# Subtle limits to connectivity revealed by outlier loci within two divergent metapopulations of the deep-sea hydrothermal gastropod Ifremeria nautilei

Adrien Tran Y. Lu, Stephanie Ruault, Claire Daguin-Thiébaut, Jade Castel, Nicolas Bierne, Thomas Broquet, Patrick Wincker, Aude Perdereau, Sophie Arnaud-Haond, Pierre-Alexandre Gagnaire, et al.

▶ **To cite this version:**

Adrien Tran Y. Lu, Stephanie Ruault, Claire Daguin-Thiébaut, Jade Castel, Nicolas Bierne, et al.. Subtle limits to connectivity revealed by outlier loci within two divergent metapopulations of the deep-sea hydrothermal gastropod Ifremeria nautilei. Molecular Ecology, 2022, 31 (10), pp.2796-2813. 10.1111/mec.16430 . hal-03669635

**HAL Id: hal-03669635**
**https://hal.umontpellier.fr/hal-03669635v1**

Submitted on 28 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Subtle limits to connectivity revealed by outlier loci within two divergent metapopulations of the deep-sea hydrothermal gastropod *Ifremeria nautilei*

**Adrien Tran Lu Y[1*], Stéphanie Ruault[2], Claire Daguin-Thiébaut[2], Jade Castel[2], Nicolas Bierne[1], Thomas Broquet[2], Patrick Wincker[5], Aude Perdereau[5], Sophie Arnaud-Haond[3], Pierre-Alexandre Gagnaire[1], Didier Jollivet[2], Stéphane Hourdez[4], François Bonhomme[1]**

1:ISEM, Institut des Sciences de l'Evolution, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France,

2: Sorbonne Université, CNRS, UMR 7144, 'Dynamique de la Diversité Marine' (DyDiv) Lab, Station biologique de Roscoff, Place G. Teissier, 29680 Roscoff, France

3: MARBEC, Marine Biodiversity Exploitation and Conservation, Univ Montpellier, CNRS, IFREMER, IRD, Sète, France

4: Sorbonne Université, CNRS, UMR 8222, Laboratoire d'Ecogéochimie des Environnements Benthiques, Observatoire Océanologique de Banyuls, Avenue Pierre Fabre, 66650, Banyuls-sur-Mer, France

5: Génomique Métabolique, Génoscope, Institut de Biologie François Jacob, CEA, CNRS, Université Évry, Université Paris-Saclay, Évry, France

***Corresponding author**: adrien.tran-lu-y@umontpellier.fr

## Keywords:

**Genetic connectivity, Demographic inference, ddRAD-seq, Hydrothermal vents, Western Pacific, Outlier detection**

**Abstract**. Hydrothermal vents form archipelagos of ephemeral deep-sea habitats that raise interesting questions about the evolution and dynamics of the associated endemic fauna, constantly subject to extinction-recolonization processes. These metal-rich environments are coveted for the mineral resources they harbor, thus raising recent conservation concerns. The evolutionary fate and demographic resilience of hydrothermal species strongly depend on the degree of connectivity among and within their fragmented metapopulations. In the deep sea, however, assessing connectivity is difficult and usually requires indirect genetic approaches. Improved detection of fine-scale genetic connectivity is now possible based on genome-wide screening for genetic differentiation.

Here, we explored population connectivity in the hydrothermal vent snail *Ifremeria nautilei* across its species range encompassing five distinct back-arc basins in the Southwest Pacific. The global analysis, based on 10 570 single nucleotide polymorphism (SNP) markers derived from double digest restriction-site associated DNA sequencing (ddRAD-seq), depicted two semi-isolated and homogeneous genetic clusters. Demo-genetic modeling suggests that these two groups began to diverge about 70 000 generations ago, but continue to exhibit weak and slightly asymmetrical gene flow. Furthermore, a careful analysis of outlier loci showed subtle limitations to connectivity between neighboring basins within both groups. This finding indicates that migration is not strong enough to totally counterbalance drift or local selection, hence questioning the potential for demographic resilience at this latter geographical scale. These results illustrate the potential of large genomic datasets to understand fine-scale connectivity patterns in hydrothermal vents and the deep sea.

# Introduction

Understanding the connectivity of populations is a central issue for evolutionary ecology, conservation and management (Cayuela et al., 2018). Direct approaches such as population monitoring or mark-recapture experiments are rarely applicable in marine environments, because many marine species have large population sizes and high dispersal capabilities due to their minute pelagic propagules. These characteristics are likely to reduce the ability of population genetics to assess population connectivity at local and regional scales, except in situations where there is sufficient genetic differentiation or where a large fraction of the population can be sampled (Jones et al., 2005; Pinsky et al., 2010). Deep-sea hydrothermal ecosystems have attracted much attention since their discovery in the late 1970s (Dover et al., 2001; Tunnicliffe et al., 1998) due to the oasis-like distribution of these unique and chaotic habitats harboring rich and endemic fauna. Hydrothermal environments are mostly found in tectonically active areas, such as mid-oceanic ridges, where neighboring vents are often separated by tens of meters to hundreds of kilometers, resulting in an almost linear, but fragmented and unstable distribution of vent communities (Chevaldonné et al., 1997; Hannington et al., 2011; Jollivet et al., 1999; Vrijenhoek, 2010).

Studies of slow (*e.g.* Mid-Atlantic Ridge) and fast (*e.g.* East Pacific Rise) spreading ridges have shown that number of species are able to maintain gene flow over thousands of kilometers (Breusing et al., 2016; Craddock et al., 1995; Hurtado et al., 2004; Jollivet et al., 1995; Teixeira et al., 2011, 2012; Won et al., 2003; Yahagi et al., 2019). Many vent invertebrates possess long-range planktonic larvae that can rapidly (re)colonize newly formed sites (Mullineaux et al., 2010). This high larval dispersal capacity leads to local colonization processes following a stepping-stone mechanism of exchanges, or to the formation of more complex metapopulation

74 dynamics where local extinctions and migration may vary greatly according to the geotectonic
75 context of the venting sites (Audzijonyte & Vrijenhoek, 2010; Jollivet et al., 1999; Vrijenhoek,
76 1997, 2010). Population connectivity can, however, be severely hampered by physical barriers
77 to larval dispersal such as transform faults, diverging ocean currents or microplates (Johnson et
78 al., 2006; Plouviez et al., 2009; Plouviez, Schultz, et al., 2013).

79 The hydrothermal ecosystems found in the Southwest Pacific are mainly associated with the
80 formation of back-arc basins (BABs). BABs result from complex subduction processes between
81 several plates, leading to a discontinuous and nonlinear distribution of venting sites. Hence, the
82 question arises as to the degree of connectivity between populations inhabiting these BABs,
83 noticeably to address the issue of their resilience with respect to deep-sea mining projects
84 (Gena, 2013; Niner et al., 2018). In this context, only a few studies to date have focused on
85 understanding the general patterns of spatial genetic connectivity in ecologically vulnerable
86 hydrothermal species. For instance, Thaler et al. (2011) showed that the gastropod *Ifremeria*
87 *nautilei* is genetically differentiated between the Manus and North-Fiji/Lau basins. Similar
88 results have been reported in other species, such as the limpet *Lepetodrilus schrolli* (Plouviez
89 et al., 2019), the shrimp *Rimicaris variabilis* and the squat lobster *Munidopsis lauensis* (Thaler
90 et al., 2014). Moreover, this latter species is characterized by additional intra-basin structuring.
91 In contrast, the limpet *Shinkailepas tollmani* does not show any differentiation at any of these
92 scales (Yahagi et al., 2020). However, due to the use of only a limited number of markers, none
93 of these studies have reached the resolution necessary for the fine-scale assessment of
94 connectivity in these species.

95 During the last decade, the development of next-generation sequencing (NGS) and associated
96 techniques have increased the quantity and accessibility of population genomic data,
97 particularly in non-model species. Analyzing these large datasets with thousands of markers

along the entire genome using demo-genetic inference methods helps reveal the complex demographic histories of species (Excoffier et al., 2013; Feutry et al., 2020; Gutenkunst et al., 2009; Rougeux et al., 2017; Tine et al., 2014). NGS datasets also give access to unprecedented statistical power to detect non-neutral genetic variation (outlier loci) that can potentially provide finer scale spatial information on connectivity, dispersal and possibly local adaptation (Gagnaire et al., 2015; Milano et al., 2014; Wyngaarden et al., 2017). This information can potentially help distinguish situations of genetic connectivity—whereby local populations are demographically independent (*i.e.* mainly replenished by local propagules) but long-range gene flow mediated by a small number of propagules is sufficient to ensure the circulation of genetic variation among them—from situations of demographic connectivity where a substantial fraction of local population size is made up of immigrants (Lowe & Allendorf, 2010). The consequences of this difference in population connectivity in terms of resilience to local extinction or habitat destruction are quite obvious, with prompt recolonization being expected only in the second situation.

The aim of the present study was to elucidate fine-scale population structure and connectivity using high-throughput double-digest restriction-site associated DNA (ddRAD) sequencing on *Ifremeria nautilei*, whose known distribution covers the Southwest Pacific, from the Manus BAB in Papua New Guinea to the Lau BAB off the Tonga Islands. This hydrothermal gastropod of family Provannidae harbors chemoautotrophic symbiotic bacteria in its gills to produce organic matter and forms dense aggregations around diffuse fluid venting at temperatures lower than 15°C (Borowski et al., 2002; Windoffer & Giere, 1997). The species is gonochoric with a nearly equal sex ratio and females brood their progeny in a metapodial pouch until the lecithotrophic embryos (several thousands of similar size) reach a specific and unique pre-veliger stage known as Warèn's larva (Reynolds et al., 2010; Warèn & Bouchet, 1993) . The

gastropod reproduces via internal fertilization leading to a patchwork of brooding and nonbrooding females throughout the year due to asynchronous spawning. Brooding is a reproductive trait that usually limits the dispersal ability of species; however, the lifespan of *I. nautilei* pelagic larvae is not known, nor is the maximum distance veligers can travel prior to settlement on a new venting site. Constituting a large portion of the biomass and harboring other species such as *S. tollmanni* and *L. schrolli*, *I. nautilei* is a keystone species important for these deep-sea ecosystems. Furthermore, it is classified as endangered by the IUCN (https://www.iucnredlist.org/species/145380421/145380604) along with some other hydrothermal species. Therefore, studying the connectivity patterns of *I. nautilei* is a flagship endeavor to assess the potential impact of deep-sea mining on this keystone species and its associated fauna.

For this study, extensive sampling was carried out on 29 sites in 17 hydrothermal vent fields across five basins distributed over 5000 km in the Southwest Pacific.

# Materials and Methods

*Sample collection and DNA extraction*

A hierarchical sampling plan was deployed on board the French oceanographic vessel RV *L'Atalante* during the ChubacArc 2019 oceanographic cruise using the remotely operated underwater vehicle (ROV) *Victor 6000*. A total of 684 individuals were collected in the Southwestern Pacific Ocean from 29 sampling sites distributed among 17 vent localities or hydrothermal fields across four BABs and one volcanic submarine area (Futuna), spanning the entire known geographical distribution range of *I. nautilei* (Figure 1, SI Table S1). This

sampling includes samples from the newly discovered active site La Scala in the Woodlark basin (Boulart et al., in press)

In addition, 27 individuals were added from collections of previous oceanographic cruises, with 22 samples obtained by S. Hourdez during the Lau basin 2009 oceanographic cruise with few at the now-extinct Kilo Moana site and 5 samples obtained during the Manus basin 2009 cruise kindly provided by C. L. Van Dover (SI Table 1). Altogether, a total of 456 unique samples were used and analyzed in this study, of which 362 remained after filtering the sequence dataset (SI Table 2).

Once on board, the snails were dissected and various tissues were preserved in EtOH or frozen at -80°C. Genomic DNA was extracted from fresh foot tissue to limit DNA contamination by symbionts hosted in the gills. A fraction of the tissue samples was stored in 90% EtOH for backup. Samples from C. L. Van Dover's collection were preserved in 90% EtOH, and those from S. Hourdez were kept frozen at -80°C. Genomic DNA was extracted using the NucleoSpin Tissue kit according to the manufacturer's protocol (Macherey-Nagel); some samples were extracted using a standard phenol-chloroform method.

*Preparation and sequencing of ddRAD libraries.*

DNA extracts were visualized on 0.8% agarose gels and each concentration was standardized to between 10 and 50 ng.µl$^{-1}$ after a fluorometric quantification with the QuantiFluor dsDNA system (Promega). Individual double-digest restriction-site associated DNA (ddRAD) *Pst*I-*Mse*I libraries were prepared following (Brelsford et al., 2016) after modifications detailed in (Thiébaut et al., 2021). Five pooled libraries were prepared with a combination of four to eight Illumina indexes and 24 barcodes per index, multiplexing a total of 486 samples, including 27 pairs of replicates for quality control, representing 456 individuals. The sequencing effort was

sized to produce on average $3\times10^6$ read pairs per individual. Each genomic pool was sequenced on one lane of a HiSeq4000 Illumina sequencer (paired end, 150 bp) at the Genoscope sequencing facility (Centre National de Séquençage, Evry, France).

*De novo ddRAD-tag assembly, SNP calling and filtering*

Fastqc (V.0.11.9) was used only for quality control of raw reads, no filters were applied on them. Individual reads were demultiplexed using the "Processradtag" pipeline in Stacks (V.2.52) (Rochette et al., 2019). Due to the lack of a reference genome for *I. nautilei*, reads were assembled *de novo* using each Stacks module one by one (ustacks, cstacks, sstacks, tsv2bam, gstacks and populations). To identify the most appropriate assembly parameters, we followed previously published recommendations (Mastretta-Yanes et al., 2015; Paris et al., 2017) (See SI and SI Figures 1-5 for details). Briefly, we used the genotyping error rate between replicates, the number of variants (SNP), polymorphic loci (ddRAD-tags) and nucleotide diversity ($\pi$ estimated in Stacks) as a function of the assembly parameters (m and M = n) determined with a subset of individuals covering all basins and localities (n = 84). The selected parameters were as follows: m = 4 (the minimum number of reads to assemble a stack), M = 8 (the maximum number of mismatches between putative alleles within individuals), n = 8 (the maximum number of mismatches between putative loci within the catalog of individuals) and R = 0.8 (the minimum percentage of individuals sharing a locus across all populations in the "populations" module).

After *de novo* assembly, several filters were applied using VCFtools (V.0.1.16) (Danecek et al., 2011) to reduce missing data and to account for potential paralogs (see SI Table 3). Briefly, we removed 1 of each of the 27 replicated individuals with the highest value of missing data. Then, we excluded SNPs with heterozygosity > 0.6, SNP and only individuals with less than 10%

missing data were kept. Variants with a mean coverage higher than 80X were excluded. Using VCFtools, we excluded loci with a minor allele frequency (MAF) lower than 5% (alternative allele), followed by those that deviated significantly from Hardy-Weinberg equilibrium (*p*-value $\leq 0.05$). We then kept only one randomly chosen SNP per ddRAD-tag to avoid short distance linkage disequilibrium between SNPs. PGDSpider (V.2.1.1.5) (Lischer & Excoffier, 2012) was used to convert the final VCF into the formats required for subsequent analyses.

*Population structure and diversity*

Principal component analysis (PCA) was first performed on the final dataset with the R package SNPrelate (V.1.24.0) (Zheng et al., 2012) . Pairwise fixation indices ($F_{ST}$) were calculated in Arlequin (V.3.5.2.2) (Excoffier & Lischer, 2010). AMOVA (Excoffier et al., 1992) was performed with 10 000 permutations of genotypes between populations by considering hierarchical geographic structure of localities within basins (See SI for parameters). Co-ancestry analyses were performed through ADMIXTURE (V.1.3.0) (Alexander & Lange, 2011) with 10 independent runs for $K = 1$ to 5. The best $K$ value was selected by using the cross-validation error as recommended by the authors. Runs of ADMIXTURE were grouped using CLUMPAK (Kopelman et al., 2015), graphical visualizations of the results were plotted using library ggplot2 (V.3.3.3) in R (V.4.0.1). TreeMix (V.1.13) (Pickrell & Pritchard, 2012) was performed with 10 independent runs with migration events ranging from 0 to 5. The optimal number was selected according to the log-likelihood of each model. $F_3$ admixture tests (Reich et al., 2009) were done using the THREEPOP programs implemented in TreeMix (V.1.13) package with default values.

To detect potential kinship, SNPrelate was used to compute identity-by-state between pairs of individuals. We used this approach to (1) minimize the risk of labeling error/exchange during

the process of library construction and sequencing and (2) infer the level of kinship structure between non-replicated individuals because the existence of undetected underlying kinship structure can distort the population structure estimated by the gene genealogies.

Observed heterozygosity ($H_o$), expected heterozygosity ($H_e$), heterozygote deficiency ($F_{IS}$), nucleotide diversity ($\pi$) and raw nucleotide divergence ($D_{xy}$) were calculated with the population module of Stacks using all sites from all ddRAD-tags in the final dataset.

*Demo-genetic history of the species*

The demographic history of the species targeting past and present gene flow between metapopulation clusters was inferred using a modified version of ∂a∂i (V.2.1) (Diffusion Approximations for Demographic Inference; Gutenkunst et al., 2009), with a dual annealing optimization function. This software simulates the joint allele frequency spectrum (JAFS) of two (or more) interacting populations according to different demo-genetic scenarios. Here, we considered 28 distinct scenarios built according to the population models used in Rougeux et al. (2017) with very few modifications, detailed below.

Basically, all these models derive from four basic models representing strict isolation (SI), isolation with migration (IM), ancient migration (AM), and secondary contact (SC). Each of them consists of an ancestral population of $N_{anc}$ size that splits into two sister populations of effective size $N_1$ and $N_2$ during time Ts for the (SI) and Tsm for the (IM) model, Tam+Ts for the (AM) model and Ts+Tsc for the (SC) model, where Ts is the time spent since the split of the two populations without migration, Tsm, the time spent since the split of the two populations with migration, Tam, the duration of the ancient migration period after the split of the ancestral population and before the emergence of strict isolation (Ts) and Tsc the duration of a secondary

contact after a period Ts of strict isolation. Directional migration between populations is allowed at rates $m_{12}$ and $m_{21}$ from population 2 to population 1 and vice versa.

Further complexity was introduced as in Rougeux et al. (2017), by adding several processes occurring after the split, such as population expansion or contraction (G), the effect of linked selection reducing the effective population size of loci over a certain fraction of the genome (2N) and the effect of semipermeable genetic barriers (*i.e.* partial reproductive isolation) reducing the effective migration rate of loci over a certain fraction of the genome (2m). Furthermore, to dissociate the effect of the effective population size (genetic drift) and migration (gene flow), we only allowed the growth parameters (G) to vary during the migration phase of each model. Graphical representation of the four basic models and the three more complex models are displayed in SI Figure 7.

For the input dataset, we considered the two metapopulations defined by the global analyses (see Results), which corresponded to Manus/Woodlark and North-Fiji/Futuna/Lau. We used the folded joint allele frequency spectrum (folded JAFS), because no external group was available to identify the allelic ancestral states. All models were fitted independently of the dataset using dual-annealing optimization and run 10 times independently each to check convergence. Model comparisons were based on the Akaike information criterion (AIC). Using the best selected models, we then converted demographic parameters into biological units. In the absence of precise information on mutation rate and generation time for this species, we used $10^{-8}$ as the mutation rate per site per generation. This widely used value falls within the range proposed by (Lynch, 2010), although admittedly the real value may be much larger or much smaller, as recalled in the Discussion. Parameters estimated using ∂a∂i are scaled from the ancestral effective population size ($N_{anc}$), which was estimated using the following formula:

$$N_{anc} = \frac{\theta}{(4 \times \mu \times L)}$$

where $L$ represents the total length of the DNA sequence used in ∂a∂i:

$$L = \frac{z \times y \times 275}{x}$$

where $z$ represents the number of SNPs used, $y$ the number of RAD-tags of 275 bp, and $x$ the initial number of SNPs ($z$ = 17 365, $y$ = 17 365 and $x$ = 250 502, $L$ = 331 032).

Estimated times were calculated in units of $2*N_{anc}$ generations and the migration parameters ($m_{12}$ and $m_{21}$) were divided by $2*N_{anc}$ to obtain the number of migrants in each population per generation. The standard deviations were estimated using the Fisher information matrix (FIM) method implemented in ∂a∂i.

*Outlier loci and detection of fine-scale structure*

To test whether fine-scale genetic structure exists within each genetic cluster defined as a result of the global analyses described in the preceding paragraph (see Results for details), we used several genome-scanning methods to identify candidate outlier SNPs (*i.e.* loci showing higher or lower levels of differentiation than expected under assumed neutrality). Such loci may be informative about fine-scale population structure and connectivity patterns (Gagnaire et al., 2015). Four different outlier detection approaches were used. The rationale behind this multiple testing is that these methods operate with somewhat different underlying assumptions or test statistics and are known to have varying discriminatory power depending on the situations to which they are applied (Villemereuil et al., 2014). Outlier loci were selected according to statistical thresholds ($p$-value ≤ 0.05 and 0.01) in each software package, while checking that candidate loci outnumbered the number of loci expected to fall outside the distribution by chance only (false positives). Furthermore, to focus on the relevant scale and avoid the detection

of false positives due strong geographic structuring, these programs were run independently on each Manus/Woodlark and North-Fiji/Futuna/Lau metapopulation previously defined in the global analyses, while considering populations either by basin or by locality within these groups.

Four methods were used. (1) BayeScan (V.2.0) (Foll & Gaggiotti, 2008) detects potential outlier loci by using differences in allele frequencies under a simple island model in a Bayesian framework. Five independent runs were performed with the default settings. (2) PCAdapt (V.4.3.3) (Luu et al., 2017) uses the correlation of SNPs with the first principal components of the PCA to detect outliers by computing a Mahalanobis distance between their z-score on each PC. (3) Arlequin (V.3.5.2.2) (Excoffier & Lischer, 2010) detects outlier SNPs under a non-hierarchical finite island model integrating $F_{ST}$ and heterozygosity through 20 000 coalescence simulations of the neutral distribution with 100 demes each. (4) The core model of Baypass (V.2.1) (Gautier, 2015) based on a hierarchical Bayesian model in which loci that are more differentiated than expected under a non-equilibrium drift model are identified through the distribution of a statistic similar to $F_{ST}$ corrected to account for demographic history. Baypass was run five times independently with default settings under the core model.

PCA, ADMIXTURE and $F_3$ tests were then performed on the various outlier subsets to explore the information they convey.

In addition, outlier loci identified at the threshold ( $p \leq 0.05$) were first blasted (BLASTN, E-value threshold: $10^{-5}$) against the *Alviniconcha boucheti* transcriptome, which was previously assembled using rnaSPAdes (V.3.13.1) (Bankevich et al., 2012) (cf. Castel et al., in prep). Transcript hits with a size greater than 300 bp were subsequently blasted (BLASTX, E-value

threshold: $10^{-5}$) against the NCBI UniProtKB/Swiss-Prot database using the software Geneious Prime® 2021.2.2.

# Results

De novo *assembly and data filtering*

*De novo* assembly resulted in a dataset of 38 608 ddRAD-tags with a mean coverage of 14X for 486 samples. The mean genotyping error rate was 0.48% and the maximum value was 1.06% from all pairs of replicates. These ddRAD-tags contained 649 106 SNPs. Following SNP filtering, the final dataset resulted in a VCF file containing 362 individuals with 10 570 unlinked bi-allelic variants with an individual mean coverage of 17.7X and a maximum of 10% of missing data per individual and variant.

*Population structure analyses considering the global dataset*

A PCA was performed to explore the level of population structure over the five western Pacific BABs (Figure 2, A). This analysis showed a very clear geographical separation with two distinct clusters, one corresponding to the Manus and Woodlark basins and the other to the North-Fiji, Futuna and Lau basins. The first component (PC1) explained 26.03% of the total variance; the second one carried only 0.03% of the total variance (Figure 2 and SI Figure 8). This pattern was consistent with the AMOVA results (Table 1), which also showed a strong and significant genetic differentiation between Manus/Woodlark and North-Fiji/Futuna/Lau, but no differentiation among basins and localities within these two groups (between Manus/Woodlark and North-Fiji/Futuna/Lau, $F_{ST}$ = 0.387, $p$-value < 0.001, inter-basins within Manus/Woodlark and North-Fiji/Futuna/Lau, $F_{CT}$ = -0.050, NS). In addition, the between-basins pairwise $F_{ST}$

(Table 2) were only significant between Manus/Woodlark and North-Fiji/Futuna/Lau pairs. No significant pairwise $F_{ST}$ values were observed between localities within either of the two groups Manus/Woodlark and North-Fiji/Futuna/Lau (SI Table 3).

This finding is also consistent with the ADMIXTURE (Figure 2 B & SI Figures 9–10) clustering results and strongly supports the same two distinct clusters ($K$ = 2) with very few individuals showing very low percentages of mixed ancestry (from 0.1% to 3%). The identity-by-state distribution (SI Figure 11) did not show evidence of any internal structure due to kinship.

TreeMix analyses showed an optimal number of two migration events, whereas additional events did not improve the likelihood (SI Figure 12 A). Displaying the first migration event showed a very low migration weight from Manus/Woodlark towards North-Fiji (SI Figure 12 B), and adding the second migration event indicated a very slight differentiation between Woodlark and Manus (SI Figure 12 C). The $F_3$ statistics showed a significant admixture signal, with source populations from each genetic cluster only when North-Fiji was chosen as the focal population (SI Figure 13).

The estimated genetic diversity of the populations considering all DNA positions of the 10 570 ddRAD-tags was bimodal, with slightly higher nucleotide diversity in Manus/Woodlark ($\pi$ = 0.00572) compared with North-Fiji/Futuna/Lau ($\pi$ = 0.00535), regardless of the populations being considered by geographic basin or by genetic cluster (SI Figure 14). The raw nucleotide divergence ($D_{xy}$) between the two genetic clusters was estimated to be 0.0136.

Hence, the analyses of the complete SNP dataset of *I. nautilei* indicate the co-occurrence of two quasi-panmictic metapopulations, one associated with the Manus/Woodlark basins and the other with the North-Fiji/Futuna/Lau basins, on either side of the Solomon Islands/New

Hebrides arc. Thus, these two metapopulations are both sufficiently homogeneous and differentiated from each other to be analyzed using ∂a∂i demo-genetic inference, which aims at summarizing the global genome-wide history of divergence/contact between them over a long period of time.

*Inference of demographic history and gene flow*

*1.Model comparisons*

The folded JAFS in Figure 3 (A) shows how allele frequencies are shared between the Manus/Woodlark and North-Fiji/Futuna/Lau metapopulations. The ∂a∂i framework can fit population models on the observed dataset and compares them based on their AIC values (Figure 4).

Among the four simplest models (SI, AM, IM and SC), SC was significantly the best fitting model. Increasing complexity by adding the parameters G, 2m, 2N independently improved the AIC values regardless of the basic model used. However, capturing the effect of linked selection (2N) explained the data much better than models with population growth (G) and heterogeneous gene flow (2m) only (Figure 4).

Conversely, the combination of these parameters led to only a slight improvement in the AIC values. Nevertheless, models including the effect of linked selection (2N) were still better than the other models (Figure 4).

Hence, considering all models together, the best ones were those that took all parameters (2N, 2m, G) into consideration, followed by models with only 2N+G (Figure 4). Moreover, for the 2N+2m and 2N+2m+G models, the proportion 1-P of the genome that evolves under restricted migration in 2m models amounted to 0.52–0.56 for the best AIC simulation among all runs, meaning that the proportion of the genome that evolves under a reduced effective migration

rate (barrier loci) may be quite substantial. With the increasing number of population parameters, the secondary contact hypothesis was no longer the best evolutionary scenario explaining our genetic dataset: the models IM+2N+2m+G, SC+2N+2m+G, AM+2N+2m+G and AM+2N+G, IM+2N+G, SC+2N+G showed very similar AICs ($\Delta_{AIC} \leq 10$, Figure 4).

*2.Inferences of model parameter values*

According to the best models based on AIC, the two metapopulations may have diverged due to early (AM), late (SC) or constant (IM) gene flow and it is difficult to distinguish among these three possibilities. However, these models have some interesting features in common. First, although the standard deviations (SDs) are rather large, the effective population sizes of the two derived populations estimated since the split ($N_1$ & $N_2$) indicate a demographic expansion (b1 and b2 > 1), regardless of the model, including a temporal change in population size (G). Second, the local effect of selection at linked sites seems to affect a very large proportion of loci (Q = 0.99) with a small value of *hrf* (Hill-Roberston factor = ~ 0.02) (Table 3). Third, the number of contemporary migrants (estimated by $(N_1*b1*m_{12})/2$ and $(N_2*b2*m_{21})/2$) shows asymmetrical, but weak flow between the two metapopulations, slightly higher from North-Fiji/Futuna/Lau to Manus/Woodlark (4.2–4.6) than in the opposite direction (2.9–3.2). Fourth, nearly half of loci show a restricted migration rate ($0.52 \leq 1\text{-}P \leq 0.56$). Fifth, the total estimated divergence time expressed as the number of generations are very similar, regardless of the model (Tsm, Tam+Ts, Ts+Tsc) and estimated to be between 66 951 and 70 295 generations.

*Outlier loci and detection of fine-scale structure*

Despite the absence of geographic structuring within each metapopulation depicted in the global analysis, several outlier loci were identified in each metapopulation at the thresholds of $p \leq 0.05$ and $p \leq 0.01$ (SI Table 4).

BayeScan identified much fewer candidate outliers than expected by chance only, and hence was not considered further. PCAdapt was largely out of its working range because it searches for loci that exceed the possible differentiation level captured by the very first principal component as opposed to the second-order axes. However, all axes, except axis 1 which separates the two geographic metapopulations, belong to this second category due to the lack of internal structure. Hence, all the second-order axes primarily captured noise, and they were unable to reveal additional structuring (SI Figure 15,16). Therefore, Baypass and Arlequin were the only two methods considered further. To increase the probability of considering true positives only, we only kept loci identified in both approaches (predicted by the intersection depicted in the Venn diagram in Figure 5 at the thresholds $p \leq 0.05$ and $p \leq 0.01$ in SI Figure 17). Only 458 and 223 outlier loci were shared between the methods Baypass and Arlequin in Manus/Woodlark and North-Fiji/Futuna/Lau at the threshold of $p \leq 0.05$ respectively.

PCA based on these different sets of outlier loci helped to visualize their contribution to the internal heterogeneity of each metapopulations (Figure 5 and SI Figure 18). Interestingly, although outliers were defined within each metapopulation, the inter-metapopulation differentiation was retrieved in all cases. Nevertheless, a clear signal of differentiation was highlighted within both regions. Individuals from the Manus and Woodlark basins showed clear genetic differentiation with no overlap on PC2 (Figure 6 A). Individuals from North-Fiji were slightly pulled towards Manus/Woodlark individuals based on PC1 (Figure 6 B), but they were also shifted on PC2 when outliers were considered at the threshold of $p \leq 0.05$.

The ADMIXTURE analyses based on the outlier SNPs datasets with the threshold of $p \leq 0.05$ displayed optimal $K$ values at $K$=2. With North-Fiji/Futuna/Lau outliers, North-Fiji displayed an admixture proportion from Manus/Woodlark ranging between 5 and 15% (Figure 6 C). For Manus/Woodlark outliers, we also found an admixture proportion from North-Fiji/Futuna/Lau

418 in Woodlark (Figure 6 D). However, very similar values of cross-validation errors were
419 obtained with $K$=3 (SI Figure 19,20,21), North-Fiji and Woodlark each being individualized as
420 the third cluster in their respective runs.

421 The $F_3$ *statistics* calculated based on outliers only did not provide any additional information
422 (SI Figure 22).

423 When blasted onto the *A. boucheti* transcriptome, 30% of outlier loci identified at the threshold
424 of $p \leq 0.05$ matched with the coding sequences of transcribed regions. This number was greater
425 than expected by chance from randomly picked ddRAD loci along the *I. nautilei* genome. Half
426 of these 30% of outlier loci (129 transcripts) had annotations on the protein database. Among
427 these annotations, many involved genes that encode for DNA/RNA replication and repair
428 enzymes, transmembrane carriers and synapse/microtubule biosynthesis, but also genes
429 involved in the exocytosis/endocytosis regulation, and more especially the GTPase regulation
430 pathway (SI Table 7). In addition, two genes involved in spermatogenesis were also detected.

431

432

433

434 # Discussion

435 Previous work by Thaler et al. (2011) using microsatellites and mitochondrial *cox1* sequences
436 demonstrated that the southwestern Pacific deep-sea hydrothermal vent gastropod *Ifremeria.*
437 *nautilei* is genetically structured into two distinct populations from Manus and North-Fh-
438 Fiji/Lau BABs. Our study extends these previous results to a finer scale, owing to our larger,
439 nested sampling design that includes the newly discovered La Scala vent field in the Woodlark

basin (Boulart et al., in press), the Futuna volcanic arc (Konn et al., 2016) and the newly discovered northernmost Mangatolo site at the entrance of the Lau basin. Using a 10 570 SNP genome-wide dataset—unprecedented for a hydrothermal species—, we confirm that *I. nautilei* is structured into two loosely connected metapopulations corresponding to two BAB ensembles. These ensembles display an almost complete internal genetic homogeneity; however, our analyses of outlier loci nevertheless revealed fine-scale differentiation among basins within each metapopulation. We discuss below the possible implications of these results in terms of larval dispersal and demographic connectivity and ultimately their consequences on the resilience of hydrothermal communities.

*Long-term gene flow and history of differentiation*

One metapopulation comprises the Manus and Woodlark basins (*i.e.* the Manus/Woodlark BAB) west of the Salomon/New Hebrides arc, whereas the other extends east of it with the North-Fiji basin, Futuna volcanic arc and Lau basin (*i.e.* the North-Fiji/Futuna/Lau BAB). The genetic divergence between the two metapopulations is relatively strong (average $F_{ST}$ = 0. 387, $p \leq 0.001$, $D_{xy}$ = 0.0136), but each of these two ensembles appears to be panmictic ($F_{CS}$ = -0.05, NS, SI Table 3). The demo-genetic inferences gleaned from ∂a∂i suggest that the two metapopulations diverged with only a brief period of isolation (Ts was found to vary between 0.001 and 0.443 in the AM and SC models), although the existence of constant gene flow (IM) could not be formally excluded. The incorporation of several other demographic parameters (2N, 2m, G) produced a clear improvement in model fit. Considering each parameter independently, the effect of linked selection (2N) had a much greater influence on AIC than the other two parameters (2m and G), suggesting that a non-negligible proportion of loci may be influenced by linked selection . For the best models (2N+2m+G), this proportion approaches Q = 0.99 (which seems to be unrealistic), whereas only half of the markers appear to be under the

influence of heterogeneous migration ($0.52 \leq 1\text{-}P \leq 0.56$). Nevertheless, these models are very close to the 2N+G models ($\Delta_{AIC} \leq 10$), which estimate a proportion of loci under linked selection ($0.56 \leq Q \leq 0.58$) and do not take into account the effect of heterogeneous migration. Disentangling these two effects is thus difficult and suggests that there are many genomic regions, possibly with lower recombination rates, where background selection and possibly selective sweeps have accelerated the rate of lineage sorting during divergence (Rougeux et al., 2017). This strong bimodality between two classes of loci affected or not by linked selection is also captured by the distribution of $F_{ST}$, which shows a clear trough and then a peak around 0.15–0.2 (SI Figure 23). However, this bimodality reduces the ability to distinguish between the isolation-with-migration, the secondary contact, or the ancient migration scenarios in the more complex models (IM+2N+2m+G, AM+2N+2m+G and SC+2N+2m+G).

Considering an average DNA mutation rate of $10^{-8}$/site/generation, we estimated the time for the onset of divergence between the two metapopulations to be 60 000–70 000 generations (but admittedly this could as well be 10 times greater if the mutation rate is 10 times smaller). The generation time of *I. nautilei* is still unknown. Nevertheless, most hydrothermal species display an *r*-strategy suggesting short generation times (1-2 years) as an adaptation to the unstable and ephemeral nature of their habitat (Tyler & Young, 1999). Hence, we can suppose that the two populations started to diverge between 60 and 140 thousand ago (*kya*) for a mutation rate per site and per generation of $10^{-8}$ and 10 times more with a mutation rate of $10^{-9}$. However, these estimates correspond to discrete non-overlapping generations and the reproduction of older cohorts may increase the equivalent generation time and, as a result, the divergence estimates. These values may be tentatively compared with estimates from the *cox1* sequences in Boulart et al. (in press) (net divergence 0.615% estimated on all sites). This latter value would amount to ~ 0.550 million years ago (*mya*) considering the widely used divergence rate of 1.4%/million

488 years (*myr*) for mitochondrial DNA (Knowlton & Weigt, 1998), but can reach 1.2 *mya* ,
489 depending on the average mitochondrial substitution rate considered for vent species (0.2-
490 0.3%/*myr* (Chevaldonné et al., 2002; Breusing et al., 2020; Castel et al. in prep.) . Although
491 these estimates are notoriously highly variable and error-prone (see for instance Breusing et al.,
492 2020), divergence time could range between 0.5 and 1 *mya*. This estimate is in rough agreement
493 with Martinez & Taylor, (1996) who showed that the center of the Manus BAB started to spread
494 quite recently (~ 0.78 *mya*), suggesting that hydrothermal vents within the spreading center may
495 be younger than this estimate. Thus, it cannot be excluded that the divergence history of *I.*
496 *nautilei* is relatively recent and not linked to the formation of BABs, but instead to regional
497 modifications of surface and deep-sea currents during previous glacial maxima in relation to
498 the extension of the Antarctic ice sheet which culminated around 0.126 *mya* (Barrows et al.,
499 2011; Joy et al., 2014).

500 In addition to these divergence time estimates, the models allowed us to quantify the existence
501 of an ongoing bidirectional and asymmetrical gene flow, with migration from North-
502 Fiji/Futuna/Lau to Manus/Woodlark being higher than in the opposite direction. Despite this
503 slight asymmetry, a genetic influence of the Manus/Woodlark metapopulation was detected in
504 North-Fiji, which shows foreign alleles coming from the former rather than from the
505 Lau/Futuna populations (also observed at mtDNA *cox1* gene in Thaler et al., 2011 and Boulart
506 et al., (in press), but not the other way around (*i.e.* influence of North-Fiji/Futuna/Lau on
507 Woodlark, but see below). This result is consistent with the geography of the region, because
508 North-Fiji and Woodlark are the closest BABs between the two metapopulations. Connectivity
509 through larval dispersal between these two BAB ensembles has been tested by Mitarai et al.
510 (2016) who simulated larval dispersal through entrainment of particles by oceanic currents
511 prevailing at depths of 1000 m in the western Pacific. That study inferred a weak stepping-stone

connection through a long planktonic larval duration (PLD of 170 days), provided that active hydrothermal sites in the Solomon and New Hebrides/Vanuatu arcs act as a relay. Such fields are known to exist, mostly associated with seamounts such as Nifonea, Tinakula or Stanton along the New Hebrides/Vanuatu arc (McConachy et al., 2005; Schmidt et al., 2017). The larval dispersal model developed by Mitarai et al. (2016) suggests a scenario where dispersal is mainly oriented from east to west: a situation also depicted in this region by Yearsley & Sigwart (2011) for a non-hydrothermal species at several depths (800 and 1400 m) and with various PLD lengths (27–151 days). However, when looking at surface countercurrents between Manus/Woodlark and North-Fiji/Futuna/Lau, Ganachaud et al. (2014) found surface currents oriented mainly from west to east through Solomon Islands and Vanuatu waters.

Cases of asymmetrical bidirectional gene flow between two metapopulations have also been found in two other hydrothermal gastropod species occurring sympatrically with *I. nautilei*, *L. schrolli* (Plouviez et al., 2019) and *A. boucheti* (Breusing et al., 2021). But, in contrast to *I. nautilei*, the predominant gene flow is oriented eastward from Manus to Lau, ($m_{M->L}$ = 0.625, $m_{L->M}$ = 0.1725 for *L. schrolli* and $m_{M->L}$ = 12, $m_{L->M}$ = 2.6 for *A. boucheti*).

Similarly to *I. nautilei*, *L. schrolli* is considered to possess lecithotrophic larvae (Berg, 1985; Craddock, Lutz, & Vrijenhoek, 1997; Tyler et al., 2008)). As for *A. boucheti,* its larval stage remains unknown, although both its morphology (Warèn & Bouchet, 1993) and the eDNA detection of *Alviniconcha* larvae close to the surface suggest planktonotrophy (Sommer et al., 2017). Provided that our estimates reflect ongoing migration, we hypothesize that *I. nautilei* larvae may be influenced by deep as well as surface currents, which could explain bidirectional gene flow, one direction being slightly stronger than the other. This asymmetry suggests vertical

migration of larvae. However, further investigations including oceanographic modeling and laboratory experiments are needed to address this hypothesis. For example, larvae of the hydrothermal gastropod *Shinkailepas myojinensis* (Yahagi et al., 2017) are able to migrate through the water column, and there is evidence of hydrothermal species' larvae in near-surface waters (Arellano et al., 2014, Sommer et al., 2017). Nevertheless, although many unknowns remain, our results indicate that *I. nautilei* has a complex dispersal strategy and pattern.

*Fine-scale population structure and connectivity*

The high homogeneity of the two clearly distinct *I. nautilei* metapopulations necessarily entails that the intra-metapopulation migration (*i.e.* inter-localities within each BAB and inter-BABs within each metapopulation) is strong or extremely recent. Moreover, no kinship-related structure was detected in the SNP dataset, indicating that there is either no self-recruitment even though females brood their larvae to the trochophoran stage, or that population sizes are so large that the probability of detecting potential kin is too small (Table 4). Consequently, genetic connectivity within each metapopulation appears to be high, with evenly distributed polymorphisms among sampled sites despite the patchy distribution of hydrothermal vents and the inter-site distances, which may vary from hundreds of meters to more than a thousand kilometers within each metapopulation. This genetic connectivity therefore suggests that *I. nautilei* larvae are able to disperse within the range of each metapopulation after spawning.

The question is now whether this genetic homogeneity of each metapopulation arises from demographic connectivity (*i.e.* recruitment at one site being strongly influenced by the exportation of propagules from other sites) or is due to sporadic/rare larval exchanges able to counterbalance very attenuated genetic drift due to large local population sizes. The mechanism behind the observed genetic homogeneity has strong implications in terms of conservation

biology, because demographic connectivity can play a crucial role in the resilience of populations faced with local extinction potentially exacerbated by deep-sea mining. The global analysis relying on a panel of primarily neutral markers indicates no differentiation at the metapopulation scale, but—as advocated by Gagnaire et al., (2015)— a few loci markers potentially undergoing direct, or indirect selective pressures may locally harbor distinct allele frequency in the recipient population. This pattern can be explained by local selection for foreign alleles that are better adapted or less loaded by deleterious mutations than resident ones, or by resolving intrinsic asymmetrical incompatibilities between divergent genomes (Simon et al., 2021) creating local soft sweeps through linked selection. These processes result in enhanced local introgression of certain marker loci, a common pattern observed in blue mussels (Fraïsse et al., 2016) or European sea bass (Robinet et al., 2020), for example. These markers will appear as $F_{ST}$ outliers that may indicate recent dispersal events.

Our outlier analyses indeed suggest introgression of some loci. In Figure 6 B and SI Figure 18 B, individuals from the North-Fiji basin seem to be closer to Manus/Woodlark than Lau/Futuna on PC1, which may correspond to introgression in some of the outlier loci. An introgression pattern was confirmed by the $F_3$ tests performed with the North-Fiji basin as the focal populations (significant negative value of the $F_3$ statistic, SI Figure 13) and the ADMIXTURE analyses (Figure 6 C). These results indicate that some alleles found at high frequency in North-Fiji individuals are the consequence of long-range migration from Manus/Woodlark. Interestingly, with Manus/Woodlark outliers, although not visually detectable on the PCA (Figure 6 A and 18 A), Woodlark individuals exhibit some level of admixture from the North-Fiji/Futuna/Lau metapopulation (Figure 6 D). This low-level admixture corroborates our inference of ongoing bidirectional gene flow. However, it is not yet clear as to why its impact appears stronger in populations of the North-Fiji basin, against the predominant direction

according to our ∂a∂i inferences. Although we are unable to determine the precise mechanisms behind these frequency changes, these alleles have not diffused from North-Fiji to Lau/Futuna, indicating a subtle—but real—limitation in connectivity between the former and the latter. The same reasoning applies for the traces of admixture detected in Woodlark that appear to have not diffused to Manus.

Another kind of differentiation depicted by outlier loci seems to be explained by intra-metapopulation divergence. The question arises as to the origin of these slight divergences on a PC axis orthogonal to the main inter-metapopulation divergence, which does not necessarily proceed from gene flow between differentiated populations as described above. Allele frequency differences for outlier loci between Manus and Woodlark are detectable on PC2 (Figure 6 A, SI Figure 18 A, ADMIXTURE $K$ = 3, SI Figure 20). The same question applies to the eastern North-Fiji/Futuna/Lau metapopulation, with differences between North-Fiji and Futuna/Lau (Figure 6 B, ADMIXTURE $K$ = 3, SI Figure 21). This pattern can be due to any combination of drift and/or selection. Local selection may result from major differences in depth or vent fluid composition. The fact that the fraction of outliers mapping on transcribed regions is greater than by chance and targets a few metabolic/regulatory pathways suggests their possible involvement in local adaptation to depth or different fluid chemistry, but this remains to be investigated (Jennings et al., 2013; Liu et al., 2021). In the absence of high demographic connectivity required to ensure the interdependency of local populations, this local differentiation can remain detectable for several generations before being shuffled among all metapopulation demes.

These subtle limitations in connectivity between basins can be associated with abyssal plains, which may limit gene flow in a disconnected ridge system such as that found at the regional scale of these BABs. Physical barriers in other parts of the world, such as transform faults and

606 microplates, have already been shown to greatly impede the effective migration of deep-sea
607 vent species at a much more restricted spatial scale (Johnson et al., 2006; Plouviez et al., 2009;
608 Plouviez, et al., 2013). However, regarding the populations of the Futuna volcanic arc and Lau
609 basin, our in-depth scrutiny of outliers did not reveal any sign of genetic differentiation. Hence,
610 the hypothesis of demographic correlation between these two regions cannot be rejected,
611 although we cannot infer with certitude the directionality of the exchanges.

612 # Conclusions

613 Overall, our analyses revealed a clear genetic differentiation of *Ifremeria nautilei* populations
614 between the Manus/Woodlark and the North-Fiji/Futuna/Lau BABs, with very high gene flow
615 within each of these two metapopulations as well as higher genetic diversity in
616 Manus/Woodlark. Despite an in-depth scrutiny of genome-wide genetic variation, no
617 geographic substructure was detected between or within localities sampled within each
618 individual ridge system. This genetic connectivity probably indicates high local
619 (re)colonization capacity for this hydrothermal vent species due to the ephemeral nature of
620 active sites in this region, at least at the scale of a given back-arc basin.

621 However, our outlier analyses revealed that this genetic connectivity does not necessarily
622 equate with demographic connectivity at the larger inter-basin intra-metapopulation scale. The
623 specific investigation of outlier loci illustrates how a few loci in a large genome-wide dataset
624 can carry useful information about actual barriers to dispersal in high gene flow species. Deep-
625 sea mining holds the potential to exacerbate dispersal barriers and limit population resilience,
626 because if a large proportion of the vent habitat is destroyed locally, population rescue from
627 other basins will be restricted.

628 Furthermore, our demographic simulations indicated a long period of divergence during the
629 Quaternary period (several tens of thousands of generations) associated with restricted long-
630 range gene flow over a large fraction of the genome. Although the effects of linked selection
631 and reduced migration (barrier loci) are not clearly distinguishable, our results suggest that the
632 effect of the latter is less pronounced. This interpretation agrees with the fact that the global
633 divergence among the two metapopulations is still quite low (net nuclear nucleotide divergence,
634 0.81%). This divergence perhaps reflects the very beginning of an ongoing speciation process,
635 where a handful of barrier loci may already exist and at the same time overall genetic
636 differentiation is not hampered by weak contemporary and asymmetrical gene flow between
637 metapopulations.
638
639
640
641
642
643
644
645
646
647
648

**Acknowledgments**

We are deeply grateful to the R/V *L'Atalante* crew as well as the ROV *Victor 6000* crew during the ChuBacArc 2019 cruise without whom nothing would have been possible. Many thanks also to Cindy L. Van Dover for sharing some Manus 2008 samples and to John Parianos and Paul Lahari from Nautilus Minerals group for sharing information on their Solwara mining prospects in the Manus basin. W. We also thank Gwenn Tanguy for advice and, and access to the Biogenouest Genomer Platform at the Station Biologique de Roscoff. Data were stored and analyzed at the Biogenouest AbiMS Bioinformatics platform, which provided data storage and computing resources, as well as at the Montpellier Bioinformatics Biodiversity platform (LabEx CeMEB). We also thank Khalid Belkhir and Christelle Fraïsse for their debugged version of ∂a∂i. We are most grateful to four anonymous reviewers whose detailed and insightful comments significantly improved the quality and clarity of the manuscript.

Ship time was supported by the French Oceanographic Fleet program (CHUBACARC cruise doi 10.17600/18001111 to D. J. and S. H.). Travel expenses of ChuBacArc participants and logistic expenses were funded by CNRS-Institut Ecologie et Environnement. Lab work and A. T. L. Y. PhD fellowship were funded through the Agence Nationale de la Recherche ANR 'CERBERUS' (contract number ANR-17-CE02-0003 to S. H.). Sequencing was integrated in the eDNAbyss project (contract AP2016-228 to S. A. H.) funded by France Génomique (ANR-10-INBS-09) and Genoscope-CEA.

# References


Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, *12*(1), 246. https://doi.org/10.1186/1471-2105-12-246

Arellano, S. M., Van Gaest, A. L., Johnson, S. B., Vrijenhoek, R. C., & Young, C. M. (2014). Larvae from deep-sea methane seeps disperse in surface waters. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1786), 20133276. https://doi.org/10.1098/rspb.2013.3276

Audzijonyte, A., & Vrijenhoek, R. C. (2010). When gaps really are gaps : Statistical phylogeography of hydrothermal vent invertebrates. *Evolution*, *64*(8), 2369-2384. https://doi.org/10.1111/j.1558-5646.2010.00987.x

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes : A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455-477. https://doi.org/10.1089/cmb.2012.0021

Barrows, T. T., Hope, G. S., Prentice, M. L., Fifield, L. K., & Tims, S. G. (2011). Late Pleistocene glaciation of the Mt Giluwe volcano, Papua New Guinea. *Quaternary Science Reviews*, *30*(19-20), 2676-2689. https://doi.org/10.1016/j.quascirev.2011.05.022

Berg, C. J. (1985). Reproductive strategies of mollusks from abyssal hydrothermal vent communities. *Bulletin of The Biological Society of Washington*, 185-197.

Boulart, C., Rouxel, O., Scalabrin, C., Le Meur, P., Pelleter, E., Poitrimol, C., Thiébaut, E., Matabos, M., Castel, J., Tran Lu Y, A., Michel, L.N., Cathalot, C., Cheron, S., Boissier, A.,Germain, Y., Arnaud-Haond, S., Bonhomme, F., Broquet, T., Cueff-Gauchard, V., Le Layec, V., L'Haridon, S, Mary, J., Le Port, A-S., Tasiemski, A., Kuama, D.C., Hourdez, S., Jollivet, D. (*in press*). Active hydrothermal vents in the Woodlark Basin may act as dispersing centres for hydrothermal fauna *Communications Earth & Environment.*

Borowski, C., Giere, O., Krieger, J., Amann, R., & Dubilier, N. (2002). New aspects of the symbiosis in the provannid snail *Ifremeria nautilei* from the North-Fiji back arc basin. *Cahiers de Biologie Marine*, 5. https://www.researchgate.net/publication/27262835_New_aspects_of_the_symbiosis_in_the_provannid_snail_Ifremeria_nautilei_from_the_North_Fiji_Back_Arc_Basin

Brelsford, A., Dufresnes, C., & Perrin, N. (2016). High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, *116*(2), 177-181. https://doi.org/10.1038/hdy.2015.83

Breusing, C., Biastoch, A., Drews, A., Metaxas, A., Jollivet, D., Vrijenhoek, R. C., Bayer, T., Melzner, F., Sayavedra, L., Petersen, J. M., Dubilier, N., Schilhabel, M. B., Rosenstiel, P., & Reusch, T. B. H. (2016). Biophysical and population genetic models predict the presence of "phantom" stepping stones connecting Mid-Atlantic Ridge vent ecosystems. *Current Biology*, *26*(17), 2257-2267. https://doi.org/10.1016/j.cub.2016.06.062

Breusing, C., Johnson, S. B., Tunnicliffe, V., Clague, D. A., Vrijenhoek, R. C., & Beinart, R. A. (2020). Allopatric and sympatric drivers of speciation in alviniconcha hydrothermal vent snails. *Molecular Biology and Evolution*, *37*(12), 3469-3484. https://doi.org/10.1093/molbev/msaa177

Castel, J., Hourdez, S., Pradillon, F., Daguin-Thiébaut, C., Ballenghien, M., Ruault, S., Corre, E., Tran Lu Y, A., Mary, J., Comtet, T., Gagnaire, P-A., Bonhomme, F., Bierne, N., Breusing, C., Broquet, T., & Jollivet D. (2022): A story of divergence and inter-specific exchanges in deep-

721 sea hydrothermal vent gastropods *Alviniconcha*. *Genes*.Manuscript
722 submitted for publication.
723 Cayuela, H., Rougemont, Q., Prunier, J. G., Moore, J.-S., Clobert, J., Besnard, A., & Bernatchez, L.
724 (2018). Demographic and genetic approaches to study dispersal in wild animal populations : A
725 methodological review. *Molecular Ecology*, *27*(20), 3976-4010.
726 https://doi.org/10.1111/mec.14848
727 Chevaldonné, P., Jollivet, D., Desbruyères, D., Lutz, R., & Vrijenhoek, R. (2002). Sister-species of
728 eastern Pacific hydrothermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera)
729 provide new mitochondrial COI clock calibration. *CBM - Cahiers de Biologie Marine*, *43*(3-
730 4), 367-370. https://archimer.ifremer.fr/doc/00000/895/
731 Chevaldonné, P., Jollivet, D., Vangriesheim, A., & Desbruyères, D. (1997). Hydrothermal-vent
732 alvinellid polychaete dispersal in the eastern Pacific. 1. Influence of vent site distribution,
733 bottom currents, and biological patterns. *Limnology and Oceanography*, *42*(1), 67-80.
734 https://doi.org/10.4319/lo.1997.42.1.0067
735 Craddock, C., Hoeh, W. R., Gustafson, R. G., Lutz, R. A., Hashimoto, J., & Vrijenhoek, R. J. (1995).
736 Evolutionary relationships among deep-sea mytilids (Bivalvia : Mytilidae) from hydrothermal
737 vents and cold-water methane/sulfide seeps. *Marine Biology*, *121*(3), 477-485.
738 https://doi.org/10.1007/BF00349456
739 Craddock, C., Lutz, R. A., & Vrijenhoek, R. C. (1997). Patterns of dispersal and larval development of
740 archaeogastropod limpets at hydrothermal vents in the eastern Pacific. *Journal of*
741 *Experimental Marine Biology and Ecology*, *210*(1), 37-51. https://doi.org/10.1016/S0022-
742 0981(96)02701-3
743 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E.,
744 Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project
745 Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-
746 2158. https://doi.org/10.1093/bioinformatics/btr330
747 Dover, C. L. V., Humphris, S. E., Fornari, D., Cavanaugh, C. M., Collier, R., Goffredi, S. K.,
748 Hashimoto, J., Lilley, M. D., Reysenbach, A. L., Shank, T. M., Damm, K. L. V., Banta, A.,
749 Gallant, R. M., Götz, D., Green, D., Hall, J., Harmer, T. L., Hurtado, L. A., Johnson, P., …
750 Vrijenhoek, R. C. (2001). Biogeography and Ecological Setting of Indian Ocean
751 Hydrothermal Vents. *Science*, *294*(5543), 818-823. https://doi.org/10.1126/science.1064574
752 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust
753 demographic inference from genomic and SNP data. *PLOS Genetics*, *9*(10), e1003905.
754 https://doi.org/10.1371/journal.pgen.1003905
755 Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5 : A new series of programs to perform
756 population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3),
757 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x
758 Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from
759 metric distances among DNA haplotypes : Application to human mitochondrial DNA
760 restriction data. *Genetics*, *131*(2), 479-491. https://www.genetics.org/content/131/2/479
761 Feutry, P., Devloo-Delva, F., Y, A. T. L., Mona, S., Gunasekera, R. M., Johnson, G., Pillans, R. D.,
762 Jaccoud, D., Kilian, A., Morgan, D. L., Saunders, T., Bax, N. J., & Kyne, P. M. (2020). One
763 panel to rule them all : DArTcap genotyping for population structure, historical demography,
764 and kinship analyses, and its application to a threatened shark. *Molecular Ecology Resources*,
765 *20*(6), 1470-1485. https://doi.org/10.1111/1755-0998.13204
766 Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both
767 dominant and codominant markers: A Bayesian Perspective. *Genetics*, *180*(2), 977-993.
768 https://doi.org/10.1534/genetics.108.092221
769 Fraïsse, C., Belkhir, K., Welch, J. J., & Bierne, N. (2016). Local interspecies introgression is the main
770 cause of extreme levels of intraspecific differentiation in mussels. *Molecular Ecology*, *25*(1),
771 269-286. https://doi.org/10.1111/mec.13299
772 Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., &
773 Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine

774 populations in the genomic era. *Evolutionary Applications*, *8*(8), 769--786.
775 https://doi.org/10.1111/eva.12288
776 Ganachaud, A., Cravatte, S., Melet, A., Schiller, A., Holbrook, N. J., Sloyan, B. M., Widlansky, M. J.,
777 Bowen, M., Verron, J., Wiles, P., Ridgway, K., Sutton, P., Sprintall, J., Steinberg, C.,
778 Brassington, G., Cai, W., Davis, R., Gasparin, F., Gourdeau, L., …, & Send, U. (2014). The
779 Southwest Pacific Ocean circulation and climate experiment (SPICE). *Journal of Geophysical*
780 *Research: Oceans*, *119*(11), 7660--7686. https://doi.org/10.1002/2013JC009678
781 Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-
782 specific covariates. *Genetics*, *201*(4), 1555--1579.
783 https://doi.org/10.1534/genetics.115.181453
784 Gena, K. (2013). Deep sea mining of submarine hydrothermal deposits and its possible environmental
785 impact in Manus basin, Papua New Guinea. *Procedia Earth and Planetary Science*, *6*, 226-
786 233. https://doi.org/10.1016/j.proeps.2013.01.031
787 Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the
788 joint demographic history of multiple populations from multidimensional SNP frequency data.
789 *PLOS Genetics*, *5*(10), e1000695. https://doi.org/10.1371/journal.pgen.1000695
790 Hannington, M., Jamieson, J., Monecke, T., Petersen, S., & Beaulieu, S. (2011). The abundance of
791 seafloor massive sulfide deposits. *Geology*, *39*(12), 1155--1158.
792 https://doi.org/10.1130/G32468.1
793 Hurtado, L. A., Lutz, R. A., & Vrijenhoek, R. C. (2004). Distinct patterns of genetic differentiation
794 among annelids of eastern Pacific hydrothermal vents. *Molecular Ecology*, *13*(9), 2603--2615.
795 https://doi.org/10.1111/j.1365-294X.2004.02287.x
796 Jennings, R. M., Etter, R. J., & Ficarra, L. (2013). Population differentiation and species formation in
797 the deep sea: The potential role of environmental gradients and depth. *PLOS ONE*, *8*(10),
798 e77594. https://doi.org/10.1371/journal.pone.0077594
799 Johnson, S. B., Young, C. R., Jones, W. J., Warén, A., & Vrijenhoek, R. C. (2006). Migration,
800 isolation, and speciation of hydrothermal vent limpets (gastropoda; lepetodrilidae) across the
801 Blanco transform fault. *The Biological Bulletin*, *210*(2), 140-157.
802 https://doi.org/10.2307/4134603
803 Jollivet, D., Chevaldonné, P., & Planque, B. (1999). Hydrothermal-vent alvinellid polychaete dispersal
804 in the Eastern Pacific. 2. A metapopulation model based on habitat shifts. *Evolution*, *53*(4),
805 1128--1142. https://doi.org/10.1111/j.1558-5646.1999.tb04527.x
806 Jollivet, D., Desbruyères, D., Bonhomme, F., & Moraga, D. (1995). Genetic differentiation of deep-
807 sea hydrothermal vent alvinellid populations (Annelida : Polychaeta) along the East Pacific
808 Rise. *Heredity*, *74*(4), 376--391. https://doi.org/10.1038/hdy.1995.56
809 Jones, G. P., Planes, S., & Thorrold, S. R. (2005). Coral reef fish larvae settle close to home. *Current*
810 *Biology*, *15*(14), 1314--1318. https://doi.org/10.1016/j.cub.2005.06.061
811 Joy, K., Fink, D., Storey, B., & Atkins, C. (2014). A 2 million year glacial chronology of the
812 Hatherton Glacier, Antarctica and implications for the size of the East Antarctic Ice Sheet at
813 the Last Glacial Maximum. *Quaternary Science Reviews*, *83*, 46--57.
814 https://doi.org/10.1016/j.quascirev.2013.10.028
815 Knowlton, N., & Weigt, L. A. (1998). New dates and new rates for divergence across the Isthmus of
816 Panama. *Proceedings of the Royal Society of London. Series B: Biological Sciences*,
817 *265*(1412), 2257--2263. https://doi.org/10.1098/rspb.1998.0568
818 Konn, C., Fourré, E., Jean-Baptiste, P., Donval, J. P., Guyader, V., Birot, D., Alix, A. S., Gaillot, A.,
819 Perez, F., Dapoigny, A., Pelleter, E., Resing, J. A., Charlou, J. L., & Fouquet, Y. (2016).
820 Extensive hydrothermal activity revealed by multi-tracer survey in the Wallis and Futuna
821 region (SW Pacific). *Deep Sea Research Part I: Oceanographic Research Papers*, *116*, 127--
822 144. https://doi.org/10.1016/j.dsr.2016.07.012
823 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak : A
824 program for identifying clustering modes and packaging population structure inferences across
825 K. *Molecular Ecology Resources*, *15*(5), 1179--1191. https://doi.org/10.1111/1755-
826 0998.12387

827 Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider : An automated data conversion tool for
828 connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298--299.
829 https://doi.org/10.1093/bioinformatics/btr642
830 Liu, R., Liu, J., & Zhang, H. (2021). Positive selection analysis reveals the deep-sea adaptation of a
831 hadal sea cucumber (*Paelopatides sp.*) to the Mariana Trench. *Journal of Oceanology and*
832 *Limnology*, *39*(1), 266--281. https://doi.org/10.1007/s00343-020-0241-0
833 Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity?
834 *Molecular Ecology*, *19*(15), 3038--3051. https://doi.org/10.1111/j.1365-294X.2010.04688.x
835 Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt : An R package to perform genome scans for
836 selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67--77.
837 https://doi.org/10.1111/1755-0998.12592
838 Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, *26*(8), 345--352.
839 https://doi.org/10.1016/j.tig.2010.05.003
840 Martinez, F., & Taylor, B. (1996). Backarc spreading, rifting, and microplate rotation, between
841 transform faults in the Manus Basin. *Marine Geophysical Researches*, *18*(2--4), 203--224.
842 https://doi.org/10.1007/BF00286078
843 Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015).
844 Restriction site-associated DNA sequencing, genotyping error estimation and de novo
845 assembly optimization for population genetic inference. *Molecular Ecology Resources*, *15*(1),
846 28--41. https://doi.org/10.1111/1755-0998.12291
847 McConachy, T. F., Arculus, R. J., Yeats, C. J., Binns, R. A., Barriga, F. J. A. S., McInnes, B. I. A.,
848 Sestak, S., Sharpe, R., Rakau, B., & Tevi, T. (2005). New hydrothermal activity and alkalic
849 volcanism in the backarc Coriolis Troughs, Vanuatu. *Geology*, *33*(1), 61--64.
850 https://doi.org/10.1130/G20870.1
851 Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espiñeira, M.,
852 Fiorentino, F., Garofalo, G., Geffen, A. J., Hansen, Jakob. H., Helyar, S. J., Nielsen, E. E.,
853 Ogden, R., Patarnello, T., Stagioni, M., Consortium, F., Tinti, F., & Bargelloni, L. (2014).
854 Outlier SNP markers reveal fine-scale genetic structuring across European hake populations
855 (*Merluccius merluccius*). *Molecular Ecology*, *23*(1), 118--135.
856 https://doi.org/10.1111/mec.12568
857 Mitarai, S., Watanabe, H., Nakajima, Y., Shchepetkin, A. F., & McWilliams, J. C. (2016). Quantifying
858 dispersal from hydrothermal vent fields in the western Pacific Ocean. *Proceedings of the*
859 *National Academy of Sciences*, *113*(11), 2976--2981.
860 https://doi.org/10.1073/pnas.1518395113
861 Mullineaux, L. S., Adams, D. K., Mills, S. W., & Beaulieu, S. E. (2010). Larvae from afar colonize
862 deep-sea hydrothermal vents after a catastrophic eruption. *Proceedings of the National*
863 *Academy of Sciences*, *107*(17), 7829--7834. https://doi.org/10.1073/pnas.0913187107
864 Niner, H. J., Ardron, J. A., Escobar, E. G., Gianni, M., Jaeckel, A., Jones, D. O. B., Levin, L. A.,
865 Smith, C. R., Thiele, T., Turner, P. J., Van Dover, C. L., Watling, L., & Gjerde, K. M. (2018).
866 Deep-Sea Mining With No Net Loss of Biodiversity—An Impossible Aim. *Frontiers in*
867 *Marine Science*, *5*. https://www.frontiersin.org/article/10.3389/fmars.2018.00053
868 Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space : A road map for stacks.
869 *Methods in Ecology and Evolution*, *8*(10), 1360--1373. https://doi.org/10.1111/2041-
870 210X.12775
871 Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-
872 wide allele frequency data. *PLOS Genetics*, *8*(11), e1002967.
873 https://doi.org/10.1371/journal.pgen.1002967
874 Pinsky, M. L., Jr, H. R. M., & Palumbi, S. R. (2010). Using isolation by distance and effective density
875 to estimate dispersal scales in anemonefish. *Evolution*, *64*(9), 2688--2700.
876 https://doi.org/10.1111/j.1558-5646.2010.01003.x
877 Plouviez, S., Faure, B., Guen, D. L., Lallier, F. H., Bierne, N., & Jollivet, D. (2013). A new barrier to
878 dispersal trapped old genetic clines that escaped the easter microplate tension zone of the

879 Pacific vent mussels. *PLOS ONE*, *8*(12), e81555.
880 https://doi.org/10.1371/journal.pone.0081555
881 Plouviez, S., LaBella, A. L., Weisrock, D. W., Meijenfeldt, F. A. B., von Ball, B., Neigel, J. E., &
882 Dover, C. L. V. (2019). Amplicon sequencing of 42 nuclear loci supports directional gene
883 flow between South Pacific populations of a hydrothermal vent limpet. *Ecology and*
884 *Evolution*, *9*(11), 6568--6580. https://doi.org/10.1002/ece3.5235
885 Plouviez, S., Schultz, T. F., McGinnis, G., Minshall, H., Rudder, M., & Van Dover, C. L. (2013).
886 Genetic diversity of hydrothermal-vent barnacles in Manus Basin. *Deep Sea Research Part I:*
887 *Oceanographic Research Papers*, *82*, 73--79. https://doi.org/10.1016/j.dsr.2013.08.004
888 Plouviez, S., Shank, T. M., Faure, B., Daguin-Thiébaut, C., Viard, F., Lallier, F. H., & Jollivet, D.
889 (2009). Comparative phylogeography among hydrothermal vent species along the East Pacific
890 Rise reveals vicariant processes and population expansion in the South. *Molecular Ecology*,
891 *18*(18), 3903--3917. https://doi.org/10.1111/j.1365-294X.2009.04325.x
892 Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian
893 population history. *Nature*, *461*(7263), 489--494. https://doi.org/10.1038/nature08365
894 Reynolds, K. C., Watanabe, H., Strong, E. E., Sasaki, T., Uematsu, K., Miyake, H., Kojima, S.,
895 Suzuki, Y., Fujikura, K., Kim, S., & Young, C. M. (2010). New molluscan larval form :
896 brooding and development in a hydrothermal vent gastropod, *Ifremeria nautilei*
897 (Provannidae). *The Biological Bulletin*, *219*(1), 7--11. https://doi.org/10.1086/BBLv219n1p7
898 Robinet, T., Roussel, V., Cheze, K., & Gagnaire, P.-A. (2020). Spatial gradients of introgressed
899 ancestry reveal cryptic connectivity patterns in a high gene flow marine fish. *Molecular*
900 *Ecology*, *29*(20), 3857--3871. https://doi.org/10.1111/mec.15611
901 Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2 : Analytical methods for
902 paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*,
903 *28*(21), 4737--4754. https://doi.org/10.1111/mec.15253
904 Rougeux, C., Bernatchez, L., & Gagnaire, P.-A. (2017). Modeling the multiple facets of speciation-
905 with-gene-flow toward inferring the divergence history of lake whitefish species pairs
906 (*Coregonus clupeaformis*). *Genome Biology and Evolution*, *9*(8), 2057--2074.
907 https://doi.org/10.1093/gbe/evx150
908 Schmidt, K., Garbe-Schönberg, D., Hannington, M. D., Anderson, M. O., Bühring, B., Haase, K.,
909 Haruel, C., Lupton, J., & Koschinsky, A. (2017). Boiling vapour-type fluids from the Nifonea
910 vent field (New Hebrides Back-Arc, Vanuatu, SW Pacific) : Geochemistry of an early-stage,
911 post-eruptive hydrothermal system. *Geochimica et Cosmochimica Acta*, *207*, 185--209.
912 https://doi.org/10.1016/j.gca.2017.03.016
913 Simon, A., Fraïsse, C., Ayari, T. E., Liautard-Haag, C., Strelkov, P., Welch, J. J., & Bierne, N. (2021).
914 How do species barriers decay? Concordance and local introgression in mosaic hybrid zones
915 of mussels. *Journal of Evolutionary Biology*, *34*(1), 208--223.
916 https://doi.org/10.1111/jeb.13709
917 Sommer, S. A., Van Woudenberg, L., Lenz, P. H., Cepeda, G., & Goetze, E. (2017). Vertical gradients
918 in species richness and community composition across the twilight zone in the North Pacific
919 Subtropical Gyre. *Molecular Ecology*, *26*(21), 6136--6156. https://doi.org/10.1111/mec.14286
920 Teixeira, S., Cambon-Bonavita, M.-A., Serrão, E. A., Desbruyéres, D., & Arnaud-Haond, S. (2011).
921 Recent population expansion and connectivity in the hydrothermal shrimp *Rimicaris exoculata*
922 along the Mid-Atlantic Ridge. *Journal of Biogeography*, *38*(3), 564--574.
923 https://doi.org/10.1111/j.1365-2699.2010.02408.x
924 Teixeira, S., Serrão, E. A., & Arnaud-Haond, S. (2012). Panmixia in a fragmented and unstable
925 environment: The hydrothermal shrimp *Rimicaris exoculata* disperses extensively along the
926 Mid-Atlantic Ridge. *PLOS ONE*, *7*(6), e38521. https://doi.org/10.1371/journal.pone.0038521
927 Thaler, A. D., Plouviez, S., Saleu, W., Alei, F., Jacobson, A., Boyle, E. A., Schultz, T. F., Carlsson, J.,
928 & Dover, C. L. V. (2014). Comparative population structure of two deep-sea hydrothermal-
929 vent-associated decapods (*Chorocaris sp. 2* and *Munidopsis lauensis*) from Southwestern
930 Pacific Back-Arc basins. *PLOS ONE*, *9*(7), e101345.
931 https://doi.org/10.1371/journal.pone.0101345

Thaler, A. D., Zelnio, K., Saleu, W., Schultz, T. F., Carlsson, J., Cunningham, C., Vrijenhoek, R. C., & Van Dover, C. L. (2011). The spatial scale of genetic subdivision in populations of *Ifremeria nautilei*, a hydrothermal-vent gastropod from the southwest Pacific. *BMC Evolutionary Biology*, *11*(1), 372. https://doi.org/10.1186/1471-2148-11-372

Thiébaut, C. D., Ruault, S., Roby, C., Broquet, T., Viard, F., & Brelsford, A. (2021, septembre 21). *Construction of individual ddRAD libraries*. Protocols.Io. https://www.protocols.io/view/construction-of-individual-ddrad-libraries-bv4tn8wn

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., …,& Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*(1), 5770. https://doi.org/10.1038/ncomms6770

Tunnicliffe, V., McArthur, A. G., & McHugh, D. (1998). A biogeographical perspective of the deep-sea hydrothermal vent fauna. Dans J. H. S. Blaxter, A. J. Southward, & P. A. Tyler (Éds.), *Advances in Marine Biology* (Vol. 34, p. 353--442). Academic Press. https://doi.org/10.1016/S0065-2881(08)60213-8

Tyler, P. A., Pendlebury, S., Mills, S. W., Mullineaux, L., Eckelbarger, K. J., Baker, M., & Young, C. M. (2008). Reproduction of gastropods from vents on the East Pacific Rise and the Mid-Atlantic Ridge. *Journal of Shellfish Research*, *27*(1), 107--118. https://doi.org/10.2983/0730-8000(2008)27[107:ROGFVO]2.0.CO;2

Tyler, P. A., & Young, C. M. (1999). Reproduction and dispersal at vents and cold seeps. *Journal of the Marine Biological Association of the United Kingdom*, *79*(2), 193--208. https://doi.org/10.1017/S0025315499000235

Villemereuil, P. de, Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models : When and how much should we trust them? *Molecular Ecology*, *23*(8), 2006--2019. https://doi.org/10.1111/mec.12705

Vrijenhoek, R. C. (1997). Gene flow and genetic diversity in naturally fragmented metapopulations of deep-sea hydrothermal vent animals. *Journal of Heredity*, *88*(4), 285--293. https://doi.org/10.1093/oxfordjournals.jhered.a023106

Vrijenhoek, R. C. (2010). Genetic diversity and connectivity of deep-sea hydrothermal vent metapopulations. *Molecular Ecology*, *19*(20), 4391--4411. https://doi.org/10.1111/j.1365-294X.2010.04789.x

Warèn, A., & Bouchet, P. (1993). New records, species, genera, and a new family of gastropods from hydrothermal vents and hydrocarbon seeps. *Zoologica Scripta*, *22*(1), 1--90. https://doi.org/10.1111/j.1463-6409.1993.tb00342.x

Windoffer, R., & Giere, O. (1997). Symbiosis of the hydrothermal vent gastropod *Ifremeria nautilei* (provannidae) with endobacteria-structural analyses and ecological considerations. *The Biological Bulletin*, *193*(3), 381-392. https://doi.org/10.2307/1542940

Won, Y., Hallam, S. J., O'Mullan, G. D., & Vrijenhoek, R. C. (2003). Cytonuclear disequilibrium in a hybrid zone involving deep-sea hydrothermal vent mussels of the genus Bathymodiolus. *Molecular Ecology*, *12*(11), 3185-3190. https://doi.org/10.1046/j.1365-294X.2003.01974.x

Wyngaarden, M. V., Snelgrove, P. V. R., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R. E., & Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*, using RADseq-derived SNPs. *Evolutionary Applications*, *10*(1), 102-117. https://doi.org/10.1111/eva.12432

Yahagi, T., Fukumori, H., Warén, A., & Kano, Y. (2019). Population connectivity of hydrothermal-vent limpets along the northern Mid-Atlantic Ridge (Gastropoda : Neritimorpha: Phenacolepadidae). *Journal of the Marine Biological Association of the United Kingdom*, *99*(1), 179-185. https://doi.org/10.1017/S0025315417001898

Yahagi, T., Thaler, A. D., Dover, C. L. V., & Kano, Y. (2020). Population connectivity of the hydrothermal-vent limpet *Shinkailepas tollmanni* in the Southwest Pacific (Gastropoda : Neritimorpha: Phenacolepadidae). *PLOS ONE*, *15*(9), e0239784. https://doi.org/10.1371/journal.pone.0239784

Yahagi, T., Watanabe, H. K., Kojima, S., & Kano, Y. (2017). Do larvae from deep-sea hydrothermal vents disperse in surface waters? *Ecology*, *98*(6), 1524-1534. https://doi.org/10.1002/ecy.1800

Yearsley, J. M., & Sigwart, J. D. (2011). Larval transport modeling of deep-sea invertebrates can aid the search for undiscovered populations. *PLOS ONE*, *6*(8), e23063. https://doi.org/10.1371/journal.pone.0023063

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326-3328. https://doi.org/10.1093/bioinformatics/bts606

## Data Accessibility

Individual fastq files are available at the European Nucleotide Archive (study accession number PRJEB47533). SNP data (VCF) and associated metadata are available at Dryad : https://doi.org/10.5061/dryad.ffbg79cwq (preview https://datadryad.org/stash/share/WfF-hiYEO6nnKdwjG_E78-dA67mmAgnJ4i4Q6lL9-JU) Scripts (R, *∂a∂i*) are available in a public Github repository (https://github.com/Atranluy/Scripts-Ifremeria).

## Author's contribution:

D. J. and S. H. designed the CHUBACARC and CERBERUS projects, F. B. supervised the genetic work. A. T. L. Y., S. R., C. D. T., J. C., P. W. and A. P. performed laboratory work. A. T. L. Y. performed bioinformatics statistical analyses with the contribution of F. B., D. J., P. A. G., N. B. and T. B. A. T. L. Y., F. B. wrote the manuscript with feedback of T. B., D. J., P. A. G., N. B., S. A. H. and C. D. T. All authors approved the manuscript.

# Tables

*Table 1: Analysis of molecular variance (AMOVA) on the final dataset with 10 000 permutations (***: p<0.001, ** : p<0.01 *: p<0.05)*

| | | |
|---|---|---|
| Manus/Woodlark vs. North-Fiji/Futuna/Lau | **0.38773*** ** | $F_{ST}$ |
| Basins in M/W and NF/F/L | -0.05 | $F_{CT}$ |
| Localities within Basins | -0.00011 | $F_{SC}$ |
| Individuals within Localities | -0.05084 | $F_{IS}$ |

*M/W: Manus/Woodlark, NF/F/L: North-Fiji/Futuna/Lau*

*Table 2: Pairwise (between basins) $F_{ST}$ matrix on the final dataset with 10 000 permutations after Bonferroni correction (***: p<0.001, **: p<0.01, *p<0.05).*

| | **Lau** | **Futuna** | **North-Fiji** | **Manus** | **Woodlark** |
|---|---|---|---|---|---|
| **Lau** | 0.00000 | | | | |
| **Futuna** | -0.00040 | 0.00000 | | | |
| **North-Fiji** | 0.00029* | -0.00004 | 0.00000 | | |
| **Manus** | 0.38350*** | 0.38275*** | 0.37651*** | 0.00000 | |
| **Woodlark** | 0.39986*** | 0.39656*** | 0.38647*** | -0.00016 | 0.00000 |

*Table 3: Parameters estimated from ∂a∂i for the three best models (IM2N2mG, SC2N2mG and AM2N2mG*) with their standard deviations (SD) estimated using a Fisher information matrix. (*isolation with migration (IM), secondary contact (SC), ancient migration (AM) and with parameters describing effective population size (2N), migration rate (2m) and population growth (between basins))*

| Parameter | IM+2N+2m+G | SD | SC+2N+2m+G | SD | AM+2N+2m+G | SD |
|---|---|---|---|---|---|---|
| **$N_1$ (NF/F/L)** | 0.435 | 0.187 | 0.913 | 0.573 | 0.390 | 0.127 |
| **$N_2$ (M/W)** | 0.411 | 0.157 | 0.840 | 0.573 | 0.356 | 0.119 |
| **b1** | 30.410 | 13.176 | 16.947 | 8.127 | 34.367 | 12.399 |
| **b2** | 25.097 | 8.683 | 13.714 | 8.451 | 29.288 | 10.523 |
| **hrf** | 0.023 | 0.006 | 0.021 | 0.006 | 0.022 | 0.006 |
| **Ts** | | | 0.443 | 0.527 | 0.001 | 0.024 |
| **Tsm/Tsc/Tam** | 1.631 | 0.280 | 1.470 | 0.379 | 1.681 | 0.335 |
| **m12** | 0.444 | 0.147 | 0.422 | 0.145 | 0.461 | 0.137 |
| **m21** | 0.825 | 0.192 | 0.810 | 0.261 | 0.825 | 0.198 |
| **me12** | 0.038 | 0.020 | 0.038 | 0.024 | 0.039 | 0.018 |
| **me21** | 0.283 | 0.063 | 0.270 | 0.088 | 0.300 | 0.070 |
| **P** | 0.483 | 0.103 | 0.439 | 0.093 | 0.471 | 0.095 |
| **Q** | 0.990 | 0.136 | 0.990 | 0.185 | 0.990 | 0.136 |
| **Theta** | 271.772 | 31.55 | 243.283 | 64.91 | 273.222 | 32.988 |

*N represents the population size of each population; **b**, the population growth factor; **hrf**, the Hill-Robertson factor; **Ts**, the time of strict divergence; **Tm/Tsc/Tam**, the time of divergence with migration; **m12**, represents the unrestricted migration rate from the population 2 towards population 1; **me12**, the restricted migration rate (e.g. barrier loci) from population 2 towards population 1; **Q**, the proportion of loci that are under the effect of linked selection (i.e. Hill-Robertson effect); **P**, the proportion of loci that have unconstrained migration; M/W, Manus/Woodlark; NF/F/L, North-Fiji/Futuna/Lau*

*Table 4: Estimates of the effective number of migrants($N_m$) exchanged per generation between metapopulations, total time of divergence since the population split and effective population size ($N_e$) for three demographic models (isolation with migration (IM), ancient migration (AM), and secondary contact (SC).*

| | $N_m$2->1 (M/W->NF/F/L) | $N_m$1->2 (NF/F/L-> M/W) | T (in generations) | $N_1$ (NF/F/L) | $N_2$ (M/W) |
|---|---|---|---|---|---|
| **IM+2N+2m+G** | 2.935 | 4.255 | 66 951 | 271 506 | 211 708 |
| **SC+2N+2m+G** | 3.265 | 4.665 | 70 295 | 284 279 | 211 653 |
| **AM+2N+2m+G** | 3.09 | 4.305 | 69 380 | 276 561 | 215 142 |

*Metapopulation M/W, Manus/Woodlark; metapopulation NF/F/L, North-Fiji/Futuna/Lau.*

# Figures

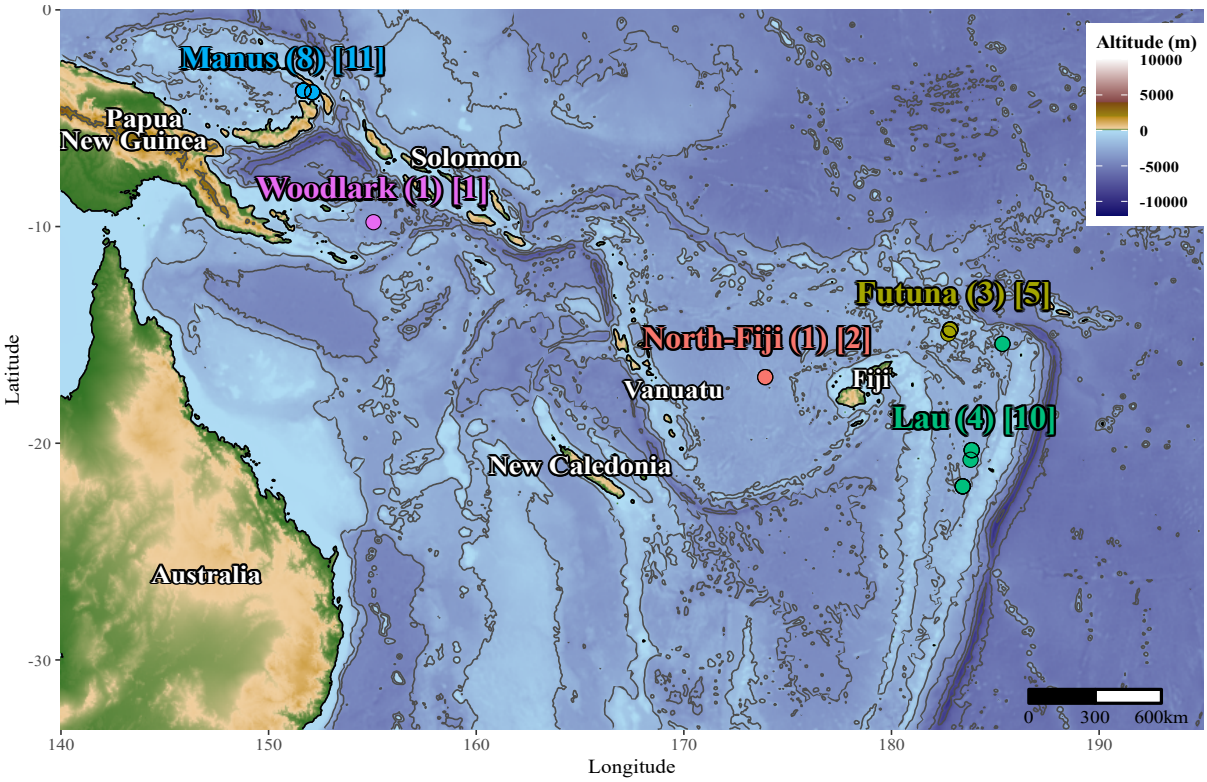


*Figure 1: Colors: Back-arc basins. Sampling map of Ifremeria nautilei in the Southwestern Pacific Ocean. The number of localities is given in parentheses and the total number of sampled sites in brackets.*

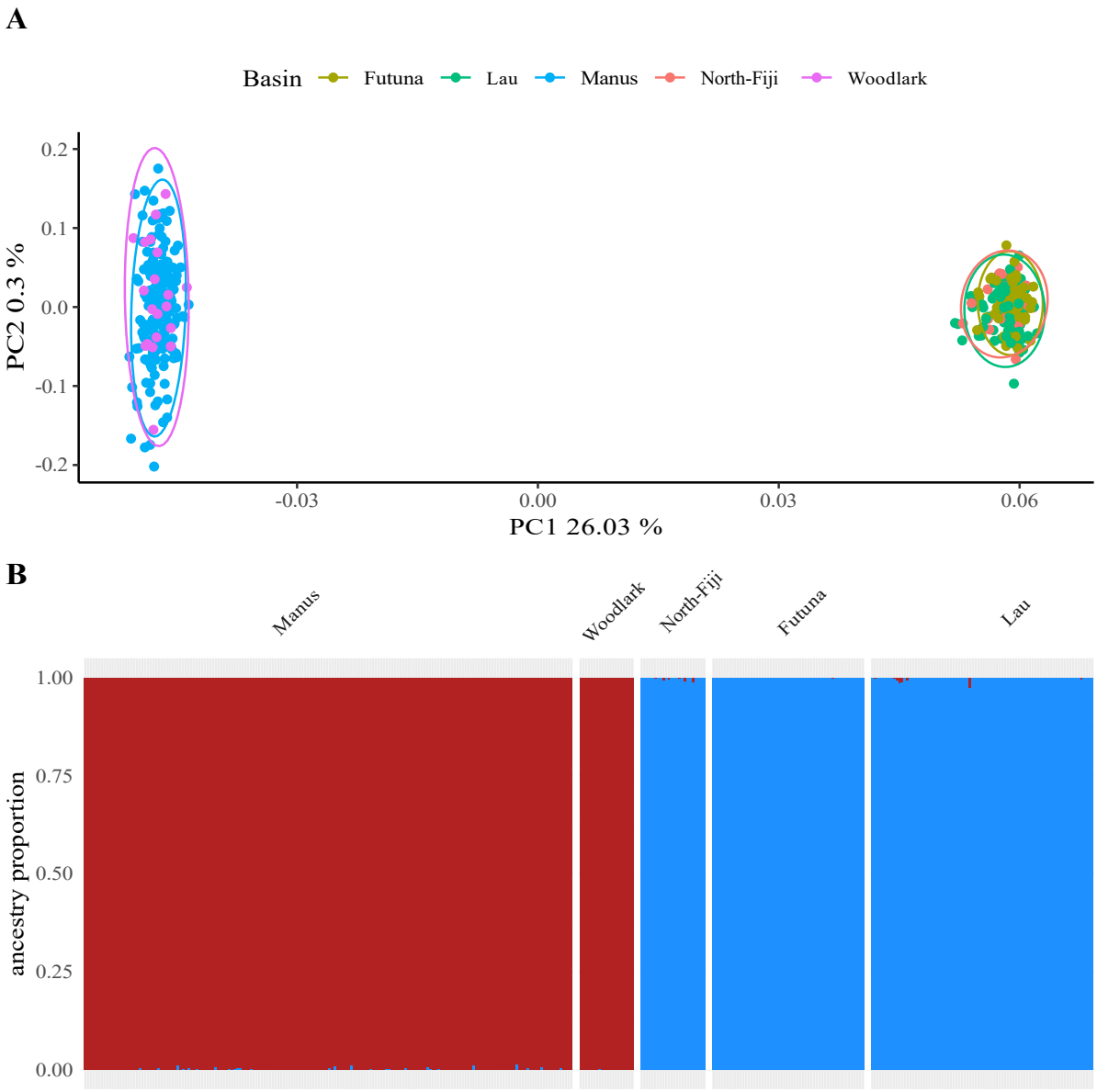


*Figure 2: (**A**) Principal component analysis plot of 362* Ifremeria nautilei *individuals from five hydrothermal basins scored at 10 570 SNPs, open circles represent the multivariate normal distribution of each group (basins) at 95%. (**B**) ADMIXTURE plot for each individual with their ancestry proportions obtained on the final dataset for the best K (K=2). Individuals are grouped according to their basin of origin.*

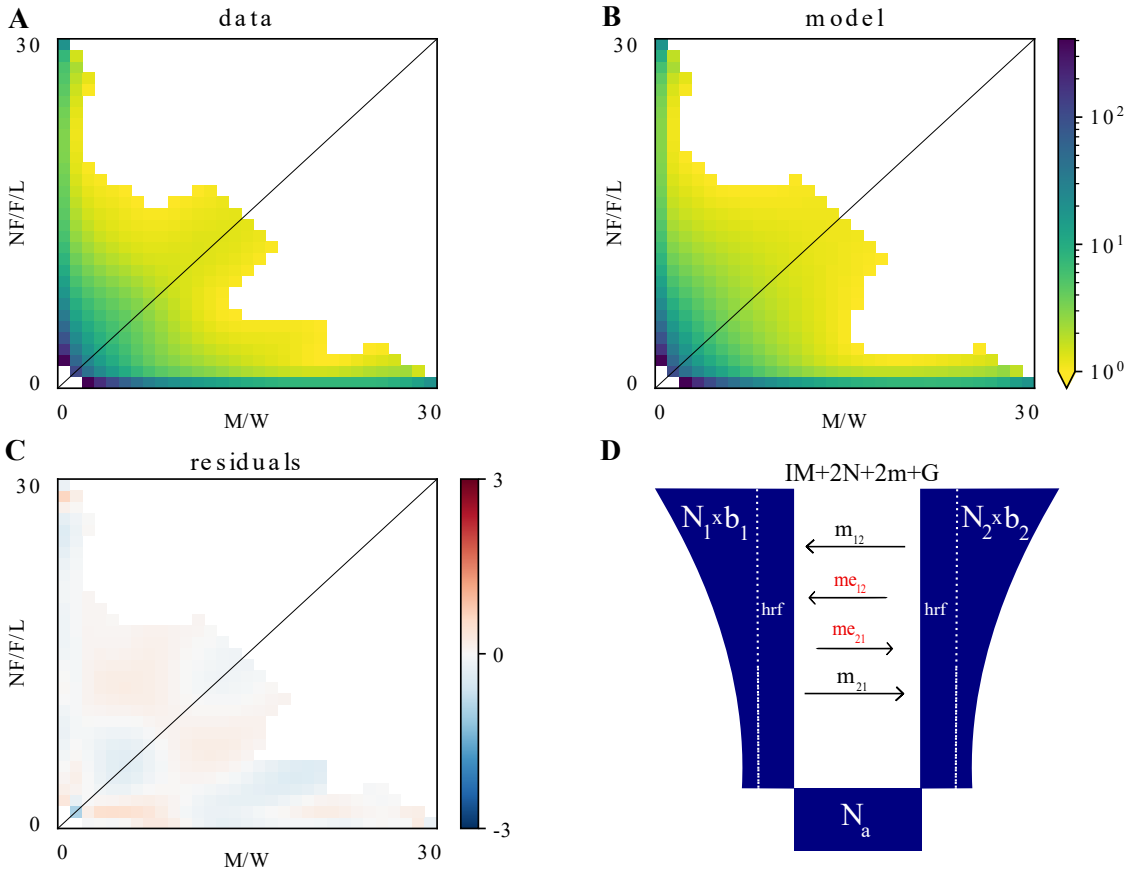


*Figure 3: (**A**) Joint allele frequency spectrum (JAFS) between the Manus/Woodlark (M/W) and North Fiji/Futuna/Lau (NF/F/L) basin systems. (**B**) Simulated JAFS under the IM2N2mG model (see Figure 4), the log scale indicates the density of SNPs in each frequency class. (**C**) Residuals of the fit of the simulated model on the data. (**D**) Representation of the fitted model. (**N** represents population size; **b**, population growth factor; **hrf**, the Hill-Robertson factor, which simulates linked selection; **m**, unrestricted migration rate; **me**, restricted migration rate, which simulates barrier loci)*

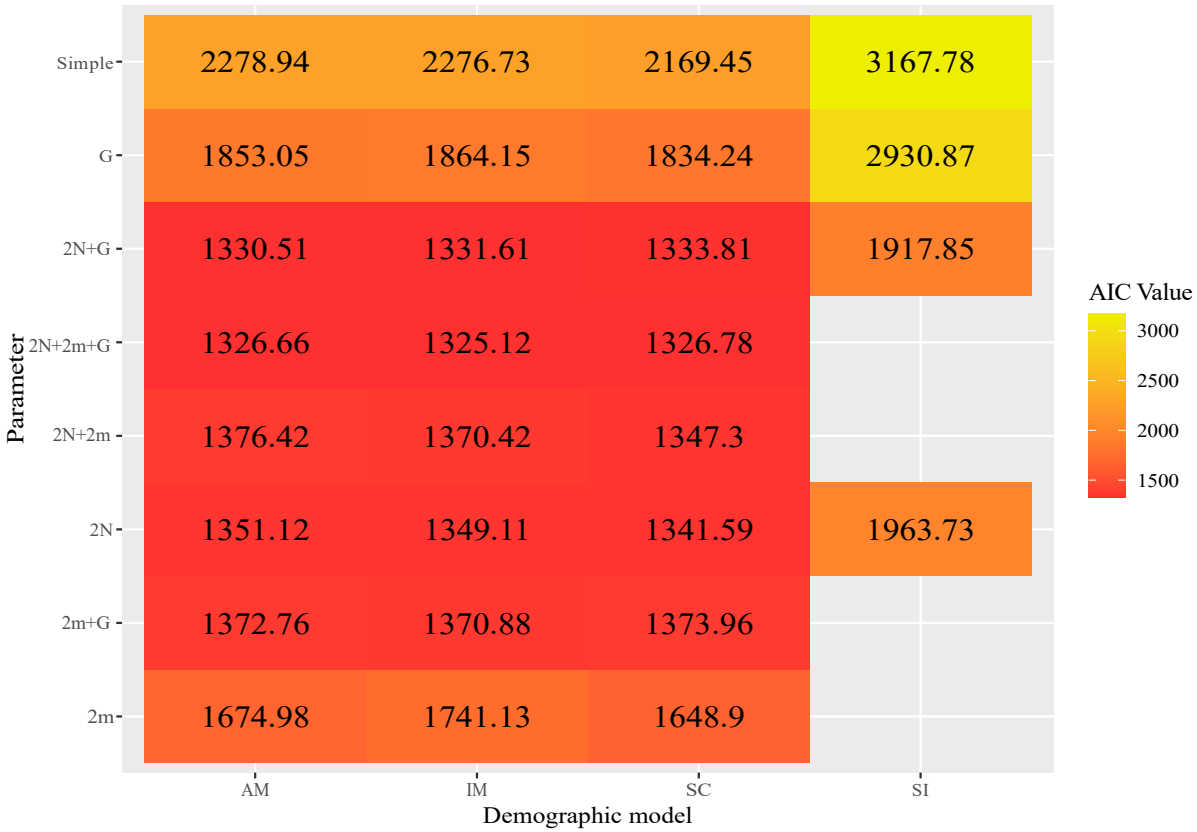


*Figure 4: Heat-map of the best Akaike information criterion (AIC) value for each parameter combination (population expansion or contraction (G), effect of linked selection (2N) and heterogeneous migration (2m)) and demographic model (strict isolation (SI), isolation with migration (IM), ancient migration (AM), and secondary contact (SC)).*

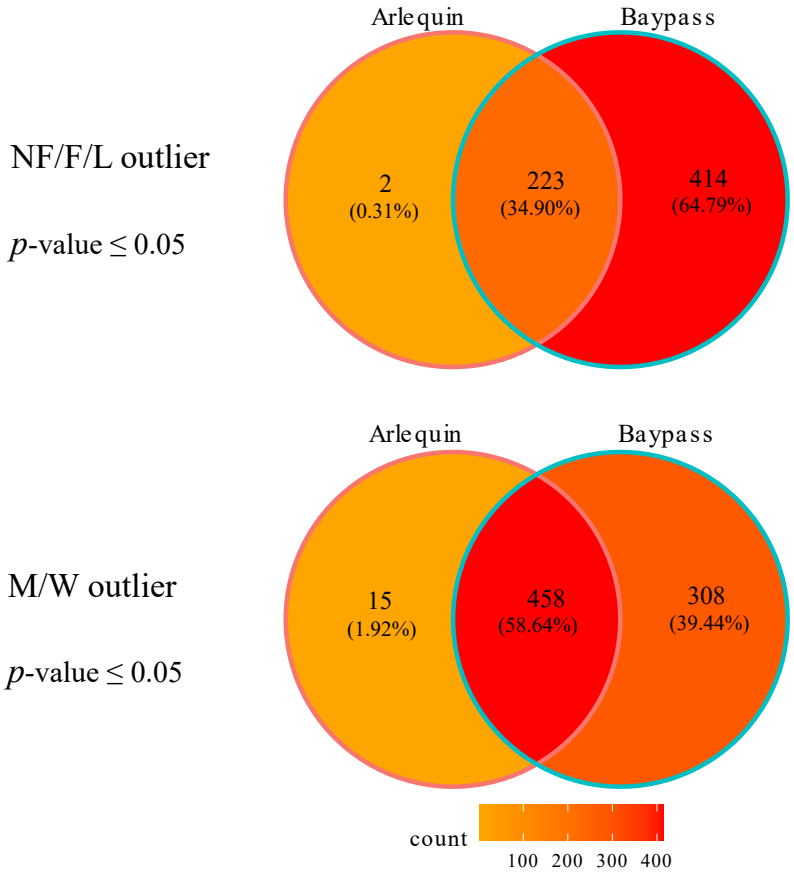


*Figure 5: Venn diagram of shared outlier loci identified in Arlequin and Baypass with a* p-*value less than or equal to 0.05 within each metapopulation (Manus/Woodlark, M/W and North Fiji/Futuna/Lau, NF/F/L) independently.*

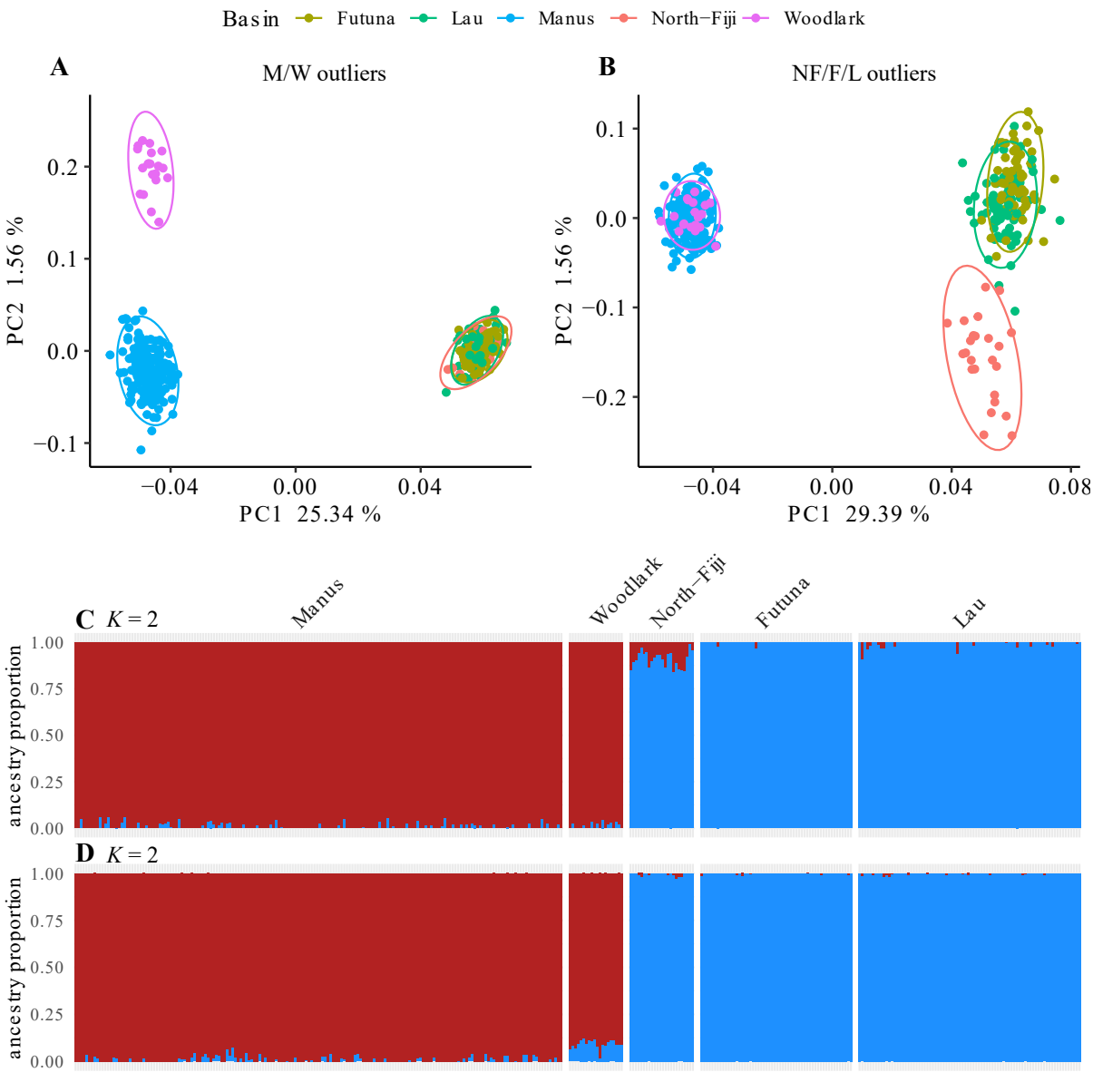


*Figure 6: Principal component analysis on all individuals with the outlier loci found in both Arlequin and Baypass at a 0.05 p-value threshold. (A) Outlier loci detected in Manus/Woodlark (M/W). (B) Outlier loci detected in North-Fiji/Futuna/Lau (NF/F/L). Plot of ancestry proportion inferred with Admixture at K = 2 on all individual at the 0.05 p-value threshold, for (C) North-Fiji/Futuna/Lau outliers and (D) for Manus/Woodlark outliers.*