR.

provinces through the Isthmus of Panama as well as the hypothesized connectivity between the Eastern Pacific and the Southern Atlantic through the Circumpolar Current (Van Dover et al. 2002).

### *Network Consistency with Tectonic History*

The star-like topology of the network (Fig. 2), characterized by a central position of WP (formally reflected in its high betweenness centrality score in the network), results in the lack of direct connection between EPR and NP at the percolation threshold, which suggests a more recent common history of WP with each of these provinces independently.

The topology derived here shows some similarity to the early history of the plate tectonic and spreading ridge system in the Pacific during the Mesozoic and its evolution over more than 100 million years (see Fig. 3). Although this consistency does not necessarily involve a causal relationship, it lends weight to the hypothesis that the current biogeographic structure of hydrothermal vent fauna may have been influenced by the tectonic dynamics of plates over millions of years (Hessler and Lonsdale 1991; Tunnicliffe et al. 1996). Indeed, the double network connection of EPR and NP exclusively through WP, which also appeared when using the community detection method of Rosvall and Bergstrom (2008), Table S2, is consistent with the history of the Pacific ridge since the early Cretaceous (∼150 Ma), and the split of the protoridges triple junction during the end of the Cretaceous (∼70 Ma) and Paleocene (∼60 Ma; Fig. 3) (Smith 2003). The branching of EPR on WP observed in the network (Fig. 2) is also consistent with the early Cretaceous position of the three ridges (Izanagi–Farallon, Pacific–Izanagi, and Pacific–Farallon), connecting the WP and the South Eastern Pacific (Fig. 3a), until their disconnection by a fracture zone in the Late Cretaceous (Fig. 3c). Finally, the network connection of NP to WP can also be considered in parallel with the early history of the Juan de Fuca Ridge, which appeared during the Paleocene and was connected to WP and the future EPR at the time when the ridge triple junction splits the Pacific, Kula, and Farallon plates.

It is obviously impossible to test this hypothesis experimentally, so the direct influence of the tectonic plate history on the present nature and distribution of hydrothermal communities, although consistent with the topology of the network built here, remains speculative. Nevertheless, it deserves to be examined in further studies, such as through the construction of complete phylogenies of some of the most widespread genera, with appropriate markers allowing the use of relaxed molecular clocks.

### *Hubs, Connectivity, and Centrality*

The central position of WP in the network as an obligatory relay between EPR and the other provinces (Fig. 2), contrasts with the hypothesis of a center of dispersion in EPR (Van Dover and Hessler 1990; Van Dover 2000; Bachraty et al. 2009). This may be due to the use of the Jaccard's distance as, in contrast to other distance metrics such as the dispersal direction coefficient (Bachraty et al. 2009), for example, which comes with the underlying hypothesis that richest provinces disperse more, Jaccard distance does not impose an underlying hypothesis linking richness to present or past connectivity. The EPR appears as the richest province in the database, both in terms of species and genera. Yet, rather than a high divergence rate and dispersal towards adjacent provinces, this richness may reflect the combination of different sampling strategies among provinces and the homogeneous topology of EPR.

The homogeneity and high connectivity in EPR is illustrated by its robustness in the network. As the threshold distance declines, the cohesion of each biogeographic province collapses in the sequence: WP, MAR, NP, EPR (Fig. 1), with EPR still strongly clustered long after the other provinces collapse (Fig. 1a). This homogeneity is illustrated by its much larger average *connectivity degree* (See Materials and Methods section; $k = 7.81$ for $n = 17$ vents observed; Table 1) that reveals the relatively low level of endemism within this province (Fig. S1). This may be the result of the specific regional properties of the EPR fields (Fig. 4) that exhibit a high accretion rate (rate of creation of new oceanic lithosphere) inducing a higher number of vents (Baker et al. 1995), a factor associated with a homogeneous depth is likely to facilitate the faunal dispersion. The relative homogeneity of EPR is also reflected by a rather low genetic structure and significant gene flow estimates for several taxa studied across EPR, the area where the most population genetic studies have been performed to date (Chevaldonné et al. 1997; Vrijenhoek 1997; Jollivet et al. 1999; Hurtado et al. 2004; Matabos et al. 2008); a result in line with the relative ubiquity of taxa in this province (Fig. S2). In contrast, the WP province shows a much higher level of endemism, as shown by a much lower average *connectivity degree* ($k = 4.53$), despite the greater number of referenced vent fields ($n = 26$). Moreover, the EPR has been the specific target of numerous cruises, allowing the examination of extinction and recolonization processes (Lutz et al. 1994), as well as the discovery and description of taxa of distinct succession stages. As a result, despite its apparently high richness in comparison with some previous studies (Van Dover 1995; Van Dover 2000; Bachraty et al. 2009), EPR does not emerge as an especially central cluster of nodes (province) in this network, which instead supports a central position for WP.

This central position of WP, both at the scale of the Pacific Ocean and worldwide, combined with its extremely high endemism, suggests a more ancient history and a central role in the biogeography of hydrothermal vents. WP is the region of the world where the basins exhibit the highest bathymetric variability of vents (Fig. 4), compared with mid-ocean ridges (Matabos et al. 2008), and where closely located vents and seeps coexist most frequently (Sibuet and Olu 2002). There

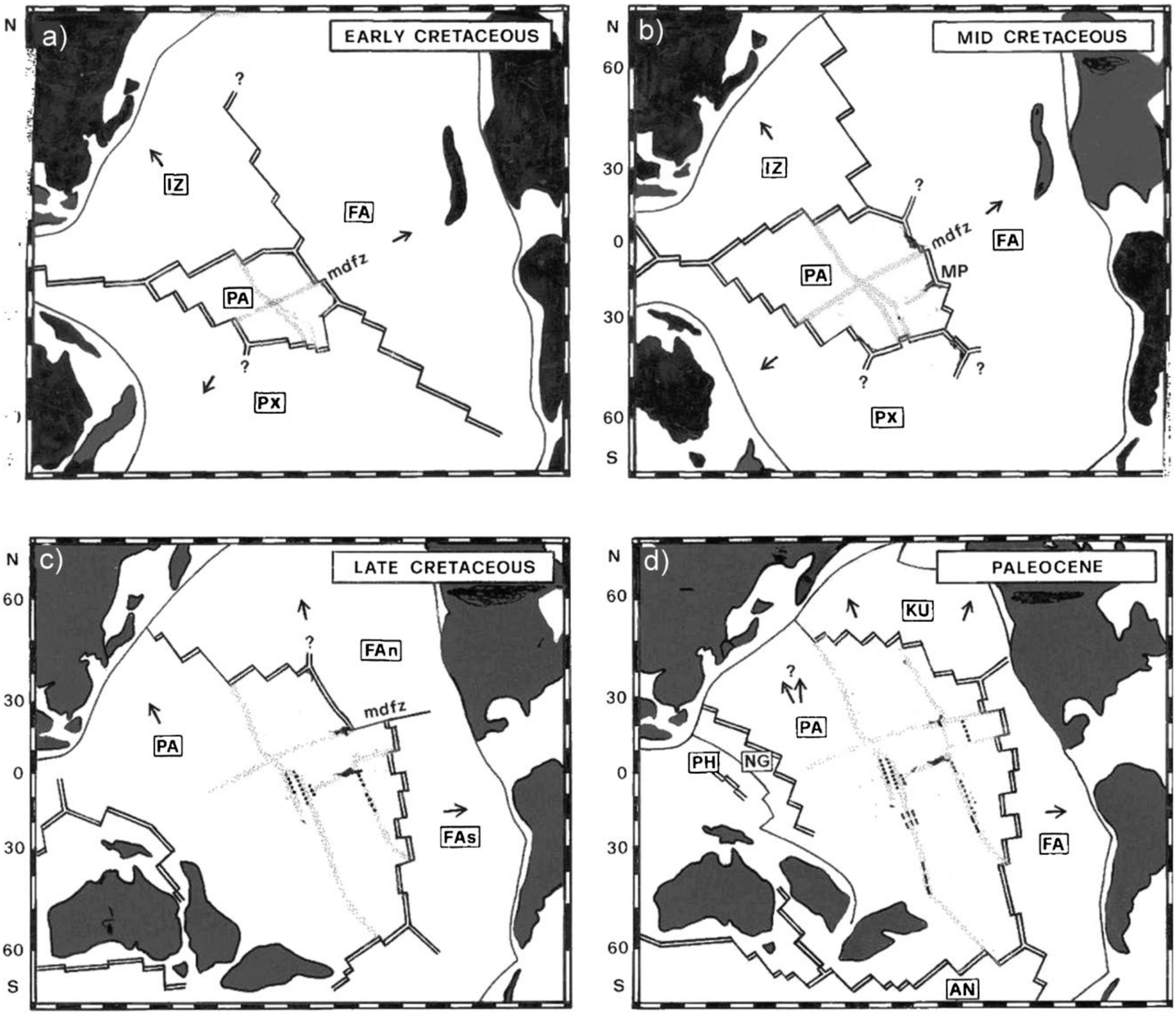

FIGURE 3. Evolution of the Pacific Basin from Early Cretaceous through Paleocene (modified from Smith (2003)) Abbreviations: AN = Antarctic; FA = Farallon (FAn = north, FAs = south); IZ = Izanagi; KU = Kula; NG = North New Guinea; PA = Pacific; PH = Philippine; PX = Phoenix. mdfz = Mendocino fracture zone. a) Early Cretaceous (140–130 Ma), b) Mid-Cretaceous (130–100 Ma), c) Late Cretaceous (approximately 83 Ma immediately prior to the formation of the Kula plate), d) Paleocene (approximately 60 Ma): The Kula plate has formed from the North Farallon plate.

is recent evidence suggesting that the colonization of deep chemosynthetic ecosystems may have followed a stepping stone model from the coastal area, through sunken wooden debris and large animal carcasses, to the cold seeps and finally the hydrothermal vents (Van Dover et al. 2002). Such findings agree with inferences drawn by Little and Vrijenhoek (2003) on fossil and molecular evidence. Observations supporting this stepping stone hypothesis were first reported on the Mytilidae (Distel et al. 2000; Samadi et al. 2007). Altogether, 1) the heterogeneity of WP field depth and their close proximity to seeps (active margins), fallen wood (tropical islands), and the coastal zone, 2) their pattern of within-region endemism and ancient divergence among taxa on this old slow-spreading ridge, and 3) their central position in the topology of the network reported here, support the hypothesis of a colonization or recolonization of hydrothermal vents through WP, followed by a subsequent worldwide dispersion that was possibly influenced by plate tectonics and the associated evolution of spreading ridge systems since the early Mesozoic (Desbruyères, Hashimoto, et al. 2006). This hypothesis could also be worthy of future testing using phylogenetic reconstructions.

### *Origin and Ageing of Communities*

The reconstruction of the aging and history of vent taxa to date has been based either on fossils or molecular clocks, which produced contradictory estimates (Little and Vrijenhoek 2003). The oldest fossils from the

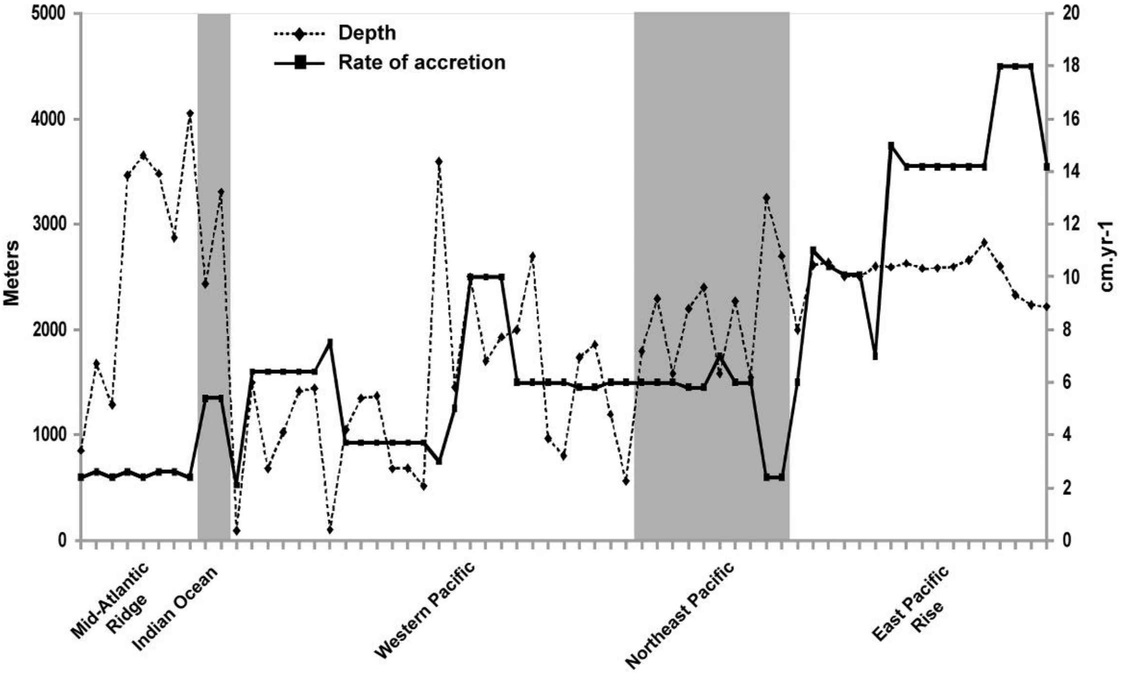

FIGURE 4. Codistribution of depth (dotted line) and rate of accretion (solid line) among the 63 fields. Depth scale is represented on the left $y$-axis and rate of accretion on the right $y$-axis. The field order is similar to the one in Table S3 and corresponds to the successive biogeographic provinces, separated by contrasted background on the graph.

Paleozoic (~450 Ma) and the high endemism of hydrothermal communities led to an "antiquity hypothesis" (Newman 1985). Molecular data, in contrast, supported a much more recent history, even when considering a slow molecular clock calibrated on the Farallon–Pacific Ridge disruption under the North American Plate (Chevaldonné et al. 2002). Reconsidering the evidence from both molecular and fossil data, Little and Vrijenhoek (2003) suggested that present hydrothermal fauna were of intermediate age and may have originated in the Mesozoic (−150 Ma). The network topology described here lends support to their suggestions, showing consistency between the evolution of oceanic ridges since the early Cretaceous and the topology of the network that illustrates the nature and connections among biogeographic provinces. It is also in agreement with a role of shallow hydrothermal vents and close cold seeps in the (re)colonization of hydrothermal vents. In line with the geographic location of most fossil records (Little 2002), the network suggests a high centrality of the WP region where (re)colonization events may have mainly taken place. The hypothesized central position of WP and its possible role in the original history of present day hydrothermal vent fauna derived from network analysis also require further investigation, which could be done through comprehensive phylogenies of different genera, to test for whether fauna from WP have the basal position that would be expected from such an ancestral role.

The application of network analysis to examine hydrothermal vent biogeography reported here provides evidence of the potential of this approach to address complex biogeography problems. First, network analysis allows the discrimination of clusters of communities, or biogeographic provinces and is genuinely obtained without introducing additional information, such as a priori assumptions on their clustering based on their geographic location. Second, this holistic approach allows illustrating the connectivity between provinces with a rather flexible geometry, free of the constraint of the dichotomous relationship imposed by branching rules in classical trees. Third, network analysis provides tools developed to derive diagnostics on the dynamics properties of the network, such as *betweenness centrality*, which allows particular regions or provinces to be pointed out that may have had a central role in the evolution and dispersion of fauna and, therefore, be of major relevance for biogeographical analyses. The application of network analysis, already used across a wide range of fields from social science to ecology, and more recently in population genetics (Rozenfeld et al. 2008; Fortuna et al. 2009), therefore opens a promising avenue for the investigation of biogeographic questions. This work takes a first step that we hope may encourage the use of network analysis, not only with the method presented here but also others, such as community detection approaches (Fortunato 2010), among the rapidly broadening and improving range of network analysis methods.

## SUPPLEMENTARY MATERIAL

Supplementary material, including data files and/or online-only appendices, can be found at http://www.sysbio.oxfordjournals.org/.

## FUNDING

This work was supported by the EU-funded HERMIONE (“Hotspot ecosystem research and man’s impact on European seas”) project, part of the Environment Theme of the Seventh Framework Programme (FP7); and the Project Deep Oases from the Agence nationale de la recherche (ANR06 BDV005).

## ACKNOWLEDGMENTS

We particularly thank Verena Tunnicliffe for helpful suggestions on a preliminary version of this work and Helen McCombie for English editing.

## REFERENCES

Albert R., Jeong H., Barabasi A.L. 2000. Error and attack tolerance of complex networks. Nature. 406:378–382.

Allesina S., Pascual M. 2009. Food web models: a plea for groups. Ecol. Lett. 12:652–662.

Bachraty C., Legendre P., Desbruyères D. 2009. Biogeographic relationships among deep-sea hydrothermal vent faunas at global scale. Deep Sea Res. Part 1 Oceanogr. Res. Pap. 56:1371–1378.

Baker E.T., German C.R., Elderfield H. 1995. Hydrothermal plumes over spreading-center axes: global distributions and geological inferences. In: S. Humphris, Editor. Physical, chemical, biological and geological interactions within hydrothermal systems. Geophysics Monographer 91. Washington (DC): AGU. p. 47–71.

Bascompte J., Jordano P. 2007. Plant–animal mutualistic networks: the architecture of biodiversity. Annu. Rev. Ecol. Evol. Syst. 38:567–593.

Batagelj V., Mrvar A. 2002. Pajek—analysis and visualization of large networks. Graph Drawing. 2265:477–478.

Blondel V.D., Guillaume J.L., Lambiotte R., Lefebvre E. 2008. Fast unfolding of communities in large networks. J. Stat. Mech. 10:P10008.

Chevaldonné P., Jollivet D., Desbruyères D., Lutz R.A., Vrijenhoek R.C. 2002. Sister-species of eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial COI clock calibration. Cah. Biol. Mar. 43:367–370.

Chevaldonné P., Jollivet D., Vangriesheim A., Desbruyères D. 1997. Hydrothermal-vent alvinellid polychaete dispersal in the eastern Pacific. 1. Influence of vent site distribution, bottom currents, and biological patterns. Limnol. Oceanogr. 42:67–80.

Corliss J.B., Dymond J., Gordon L.I., Edmond J.M., von Herzen R.P., Ballard R.D., Green K., Williams D., Bainbridge A., Crane K., van Andel T.H. 1979. Submarine thermal springs on the Galápagos Rift. Science. 203:1073–1083.

Desbruyères D., Hashimoto J., Fabri M.C. 2006. Composition and biogeography of hydrothermal vent communities in Western Pacific back-arc basins. Geophys. Monogr. Series. 166:215–234.

Desbruyères D., Segonzac M., Bright M. 2006. Handbook of deep-sea hydrothermal vent fauna. Linz (Austria): Landesmuseen.

Distel D.L., Baco A.R., Chuang E., Morrill W., Cavanaugh C., Smith C.R. 2000. Do mussels take wooden steps to deep-sea vents? Nature. 403:725–726.

Dongen S.V. 2008. Graph clustering via a discrete uncoupling process. SIAM J. Matrix Anal. Appl. 30:121–141.

Dos Santos D.A., Fernandez H.R., Cuezzo M.G., Dominguez E. 2008. Sympatry inference and network analysis in biogeography. Syst. Biol. 57:432–448.

Fortuna M.A., Albaladejo R.G., Fernandez L., Aparicio A., Bascompte J. 2009. Networks of spatial genetic variation across species. Proc. Natl. Acad. Sci. U S A. 106:19044–19049.

Fortunato S. 2010. Community detection in graphs. Phys. Rep. 486:75–174.

Freeman L.C. 1977. Set of measures of centrality based on betweenness. Sociometry. 40:35–41.

Girvan M., Newman M.E. 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. U S A. 99:7821–7826.

Hausdorf B. 2002. Units in biogeography. Syst. Biol. 51:648–652.

Hausdorf B., Hennig C. 2003. Biotic element analysis in biogeography. Syst. Biol. 52:717–723.

Hekinian R., Francheteau J., Renard V., Ballard R.D., Choukroune P., Cheminee J.L., Albarede F., Minster J.F., Charlou J.L., Marty J.C., Boulegue J. 1983. Intense hydrothermal activity at the axis of the East Pacific Rise near 13 ° N: submersible witnesses the growth of sulfide chimney. Mar. Geophys. Res. 6:1–14.

Hessler R.R., Lonsdale P.F. 1991. Biogeography of Mariana Trough hydrothermal vent communities. Deep Sea Res. 1 Oceanogr. Res. Pap. 38:185–199.

Hurtado L.A., Lutz R.A., Vrijenhoek R.C. 2004. Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. Mol. Ecol. 13:2603–2615.

Jollivet D., Chevaldonné 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.

Krause A.E., Frank K.A., Mason D.M., Ulanowicz R.E., Taylor W.W. 2003. Compartments revealed in food-web structure. Nature. 426:282–285.

Legendre P., Legendre L. 1998. Numerical ecology. Amsterdam (The Netherlands): Elsevier Scientific Publishing Company.

Little C.T.S. 2002. The fossil record of hydrothermal vent communities. Cah. Biol. Mar. 43:313–316.

Little C.T.S., Vrijenhoek R.C. 2003. Are hydrothermal vent animals living fossils? Trends Ecol. Evol. 18:582–588.

Lutz R.A., Shank T.M., Fornari D., Haymon R.M., Lilley M.D., Von Damm K.L., Desbruyères D. 1994. Rapid growth at deep-sea vents. Nature. 371:663–664.

Matabos M., Thiébaut E., Le Guen D., Sadosky F., Jollivet D., Bonhomme F. 2008. Geographic clines and stepping-stone patterns detected along the East Pacific Rise in the vetigastropod *Lepetodrilus elevatus* reflect species crypticism. Mar. Biol. 153:545–563.

McArthur A.G., Tunnicliffe V. 1998. Relics and antiquity revisited in the modern vent fauna. In: Mills R.A., Harrison K., editors. Modern ocean floor processes and the geological record. London: Geological Society. p. 271–291.

McRae B.H., Beier P. 2007. Circuit theory predicts gene flow in plant and animal populations. Proc. Natl. Acad. Sci. U S A. 104:19885–19890.

Nelson G., Platnick N.I. 1981. Systematics and biogeography: cladistics and vicariance. New York: Columbia University Press.

Newman M.E. 2004. Analysis of weighted networks. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 70:056131.

Newman W.A. 1985. The abyssal hydrothermal vent invertebrate fauna: a glimpse of antiquity? Bull. Biol. Soc. Wash. 6:231–242.

Plouviez S., Shank T.M., Faure B., Daguin-Thiebaut C., Viard F., Lallier F.H., Jollivet D. 2009. Comparative phylogeography among hydrothermal vent species along the East Pacific Rise reveals vicariant processes and population expansion in the South. Mol. Ecol. 18:3903–3917.

Proulx S.R., Promislow D.E.L., Phillips P.C. 2005. Network thinking in ecology and evolution. Trends Ecol. Evol. 20:345–353.

Rosvall M., Bergstrom C.T. 2008. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. U S A. 105:1118–1123.

Rozenfeld A.F., Arnaud-Haond S., Hernandez-Garcia E., Eguiluz V.M., Matias M.A., Serrao E., Duarte C.M. 2007. Spectrum of genetic diversity and networks of clonal organisms. J. R. Soc. Interface. 4:1093–1102.

Rozenfeld A.F., Arnaud-Haond S., Hernandez-Garcia E., Eguiluz V.M., Serrao E.A., Duarte C.M. 2008. Network analysis identifies weak and strong links in a metapopulation system. Proc. Natl. Acad. Sci. U S A. 105:18824–18829.

Sales-Pardo M., Guimera R., Moreira A.A., Amaral L.A. 2007. Extracting the hierarchical organization of complex systems. Proc. Natl. Acad. Sci. USA. 104:15224–15229.

Samadi S., Quemere E., Lorion J., Tillier A., von Cosel R., Lopez P., Cruaud C., Couloux A., Boisselier-Dubayle M.C. 2007. Molecular phylogeny in mytilids supports the wooden steps to deep-sea vents hypothesis. C. R. Biol. 330:446–456.

Sibuet M., Olu K. 2002. Cold seep communities on continental margins: structure and quantitative distribution relative to geological and fluid venting patterns. In: Ocean margin systems. Berlin-Heidelberg (Germany): Springer-Verlag. p. 235–251.

Smith A.D. 2003. A reappraisal of stress field and convective roll models for the origin and distribution of cretaceous to recent intraplate volcanism in the Pacific basin. Int. Geol. Rev. 45: 287–302.

Spiess F.N., Macdonald K.C., Atwater T., Ballard R., Carranza A., Cordoba D., Cox C., Diaz Garcia V.M., Francheteau J., Guerrero J., Hawkins J., Haymon R., Hessler R., Juteau T., Kastner M., Larson R., Luyendyk B., Mcdougall J.D., Miller S., Normark W., Orcutt J., Rangin C. 1980. East Pacific Rise: hot springs and geophysical experiments. Science. 207:1421–1433.

Stauffer D., Aharony A. 1994. Introduction to percolation theory. London: Taylor & Francis.

Strogatz S.H. 2001. Exploring complex networks. Nature. 410:268–276.

Tunnicliffe V. 1997. Hydrothermal vents: a global ecosystem. JAMSTEC J. Deep Sea Res. 105–110.

Tunnicliffe V., Fowler C.M.R. 1996. Influence of sea-floor spreading on the global hydrothermal vent fauna. Nature. 379:531–533.

Tunnicliffe V., Fowler C.M.R., McArthur A.G. 1996. Plate tectonic history and hot vent biogeography. In: MacLeod C.J., Tyler P.A., Walker C.L., editors. Tectonic, magmatic, hydrothermal and biological segmentation of mid-ocean ridges. London: Geological Society. p. 225–238.

Tunnicliffe V., McArthur A.G., McHugh D. 1998. A biogeographical perspective of the deep-sea hydrothermal vent fauna. Adv. Mar. Biol. 34:353–442.

Urban D., Keitt T. 2001. Landscape connectivity: a graph-theoretic perspective. Ecology. 82:1205–1218.

Van Dover C.L. 1995. Ecology of mid-Atlantic ridge hydrothermal vents. In: Parson L.M., Walker C.L., Dixon D.R., editors. Hydrothermal vents and processes. London: Geological Society Special Publication. p. 257–294.

Van Dover C.L. 2000. The ecology of deep-sea hydrothermal vents. Princeton (NJ): Princeton University Press.

Van Dover C.L., German C.R., Speer K.G., Parson L.M., Vrijenhoek R.C. 2002. Evolution and biogeography of deep-sea vent and seep invertebrates. Science. 295:1253–1257.

Van Dover C.L., Hessler R.R. 1990. Spatial variation in faunal composition of hydrothermal vent communities on the East Pacific rise and Galapagos spreading center. In: McMurray G.R., editor. Gorda Ridge: a seafloor spreading center in the United States' exclusive economic zone. New York: Springer Verlag. p. 253–264.

Vrijenhoek R.C. 1997. Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. Heredity. 88:285–293.

Watts D.J., Strogatz S.H. 1998. Collective dynamics of 'small-world' networks. Nature. 393:440–442.

## APPENDIX

### *Glossary*

**Betweenness centrality**: the fraction of the shortest paths in the network to pass through a node.

**Clustering coefficient**: a measure of the proportion neighboring nodes that can be reached through the nodes of other neighbors, calculated as the proportion of the focal node neighbors that are themselves neighbors.

**Degree of connectivity**: number of edges that connect the focal node to other nodes.

**Edge**: a connection between interacting nodes.

**Node**: an individual element within a network.

**Shortest path**: the path that traverses the minimum number of edges between two nodes.