Archimer
http://archimer.ifremer.fr

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
[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

*: Corresponding author : Sophie Arnaud-Haond, email address : Sophie.Arnaud@ifremer.fr

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.

Keywords: Biogeography, deep sea ecology, hydrothermal vents, network analysis, systems biology

1   Biogeographic provinces are areas of animal and plant distribution that have

2   similar or shared characteristics throughout. Distinct provinces support different biomes

3   resulting from divergent ecological and/or evolutionary processes. The identification of

4   biogeographic provinces and their delimitation are important goals within biogeography,

5   but such work is seldom straightforward. This is illustrated by the diversity of approaches

6   that have been proposed for this task (Hausdorf, 2002; Hausdorf and Hennig, 2003;

7   Nelson and Platnick, 1981), which often involve a somewhat arbitrary delimitation of

8   boundaries (Hausdorf, 2002) such as the ones based on the geographical distribution of

9   sites. Approaches are needed that limit the *a priori* hypotheses about the predominant

10   factors that shape provinces, and therefore deliver more objective delineations of

11   biogeographic provinces and the pathways of connectivity between them. Network

12   analysis appears to be a promising tool, as it can help us to understand geographical

13   landscape connectivity (Dos Santos et al., 2008; McRae and Beier, 2007; Urban and

14   Keitt, 2001) and genetic relationships between populations or individuals (Rozenfeld et

15   al., 2007; Rozenfeld et al., 2008). Networks provide representations of complex datasets

16   where nodes (e.g., communities) are interconnected by links that are scaled to the

17   connectivity between them, as indicated by metrics of distances or the degree of

18   interaction between nodes, all of which depend on the system under analysis. The

19   topology of the ensuing network can then be analysed with a range of tools and models

20   developed in network theory, allowing inferences on the past or present dynamic

21   properties of the system. Network analysis provides a holistic approach, free of *a priori*

22   dataset assumptions, which is better than classical pairwise interaction analyses for

23   assessing interactions within complex datasets (Proulx et al., 2005). Therefore, network

1 analysis has been shown to be a powerful tool for understanding the behaviour of

2 biological systems composed of interacting units, with nodes representing units from

3 genes to communities, and links illustrating their interactions. Analysis of biological

4 systems as complex networks has allowed, for example, an understanding of species

5 specialization in mutualistic interactions (Bascompte and Jordano, 2007), species

6 interactions in a food web (Krause et al., 2003), and the identification of specific

7 molecules with a crucial role in the stability of specific metabolic pathways (Albert et al.,

8 2000; Sales-Pardo et al., 2007; Strogatz, 2001). At the intra-specific scale, network

9 analysis has allowed the elucidation of genetic relationships and structure among

10 individuals or populations, and provided important insights into dispersal migration

11 (Fortuna et al., 2009; Rozenfeld et al., 2007), and source-sink dynamics (Rozenfeld et al.,

12 2008). At the community level, network analysis has allowed the identification of

13 differentiated communities and biogeographic entities along a continuously-distributed

14 space, based on patterns of sympatry (Dos Santos et al., 2008).

15   Here, we use network analysis to resolve the biogeography of hydrothermal vent

16 sites discretely scattered along oceanic ridges, using the faunal composition of 63 such

17 fields (Desbruyères et al., 2006b). Hydrothermal vents occur across all oceans due to

18 volcanic activity at active deep-sea ridges, back-arc spreading centers, volcanic arcs and

19 fore-arcs, and some intraplate seamounts. Widely spaced and often ephemeral, these

20 ecosystems develop in the darkness of the oceans. They depend almost entirely on

21 prokaryote chemosynthesis, which allows the proliferation of a specialized fauna

22 exhibiting limited biodiversity but extraordinarily high biomass (Van Dover, 2000). Since

23 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 4 and 7, depending on the number of vent fields explored and the methods and the underlying hypotheses used to delineate them (Desbruyères et al., 2006a; Tunnicliffe, 1997; Van Dover et al., 2002). 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 trees, to cluster hydrothermal sites into provinces. For example, in different studies, the East Pacific Rise 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

1 similarities in hydrothermal vent assemblages are particularly important because, in

2 addition to helping delineate biogeographical provinces, they can also help elucidate the

3 links between different biogeographical provinces, which may contribute to

4 understanding their history and colonization pathways, and the dispersal patterns of the

5 fauna they contain. The extreme level of endemism across vent sites can be attributed to

6 (i) the partial colonization of limited ridge segments by some species, and (ii) the

7 divergence of species over geological time. The occurrence of a large number of common

8 genera, represented by different species across vent sites, supports the idea of vicariance

9 as one of the main drivers of the spatial heterogeneity observed in community

10 composition (Tunnicliffe et al., 1998), resulting from species divergence on both sides of

11 geographical barriers such as major ridge discontinuities or transforming faults (Plouviez

12 et al., 2009). High endemism, combined with the extremely fragmented distribution of

13 vents along Mid-oceanic Ridges, means that the Network Analysis Method (NAM)

14 proposed by Dos Santos et al. (2008), is not possible based on species; nor was it

15 necessary, as the fragmented distribution of sites offers a natural delineation of

16 geographic units (i.e., hydrothermal fields).

17 An alternative approach to examine the relationships between the hydrothermal

18 vent fauna can be taken, using network analysis to represent each of the 63 hydrothermal

19 vent fields, investigated to date, as nodes can be connected among them with a distance

20 reflecting the common genera among sites (links). These networks of hydrothermal vent

21 communities can then be formally analyzed to explore their structures; identifying groups

22 of closely-related locations that are helpful in delineating biogeographic provinces and

23 the hydrothermal vent regions connecting these provinces. Moreover, such networks can

1   be constructed without entering any information concerning the geological or geographic

2   properties of the sites, thereby allowing subsequent examination of the consistency

3   between the network structure and the independent geological or geographic properties as

4   a validation exercise. In particular, we aimed at testing i) the various delineation of

5   biogeographic provinces hitherto proposed by different authors, and ii) the hypothesis of

6   an ancient history of colonization and evolution of hydrothermal communities by

7   comparing the pattern of clustering and connectivity revealed by network analysis with

8   the spreading of oceanic ridges across geological time (Desbruyères et al., 2006a;

9   Tunnicliffe, 1997; Van Dover et al., 2002).

10

11   MATERIALS AND METHODS

12   *Dataset used in this study*

13   Hydrothermal vent fields (n=63) were selected across the world's oceans

14   according to available information on their faunal composition. A taxonomic database

15   from a recent compilation (Desbruyères et al., 2006b), was assembled from information

16   from the literature, databases on the World Wide Web (ChEss:

17   http://www.noc.soton.ac.uk/chess/database/, NOAA: http://www.pmel.noaa.gov/, NeMO:

18   http://www.pmel.noaa.gov/vents/nemo/), and researchers we contacted directly. At each

19   of the 63 hydrothermal vent fields, the presence or absence of taxa was noted among the

20   total 591 species and 331 genera described. All the taxa recorded in the biogeographic

21   provinces found by network analysis are described in Supplementary Table 1 (available

22   from http://www.sysbio.oxfordjournals.org).

23   *Networks*

1   We began analysis with a fully connected network of 63 hydrothermal vent fields

2   considered as nodes. Each link joining a pair of fields was assigned the corresponding

3   Jaccard's index distance. This assemblage distance is given by:

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

5   where a is the number of taxa (i.e., genera or species) common to two fields (for example

6   field 1 and field 2), b is the total number of taxa present in field 1 but absent from field 2,

7   and c is the total number of taxa present in field 2 but absent from field 1.

8   This coefficient ranges from 0 (both fields are identical) to 1 (the fields share no

9   items) and gives similar weight to the presence or absence of a taxon. This property

10   makes the Jaccard's index distance more appropriate for the present case because the

11   available description of these remote communities is far from exhaustive, and some

12   absences may also be due to the lack of observation of a given taxon, rather than to true

13   absences from the community. Nevertheless, when tests with Sorensen index, which

14   gives double weighting to double presence, were realized on the same dataset, results

15   yielded a similar network topology, with identical clusters. Consequently, we kept the

16   simplest index and focused on analyses at different taxonomic levels. Therefore, we

17   analyzed and compared networks of Jaccard's index at both the species and genus levels.

18   Like previous authors (Tunnicliffe et al., 1996), we consider that genus-level analysis

19   provides a more appropriate picture of the patterns of dispersal and/or vicariance between

20   hydrothermal fields, based on the high levels of endemism observed at the species level.

21   Jaccard distance interpretation in this context is that hydrothermal vent areas once

22   connected by strongest dispersal paths would have exchanged more species, resulting in a

1   lower Jaccard distance at both species and genus levels among contemporary

2   communities.

3   The similarity matrix provides a fully-connected network that does not allow

4   recognition of biogeographic provinces. Various approaches have been developed to

5   extract biogeographic provinces from networks, including those based on community

6   detection (Allesina and Pascual, 2009; Krause et al., 2003; Newman, 2004) and

7   percolation theory (Rozenfeld et al., 2008). A percolation approach was used to break

8   down the fully-connected network into discrete clusters of biogeographic provinces

9   because these are better able to handle continuous link weights (Jaccard's index distance).

10 Moreover, Rozenfeld *et al.* (2008) successfully applied the percolation distance approach

11 to resolve gene flow within species, based on a fully-connected network of populations.

12 We transposed this method to the community level to resolve the biogeographic

13 provinces emerging from historical disruption of gene flow through ancient vicariant

14 events.

15   The percolation method is based on the analysis of a network built only with links

16 illustrating the minimum Jaccard distance necessary to maintain the connexion across

17 most components of the system. At this distance precisely, the network still depicts a

18 single, giant cluster composed of sub-clusters linked through primary connections. Below

19 this distance, the network collapses into disconnected sub-clusters, and the pattern of

20 global connectivity is lost. This critical threshold distance is also named percolation

21 distance (Dp); (Stauffer and Aharony, 1994). For a finite system, this point is derived by

22 calculating the average cluster size of all clusters excluding the largest,

23 $\langle S \rangle^* = \frac{1}{N} \sum_{s \langle S \max} s^2 n_s$,

1   as a function of the last threshold distance value beyond which links were removed. N is

2   the total number of nodes not included in the largest cluster and n_s is the number of

3   clusters containing s nodes. Smax is the size of the largest cluster. The Dp is then

4   identified heuristically in the transitional region characterized by a strong decrease in

5   <S>*. Inside this transitional stage, the single, giant cluster network is found at Dp=0.84.

6   The network topology and its characteristics were analyzed at this percolation distance,

7   meaning that the links retained in the network analyzed are the ones corresponding to

8   ecological distances lower or equal to Dp, while all links beyond that value were

9   discarded. Additionally, the network was also explored at different distance thresholds

10 around this percolation point, in order to assess the consistency of its topology and the

11 inferred properties and interpretation. The network characteristics (see Glossary) were

12 quite robust to variability with the applied threshold distance Dp: major clusters, nodes

13 and paths remained unchanged indicating in the robustness of these key properties to

14 uncertainty in the percolation threshold and, therefore, supporting the reliability of the

15 network for displaying the biogeographic structure of the communities.

16   In order to assess the respective influence of the three main phyla that encompass

17 more than 80 % of the genera (Annelida, Mollusca and Arthropoda, which represent 20.5

18 %, 25 % and 37 %, respectively) on the global network topology, three sub-networks

19 were then built based on genera within these three major phyla.

20

21   *Estimation of the global and local properties of the network*━ The *connectivity*

22 *degree*, k_i of a given node *i*, is the number of other nodes linked to it (i.e., the number of

23 neighbor nodes). We named E_i as the number of links existing among the neighbors of

1 node i. This quantity takes values between 0 and $E_i^{(max)} = k_i(k_i-1)/2$, which is the case in a

2 fully connected neighborhood. This value is used to calculate the *clustering coefficient C_i*

3 of node i, defined as:

4 $$C_i = \frac{E_i}{E_i^{(max)}} = \frac{2E_i}{k_i(k_i-1)},$$

5 C_i quantifies how close the node i and its neighbors are to being a clique (See

6 Supplementary Fig. 1a for an illustrative diagram). The *clustering coefficient* (Watts and

7 Strogatz, 1998) of the whole network $<CC>$ is defined as the average of all the individual

8 clustering coefficients in the system. C_i values vary between 0 and 1. The *clustering*

9 *coefficient* helps us to understand how nodes are organized into clusters within the system

10 as a whole. In order to test the significance of $<CC>$, its value was compared to the

11 average value $<CC_o>$ of 10 000 random simulations of our dataset. To perform this test,

12 we randomly rewired our network with the same number of links present at a given

13 threshold (here percolation), thereby yielding random networks with the same overall

14 structure of taxa presence and absence.

15 The *betweenness centrality* (Freeman, 1977) of node i, $bc(i)$, counts the fraction of

16 shortest paths between pairs of nodes that pass through node i. Let σ_{st} denote the number

17 of shortest paths connecting nodes s and t, and $\sigma_{st}(i)$ the number of those passing through

18 the node i; then,

19 $$bc(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}.$$

20 The *betweenness centrality* determines the relative importance of a node within the

21 network as an intermediary in the flow of information and its vulnerability to

22 fragmentation (See Supplementary Fig. 1b for an illustrative schema).

1   Networks were visualized and analysed using Pajek software (Batagelj and Mrvar,

2   2002). Random simulations were compiled by C++ scripts.

3

4   *Comparative analysis*— A set of four methods based on community detection

5   were additionally tested on this dataset to infer their potential for biogeographic analysis

6   on our dataset (Blondel et al., 2008; Dongen, 2008; Girvan and Newman, 2002; 2008).

7   See Supplementary Table 2 for details on these methods.

8   RESULTS

9   *Delineation of Biogeographic Provinces*

10   In order to reveal the clustering and interconnectivity of hydrothermal vents, we

11   built networks with data from the 63 fields (nodes) referenced in the database (See

12   Supplementary Table 3), connected by links (edges) on the basis of their faunal distance

13   in terms of genera, as assessed with Jaccard's index in an adjacency matrix (Legendre

14   and Legendre, 1998). Genera were chosen, rather than species, because of the extremely

15   high level of endemism at the species level (95.3 %, compared with 76.2 % for genera;

16   see Supplementary Fig. 2), indicating that biogeographic history is ancient and likely to

17   be better reflected by the distribution of genera (Tunnicliffe and Fowler, 1996).

18   The Jaccard's distances were, as expected, lower for neighbouring vent sites but

19   the resulting network had a percolation value at a Jaccard distance of Dp=0.84, below

20   which distinct groups of vent sites emerged suggesting distinct biogeographic regions

21   (Fig. 1). In particular, the network topology clearly shows the existence of 5 strong

22   clusters of hydrothermal vents (Fig. 1 and Fig. 2, see Table 1 for the network properties),

23   even though the data used contained no geographic information and the process did not

1	involve any subjective dissection like those needed for a 'classic' dichotomic UPGMA

2	tree (See Supplementary Fig. 3). The ensuing clusters of hydrothermal vents correspond

3	to five well-defined regional provinces, coherent both at the genus (Fig. 2) and species

4	(Supplementary Fig. 4 and 5) levels, in agreement with Tunnicliffe *et al.* (1998). At the

5	species level, the network topology looses its integrity at a Dp value higher than the one

6	of the genus level (0.95 vs 0.84). This is due to the high species endemism; the much

7	lower number of shared species than genera among fields results in a higher threshold

8	below which the connectivity across the entire network breaks down (Supplementary Fig.

9	5).

10	 Analyses performed at the genus level showed that this hierarchical structure is

11	supported by a significantly higher average clustering coefficient, $\langle CC \rangle$=0.52, than

12	expected by chance (i.e., by randomly rewiring the links ($\langle CC_o \rangle$=0.17 with σ_o=0.01, after

13	10 000 random simulations), thus illustrating the modularity of the network composed of

14	clusters of hydrothermal fields, with greater internal interconnection than would be

15	expected by chance. Moreover, the recognition of the West Pacific (WP), Northeast

16	Pacific (NP), East Pacific Rise (EPR), and Mid-Atlantic Ridge (MAR) as provinces is

17	confirmed by their individual $\langle CC \rangle$ values, which are higher than the average (Table 1).

18	This highlights the strong endemism of hydrothermal vent fauna among the provinces

19	defined by the network clusters (Fig. 2).

20	*Connectivity among Provinces*

21	 In addition to an objective identification of biogeographic provinces and sub-

22	provinces, network analysis offers another fundamental advantage: the study of

23	biogeographic patterns in hydrothermal fields as the patterns of connectivity among fields

1   and provinces can suggest past and present dynamics of information flow reflecting

2   species divergence. Putative pathways can be assessed from the connections among

3   biogeographic provinces in the network topology, identifying regions that possibly play

4   or have played a role in the foundation and/or connectivity of the system. The first

5   observation that emerges when these 'inter-cluster' links are considered is the central

6   position of the WP province. WP is intermediate among all other provinces, which

7   themselves share no links (Fig. 2). An accurate diagnostic for this analysis is the

8   *betweenness centrality* (see Materials and Methods), which estimates the relative

9   importance of a node in relaying the flow of information through the system,

10 corresponding here to past gene flow, which will result in common genera across fields

11 represented in the network (Fig 2 and Supplementary Table 4). Here, the two strongest

12 *betweenness centrality* values indeed correspond to fields in the WP region: the Mariana

13 Trough and the Pacmanus system, which are present in 28 % and 26 %, respectively, of

14 all shortest paths among fields. As noted above, the MAR province has no direct link

15 with EPR, so that the Eastern Pacific and Atlantic Oceans are only connected in the

16 network through the West Pacific and Indian Oceans.

17   The taxon-specific networks represented in Supplementary Fig. 6 show slightly

18 lower percolation threshold than the global network (Dp = 0.84) revealing a higher

19 overall proximity of communities of Annelida (Dp = 0.68), Arthropoda (Dp = 0.77), and

20 Mollusca (Dp = 0.75) but generally support the number and identity of the biogeographic

21 provinces revealed by the global network. These patterns are clearly similar for Annelida

22 and Arthropoda (Supplementary Fig. 6a and 6b) with the same five provinces

23 distinguished. For Mollusca (Supplementary Fig. 6c), WP, MAR, and EPR provinces still

1   appear as a single cluster with NP divided into two clusters connected either to WP or to

2   EPR, and IO disconnected.

3

4   *Comparative Analysis*

5   Among the methods tested based on community detection, only the method from

6   Rosvall and Bergstrom (2008) showed resolution and cluster discrimination, while no

7   coherent clusters emerged with the other methods (see Supplementary Table 1). The

8   analysis with the Rosvall and Bergstrom (2008) method shows a lower resolution than the

9   percolation approach, while still supporting the discrimination of NP and EPR, and their

10   sequential branching with and through the cluster including WP (merged with MAR and

11   IO).

12   DISCUSSION

13   The overall picture of hierarchical structure and connectivity among vent sites

14   depicted by this network analysis allows the definition of five biogeographic provinces at

15   the percolation threshold distance of the network, which is defined on the basis of the

16   global analysis of taxonomic similarity among sites. Compared to previous studies based

17   on other approaches that rely on a more 'arbitrary' identification of clusters, some

18   previously-recognized distinct provinces appear to be compounded into a single one here,

19   while others are well differentiated. In fact, the splitting of EPR or MAR proposed in

20   previous studies (Tunnicliffe, 1997; Van Dover et al., 2002) is not supported here for

21   MAR (Fig. 1) which corresponds, at best, to slightly differentiated sub-provinces for

22   EPR, disconnecting at a threshold far lower than that at which the other provinces are

23   identified (just before the threshold of 0.60; Fig. 1a). The integrity of the IO province

1   remains questionable due to its extremely weak representation in the data gathered so far,

2   due to a lack of scientific cruises in this area. The two fields explored in IO do however

3   remain strongly linked together, which, coupled with the robust intermediate position of

4   IO between MAR and WP (Fig. 2), strongly supports the existence of at least one discrete

5   province in this region.

6   The central position of WP in all connection paths between provinces was

7   supported both by the global analysis at the genus and species (Supplementary Fig. 4)

8   levels, by the analysis specific to each of the three main phyla (Annelida, Mollusca, and

9   Arthropoda). The network topology derived here calls into question the hypothesis of a

10   major ancestral pathway between the EPR and MAR provinces through the Isthmus of

11   Panama as well as the hypothesized connectivity between the Eastern Pacific and the

12   Southern Atlantic through the Circumpolar Current (Van Dover et al., 2002).

13

14   *Network Consistency with Tectonic History*

15   The star-like topology of the network (Fig. 2), characterised by a central position

16   of WP (formally reflected in its high betweenness-centrality score in the network), results

17   in the lack of direct connection between EPR and NP at the percolation threshold, which

18   suggests a more recent common history of WP with each of these provinces

19   independently.

20   The topology derived here shows some similarity to the early history of the plate

21   tectonic and spreading ridge system in the Pacific during the Mesozoic and its evolution

22   over more than 100 million years (see Fig. 3). Although this consistency does not

23   necessarily involve a causal relationship, it lends weight to the hypothesis that the current

1   biogeographic structure of hydrothermal vent fauna may have been influenced by the

2   tectonic dynamics of plates over millions of years (Hessler and Lonsdale, 1991;

3   Tunnicliffe et al., 1996). Indeed, the double network connection of EPR and NP

4   exclusively through WP, which also appeared when using the community detection

5   method of Rosvall and Bergstrom (2008, Supplementary Table 2), is consistent with the

6   history of the Pacific ridge since the early Cretaceous (~150 Ma), and the split of the

7   protoridges triple junction during the end of the Cretaceous (~70 Ma) and Paleocene (~60

8   Ma; Fig. 3) (Smith, 2003). The branching of EPR on WP observed in the network (Fig.

9   2) is also consistent with the early Cretaceous position of the three ridges (Izanagi-

10   Farallon, Pacific-Izanagi, and Pacific-Farallon), connecting the West and the South

11   Eastern Pacific (Fig. 3a), until their disconnection by a fracture zone in the Late

12   Cretaceous (Fig. 3c). Finally, the network connection of NP to WP can also be considered

13   in parallel with the early history of the Juan de Fuca Ridge, which appeared during the

14   Paleocene and was connected to WP and the future EPR at the time when the ridge triple

15   junction splits the Pacific, Kula, and Farallon plates.

16   It is obviously impossible to test this hypothesis experimentally, so the direct

17   influence of the tectonic plate history on the present nature and distribution of

18   hydrothermal communities, while consistent with the topology of the network built here,

19   remains speculative. Nevertheless, it deserves to be examined in further studies, such as

20   through the construction of complete phylogenies of some of the most widespread genera,

21   with appropriate markers allowing the use of relaxed molecular clocks.

22

1   *Hubs, Connectivity and Centrality*

2   The central position of WP in the network as an obligatory relay between EPR

3   and the other provinces (Fig. 2), contrasts with the hypothesis of a center of dispersion in

4   EPR (Bachraty et al., 2009; Van Dover, 2000; Van Dover and Hessler, 1990). This may

5   be due to the use of the Jaccard's distance as, in contrast to other distance metrics such as

6   the dispersal direction coefficient (Bachraty et al., 2009), for example, which comes with

7   the underlying hypothesis that richest provinces disperse more, Jaccard distance does not

8   impose an underlying hypothesis linking richness to present or past connectivity. The

9   EPR appears as the richest province in the database, both in terms of species and genera.

10   Yet, rather than a high divergence rate and dispersal towards adjacent provinces, this

11   richness may reflect the combination of different sampling strategies among provinces

12   and the homogeneous topology of EPR.

13   The homogeneity and high connectivity in EPR is illustrated by its robustness in

14   the network. As the threshold distance declines, the cohesion of each biogeographic

15   province collapses in the sequence: WP, MAR, NP, EPR (Fig. 1), with EPR still strongly

16   clustered long after the other provinces collapse (Fig. 1a). This homogeneity is illustrated

17   by its much larger average *connectivity degree* (See Methods; k = 7.81 for n = 17 vents

18   observed; Table 1) that reveals the relatively low level of endemism within this province

19   (Supplementary Fig. 2). This may be the result of the specific regional properties of the

20   EPR fields (Fig. 4) that exhibit a high accretion rate (rate of creation of new oceanic

21   lithosphere) inducing a higher number of vents (Baker et al., 1995), a factor which,

22   associated with a homogeneous depth is likely to facilitate the faunal dispersion. The

23   relative homogeneity of EPR is also reflected by a rather low genetic structure and

1 significant gene flow estimates for several taxa studied across EPR, the area where the

2 most population genetic studies have been performed to date (Chevaldonné et al., 1997;

3 Hurtado et al., 2004; Jollivet et al., 1999; Matabos et al., 2008; Vrijenhoek, 1997); a

4 result in line with the relative ubiquity of taxa in this province (Supplementary Fig. 2). In

5 contrast, the WP province shows a much higher level of endemism, as shown by a much

6 lower average *connectivity degree* (k=4.53), despite the greater number of referenced

7 vent fields (n=26). Moreover, the EPR has been the specific target of numerous cruises,

8 allowing the examination of extinction and recolonization processes (Lutz et al., 1994), as

9 well as the discovery and description of taxa of distinct succession stages. As a result,

10 despite its apparently high richness in comparison with some previous studies (Bachraty

11 et al., 2009; Van Dover, 1995; Van Dover, 2000), EPR does not emerge as an especially

12 central cluster of nodes (province) in this network, which instead supports a central

13 position for WP.

14   This central position of WP, both at the scale of the Pacific Ocean and worldwide,

15 combined with its extremely high endemism, suggests a more ancient history and a

16 central role in the biogeography of hydrothermal vents. WP is the region of the world

17 where the basins exhibit the highest bathymetric variability of vents (Fig. 4), compared

18 with mid-ocean ridges (Matabos et al., 2008), and where closely located vents and seeps

19 coexist most frequently (Sibuet and Olu, 2002). There is recent evidence suggesting that

20 the colonization of deep chemosynthetic ecosystems may have followed a stepping stone

21 model from the coastal area, through sunken wooden debris and large animal carcasses,

22 to the cold seeps and finally the hydrothermal vents (Van Dover et al., 2002). Such

23 findings agree with inferences drawn by Little and Vrijenhoek (2003) on fossil and

1   molecular evidence. Observations supporting this stepping stone hypothesis were first

2   reported on the mythilidae (Distel et al., 2000; Samadi et al., 2007). Altogether, (i) the

3   heterogeneity of WP field depth and their close proximity to seeps (active margins),

4   fallen wood (tropical islands), and the coastal zone, (ii) their pattern of within-region

5   endemism and ancient divergence among taxa on this old, slow-spreading ridge, and (iii)

6   their central position in the topology of the network reported here, support the hypothesis

7   of a colonization or recolonization of hydrothermal vents through WP, followed by a

8   subsequent worldwide dispersion that was possibly influenced by plate tectonics and the

9   associated evolution of spreading ridge systems since the early Mesozoic (Desbruyères et

10   al., 2006a). This hypothesis could also be worthy of future testing using phylogenetic

11   reconstructions.

12

13   *Origin and Ageing of Communities*

14   The reconstruction of the ageing and history of vent taxa to date has been based

15   either on fossils or molecular clocks, which produced contradictory estimates (Little and

16   Vrijenhoek, 2003). The oldest fossils from the Paleozoic (~450 Ma) and the high

17   endemism of hydrothermal communities led to an 'antiquity hypothesis' (Newman,

18   1985). Molecular data, in contrast, supported a much more recent history, even when

19   considering a slow molecular clock calibrated on the Farallon-Pacific Ridge disruption

20   under the North-American Plate (Chevaldonné et al., 2002). Reconsidering the evidence

21   from both molecular and fossil data, Little and Vrijenhoek (2003) suggested that present

22   hydrothermal fauna were of intermediate age and may have originated in the Mesozoic (-

23   150 Ma). The network topology described here lends support to their suggestions,

1   showing consistency between the evolution of oceanic ridges since the early Cretaceous

2   and the topology of the network that illustrates the nature and connections among

3   biogeographic provinces. It is also in agreement with a role of shallow hydrothermal

4   vents and close cold seeps in the (re)colonization of hydrothermal vents. In line with the

5   geographic location of most fossil records (Little, 2002), the network suggests a high

6   centrality of the WP region where (re)colonization events may have mainly taken place.

7   The hypothesized central position of WP and its possible role in the original history of

8   present day hydrothermal vent fauna derived from network analysis also require further

9   investigation, which could be done through comprehensive phylogenies of different

10 genera, to test for whether fauna from WP have the basal position that would be expected

11 from such an ancestral role.

12

13   The application of network analysis to examine hydrothermal vent biogeography

14 reported here provides evidence of the potential of this approach to address complex

15 biogeography problems. First, network analysis allows the discrimination of clusters of

16 communities, or biogeographic provinces and is genuinely obtained without introducing

17 additional information, such as *a priori* assumptions on their clustering based on their

18 geographic location. Second, this holistic approach allows illustrating the connectivity

19 between provinces with a rather flexible geometry, free of the constraint of the

20 dichotomous relationship imposed by branching rules in classical trees. Third, network

21 analysis provides tools developed to derive diagnostics on the dynamics properties of the

22 network, such as *betweenness centrality*, which allows particular regions or provinces to

23 be pointed out that may have had a central role in the evolution and dispersion of fauna

1 and, therefore, be of major relevance for biogeographical analyses. The application of

2 network analysis, already used across a wide range of fields from social science to

3 ecology, and more recently in population genetics (Fortuna et al., 2009; Rozenfeld et al.,

4 2008), therefore opens a promising avenue for the investigation of biogeographic

5 questions. This work takes a first step that we hope may encourage the use of network

6 analysis, no only with the method presented here but also others, such as community

7 detection approaches (Fortunato, 2010), among the rapidly broadening and improving

8 range of network analysis methods.

9

10

18

REFERENCES

Albert, R., H. Jeong, and A. L. Barabasi. 2000. Error and attack tolerance of complex networks. Nature 406:378-82.

Allesina, S., and M. Pascual. 2009. Food web models: a plea for groups. Ecol Lett 12:652-62.

Bachraty, C., P. Legendre, and D. Desbruyères. 2009. Biogeographic relationships among deep-sea hydrothermal vent faunas at global scale. Deep Sea Research Part I: Oceanographic Research Papers 56:1371-1378.

Baker, E. T., C. R. German, and H. Elderfield. 1995. Hydrothermal plumes over spreading-center axes: global distributions and geological inferences. Pages 47-71 *in* Seafloor Hydrothermal Systems (R. A. Z. S.E. Humphris, L.S. Mullineaux and R.E. Thomson, ed.) American Geophysical Union.

Bascompte, J., and P. Jordano. 2007. Plant-Animal Mutualistic Networks: The Architecture of Biodiversity. Annual Review of Ecology, Evolution, and Systematics 38:567-593.

Batagelj, V., and A. Mrvar. 2002. Pajek - Analysis and visualization of large networks. Graph Drawing 2265:477-478.

Blondel, V. D., J. L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics-Theory and Experiment:-.

Chevaldonné, P., D. Jollivet, D. Desbruyères, R. A. Lutz, and R. C. Vrijenhoek. 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., D. Jollivet, A. Vangriesheim, and D. Desbruyères. 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.

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

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

Desbruyères, D., M. Segonzac, and M. Bright. 2006b. Handbook of deep-sea hydrothermal vent fauna. Landesmuseen, Linz.

Distel, D. L., A. R. Baco, E. Chuang, W. Morrill, C. Cavanaugh, and C. R. Smith. 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 Journal on Matrix Analysis and Applications 30:121-141.

Dos Santos, D. A., H. R. Fernandez, M. G. Cuezzo, and E. Dominguez. 2008. Sympatry inference and network analysis in biogeography. Syst Biol 57:432-48.

Fortuna, M. A., R. G. Albaladejo, L. Fernandez, A. Aparicio, and J. Bascompte. 2009. Networks of spatial genetic variation across species. Proc Natl Acad Sci U S A 106:19044-9.

Fortunato, S. 2010. Community detection in graphs. Physics Reports-Review Section of Physics Letters 486:75-174.

Freeman, L. C. 1977. Set of Measures of Centrality Based on Betweenness. Sociometry 40:35-41.

Girvan, M., and M. E. Newman. 2002. Community structure in social and biological networks. Proc Natl Acad Sci U S A 99:7821-6.

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

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

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

Hessler, R. R., and P. F. Lonsdale. 1991. Biogeography of Mariana Trough Hydrothermal Vent Communities. Deep-Sea Research Part a-Oceanographic Research Papers 38:185-199.

Hurtado, L. A., R. A. Lutz, and R. C. Vrijenhoek. 2004. Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrothermal vents. Mol Ecol 13:2603-15.

Jollivet, D., P. Chevaldonné, and B. Planque. 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., K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor. 2003. Compartments revealed in food-web structure. Nature 426:282-5.

Legendre, P., and L. Legendre. 1998. Numerical Ecology. Elsevier Scientific Publishing Company.

Little, C. T. S. 2002. The fossil record of hydrothermal vent communities. Cahiers de Biologie Marine 43:313-316.

Little, C. T. S., and R. C. Vrijenhoek. 2003. Are hydrothermal vent animals living fossils? Trends in Ecology & Evolution 18:582-588.

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

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

McArthur, A. G., and V. Tunnicliffe. 1998. Relics and antiquity revisited in the modern vent fauna. Pages 271-291 *in* Modern Ocean Floor Processes and the Geological Record (K. Mills R.A. & Harrison, ed.) Geological Society, London.

McRae, B. H., and P. Beier. 2007. Circuit theory predicts gene flow in plant and animal populations. Proc Natl Acad Sci U S A 104:19885-90.

Nelson, G., and N. I. Platnick. 1981. Systematics and Biogeography: Cladistics and Vicariance, New York.

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? Bulletin of the Biological Society of Washington 6:231-242.

Plouviez, S., T. M. Shank, B. Faure, C. Daguin-Thiebaut, F. Viard, F. H. Lallier, and D. Jollivet. 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., D. E. L. Promislow, and P. C. Phillips. 2005. Network thinking in ecology and evolution. Trends in Ecology & Evolution 20:345-353.

Rosvall, M., and C. T. Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A 105:1118-23.

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

Rozenfeld, A. F., S. Arnaud-Haond, E. Hernandez-Garcia, V. M. Eguiluz, E. A. Serrao, and C. M. Duarte. 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., R. Guimera, A. A. Moreira, and L. A. Amaral. 2007. Extracting the hierarchical organization of complex systems. Proc Natl Acad Sci U S A 104:15224-9.

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

Sibuet, M., and K. Olu. 2002. Cold Seep Communities on Continental Margins: Structure and Quantitative Distribution Relative to Geological and Fluid Venting

Patterns. Pages 235-251 *in* Ocean MArgin Systems (Springer-Verlag, ed.), Berlin Heidelberg.

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. International Geology Review 45:287-302.

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

Stauffer, D., and A. Aharony. 1994. Introduction to Percolation Theory Taylor & Francis, London.

Strogatz, S. H. 2001. Exploring complex networks. Nature 410:268-76.

Tunnicliffe, V. 1997. Hydrothermal vents: A global ecosystem. JAMSTEC Journal of Deep-Sea Research:105-110.

Tunnicliffe, V., and C. M. R. Fowler. 1996. Influence of sea-floor spreading on the global hydrothermal vent fauna. Pages 531-533 *in* Nature.

Tunnicliffe, V., C. M. R. Fowler, and A. G. McArthur. 1996. Plate tectonic history and hot vent biogeography. Pages 225-238 *in* Tectonic, Magmatic, Hydrothermal and Biological Segmentation of Mid-Ocean Ridges (C. J. MacLeod, P. A. Tyler, and C. L. Walker, eds.). Geological Society.

Tunnicliffe, V., A. G. McArthur, and D. McHugh. 1998. A biogeographical perspective of the deep-sea hydrothermal vent fauna. Advances in Marine Biology 34:353-442.

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

Van Dover, C. L. 1995. Ecology of Mid-Atlantic Ridge hydrothermal vents. Pages 257-294 *in* Hydrothermal vents and processes (L. M. Parson, C. L. Walker, and D. R. Dixon, eds.). Geological Society Special Publication, London.

Van Dover, C. L. 2000. The Ecology of Deep-Sea Hydrothermal Vents. Princeton University Press, Princeton New Jersey.

Van Dover, C. L., C. R. German, K. G. Speer, L. M. Parson, and R. C. Vrijenhoek. 2002. Evolution and biogeography of deep-sea vent and seep invertebrates. Science 295:1253-7.

Van Dover, C. L., and R. R. Hessler. 1990. Spatial variation in faunal composition of hydrothermal vent communities on the East Pacific rise and Galapagos spreading center. Pages 253-264 *in* Gorda Ridge : A seafloor spreading center in the United States' exclusive economic zone (G. R. McMurray, ed.) Springer Verlag.

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., and S. H. Strogatz. 1998. Collective dynamics of 'small-world' networks. Nature 393:440-2.
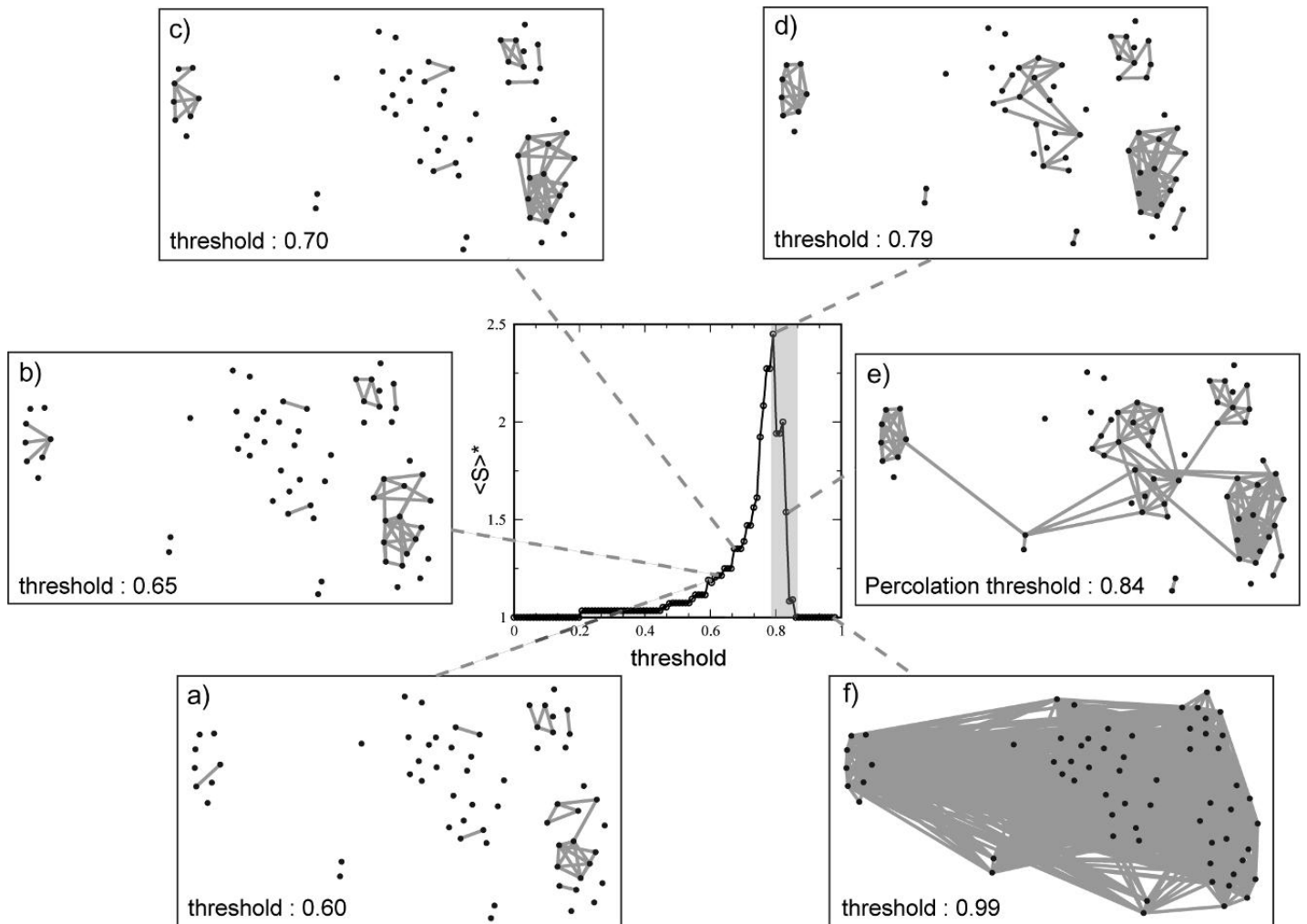
Figures

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.
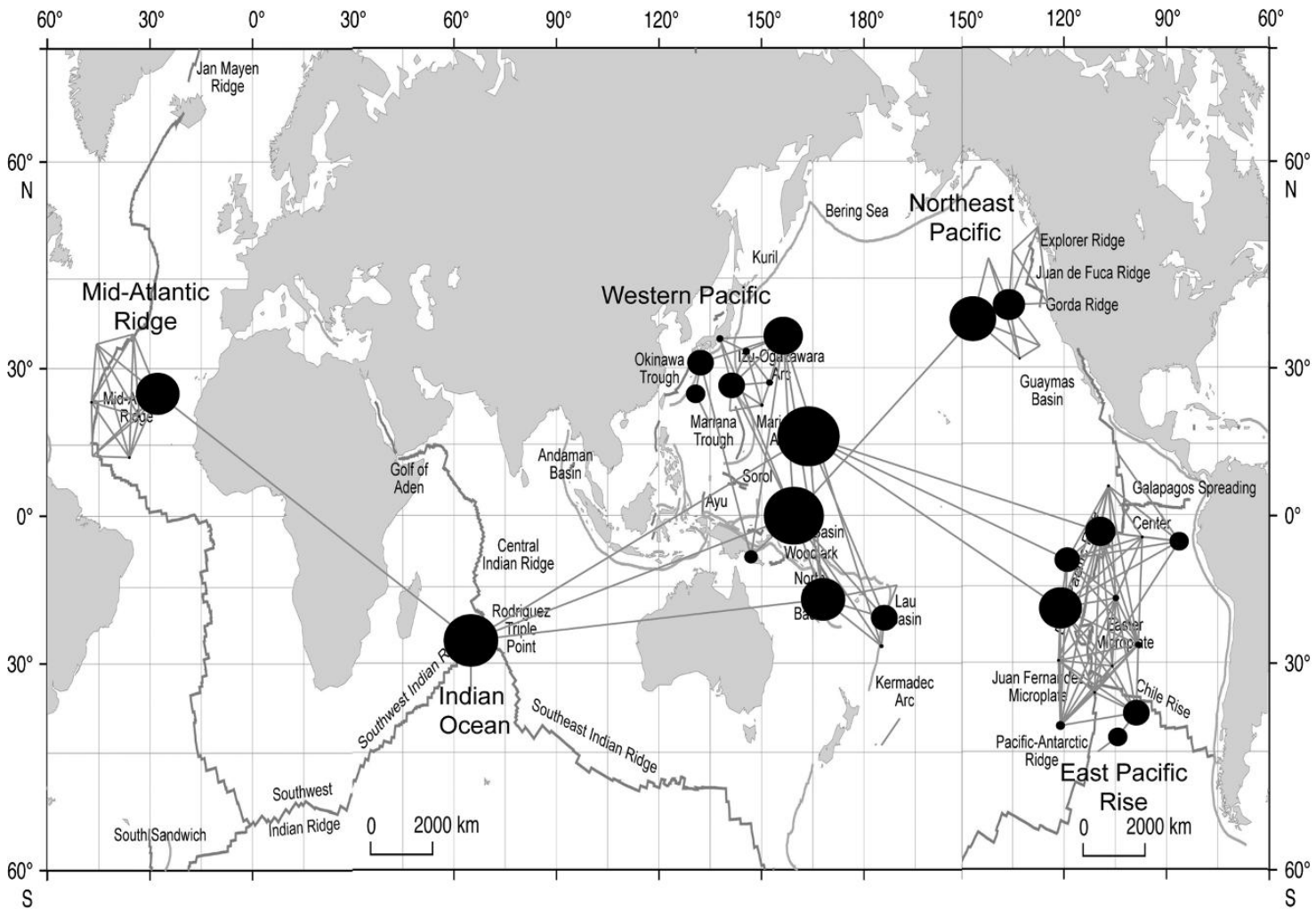
FIGURE 2. Global network of hydrothermal vent fauna diversity built on the basis of Jaccard's distance among fields and represented here at the percolation threshold (Dp = 0.84. See Materials and Methods section). Circle size represents the betweenness centrality values of the corresponding field. Five provinces are highlighted by this network analysis: MAR, IO, WP, NP, and EPR.
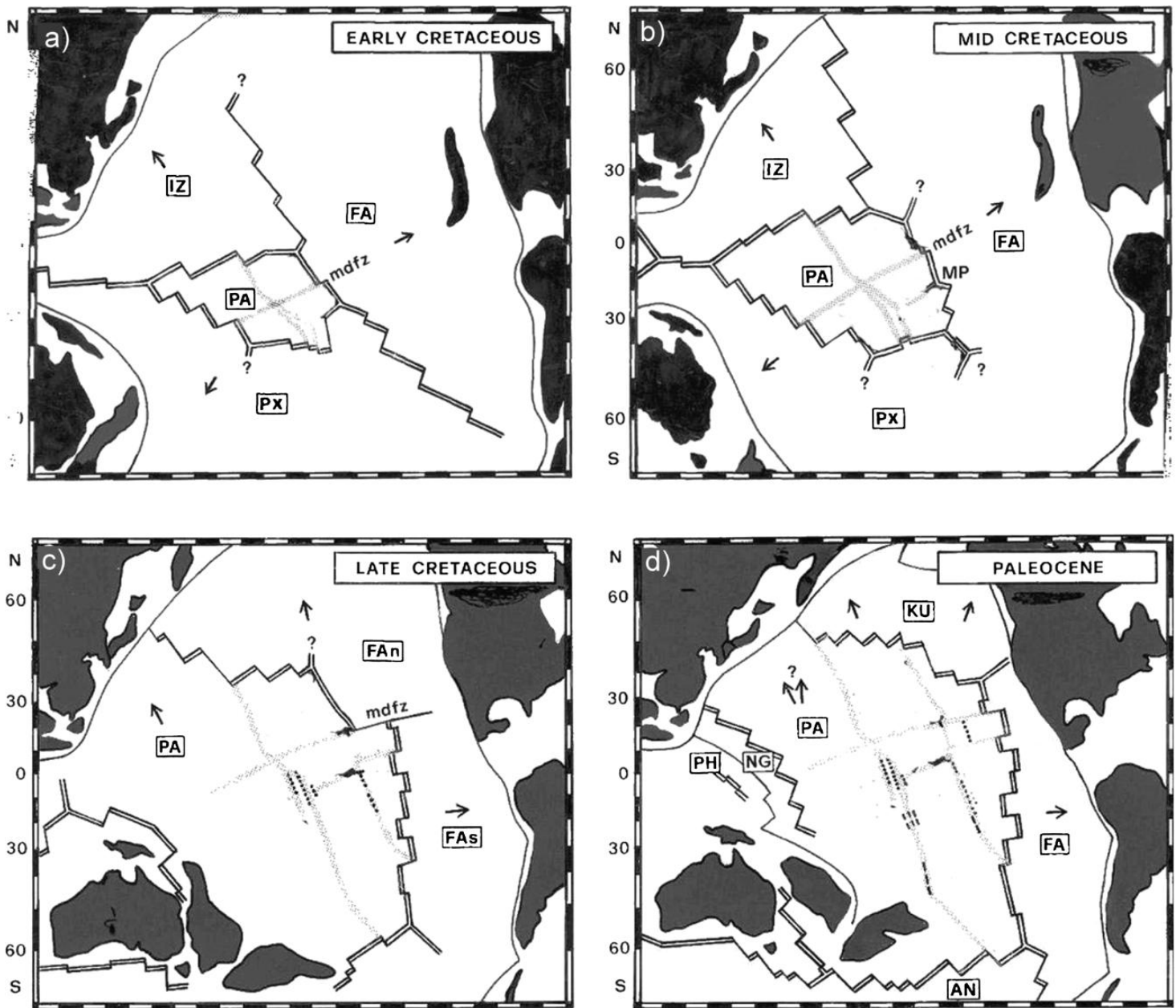
FIGURE 3. Evolution of the Pacific Basin from Early Cretaceous through Paleocene (modified from Smith (2003)) Abbreviations: AN = Antarctic; FA = Farallon (FAn = north, FAs = south); IZ = Izanagi; KU = Kula; NG = North New Guinea; PA = Pacific; PH = Philippine; PX = Phoenix. mdfz = Mendocino fracture 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.
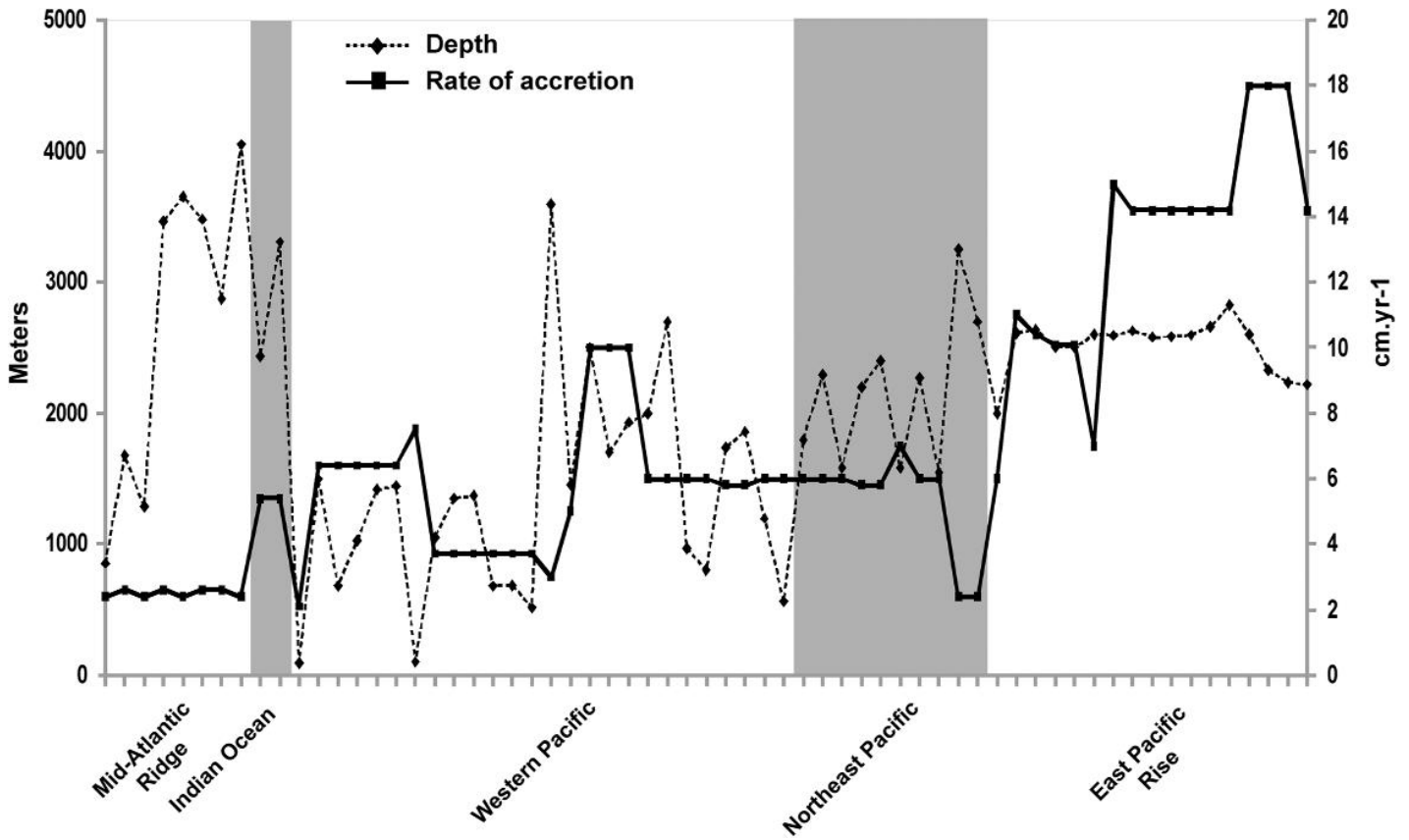
FIGURE 4. Codistribution of depth (dotted line) and rate of accretion (solid line) among the 63 fields. Depth scale is represented on the left y-axis and rate of accretion on the right y-axis. The field order is similar to the one in Table S3 and corresponds to the successive biogeographic provinces, separated by contrasted background on the graph.

Supplementary Materials

Supplementary Figures Captions

Supplementary Figure 1. Illustrations of network properties. a) Example of Clustering coefficient evolution on an undirected graph of 4 nodes. The local clustering coefficient of the black node is given for each graphs according to the links relying its neighbors (grey nodes connected by grey links). The clustering coefficient C is the proportion of existing connections between grey nodes compared with the total number of all possible connections. When C=0, there is no connection between the grey nodes among the 3 potential ones (grey dash lines). For C=1/3, there is only one connection on the 3 possible. When all the grey nodes are interconnected to each other, C=1. b) Example of Betweenness centrality (BC) values on an undirected graph of 13 nodes. The local betweenness centrality value of a node is calculated as the proportions of shortest paths between other node that go through it. Here, the BC values are 0 (BC=0/66), 0.17(BC=11/66) and 0.91(BC=60/66) according to white, grey and black nodes respectively.

Supplementary Figure 2. Province endemism per genera a) and species b). n is the total number of entities observed inside each of the provinces symbolized by a chart. For each of the chart, light grey represents the percentage of endemic entities only observed in one field, grey is the percentage of entities shared inside the province and dark grey corresponds to the percentage of entities observed outside the provinces.

Supplementary Figure 3. Example of classical Multivariate exploratory techniques: Cluster analysis. A dichotimic diagram tree was built from the same distance matrix
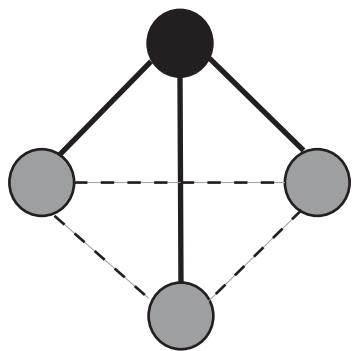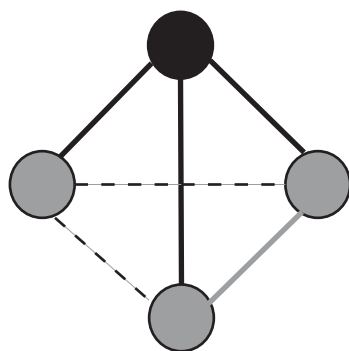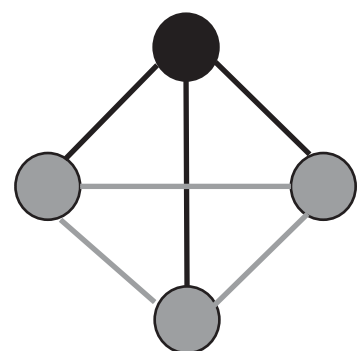
used for network analysis. Each of the 63 fields is plot on the left of the diagram. The linkage distance corresponds to Jaccard's distance. Colored branches match to the biogeographical provinces found with network analysis. Red for the Indian Ocean, yellow for the West-Pacific, brown for the Mid-Atlantic Ridge, green for the Northeast-Pacific and Blue for the East Pacific Rise. Black branches are the one that are connected aside the major cluster of the tree. It can be noted that some of the identified provinces are here "polyphyletic", therefore the difficulties in defining them on the basis of that kind analysis.

Supplementary Figure 4. World-wide network of the hydrothermal vents fauna diversity (Species Level). Network built on the basis of Jaccard distance among fields and represented here at the percolation threshold (Dp=0.95). Circle size illustrates the betweenness centrality values of the corresponding field. The same five provinces found at the genera level are highlighted by this network analysis: the Mid-Atlantic Ridge, the Indian Ocean, the Western-Pacific, the Northeast-Pacific and the East Pacific Rise.
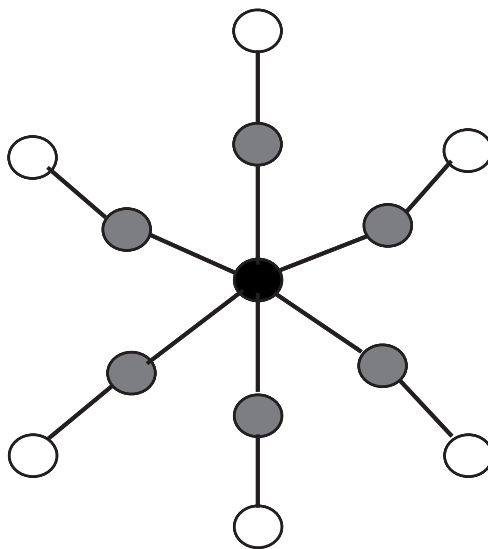
Supplementary Figure 5. Evolution of network topology according to the average cluster size imposed by the Jaccard's distance at the species level. Six thresholds values were selected to visualize networks in function of the average cluster size excluding the largest one. The critical point is the Percolation Threshold. It is located in the transitional region (grey area). It is reached just before all the clusters get disconnected. In this case, the critical value is 0.95.

Supplementary Figure 6. World-wide networks of the three major phylum hydrothermal vents fauna diversity at the Genera Level. a) Annelida, b) Arthropoda and c) Mollusca. Networks built on the basis of Jaccard distance among fields and represented here at their Percolation threshold. Circle size illustrates the betweenness centrality values of the corresponding field.

a)

C=0   C=1/3   C=1

b)

a)

All Fields (n=331)

Mid-Atlantic Ridge (n=82)

Indian Ocean (n=14)

West Pacific (n=118)

Northeast Pacific (n=60)

East Pacific Rise (n=170)

b)

All Fields (n=591)

Mid-Atlantic Ridge (n=103)

Indian Ocean (n=15)

West Pacific (n=183)

Northeast Pacific (n=67)

East Pacific Rise (n=254)

Tree Diagram for 63 HV Fields
Unweighted pair-group average
Dissimilarities from matrix

| Field9 |
| Field10 |
| Field24 |
| Field26 |
| Field29 |
| Field34 |
| Field15 |
| Field18 |
| Field20 |
| Field19 |
| Field22 |
| Field27 |
| Field28 |
| Field33 |
| Field11 |
| Field23 |
| Field13 |
| Field1 |
| Field2 |
| Field3 |
| Field5 |
| Field6 |
| Field7 |
| Field4 |
| Field37 |
| Field41 |
| Field45 |
| Field46 |
| Field38 |
| Field39 |
| Field44 |
| Field42 |
| Field47 |
| Field48 |
| Field51 |
| Field49 |
| Field52 |
| Field50 |
| Field53 |
| Field54 |
| Field56 |
| Field57 |
| Field58 |
| Field60 |
| Field55 |
| Field59 |
| Field35 |
| Field36 |
| Field61 |
| Field62 |
| Field63 |
| Field12 |
| Field14 |
| Field17 |
| Field16 |
| Field25 |
| Field30 |
| Field21 |
| Field43 |
| Field8 |
| Field32 |
| Field31 |
| Field40 |

0.0   0.2   0.4   0.6   0.8   1.0

Linkage Distance

c) threshold : 0.85

d) threshold : 0.90

b) threshold : 0.80

e) Percolation threshold : 0.95

a) threshold : 0.60

f) threshold : 0.99

a) Dp = 0.68

b) Dp = 0.77

c) Dp = 0.75

Table 1 Network properties of identified provinces at the percolation threshold

	Provinces				
	Mid-Atlantic Ridge	Indian Ocean	West Pacific	Northeast Pacific	East Pacific Rise
n	8	2	26	10	17
Cluster size	7	2	17	8	16
BC max/$\langle BC \rangle$	0.14/0.017*	0.21/0.107*[†]	0.28/0.038***	0.16/0.023**	0.13/0.021***
$\langle k \rangle$	5.57	3.00	4.53	3.88	7.81
$\langle CC \rangle$	0.92	-	0.68	0.88	0.81

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, $\langle k \rangle$ is the average connectivity degree, BC max is the maximum value of betweenness centrality and $\langle CC \rangle$ is the average value of clustering coefficient. *p-values *<0.1, **<0.05, ***<0.001.* [†] the mean value for IO is based on only two sampled fields, and is therefore subject to high variance.

used for network analysis. Each of the 63 fields is plot on the left of the diagram. The linkage distance corresponds to Jaccard's distance. Colored branches match to the biogeographical provinces found with network analysis. Red for the Indian Ocean, yellow for the West-Pacific, brown for the Mid-Atlantic Ridge, green for the Northeast-Pacific and Blue for the East Pacific Rise. Black branches are the one that are connected aside the major cluster of the tree. It can be noted that some of the identified provinces are here "polyphyletic", therefore the difficulties in defining them on the basis of that kind analysis.

Supplementary Figure 4. World-wide network of the hydrothermal vents fauna diversity (Species Level). Network built on the basis of Jaccard distance among fields and represented here at the percolation threshold (Dp=0.95). Circle size illustrates the betweenness centrality values of the corresponding field. The same five provinces found at the genera level are highlighted by this network analysis: the Mid-Atlantic Ridge, the Indian Ocean, the Western-Pacific, the Northeast-Pacific and the East Pacific Rise.

Supplementary Figure 5. Evolution of network topology according to the average cluster size imposed by the Jaccard's distance at the species level. Six thresholds values were selected to visualize networks in function of the average cluster size excluding the largest one. The critical point is the Percolation Threshold. It is located in the transitional region (grey area). It is reached just before all the clusters get disconnected. In this case, the critical value is 0.95.

Supplementary Figure 6. World-wide networks of the three major phylum hydrothermal vents fauna diversity at the Genera Level. a) Annelida, b) Arthropoda and c) Mollusca. Networks built on the basis of Jaccard distance among fields and represented here at their Percolation threshold. Circle size illustrates the betweenness centrality values of the corresponding field.

Supplementary Table 1. Number of genera per phylum in each identified provinces

	Provinces					
	Mid-Atlantic Ridge	Indian Ocean	West-Pacific	Northeast Pacific	East-Pacific Rise	All Fields
Arthropoda	40	7	38	17	54	122
Mollusca	17	5	41	20	39	83
Annelida	8	1	23	21	45	67
Chordata	4	0	7	0	8	16
Cnidairia	6	1	2	0	8	15
Porifera	2	0	4	0	5	9
Echinodermata	2	0	1	0	4	6
Nematoda	1	0	2	0	3	5
Foraminiferida	1	0	0	0	1	2
Ciliophora	0	0	0	1	0	1
Nemertea	0	0	0	1	0	1
Acanthocephala	0	0	0	0	1	1
Hemicordata	0	0	0	0	1	1
macnabi	0	0	0	0	1	1
Chaethognata	1	0	0	0	0	1
genera number	82	14	118	60	170	331

Supplementary Table 2. Comparative analyses of different communities detection method.

Title of the method	Description	Results	Reference
Maps of information flow reveal community structure in complex networks	Use of probability flow of random walks on a network as a proxy for information flows in the real system and decompose the network into modules by compressing a description of the probability	3 clusters but congruent with the percolation network topology: Cluster1: MAR-IO-WP Cluster2: NP Cluster3: EPR Preferential connections order: Cluster1-Cluster2> Cluster1-Cluster3> Cluster2-Cluster3	**Rosvall and Bergstrom, 2008[a]**
Fast unfolding of communities in large network	Based on two iterative phases repeat until no increase of modularity: First, modularity optimization. Second, communities aggregation.	1 single giant cluster, unresolved	**Blondel et al., 2008[b]**
Graph Clustering Via a Discrete Uncoupling Process	A Markov Cluster Algorithm that simulates (stochastic) flow in graphs based on expansion and inflation. Eventually, iterating expansion and inflation results in the separation of the graph into different segments (clusters)	Unresolved biogeographic provinces: Cluster1: WP-NP-EPR-MAR Cluster2: IO-WP Cluster3: WP Cluster4: WP	**Dongen, 2008[c]**
Community structure in social and biological networks	Based on edge "betweenness" centrality to find community boundaries	1 single "giant" cluster only resolved when analyzing at the percolation point where the same clusters/provinces were identified as the ones in the simple percolation analysis	**Girvan and Newman, 2002[d]**

[a] Rosvall, M., and C. T. Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci U S A 105:1118-23.

[b] Blondel, V. D., J. L. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics-Theory and Experiment

[c] Dongen, S. V. 2008. Graph Clustering Via a Discrete Uncoupling Process. SIAM Journal on Matrix Analysis and Applications 30:121-141

[d] Girvan, M., and M. E. Newman. 2002. Community structure in social and biological networks. Proc Natl Acad Sci U S A 99:7821-6

Supplementary Table 3. Hydrothermal vents fields used in this study

Field (abbreviation)	Ridge	Latitude deg-min	Longitude deg-min	Field number on map	Depth meter	Rate of accretion cm.yr^{-1}
Menez Gwen (MG)	Mid-Atlantic Ridge	37°51N	31°31W	1	853	2.4
Lucky Strike (LS)	Mid-Atlantic Ridge	37°17N	32°16W	2	1675	2.6
Rainbow (Rb)	Mid-Atlantic Ridge	37°30N	33°54W	3	1285	2.4
Broken Spur (BS)	Mid-Atlantic Ridge	29°10N	43°10W	4	3463	2.6
TAG (TAG)	Mid-Atlantic Ridge	26°08N	44°49W	5	3653	2.4
Snake Pit (SP)	Mid-Atlantic Ridge	23°23N	44°58W	6	3480	2.6
Logatchev (Log)	Mid-Atlantic Ridge	14°45N	44°58W	7	2875	2.6
Ashadze 1 (A1)	Mid-Atlantic Ridge	12°58N	44°52W	8	4050	2.4
Kairei Vent Field (KVF)	Central Indian Ridge	25°19S	70°02E	9	2438	5.4
Edmond Vent Field (EVF)	Central Indian Ridge	23°52S	69°35E	10	3305	5.4
Kagoshima Bay (KB)	Japan Ridge	31°39N	130°48E	11	96	2.1
Hatoma Knoll (HK)	Okinawa Trough	28°24N	123°50E	12	1495	6.4

Minami-Ensei Knoll (MEK)	Okinawa Trough	28°24N	127°38E	13	680	6.4
North Iheya Knoll (NIK)	Okinawa Trough	27°47N	126°54E	14	1030	6.4
Iheya Ridge (IR)	Okinawa Trough	27°33N	126°59E	15	1415	6.4
Izena Cauldron (IC)	Okinawa Trough	27°16N	127°05E	16	1445	6.4
Northeastern Taiwan (NET)	Okinawa Arc	24°50N	121°60E	17	100	7.5
Mokuyo Seamount (MS)	Izu-Ogasawara Arc	28°34N	140°39E	18	1050	3.7
Myojin Knoll (MK)	Izu-Ogasawara Arc	32°07N	139°51E	19	1350	3.7
Suiyo Seamount (SS)	Izu-Ogasawara Arc	28°34N	140°39E	20	1367	3.7
Sumisu Caldera (SC)	Izu-Ogasawara Arc	31°28N	140°04E	21	680	3.7
Kaikata Seamount (KS)	Izu-Ogasawara Arc	26°42N	141°04E	22	685	3.7
Nikko Seamount (NS)	Izu-Ogasawara Arc	23°06N	142°40E	23	515	3.7
Mariana	Mariana Back-Arc	18°12N	144°42E	24	3595	3

Trough (MT)	Basin					
Edison Seamount (EDS)	Tabar-Feni Volcanic Fore Arc	03°01S	152°03E	25	1450	5
Vienna Woods (VW)	Manus Back-Arc Basin	03°09S	150°16E	26	2500	10
Pacmanus Complex (PC)	Manus Back-Arc Basin	03°43S	151°40E	27	1700	10
Desmos Cauldron (DC)	Manus Back-Arc Basin	03°41S	151°52E	28	1930	10
White Lady (WL)	North-Fiji Back-Arc Basin	16°59S	173°55E	29	2000	6
Mussel Valley (MV)	North-Fiji Back-Arc Basin	18°49S	173°29E	30	2700	6
Loihi Seamount (LOS)	Intra-Plate Seamount	18°57N	155°16W	31	969	6
Vailulu'u Seamount (VS)	Intra-Plate Seamount	14°13S	169°04W	32	800	6
Vaï Lili (VL)	Lau Back-Arc Basin	22°13S	176°37W	33	1736	5.8
Hine Hina (HH)	Lau Back-Arc Basin	22°32S	176°43W	34	1860	5.8
Brothers Seamount	Kermadec Arc	34°52S	179°04E	35	1197	6

(BRS)						
Rumble V Seamount (RVS)	Kermadec Arc	36°08S	178°11E	36	561	6
Magic Mountain (MM)	Explorer	49°46N	130°15W	37	1797	6
Middle Valley (MIV)	Juan de Fuca	48°27N	128°42W	38	2291	6
Endeavour Segment (ES)	Juan de Fuca	48°00N	129°04W	39	1580	6
Coaxial Segment (CAS)	Juan de Fuca	46°19N	129°42W	40	2200	5.8
Axial-Volcano (AV)	Juan de Fuca	45°55N	129°59W	41	2400	5.8
Axial Volcano Coaxial Seamount (AVCASM)	Juan de Fuca	45°59N	130°02W	42	1580	7
Axial Volcano Ashes Vent Field (AVAVF)	Juan de Fuca	45°56N	130°00W	43	2270	6
Cleft Segment (CS)	Juan de Fuca	44°39N	130°15W	44	1547	6
Sea Cliff Vent Field (SCVF)	Gorda	42°15N	126°42W	45	3250	2.4

Escabana Trough (ET)	Gorda	41°00N	127°29W	46	2700	2.4
Guaymas Basin (GB)	Gulf of California	27°00N	111°24W	47	2000	6
21°N (21N)	Northern East- Pacific Rise	20°50N	109°06W	48	2615	11
13°N (13N)	Northern East- Pacific Rise	12°50N	103°57W	49	2635	10.4
11°N (11N)	Northern East- Pacific Rise	11°23N	103°46W	50	2504	10.1
9°N (9N)	Northern East- Pacific Rise	09°49N	104°17W	51	2500	10.1
Galapagos Spreading Center (GSC)	Galapagos	00°48N	86°10W	52	2601	7
7°S (7S)	Southern East- Pacific Rise	07°21S	107°47W	53	2594	15
14°S (14S)	Southern East- Pacific Rise	13°58S	112°28W	54	2629	14.2
17°S (17S)	Southern East- Pacific Rise	17°24S	113°12W	55	2580	14.2
17°24S (1724S)	Southern East- Pacific Rise	17°25S	113°12W	56	2583	14.2

17°34S (1734S)	Southern East- Pacific Rise	17°34S	113°14W	57	2595	14.2
18°S (18S)	Southern East- Pacific Rise	18°24S	113°24W	58	2658	14.2
21°S (21S)	Southern East- Pacific Rise	21°25S	114°17W	59	2825	14.2
23°S (23S)	Southern East- Pacific Rise	23°32S	115°34W	60	2598	18
31°S (31S)	Southern East- Pacific Rise	31°09S	111°55W	61	2330	18
Saguaro Field (SF)	Southern East- Pacific Rise	31°51S	112°02W	62	2235	18
German Flats (GF)	Southern East- Pacific Rise	37°47S	110°55W	63	2216	14.2

Supplementary Table 4. Key fields in relaying information

	Provinces								
	Mid-Atlantic Ridge	Indian Ocean	West Pacific			Northeast Pacific	East-Pacific Rise		
Key Fields	3	9	24	27	29	41	50	51	55
BC_value/	0.14/	0.21/	0.28/	0.26/	0.14/	0.16/	0.07/	0.05/	0.13/
Provinces_connection	IO	MA&WP	IO&EPR	IO&NP	IO	WP	WP	WP	WP

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