Exploring the microdiversity within marine bacterial taxa: Towards an integrated biogeography in the Southern Ocean

Schwob G[1,5]*, Segovia NI[4,5], González-Wevar CA[2,5], Cabrol L[3,5], Orlando J[1] and Poulin E[1,5]

(1) Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

(2) Instituto de Ciencias Marinas y Limnológicas & Centro Fondap IDEAL, Universidad Austral de Chile, Facultad de Ciencias, Valdivia, Chile

(3) Mediterranean Institute of Oceanography (MIO) UM 110, CNRS, IRD, Aix Marseille University, Univ Toulon, Marseille, France

(4) Universidad Católica del Norte, Coquimbo, Chile

(5) Instituto de Ecología y Biodiversidad, Santiago, Chile

* Corresponding author: guillaume.schwob@uchile.cl

Abstract

The phylogeography traditionally correlates the genetic relationships among individuals within a macroorganism species, to their spatial distribution. Most microbial phylogeographic studies so far have been restricted to narrow geographical regions, mainly focusing on isolated strains, either obtained by culture or single-strain natural enrichments. However, the laborious culture-based methodology imposes a low number of studied individuals, leading to poor resolution of haplotype frequency estimation, making difficult a realistic evaluation of the genetic structure of natural microbial populations in the environment.

To tackle this limitation, we present a new approach to unravel the phylogeographic patterns of bacteria combining (i) community-wide survey by 16S rRNA gene metabarcoding, (ii) intra-

species resolution through the oligotyping method, and (iii) genetic and phylogeographic indices, as well as migration parameters, estimated from populational molecular data as traditionally developed for macroorganisms as models.

As a proof-of-concept, we applied this methodology to the bacterial genus *Spirochaeta*, classically reported as a gut endosymbiont of various invertebrates inhabiting the Southern Ocean (SO), but also described in marine sediment and in open waters. For this purpose, we centered our sampling into three biogeographic provinces of the SO; maritime Antarctica (King George Island), sub-Antarctic Islands (Kerguelen archipelago) and Patagonia in southern South America. Each targeted OTU was chaLRracterized by substantial intrapopulation microdiversity, a significant genetic differentiation and a robust phylogeographic structure among the three distant biogeographic provinces. Patterns of gene flow in *Spirochaeta* populations support the role of the Antarctic Polar Front (APF) as a biogeographic barrier to bacterial dispersal between Antarctic and sub-Antarctic provinces. Conversely, the Antarctic Circumpolar Current (ACC) appears as the main driver of connectivity between geographically distant sub-Antarctic areas such as Patagonia and Kerguelen archipelago, and between Kerguelen archipelago and maritime Antarctica. Additionnally, we found that historical processes (drift and dispersal limitation) together govern up to 86% of the spatial turnover among *Spirochaeta* populations. Overall, our approach represents a substantial first attempt to bridge the gap between microbial and macrobial ecology by unifying the way to study phylogeography. We revealed that strong congruency with macroorganisms patterns at the populational level shaped by the same oceanographic structures and ecological processes.

Keywords (3 to 10 required)

Southern Ocean, Antarctic Polar Front, Antarctic Circumpolar Current, *Abatus, Spirochaeta*, oligotyping, microdiversity, phylogeography, 16S rRNA metabarcoding, Minimum Entropy Decomposition, Microbial Conveyor Belt

Background

Biogeography has traditionally investigated the geographic distribution of macroorganisms in the Eukaryota domain. However, during the last two decades, a growing number of studies has focused on the biogeography of prokaryotic microorganisms, taking advantage of the breakthrough and the constant advances of next-generation sequencing (NGS), which allow extensive surveys of previously inaccessible microbial diversity from a wide range of ecological contexts [1]. Although long debated in the past, it is now accepted that microbes do have biogeographic patterns, repeatedly illustrated by the observation of non-random community assemblages of various prokaryotic microorganisms [2-4]. Contrary to the contemporary driving factors (*i.e.* environmental selection) that have been extensively studied [5-9], the role of historical processes –past ecological and evolutionary events– onto the present-day distribution patterns of microorganisms remains poorly investigated. Initially, the consensus was that the rapid and widespread dispersal of microbes should erase any signal of past historical events [1]. Nevertheless, it is now clear that historical processes, such as the dispersal barriers and geographic distance, might substantially contribute to microbes' biogeography instead of environmental filtering [5, 10, 11]. For instance, biogeographic regionalization, isolation, and endemism have been reported for microbes, as well as in larger organisms, and evidence a predominant effect of geographic distance over environmental variations [12, 13].

To date, most of the microbial biogeographic patterns have been depicted at the whole community level [4, 14, 15]. Nevertheless, as observed in various empirical studies, a finer taxonomic scale generally allows better detection of geographic patterns [10, 16, 17]. Moreover, the ecological processes driving the biogeographic patterns at the community-level intrinsically result from the accumulation of micro-evolutive processes, *i.e.* mechanisms contributing to the genetic composition and diversity within populations, and how they vary in space and time [10, 18]. The comprehensive description of these micro-evolutive processes requires considering intra-population diversity, as commonly applied in phylogeographic studies of macroorganisms.

Hence, microbial assembly processes need to be investigated at a finer taxonomic resolution than usually done by microbial biogeographic surveys and consider the "microdiversity" within groups [17, 18].

The oceans have been considered among the most challenging environments to test hypotheses about microbial biogeography, mainly due to the speculated transport of organisms over large distances by marine currents and the absence of perceivable marine barriers impeding potential dispersal events [19]. However, oceanic fronts separating different water masses have been recently identified as major microbial dispersal barriers [20]. The Southern Ocean (SO) is a vast region representing approximately 20% of the world ocean surface. It surrounds Antarctica, and its northern limit is the Subtropical Front [21]. Two main oceanographic structures shape the SO biogeography; the Antarctic Polar Front (APF) and the Antarctic Circumpolar Current (ACC). The APF is classically considered a harsh north-south obstacle for dispersing marine organisms due to the brutal change in water temperature and salinity [22, 23]. Phylogenetic reconstruction achieved on various vertebrate and invertebrate taxa clearly supports the role of the APF on their respective diversification processes [24-29]. Accordingly, and based on the described distribution of species, biogeographers have traditionally recognized Antarctica and sub-Antarctica as the two main biogeographic provinces of the SO, even if several provinces have been proposed within each of them [30]. Contrarily, outside the APF, the ACC is generally described as the driver of genetic connection across the sub-Antarctic zone due to its clockwise circulation [31-35]. Intraspecific genetic and phylogeographic studies of macroorganisms have demonstrated the ACC's role in connecting geographically distant sub-Antarctic provinces [32, 36-38]. The marine biota distribution in this region has been synthesized in the Biogeographic Atlas of the SO, providing updated biogeographic information of a wide range of benthic and pelagic taxa from Metazoan, macroalgae, and phytoplankton [39]. Despite being the most abundant and diverse domains on Earth, Bacteria and Archaea were not included in the SO Atlas [40]. Even when marine microbial communities have been previously characterized in the

region, their geographic distribution's underlying drivers remain poorly understood. Studies conducted at the whole community-level support (i) the role of ACC as a likely efficient mechanism of circumpolar microbial transport and dispersal [3, 41] and (ii) the role of APF as the main dispersal barrier separating Antarctic and sub-Antarctic microbial assemblages [3, 20, 42]. However, the observed biogeographic patterns' underlying evolutionary processes have not been elucidated and would intrinsically rely on bacterial populations' microdiversity.

Targeting the intraspecific diversity using NGS data requires specific computational methods to discriminate the stochastic noise caused by random sequencing errors from those associated with biologically significant diversity [10, 17, 43]. For this purpose, an algorithm called "Minimum Entropy Decomposition" (MED) relying on the oligotyping method has been proposed by Eren *et al*. [44]. This algorithm allows to identify true sequence variants (*i.e.* oligotypes) within the "Operational Taxonomic Units" (OTUs), classically defined at 97% identity of the bacterial 16S rRNA gene. This approach has already been successfully used to unravel fine-grained biogeographic patterns of bacterial microdiversity in Arctic sediments, such as variations in oligotype distribution according to spatial and environmental parameters [45]. Moreover, focusing on the sulfate-reducing genus *Desulfotomaculum* in Arctic marine sediments, Hanson *et al*. [11] showed clear biogeographic patterns –attributed to historical factors associated with past environments– were only evident at the microdiversity level achieved with the oligotyping method. However, the microevolutionary processes causing the microdiversity were not assayed, and the study did not encompass large-scale distribution among different biogeographic provinces, as it was restricted to the west coast of Spitsbergen, Svalbard in the Arctic Ocean.

In the present proof-of-concept study, we aim to bring new insights on the evolutionary processes driving microbial biogeography across different provinces of the SO by combining (i) community-wide surveying provided by the high-throughput sequencing of the 16S rRNA gene, (ii) intra-species resolution obtained through the oligotyping method implemented in the MED pipeline, and (iii) phylogeographic analysis as traditionally developed for macroorganisms as models.

Considering the SO as an outstanding idoneous frame, we investigated the geographic distribution of genetic diversity of marine bacterial taxa across three main biogeographic provinces: maritime Antarctica (King George Island, South Shetland Islands, West Antarctic Peninsula), sub-Antarctic Islands of the Indian Ocean (Kerguelen archipelago) and southern South America (Patagonia), encompassing sites separated by the APF, and others connected through the ACC.

As the contribution of geography to biological diversity patterns (*i.e.* dispersal limitation) is stronger on habitat-specialists (*i.e.* taxa found in habitat with high selective strength) [46, 47], and emphasized within homogeneous habitats distributed across large spatial scales [10, 48, 49], we focused our study on the bacterial community associated to a specific habitat: the gut of *Abatus* irregular sea urchins. The *Abatus* genus is distributed across the SO and gathers various sibling species homologous in ecology and habitat, such as *Abatus cavernosus* in southern South America, *Abatus cordatus* in Kerguelen Islands, and *Abatus agassizii* in maritime Antarctica [50-53]. Since these species lack specialized respiratory structure, they are restricted to the well-oxygenated coarse sediments found at shallow depth (typically 1 to 3 meters depth) in sheltered bays, protected from the swell [51]. Within the *Abatus* hosts, we focused on a specific micro-environment –the gut tissue– previously described to act as a selective filter of the external sediment microbiota, as illustrated by the reduction of bacterial diversity at both taxonomic and functional levels [54]. Working on the gut community with supposedly more limited dispersal capacity, through a high sequencing depth, is expected to (i) provide robust coverage of the bacterial diversity, (ii) minimize the relevance of environmental filtering between provinces, (iii) emphasize the contribution of geographic and oceanographic factors, and therefore (iv) enhance the detection of phylogeographic signals across the SO [10]. As a model taxon to explore bacterial phylogeography in the SO, we chose the *Spirochaeta* genus (phylum *Spirochaetes*). *Spirochaeta* bacteria are recognized as the most prevalent and abundant genus in the *Abatus* gut tissue [54]. Moreover, spirochaetes are classically found in marine benthic sediments [55, 56]

and, to a lesser extent, in the water column (Ocean Barcode Atlas; http://oba.mio.osupytheas.fr/ocean-atlas/). Thus, due to its ease of detection and ubiquity across biogeographic provinces, *Spirochaeta* represents an illustrative model to validate our methodology and explore marine bacteria's spatial genetic patterns, from genus to population level. We hypothesized that the strong biogeographic barrier between South America and maritime Antarctica classically observed in the literature for macroorganisms (*i.e.* vicariance process) also affects the fine-scale genetic structure and the phylogeographic patterns within *Spirochaeta* OTUs. Besides, the ACC-mediated connectivity among sub-Antarctic provinces should be reflected by a greater genetic homogeneity of *Spirochaeta* populations between South American sites and the Kerguelen Islands, rather than with maritime Antarctica. Alternatively, the potential high dispersal capacity of *Spirochaeta* taxa may result in the absence of genetic and phylogeographic structure across the SO.

Methods

Sampling collection, DNA extraction, and 16S rRNA gene library preparation

Adult *Abatus* individuals were sampled from four localities across the SO, including two sites in Patagonia, southern South America (Possession Bay, PAT1 and Puerto Deseado, PAT2), one site in Kerguelen Islands (Port-aux-Français, KER), and one site in the West Antarctic Peninsula (King George Island, KGI) (Figure 1, Table 1).

Table 1. Experimental design and sequencing data

Locality	Province	GPS coordinates	Date	Designation	Sample types	N	Nseq. (Relat. Abund.)
King George Island	Maritime Antarctica	62°12'55.3"S 58°56'43.8"W	01-2019	KGI	External sediments	8	255786 (10%)
					Gut tissue	31	563383 (22%)
Bahía Posesión	Patagonia	52°19'52.97"S 69°29'10.50"W	07-2019	PAT1	External sediments	6	271828 (11%)
					Gut tissue	15	447892 (18%)
Puerto Deseado	Patagonia	47°45'07.0"S 65°52'04.0"W	12-2016	PAT2	External sediments	*NA*	*NA*
					Gut tissue	10	470087 (18%)
Port-aux-Français	Kerguelen Island	49°21'13.32''S 70°13'8.759''E	11-2017	KER	External sediments	5	92564 (4%)
					Gut tissue	14	440498 (17%)

N: number of samples, Nseq.: total number of cleaned sequences obtained, Relat. Abund.: (relative abundance in the global dataset).

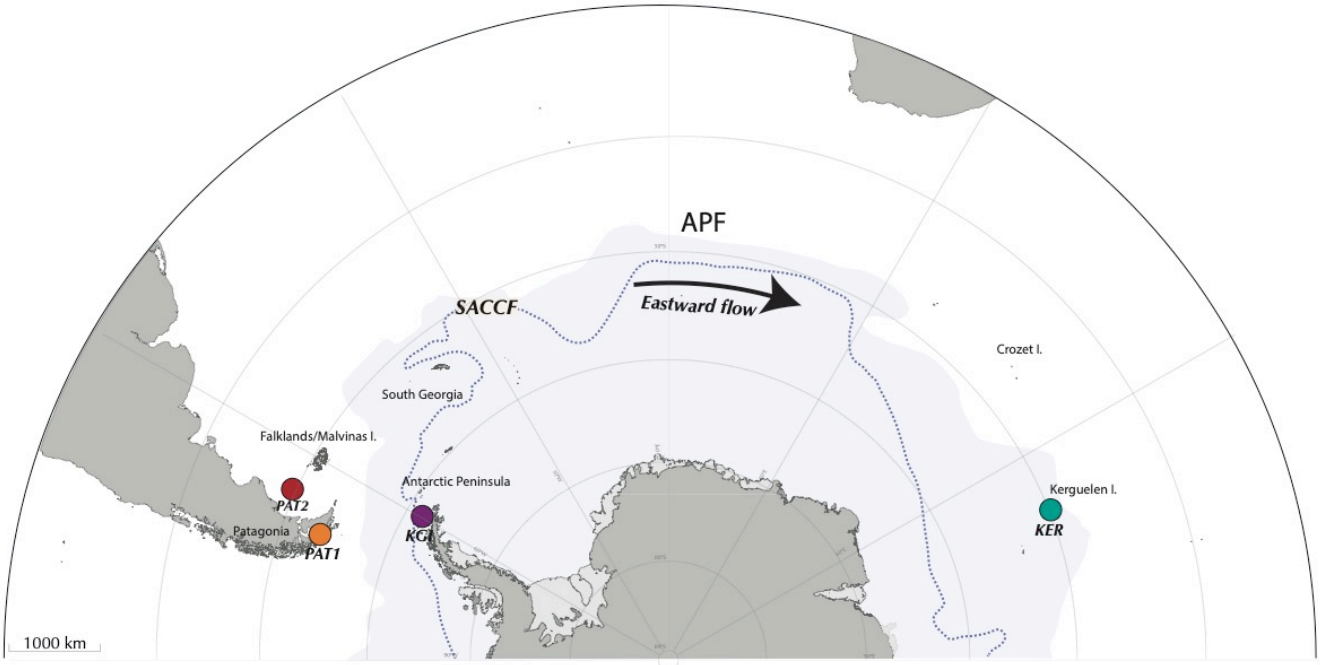

Figure 1. Sampling localities across the Southern Ocean, encompassing Possession Bay and Puerto Deseado in Patagonia (PAT1 and PAT2, respectively), King George Island in Maritime Antarctica (KGI), and Port-aux-Français in Kerguelen Islands (KER). The Antarctic Polar Front (APF) and the Southern Antarctic Circumpolar Front (SACCF) are represented.

Marine surface sediments (0 – 5 cm, referred here as "external sediment") were also sampled in each *Abatus* population's immediate vicinity as the ingested food source of the sea urchins. Due to logistic constraints, it was not possible to collect external sediment in the PAT2 site. All individuals were dissected under sterile conditions to collect the whole digestive tract minus the caecum (identified as "gut tissue") following Schwob *et al.* [54]. Gut tissue samples were gently rinsed with nuclease-free sterile water to remove the content (*i.e.* in sediment) and were then individually homogenized using mortar and pestle under a laminar-flow cabinet. Genomic DNA was extracted from the totality of the homogenized gut tissue samples using the DNeasy PowerSoil® Kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations.

A metabarcoding approach was used to assess the bacterial community composition in the external sediment and *Abatus* gut tissue samples. Briefly, extracted genomic DNA was used as the template for PCR amplification using the primers 515F 5′-GTGYCAGCMGCCGCGGTA-3' and 926R 5′-CCCCGYCAATTCMTTTRAGT-3' [57]. The PCR conditions and 16S rRNA gene library preparation were the same as described in Schwob *et al.* [54].

187 **Metabarcoding data processing**

188 External sediment and gut tissue amplicons were sequenced using the paired-end sequencing
189 technology (2 x 250 bp) on the Illumina MiSeq platform at the UWBC DNA Sequencing Facility
190 (University of Wisconsin-Madison, USA). Reads of 16S rRNA were processed using the open-
191 source software MOTHUR v1.44.0 following Schwob *et al.* [54]. Once the raw reads were
192 processed into Operational Taxonomic Units (OTUs) at 97% identity threshold, we applied a
193 filter of relative abundance at > 0.0001% as recommended by Bokulich *et al.* [58]. Following
194 this, a taxonomic classification was performed with the *classify.otu* function and the SILVA
195 database v138 implemented in MOTHUR. An OTU table of *Spirochaeta* was edited (*i.e.* all OTUs
196 assigned to the genus *Spirochaeta*), and converted into a presence/absence matrix. Bray-Curtis
197 and Unweighted Unifrac distances were calculated from the OTU presence/absence matrix and
198 used to perform a Non-Metric Multidimensional Scaling (NMDS) with the *metaMDS* function of
199 the ADE4 package [59]. The NMDS was plotted through a scatter diagram using the *s.class*
200 function implemented in ADE4. The locality's contribution to the *Spirochaeta* OTUs composition
201 in gut tissue samples was tested with permutational multivariate analysis of variance
202 (permanova), using *adonis* and *pairwise.adonis* functions implemented in *vegan* and
203 *pairwiseAdonis* R packages, respectively [60, 61].
204 A subset of the three most abundant *Spirochaeta* OTUs present in the four localities was retained
205 for further analysis (Table S2). All the sequences assigned to the selected *Spirochaeta* OTUs were
206 retrieved using the *bin.seqs* command in MOTHUR. Finally, the resulting fasta files were processed
207 independently through the Minimum Entropy Decomposition pipeline following Eren *et al.* [44].

208 **Minimum Entropy Decomposition**

209 Minimum Entropy Decomposition (MED) pipeline was used to identify nucleotidic
210 polymorphism at fine-scale resolution (> 3% identity) within the 16S rRNA gene sequences from
211 *Spirochaeta* OTUs. Briefly, MED employs the Shannon entropy algorithm to discriminate
212 biologically meaningful variations of closely related sequences from the stochastic noise caused

by random sequencing errors, focusing on informative-rich variable nucleotide positions [44, 62]. The resulting taxonomic units will be referred to as *Spirochaeta* oligotypes. Unsupervised oligotyping was carried out individually on *Spirochaeta* OTUs using the default dynamically computed threshold from which entropy is considered as zero (-m). Additionally, each identified oligotype had to have a default minimum relative abundance of 2% in the OTU sequences dataset [44]. Accumulation curves of oligotypes' richness were computed for each *Spirochaeta* OTU at a 95% confidence interval using the package iNEXT [63] in R v3.6.0 [64]. Pie charts of the relative site contributions in the total abundance of the *Spirochaeta* OTUs oligotypes were performed with the *pie* function in the package GRAPHICS in R v3.6.0.

Genetic diversity and structure of *Spirochaeta* populations

The number of oligotypes (*k*), the oligotype diversity (*H*), the number of discriminant sites (*S*) and the pairwise difference between sequences (Π) were estimated individually for each *Spirochaeta* OTU using the packages PEGAS [65] and APE v5.3.0 [66] in R v3.6.0. For comparative purposes among sites with unequal sample sizes, a composite bootstrapping script was written to rarefy the sequence datasets to the minimum number of sequences per site and repeat the rarefaction with 1,000 re-samplings. Confidence intervals at 95% of genetic diversity indices were then calculated using these iteration values. The genetic differentiation (F_{ST} and Φ_{ST}) among *Spirochaeta* populations was analyzed using the software ARLEQUIN v3.5.2 [67] with 1,000 permutations and a significance threshold at 0.05. Phylogeographic differentiation was also estimated with the nearest-neighbor statistic Snn [68], and the significance of Snn estimates was tested with a permutation test through DNASP v5.10.01 [69]. The reconstruction of oligotype networks was performed using the Median Joining method with the software Populational Analysis with Reticulate Trees v1.7.0 in PopART [70]. Oligotype abundances were normalized by the number of sequences per locality for a given OTU to improve the networks' readability.

Quantification of selection, dispersal, and drift

The relative contribution of stochastic (*i.e.* dispersal, drift) and deterministic (*i.e.* selection) processes, also referred to as historical and contemporary processes respectively, on *Spirochaeta* oligotype assembly was measured for the selected OTU by following the analytical framework described in Stegen *et al.* [71] and illustrated by Feng *et al.* [72]. In a nutshell, the approach relies on the comparison of the phylogenetic turnover between communities across samples (βMNTD; β mean nearest-taxon distance) to a null distribution of βMNTD, and denoted as the β-nearest taxon index (βNTI). The phylogenetic tree required for the βMNTD/βNTI calculations was generated using PhyML v3.0 [73], and the oligotype sequences of *Spirochaeta* previously aligned with MUSCLE [74]. βNTI values indicate that taxa between two communities are more (*i.e.* βNTI < -2) or less (*i.e.* βNTI $> +2$) phylogenetically related than expected by chance, thus suggesting that communities experience homogenizing or variable selection, respectively [75]. βNTI values ranging from -2 to +2 indicate a limited selection effect and point to dispersal limitation and ecological drift out as possible community composition drivers. To further disentangle the respective effect of these two alternatives processes, we calculated the pairwise Bray Curtis-based Raup-Crick dissimilarity index (RC_{Bray}) between sites [76], weighted by oligotype abundance [71]. For this, we used an optimized version of the initial script of Stegen *et al.* [71], developed by Richter-Heitmann [77], and available via GitHub (https://github.com/FranzKrah/raup_crick). RC_{Bray} values < -0.95 and $> +0.95$ correspond to communities that have –respectively– more or fewer taxa in common than expected by chance, and therefore indicate that community turnover is driven by homogenising dispersal ($RC_{Bray} < -0.95$) or dispersal limitation plus drift ($RC_{Bray} > +0.95$). On the contrary, RC_{Bray} values > -0.95 and $< +0.95$ are indicative of ecological drift [78]. Both βNTI and RC_{Bray} null models included 999 randomizations [71].

Testing for isolation by distance (IBD) and environment (IBE)

To disentangle the relative effect of geographic distance and abiotic environmental differences on the *Spirochaeta* oligotype composition between samples, we used the distance-based multiple matrix regression with randomization (MMRR) approach [79]. We extracted a set of 9

environmental variables for each of our sampling site from the Bio-ORACLE database [80], including pH, the means of nitrate, silicate, and phosphate concentrations, and the means at the mean depth of seawater salinity, dissolved oxygen concentration, seawater temperature, seawater temperature range and chlorophyll concentration. All environmental variables were standardized ($(x_i - \bar{x})/sd(x)$), and were then analyzed through Principal Components Analysis (PCA). As a high percentage of the variation among localities was explained by the first component (PC1, >91%), we transformed the scores of PC1 into Euclidean distance using the *vegdist* function in the *vegan* package in R to use it as the environmental distance matrix further. The longitude and latitude coordinates were converted into kilometers using the *earth.dist* function implemented in the FOSSIL package in R [81]. The geographic distances were transformed using the Hellinger method through the *decostand* function of the *vegan* package in R. The dissimilarity matrix of *Spirochaeta* oligotype composition among samples was also created from Bray-Curtis distances using the *vegdist* function of the R package *vegan*. Finally, to evaluate the relative weight of environmental and geographic distance matrices, an MMRR was performed using the R package PopGenReport [82], and the correlation coefficients and their significance were estimated based on 10,000 random permutations.

Connectivity among *Spirochaeta* populations

The amount and direction of gene flow among *Spirochaeta* populations were estimated using the coalescent-based program LAMARC v2.1.10 [83]. A total of ten runs was performed for each *Spirochaeta* OTU, consisting of likelihood searches of 20 initial and two final chains, with a minimum of 500 and 10,000 recorded trees, respectively, and sampling every 20 generations after a burn-in of 1,000 genealogies. The effective number of migrants per generation (Nm) among *Spirochaeta* populations was calculated by multiplying the Maximum Likelihood Estimates (MLE) of the mutation parameter (θ) by the migration parameter (M), both estimated through the LAMARC program. We present the mean and standard deviation of the estimated Nm values obtained from the ten runs for each *Spirochaeta* OTU.

Results

Sequencing performance and OTUbased analysis

A total of 4,184,226 raw reads was generated from the 91 samples of external sediments and gut tissues. Once processed, 2,542,038 cleaned sequences distributed into 727 OTUs were obtained from the external sediment and gut tissue samples (details provided in Table 1). Out of this condensed dataset, 425,613 sequences associated with the *Spirochaeta* genus were retrieved, representing a total of 10 OTUs.

Both Bray-Curtis and Unweighted Unifrac distance methods did not reveal any difference in *Spirochaeta* OTU composition between the Patagonian sites (PAT1 and PAT2) (Figure 2A and 2B, Additional file 1).

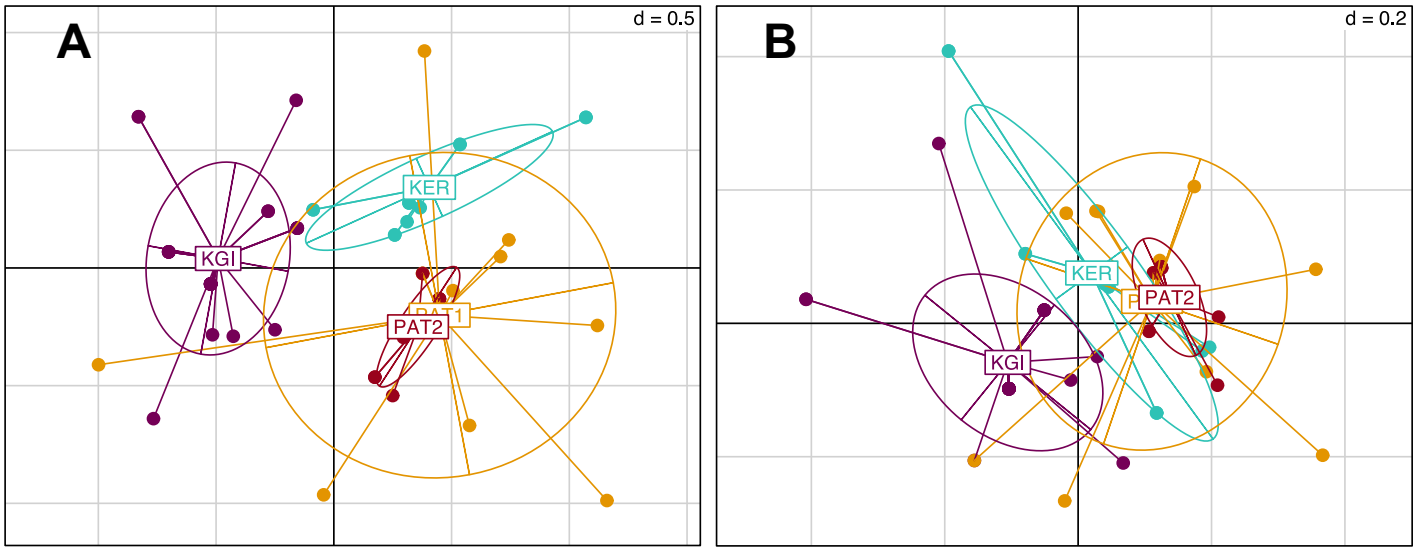

Figure 2. NMDS scatter diagram of the *Spirochaeta* OTUs composition in gut tissue samples across the localities. (A) Presence/absence matrix converted in Bray-Curtis distances, **(B)** Unweighted Unifrac distance. Colors are assigned to the locality.

Conversely, Kerguelen Islands (KER) and martime Antarctic (KGI) sites were significantly different from each other in terms of *Spirochaeta* OTU composition and with the Patagonian ones (Additional file 1).

The relative abundance analyses among the 10 *Spirochaeta* OTUs (Figure 3) showed four of them were shared among all the Southern Ocean's sampled provinces. Three OTUs (OTU6, OTU7, and OTU532) were more abundant in maritime Antarctica (KGI), three (OTU23, OTU42, and

306 OTU221) were predominantly found in the Patagonian locality PAT2, three (OTU40, OTU278,
307 and OTU561) were more abundant in the Patagonian locality PAT1, and a single one (OTU349)
308 was predominant in Kerguelen Islands (Figure 3).

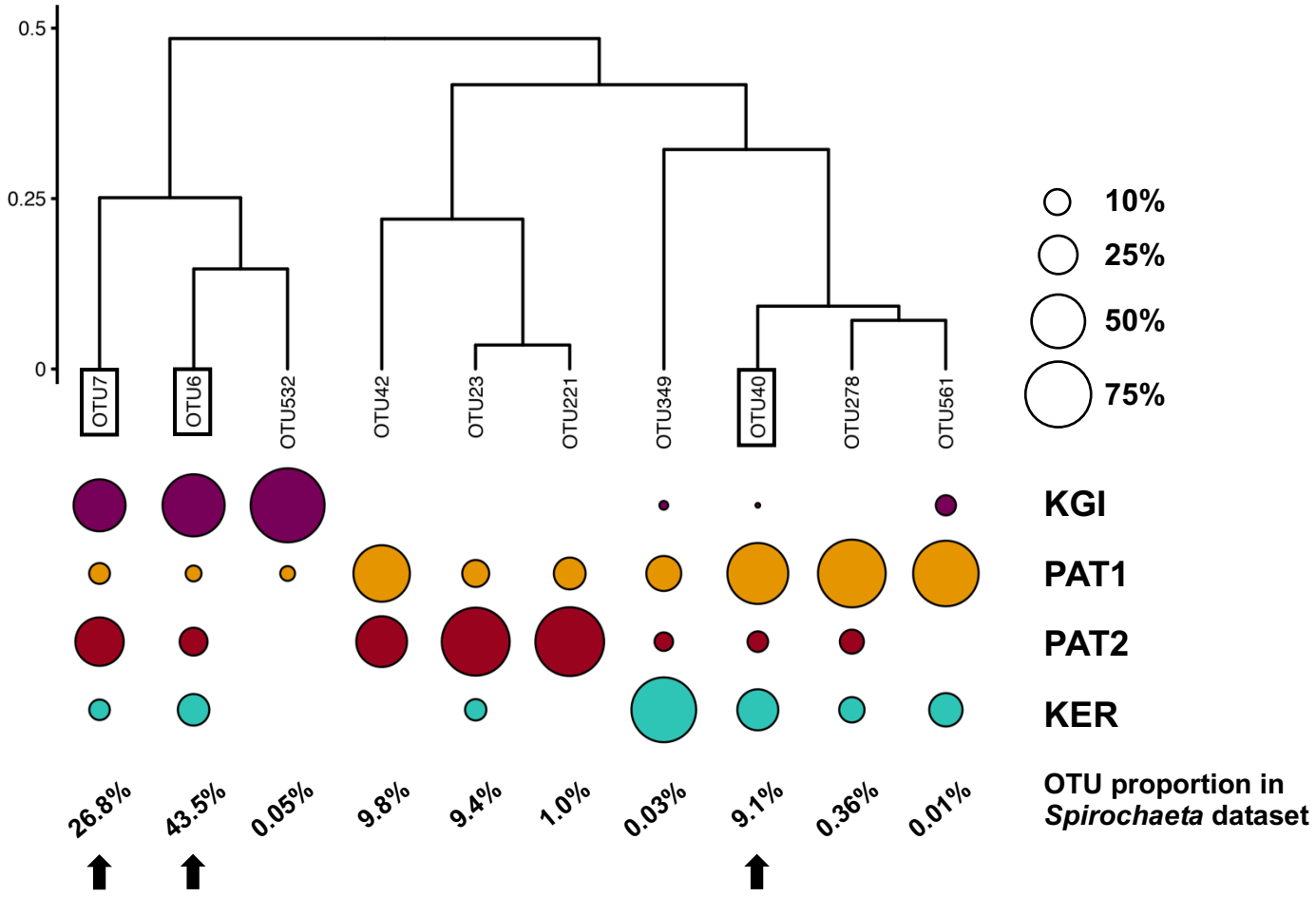

Figure 3. Clustering of *Spirochaeta* OTUs based on their relative abundance in each site.
Clustering of Bray-Curtis distances matrix using the complete linkage method. The size of the circles indicates the repartition of a given OTU sequences among the 4 localities. The percentages indicate the OTUs' proportions in the *Spirochaeta* dataset. Framed labels and black arrows indicate the selected OTUs that were selected to be process through the Minimum Decomposition Entropy pipeline (MED).

309 To test the genetic and phylogeographic structure of *Spirochaeta* at the broadest geographic scale
310 available through our dataset, we selected the three most abundant *Spirochaeta* OTUs which
311 were detected within the four localities of the dataset (*i.e.* OTU6, OTU7, and OTU40). These
312 three selected co-distributed OTUs are good targets to constitute a metapopulation, which is a
313 meaningful ecological unit of distinct local populations separated by gaps in habitats and
314 interconnected to some extent via dispersal events of individuals [84]. The relative abundance in

315 sample types, the closest sequence retrieved from Blast analysis, and the distribution of these
316 OTUs among the localities are provided in the Additionnal files 2 and 3.

317 **Microdiversity within *Spirochaeta* OTUs**

318 A total of 48, 96, and 48 oligotypes were defined for OTU6, OTU7, and OTU40, respectively
319 (Additionnal file 4). Accumulation curves of OTU6, OTU7, and OTU40 oligotypes reached
320 saturation in almost all localities indicating that the overall majority of *Spirochaeta* microdiversity
321 has been found within the analyzed samples (Additionnal file 5). Diversity indices measured as
322 *N*, *k*, *S*, *h*, and Π for each OTU in each locality are provided in Table 2.

Table 2. Summary of oligotypes number and genetic indices per OTU and per site for the three most abundant *Spirochaeta* OTUs found in all sampling localities.

OTU	Site	*N*	*k*	*S*	*H*	*Π*
OTU6	KGI	125,112	26 ± 0	24 ± 0	0.4879 ± 0.0004	0.7919 ± 0.0012
	PAT1	6,452	18 ± 0	8 ± 0	0.6701 ± 0.0000	1.5772 ± 0.0000
	PAT2	22,545	22 ± 0	12 ± 0	0.6574 ± 0.0002	1.5426 ± 0.0006
	KER	29,448	33 ± 1	29 ± 0	0.5677 ± 0.0003	0.9687 ± 0.0011
OTU7	KGI	53,645	31 ± 0	32 ± 0	0.5555 ± 0.0003	1.4443 ± 0.0021
	PAT1	7,249	69 ± 0	37 ± 0	0.8036 ± 0.0000	2.0247 ± 0.0003
	PAT2	44,509	60 ± 0	33 ± 0	0.7958 ± 0.0002	1.7336 ± 0.0011
	KER	7,021	43 ± 0	37 ± 0	0.6306 ± 0.0000	1.0261 ± 0.0000
OTU40	KGI	47	4 ± 0	4 ± 0	0.6984 ± 0.0000	1.6606 ± 0.0000
	PAT1	24,612	32 ± 0	14 ± 0	0.3741 ± 0.0007	0.4922 ± 0.0011
	PAT2	2,423*	34 ± 0	11 ± 0	0.3816 ± 0.0000	0.4903 ± 0.0000
	KER	11,017	12 ± 0	8 ± 0	0.0681 ± 0.0004	0.0913 ± 0.0006

N: number of sequences, *k*: number of oligotypes, *S*: number of polymorphic sites, *H*: genetic diversity, *Π*: mean number of pairwise diversity. The mean and standard deviation were calculated from a total of 1,000 bootstraps, performed by randomly subsampling per site a number of sequences equal to the minimum number of sequences obtained among sites for a given OTU. * In the case of OTU40, the number of sequences in the PAT2 site was used to perform the resampling.

323 The genetic diversity (*H*) ranged from 0.07 (OTU40 in KER) to 0.80 (OTU7 in PAT1) across
324 localities (Table 2). Patagonian sites exhibited higher oligotype and nucleotide diversity for OTU6
325 and OTU7 than maritime Antarctica and Kerguelen Islands localities. In contrast, the genetic

326 diversity of the OTU40 oligotypes was higher in the maritime Antarctic site and lower for the
327 Kerguelen population.

328 **Populations differentiation and phylogeographic structure of *Spirochaeta* oligotypes**

329 Independently of the OTU considered, the genetic (F_{ST}) and phylogeographic (ϕ_{ST}) structures
330 between the two closest localities (*i.e.* Patagonian sites PAT1 and PAT2) were extremely to
331 moderately low. In the case of the OTU40, the genetic diversity and frequencies of *Spirochaeta*
332 oligotypes were fully homogenous between PAT1 and PAT2, as indicated by the non-significant
333 values of F_{ST} and ϕ_{ST} comparisons (Table 3).

Table 3. Genetic (F_{ST}) and phylogeographic structure (ϕ_{ST}) of the *Spirochaeta* populations among localities.

OTU	Index	Locality	KGI	PAT1	PAT2	KER
OTU6	FST	KGI	-	0	0	0
		PAT1	0.4428	-	0	0
		PAT2	0.4417	0.0026	-	0
		KER	0.0654	0.3856	0.3829	-
	ϕST	KGI	-	0	0	0
		PAT1	0.5371	-	0	0
		PAT2	0.5360	0.0027	-	0
		KER	0.1041	0.5063	0.4933	-
OTU7	FST	KGI	-	0	0	0
		PAT1	0.3436	-	0	0
		PAT2	0.3254	0.0726	-	0
		KER	0.3902	0.0568	0.1734	-
	ϕST	KGI	-	0	0	0
		PAT1	0.4393	-	0	0
		PAT2	0.5361	0.1341	-	0
		KER	0.3967	0.1312	0.3429	-
OTU40	FST	KGI	-	0	0	0
		PAT1	0.5549	-	0.4505	0
		PAT2	0.5445	<0.0001	-	0
		KER	0.8491	0.7358	0.8614	-
	ϕST	KGI	-	0	0	0
		PAT1	0.8326	-	0.2793	0
		PAT2	0.8278	<0.0001	-	0
		KER	0.9354	0.7398	0.8652	-

Structure values are beneath each diagonal and *p*-values are above them. *p*-values were obtained through 1,000 permutations and the significance level was set < 0.05. *p*-values of 0 indicate value < 0.00001.

334 Contrarily, higher values of F_{ST} and Φ_{ST} comparisons were recorded among the three provinces
335 considered in this study (Patagonia, PAT1/PAT2; maritime Antarctica, KGI; Kerguelen Islands,
336 KER) (Table 3). Consistently, the distribution of the *Spirochaeta* oligotypes was geographically
337 discontinuous across the localities, with various province-specific oligotypes (Additionnal file 4).
338 Two exceptions were observed in the case of maritime Antarctica and Kerguelen Islands (KGI
339 and KER) comparisons for both OTU6 and OTU7, with relatively lower values of F_{ST} and Φ_{ST} (Table
340 3). The Snn test for phylogeographic structure among all sites was significant with statistic values
341 ≥ 0.5 (*i.e.* OTU6; Snn = 0.50, *p*-value < 0.0001, OTU7; Snn = 0.57, *p-value* < 0.0001, OTU40;
342 Snn = 0.50, *p*-value < 0.0001). All in all, these results indicate the existence of both genetically
343 and geographically differentiated *Spirochaeta* populations across the three biogeographic
344 provinces sampled.
345 Within the 48 oligotypes identified in the OTU6, 11 (~23%) were private to one of the three
346 provinces (Patagonia, maritime Antarctica, or the Kerguelen Islands), and more than half were
347 exclusive to Kerguelen Islands (Additionnal file 4). Maritime Antarctic and Kerguelen Islands (KGI
348 and KER) shared 27 (~66%) of their oligotypes. The Patagonian localities (PAT1/PAT2) shared 18
349 of the 23 total oligotypes (~78%) observed in this province (Figure 4, Table S3). A total of five
350 (~10%) oligotypes were broadly distributed across all localities. Oligotype network of OTU6
351 showed short genealogies and the presence of at least five dominant oligotypes. The dominant
352 oligotype in Patagonia (PAT1/PAT2), and the dominant oligotype in maritime Antarctica and
353 Kerguelen Islands (KGI/KER), were separated by a single substitution (Figure 4).
354 In the case of OTU7, the percentage of private oligotypes was almost the same as the OTU6,
355 with 21 (~22%) oligotypes exclusive to one of the provinces (Additionnal file 4). The dominant
356 oligotype was different between maritime Antarctic and Kerguelen Islands localities (KGI and
357 KER) (Figure 4). While most oligotypes from Kerguelen Islands (KER) were detected in at least one
358 of the Patagonian sites (PAT2 or PAT1) (~94%), fewer oligotypes from maritime Antarctica (KGI)

359 were observed in Kerguelen Islands (~68%) and even fewer in the Patagonian localities (~62%)

360 (Figure 4).

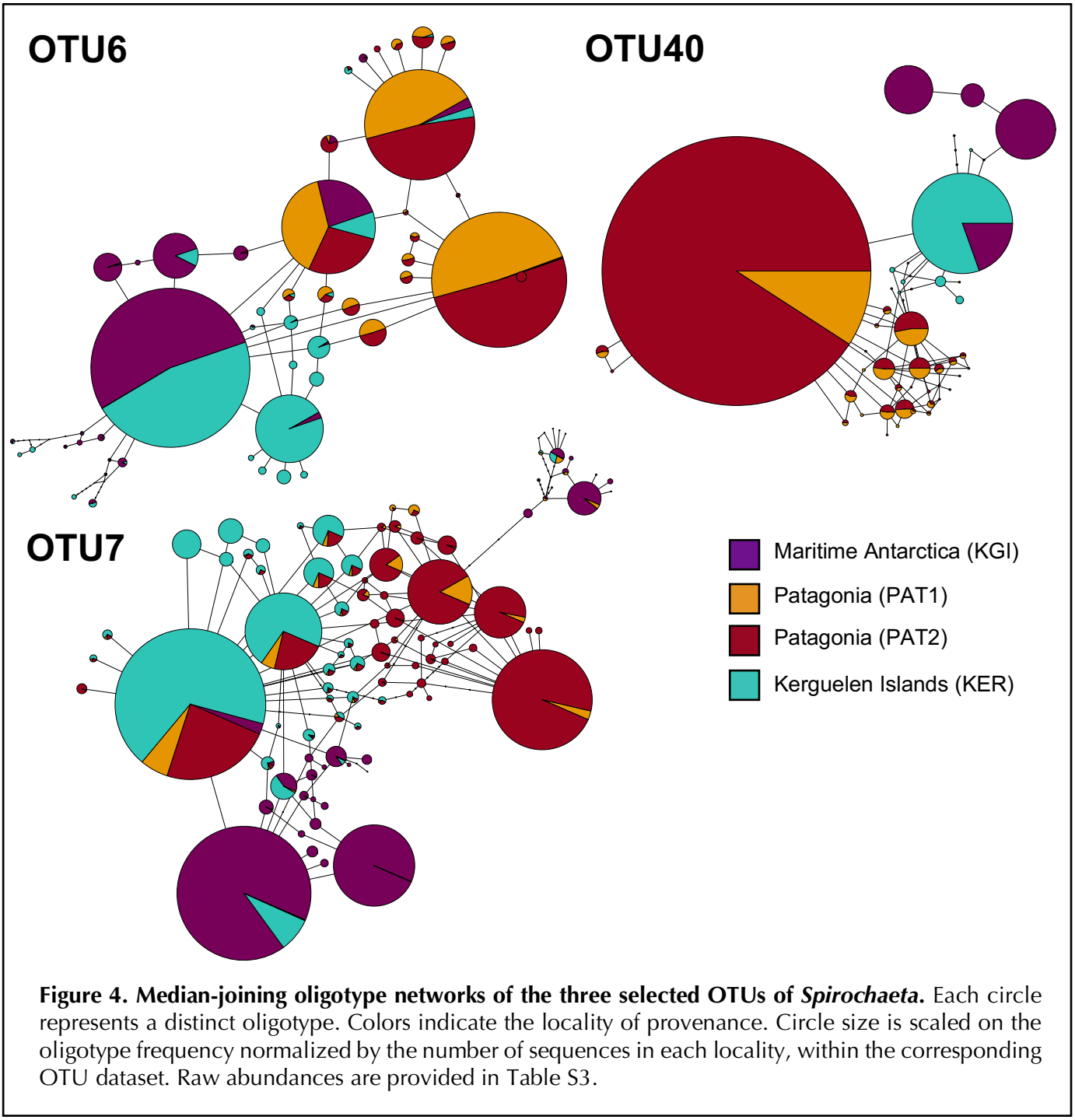

Figure 4. Median-joining oligotype networks of the three selected OTUs of *Spirochaeta*. Each circle represents a distinct oligotype. Colors indicate the locality of provenance. Circle size is scaled on the oligotype frequency normalized by the number of sequences in each locality, within the corresponding OTU dataset. Raw abundances are provided in Table S3.

361 For OTU40, we recorded a predominant group of oligotypes specific to the Patagonian sites

362 representing 65% of the oligotypes identified within the OTU40 (Additionnal file 4, Figure 4). A

363 clear separation was observed in the oligotype network between the KGI/KER and PAT1/PAT2

364 localities (Figure 4), with only 4 shared oligotypes (~8%) (Table S3). In maritime Antarctica (KGI),

three of the four oligotypes were private, whereas the dominant one from Kerguelen Islands (KER) was shared with maritime Antarctica (KGI) (~8%) (Figure 4, Additionnal file 4).

Gene flow under a migration–drift equilibrium model

All the analyzed OTUs showed high genetic similarities between the analyzed Patagonian populations (PAT1 and PAT2). Gene flow analyses identified a bidirectional pattern from PAT2 to PAT1 (effective number of migrants per generation, Nm > 4) and from PAT1 to PAT2 (Nm > 14) (Table 4).

Table 4. Effective numbers of migrants per generation (Nm) among *Spirochaeta* populations of the OTU6, OTU7 and OTU40.

OTU	From	Θ ± se	To	Nm ± se
OTU6	KGI	0.003 ± 0.0003	KER	**9.81 ± 3.14**
			PAT1	0.12 ± 0.06
			PAT2	0.14 ± 0.09
	KER	0.004 ± 0.0004	KGI	**2.63 ± 1.19**
			PAT1	0.12 ± 0.05
			PAT2	0.07 ± 0.05
	PAT1	0.003 ± 0.0004	KGI	0.14 ± 0.05
			KER	**0.43 ± 0.15**
			PAT2	**24.99 ± 7.10**
	PAT2	0.005 ± 0.0005	KGI	0.20 ± 0.07
			KER	**0.27 ± 0.16**
			PAT1	**14.57 ± 4.93**
OTU7	KGI	0.002 ± 0.0003	KER	**1.39 ± 0.60**
			PAT1	**0.64 ± 0.23**
			PAT2	0.03 ± 0.02
	KER	0.006 ± 0.0006	KGI	**0.42 ± 0.13**
			PAT1	**24.41 ± 8.23**
			PAT2	**2.00 ± 0.54**
	PAT1	0.015 ± 0.0029	KGI	0.16 ± 0.06
			KER	**0.41 ± 0.16**
			PAT2	**4.69 ± 1.82**
	PAT2	0.004 ± 0.0005	KGI	0.05 ± 0.03
			KER	0.13 ± 0.09
			PAT1	**19.95 ± 9.57**
OTU40	KGI	0.002 ± 0.0002	KER	0.09 ± 0.04
			PAT1	0.01 ± 0.01
			PAT2	0.02 ± 0.02
	KER	0.001 ± 0.0002	KGI	**0.87 ± 0.21**
			PAT1	0.02 ± 0.02
			PAT2	0.02 ± 0.02
	PAT1	0.005 ± 0.0008	KGI	0.01 ± 0.01
			KER	0.03 ± 0.02
			PAT2	**28.65 ± 10.15**
	PAT2	0.005 ± 0.0008	KGI	0.00 ± 0.00
			KER	0.06 ± 0.02
			PAT1	**15.94 ± 10.48**

Only gene flows with Nm values > 0.25 are considered as significant. Mean and standard error values were calculated from the 10 runs performed for each OTU.

The connectivity between the Patagonian and maritime Antarctic localities showed relatively low values of Nm, ranging from 0.001 (OTU40, from PAT2 to KGI) to 0.6 (OTU7, from KGI to PAT2) (Table 4). The OTU6 and OTU7 were both characterized by a substantial gene flow between maritime Antarctica and Kerguelen Islands that was stronger in the direction KGI to KER (OTU6, Nm ~ 9.8 and OTU7, Nm ~ 1.4) than in the direction KER to KGI (OTU6, Nm ~ 2.6 and OTU7, Nm ~ 0.4) (Table 4). Contrarily, an unidirectional and low gene flow from KER to KGI (Nm ~ 0.9) was recorded for the OTU40 (Table 4). Finally, the connectivity between Patagonian (PAT1 and PAT2) and Kerguelen Islands (KER) localities was illustrated by three distinct patterns; a low-intensity flow (Nm < 0.5) predominant in the direction PAT1/PAT2 to KER for the OTU6, a substantial flow (Nm > 2) predominant in the direction KER to PAT1/PAT2 for the OTU7, and an absence of connectivity (Nm < 0.03) in the case of the OTU40 (Table 4).

Contribution of contemporary selection versus historical processes in shaping the *Spirochaeta* microdiversity

For each of the three selected *Spirochaeta* OTUs, neutral ecological processes were essential in shaping the population composition turnover in *Abatus* gut membrane. According to the quantitative parsing of ecological processes, the composition of *Spirochaeta* population was mostly driven by ecological drift, ranging from 50% (OTU40) to 74% (OTU6) of turnover, followed by dispersal limitation, ranging from 12% (OTU6) to 20% (OTU40) of turnover, and homogenizing selection, ranging from 9% (OTU6) to 19% (OTU40) of turnover (Table 5).

Table 5. Quantitative parsing of ecological processes driving populations turnover within *Spirochaeta* OTUs.

	Ecological processes contributions				
***Spirochaeta* OTU**	**Homogeneous selection (%)**	**Homogenizing dispersal (%)**	**Ecological drift (%)**	**Dispersal limitation (%)**	**Variable selection (%)**
OTU6	2.7	8.8	74.0	12.1	2.3
OTU7	0.4	10.3	63.7	17.6	8.0
OTU40	0.3	19.4	49.6	21.7	9.1

According to the Stegen et al. (2013) approach, percentage refers to the percentage of pairs of communities that appear to be driven by either homogeneous selection, homogenizing dispersal, ecological drift, dispersal limitation or variable selection.

Overall, deterministic processes (*i.e.* homogeneous and variable selection) did not account for more than 10% of the populations' turnover.

The MMRR approach was used to disentangle the relative effect of geographic distance environmental abiotic differences on the *Spirochaeta* oligotype compositions between samples. The geographic distance matrix was linearly correlated to the abundance-based similarity matrix of *Spirochaeta* population composition for OTU7 and OTU40, explaining about 31% and 67% of the observed variation, respectively (Table 6).

Table 6. Multiple Matrix Regression with Randomization (MMRR) to quantify the relative effects of isolation by distance (IBD) and isolation by environment (IBE) on oligotypes assemblage within *Spirochaeta* OTUs.

Spirochaeta OTU	Model	Coefficient	t statistic	t *p*-value	F statistic	F *p*-value	R^2
OTU6	IBE	0.037	17.82	<0.001	187.32	<0.001	0.208
	IBD	-0.158	-2.59	0.078	NA	NA	NA
OTU7	IBE	0.020	12.10	<0.001	162.43	<0.001	0.165
	IBD	0.309	6.11	<0.001	NA	NA	NA
OTU40	IBE	0.022	6.45	<0.001	102.08	<0.001	0.235
	IBD	0.667	7.97	<0.001	NA	NA	NA

The first statistical test (t) individually estimates the effect of the environmental distance and the geographic distance matrices, whereas the second one (F) evaluates the global fit of the model considering both distance matrices. *p*-values are considered as significant < 0.05.

In contrast, the geographic distance did not significantly impact *Spirochaeta* oligotype composition for OTU6 (Table 6). Whatever the OTU considered, the environment distance had a significant but slight contribution (< 4% of the observed variation) to the *Spirochaeta* population composition (Table 6), and the global R^2 of the model (environmental and geographic distance) did not exceed 24%.

Discussion

In this study, we coupled 16S rRNA metabarcoding and oligotyping algorithm to reveal the microdiversity within three bacterial OTUs affiliated to the *Spirochaeta* genus, and co-distributed across Patagonia, Kerguelen Islands, and maritime Antarctic provinces of the Southern Ocean. Through this innovative approach, we identified numerous oligotypes within each of the

Spirochaeta OTUs. These oligotypes, corresponding to *Spirochaeta* sub-taxa, were characterized by contrasting geographic distribution and high levels of 16S rRNA gene similarity (> 97%). Taking advantage of the populational taxonomic resolution provided by the oligotype definition, we depicted the *Spirochaeta* biogeographic patterns across the analyzed provinces in the SO, using various tools adapted from population genetics classically applied in phylogeographic study of macroorganisms' models. Despite its low substitution rate (approximately 1% in 50 million years [85]), our study demonstrates that the 16s rRNA gene still has its advantage in effectiveness and efficiency since it offers the best compromise between an informative genetic signal, and robust screening of global microbial diversity at intra-OTU level, in a wide range of barely unknown habitats [17].

Unlike the studies with macroorganisms, which are usually more demanding in terms of individual sampling effort, we benefit here from the high sequencing depth provided by the metabarcoding of a low-diversity habitat (*i.e.* the *Abatus* gut tissue). This allows a robust coverage of the *Spirochaeta* diversity (up to 180,000 sequences per OTU), and thus high precision of the oligotypes frequencies. Our methodology echoes the "metaphylogeographic" approach recently proposed by Turon *et al*. [86] to investigate eukaryotic intraspecies diversity through COI (cytochrome c oxidase subunit I) gene amplicon-sequencing and an oligotyping-like cleaning protocol of the reads based on entropy variation. We propose to expand this concept of "metaphylogeography" to the prokaryotes since our method permits phylogeographic inferences of uncultured microbes from a wide range of habitats.

The β-diversity analysis performed at the *Spirochaeta* genus level revealed that each of the three geographic provinces might host specific *Spirochaeta* OTUs representing distinct phylogenetic lineages. We also reported a non-random distribution trend with contrasting patterns of *Spirochaeta* OTU abundances across the localities. Nevertheless, about half of the *Spirochaeta* OTUs exhibited a broad distribution encompassing Patagonia, maritime Antarctica, and the Kerguelen Islands located more than 7,000 kilometers to the east. This result suggests that despite

being mostly detected in *Abatus* gut, and to a lesser extent in marine benthic sediments, some *Spirochaeta* representatives would disperse through the SO currents. Concordantly, previous campaigns of high-throughput sequencing of the ocean water column have consistently reported the presence of free-living *Spirochaeta* OTUs in surface to mesopelagic water, away from the coastlines [87].

For each of the three assessed OTUs, the *Spirochaeta* populations were expected to be remarkably homogeneous between the two Patagonian sites due to the geographic vicinity and the absence of an evident oceanographic barrier. Indeed, most of the *Spirochaeta* oligotypes were shared between these two localities that displayed the lowest genetic and phylogeographic structure for each of the three OTUs. High levels of gene flow, illustrated by a high effective number of migrants per generation (Nm), were also recorded between the two Patagonian localities, in accordance with the homogeneity of their oligotype compositions. Similarly, low or absent differentiation patterns along the Atlantic coast of Patagonia were previously reported for marine Patagonian macroorganisms, including notothenioid fishes [88] scorched mussels [89], and pulmonate gastropods [90]presumably due to their high dispersal potential and the ecological continuum of the sampled localities that may conform a same biogeographic province connected through the equator-ward Falkland current [91, 92]. Further phylogeographic studies focusing on microbial taxa of additional sampling sites from Atlantic Patagonia should confirm the microbial biogeographic consistency of this province.

Between Patagonian and maritime Antarctic provinces, *Spirochaeta* populations exhibited strong genetic and phylogeographic structure, as illustrated by the high F_{ST} and Φ_{ST} values and the relatively low number of shared oligotypes. Additionally, low levels of gene flow were estimated between these two provinces (*i.e.* effective numbers of migrants per generation $Nm < 1$). These results corroborate our hypothesis that the APF hinders individual dispersion and genetic homogeneity among bacterial populations and suggest that the geographically structured *Spirochaeta* populations from these two provinces are genetically diverging over time [93, 94].

460 Previous studies focusing on diverse macroorganisms taxonomic groups of the SO have
461 evidenced the critical role of the APF on biogeographic patterns, as an open-ocean barrier
462 inducing a genetic break between South America and Antarctica (*e.g.* ribbon worms, [95]; brittle
463 stars, [96]; notothenioid fishes, [25, 97]; limpets [26, 29, 98]; sea urchins [24]). Regarding the
464 microbial distribution patterns, significant β-diversity differences between prokaryotes assembly
465 from both sides of the APF have been reported in the past, but most of the studies focused on
466 global community in the water column, at high taxonomic resolution (summarized in Flaviani et
467 al. [20]). Here, we extend this discontinuity in bacterial diversity to a finer taxonomic resolution
468 (*i.e.* intra-OTU), revealing province-restricted oligotypes and strong genetic and phylogeographic
469 structure between Patagonian and maritime Antarctic *Spirochaeta* populations.

470 Contrarily, and despite the substantial geographic distance separating the sub-Antarctic
471 Kerguelen Islands and the Patagonian and Antarctic sites (> 6,500 kilometers), population genetic
472 analyses suggest the existence of some level of connectivity between Kerguelen and the other
473 sites. These findings support a potential dispersion of *Spirochaeta* taxa from Patagonia and
474 maritime Antarctica to the Kerguelen Islands, and contrariwise, from the Kerguelen Islands to
475 Patagonia and maritime Antarctica. As evidenced by the numerous shared oligotypes, such
476 connectivity would maintain a sufficient gene flow among these provinces to partially counteract
477 the genetic divergence driven by selection, mutation, and genetic drift, inducing oligotypes
478 mixing, and limiting the spatial differentiation of oligotypes assembly [1] [99]. Moreover, we
479 suggest that this gene flow is not bidirectional, but governed by exclusively eastward oriented
480 dispersion routes (Figure 5), following the major and constant flow of the ACC [100]. Under this
481 scenario, *Spirochaeta* individuals from Kerguelen Islands may seed towards Patagonia following
482 the ACC eastward flow around Antarctica. Such ACC-mediated connectivity among sub-
483 Antarctic provinces (Patagonia and Kerguelen Islands) is well known in a wide range of benthic
484 macroorganisms populations, such as buoyant kelps *Durvillaea antarctica* [38] and *Macrocystis*
485 *pyrifera* [101], and several kelp-associated macroinvertebrates [32, 34, 35, 100, 102].

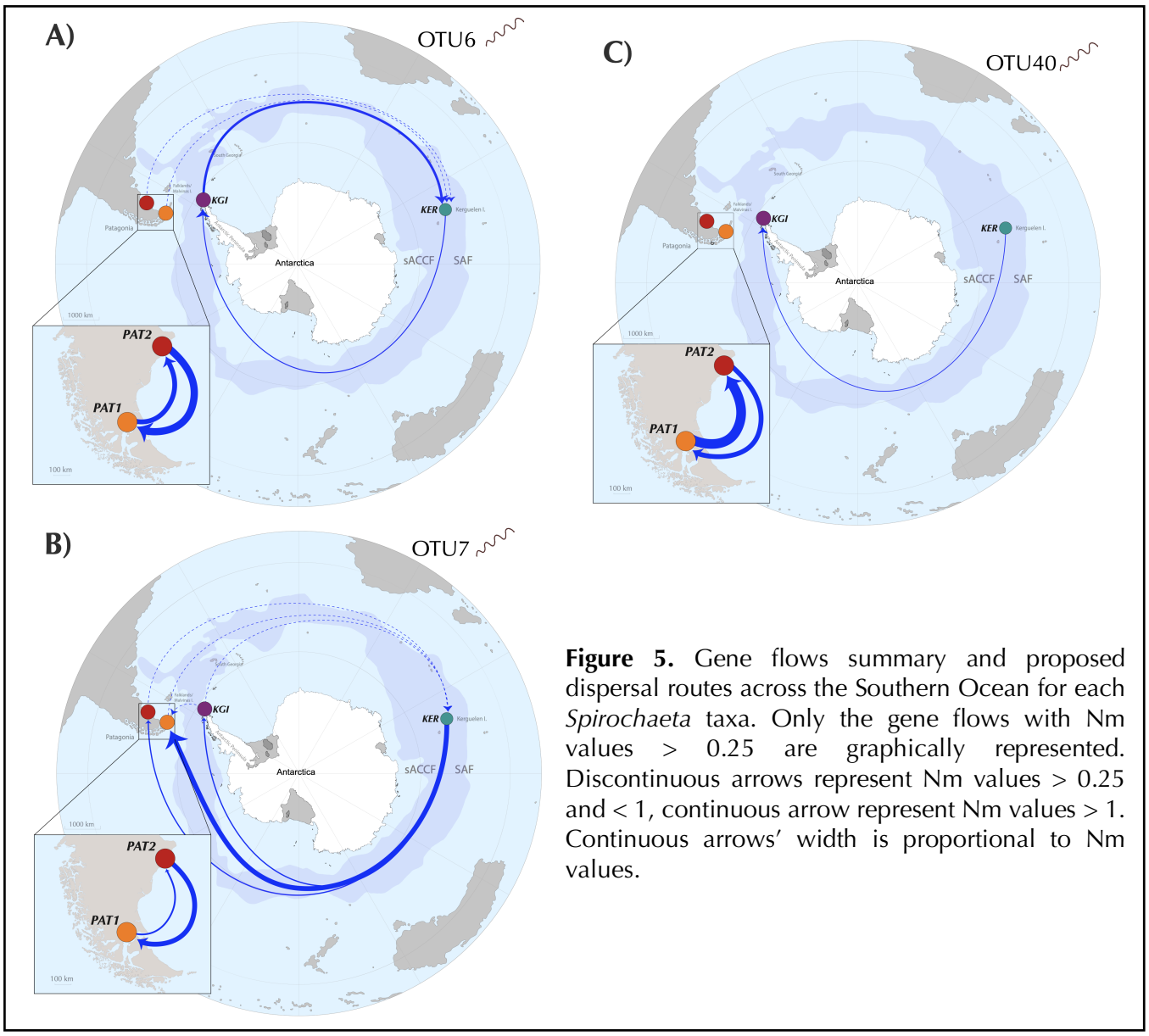

Figure 5. Gene flows summary and proposed dispersal routes across the Southern Ocean for each *Spirochaeta* taxa. Only the gene flows with Nm values > 0.25 are graphically represented. Discontinuous arrows represent Nm values > 0.25 and < 1, continuous arrow represent Nm values > 1. Continuous arrows' width is proportional to Nm values.

Occasionally, *Spirochaeta* individuals from Kerguelen Islands may also be able to reach the maritime Antarctic province. Such pattern has been recently reported for the southern bull kelp *D. antarctica*, a typical sub-Antarctic macroalgae, which is transported by rafting to as far as the West Antarctic Peninsula coasts, pushed by the circumpolar flow of the ACC or by storms leading to the occasional crossing of the APF [103].

Several studies have provided evidence of a high dispersal capacity of marine bacteria by comparing community composition mostly at high taxonomic resolution (*e.g.* class, genus, or OTU) among various water masses and oceanic regions [104-107]. Particularly, the most

494 abundant marine bacteria are supposed to migrate between adjacent regions through passive
495 transport [104]. However, this is the first time that it is evidenced in bacterial populations through
496 the microdiversity resolution, being solely suggested so far at the community level [20]. An
497 innovative conceptual framework called "Microbial Conveyor Belt" (MCB) has been proposed by
498 Mestre & Höfer [108], to emphasize that the marine microorganisms' dispersion would not
499 merely rely on passive and stochastic dispersal, but instead on the adaptation of life-history traits
500 (*e.g.* dormancy stage). These traits would allow microorganisms to successfully and recurrently
501 disperse in unfavorable habitats through specific dispersion avenues [109]. Here, we provided
502 empiric results from *Spirochaeta* population based on genetic data supporting a partial MCB in
503 the SO driven by the ACC. Unfortunately, details about the benthic *Spirochaeta* taxa's ecology
504 are scarce, with a single isolated strain from subseafloor sediment [110]. Thus, the life-history
505 traits of *Spirochaeta*, as the sporulation capacity, remain to be investigated to further understand
506 its distribution pattern in the SO. Nevertheless, in order to disperse, we propose that *Spirochaeta*
507 individuals (enriched in the digestive tract) could be released from the host gut towards the
508 surrounding benthic sediments through fecal pellets. Such enrichment of the digesta with taxa
509 from the host microbiota, as well as the presence of *Spirochaeta* within the fecal material, have
510 been demonstrated in the sea urchin species *Lytechinus variegatus* [111]. The released
511 *Spirochaeta* individuals may be resuspended in the water column through the action of one or
512 several processes such as upwelling, bioturbation by the benthic deposit-feeders, or water
513 column mixing during winter [108, 112, 113]. Once in the water column, these *Spirochaeta*
514 individuals may disperse over large geographic scales, transported through oceanographic
515 features (*e.g.* currents, punctual meteorological events) [108]. The attachment to suspended
516 particulate matter, either biotic (*e.g.* hitchhiking on zooplankton [114] and seaweed [115]) or
517 abiotic (*e.g.* microplastics [116], known to have a long-distance dispersion potential), may also
518 contribute to the bacterial spreading in the oceans [107, 108].

519 The marine prokaryote communities are usually considered widely dispersed and mainly shaped
520 by contemporary ecological processes such as environmental filtering [106, 117]. By applying
521 the ecological framework developed by Stegen et al. [71] to oligotypes data, we found contrarily
522 that ecological drift was the predominant stochastic mechanism shaping intra-populations
523 turnover within *Spirochaeta* taxa across the SO. Our previous study of the *Abatus* gut microbiota
524 showed that non-neutral processes drove the bacterial community at the OTU level in the host
525 gut tissue [54]. While deterministic processes are usually prevalent in structuring microbial
526 communities' assembly at a higher taxonomic resolution [18, 79, 118, 119], the stochastic
527 mechanisms tend to have a more significant contribution at finer taxonomic scales [105], since
528 niche overlapping and functional redundancy enhance the susceptibility of populations to drift
529 [120]. Thus, the biogeographic structure observed within *Spirochaeta* populations might result
530 from stochastic birth, death, disturbance, emigration, and immigration events rather than
531 oligotype-sorting through the biotic and abiotic environmental variations [1, 121, 122].
532 Consistently, the MMRR analysis revealed that the Isolation-by-Environment (IBE) model might
533 account for a low percentage of the *Spirochaeta* oligotypes turnover. Altogether, these results
534 validate the strategy applied in our study, that is, to focus on specialist bacterial taxa hosted in
535 sibling sea urchin species with the same habitat preferences, in order to homogenize the
536 environment, to reduce the diversity, to soften the deterministic selection driven by
537 environmental variations and to maximize the detection of neutral micro-evolutive processes
538 associated with biogeography [123].

539 By analogy with the genetic drift, whereby changes in gene frequencies occur solely by chance
540 in a population [76], our result suggests that the microdiversity observed within the *Spirochaeta*
541 taxa would be mostly generated by genetic drift without any adaptive implications. An earlier
542 study reported that microdiversity observed in the 16S rRNA gene of marine coastal *Vibrio*
543 *splendidus* isolates was ecologically neutral [124]. Nevertheless, we cannot discard that, while
544 the microdiversity within the V4-V5 of 16S rRNA gene-targeted here is likely to be acquired

545 through neutral processes [125], it may also be associated with substantial modifications in
546 niche-defining traits and functional attributes specific of the *Spirochaeta* strains, driven by
547 deterministic processes, in order to cope with local conditions [17, 126]. Further studies will
548 need to focus on other loci (*e.g.* functional genes), potentially under selection, as they are
549 expected to display a higher degree of differentiation among populations and to provide an
550 insight into the ecology of the *Spirochaeta* sub-taxa [127]. Furthermore, notwithstanding the
551 consistency of the global phylogeographic and connectivity patterns depicted across the three
552 tested OTUs, we also reported some differences according to the taxa considered, which might
553 be related to different ecotypes with distinct ecological niches or different dispersal capacity. For
554 instance, various marine bacterial taxa, such as the cyanobacteria *Synechococcus* or the *Vibrio*
555 populations, demonstrate fine-tuning of their physiology by accumulating microdiversity in
556 functional genes through duplication events, SNPs, and allelic variants [128, 129]. Alternatively,
557 these differences may also be related to the *Spirochaeta* OTU abundance, since the more
558 relatively abundant the *Spirochaeta* populations were (*i.e.* higher number of sequences retrieved
559 from the gut tissue through the metabarcoding approach), the more they tend to exhibit
560 cosmopolitan oligotypes (*i.e.* detected across each of the four localities). It is not unreasonable
561 to infer that a larger population may have more chance to migrate and successfully reach a
562 suitable habitat, while small-size populations may be more likely diluted along the dispersal route
563 with no/too few dispersive particles to establish in the new habitat [104].

564 The diversity units defined by 16S rRNA gene sequences are generally considered as insensitive
565 to diversification resulting from dispersal limitation [10]. Contrastingly, we reported that the
566 dispersal limitation was the second most crucial ecological factor driving the turnover of
567 *Spirochaeta* oligotypes, and by extension, their genetic divergence. Dispersal limitation is
568 classically considered as a historical factor since current oligotypes assemblage results from past
569 dispersal limitations [1]. Our result indicates that the potentially suitable habitats are too distant
570 [130], or inaccessible due to the existence of oceanic currents [104, 131], hence limiting the

571 homogenization of *Spirochaeta* oligotypes' frequencies across populations and allowing the
572 neutral genetic divergence of genomic regions overtime via genetic drift [10, 132]. Note that our
573 results obtained from distinct methodologies (*i.e.* the genetic differentiation and phylogeographic
574 structure, the contribution of dispersal limitation from Stegen *et al.* framework [71], and the
575 contribution of the geographic distance (IBD) from the MMRR analysis) were highly consistent
576 with each other, and across the three selected *Spirochaeta* OTUs. For instance, the OTU40 that
577 harbored the overall highest value of genetic divergence was also characterized by the highest
578 estimated contribution of geographic distance and dispersal limitation, thus supporting the
579 interrelation between genetic divergence and oligotypes population turnover, and the overall
580 consistency of the approach implemented.

581 **Conclusion**

582 Our study highlights the application of V4-V5 16S rRNA gene metabarcoding and oligotyping
583 approach as rapid, robust, and resolutive enough to unravel marine bacterial phylogeographic
584 patterns and detect genetic connectivity among the SO provinces. Taken together, the three
585 *Spirochaeta* OTUs analyzed evidence three consistent phylogeographic patterns, classically
586 observed in the studies involving benthic macroinvertebrates across the SO: (i) a high
587 populational and genetic homogeneity within the Patagonia province, (ii) a strong barrier to
588 dispersal between Patagonia and maritime Antarctica due to the APF, resulting in a high
589 differentiation of *Spirochaeta* populations, and (iii) the existence of connectivity between sub-
590 Antarctic provinces of the Kerguelen Islands and Patagonia, and from Kerguelen Islands to
591 maritime Antarctic, due to the ACC-mediated connectivity. Nevertheless, as connected as these
592 provinces are, the gene flow does not seem to be strong enough to prevent the ongoing
593 intraspecific differentiation process of the *Spirochaeta* taxa. The microdiversity of *Spirochaeta*,
594 underlying these biogeographic patterns, is essentially driven by historical processes, such as
595 ecological and genetic drift, and dispersal limitation related to the SO's oceanographic features.

596 In the future, extending this framework to other localities and taxonomic groups will contribute
597 to the comprehensive understanding of the Southern Ocean microbiota.

598 **Declaration**

599 **Ethics approval and consent to participate**

600 Not applicable

601 **Consent for publication**

602 Not applicable

603 **Availability of data and materials**

604 The datasets generated for this study can be found at the National Centre for Biotechnology
605 Information (NCBI) repository, under the following accession numbers: PRJNA658980,
606 PRJNA590493 and PRJNA659050, corresponding to the datasets of *Abatus cordatus* from
607 Kerguelen Archipelago, *Abatus agassizii* from West Antarctic Peninsula and *Abatus cavernosus*
608 from South America, respectively.

609 **Competing interests**

610 The authors declare that they have no competing interests

611 **Funding**

612 This work was financially supported by the project ANID/CONICYT PIA ACT 172065.
613 Additionally, this research was supported by the post-doctoral projects ANID/CONICYT
614 FONDECYT 3200036 (GS) and 3190482 (NS).

615 **Authors' contributions**

616 EP, JO and LC designed the study. EP and JO organized the sampling missions. EP, JO, LC and
617 GS collected samples. GS extracted the DNA, organized sequencing and managed data mining
618 and analyses. NS contributed to the gene-flow and the MMRR analyses and designed the

illustrative maps. GS, JO, NS, CGW and EP interpreted the results. GS wrote the manuscript. All authors contributed substantially to manuscript revisions. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr Peter Beerli and Dr Franz-Sebastian Krah for their technical supports. We also thank Jonathan Flores for providing *Abatus* samples from Argentinian Patagonia, and Dr Karin Gérard for managing the sampling logistic in Chilean Patagonia. We recognize the Chilean Antarctic Institute (INACH) for the logistic support during the Chilean Antarctic Expeditions (ECA 55 and 56).

Bibliography

1. Martiny JB, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, et al: **Microbial biogeography: putting microorganisms on the map.** *Nat Rev Microbiol* 2006, **4:**102-112.
2. Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N, Townsend AR, Cleveland CC, Stanish L, Knight R: **Global patterns in the biogeography of bacterial taxa.** *Environ Microbiol* 2011, **13:**135-144.
3. Wilkins D, Lauro FM, Williams TJ, Demaere MZ, Brown MV, Hoffman JM, Andrews-Pfannkoch C, McQuaid JB, Riddle MJ, Rintoul SR, Cavicchioli R: **Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics.** *Environ Microbiol* 2013, **15:**1318-1333.
4. Karimi B, Terrat S, Dequiedt S, Saby NP, Horrigue W, Lelièvre M, Nowak V, Jolivet C, Arrouays D, Wincker P: **Biogeography of soil bacteria and archaea across France.** *Science advances* 2018, **4:**eaat1808.
5. Bahl J, Lau MC, Smith GJ, Vijaykrishna D, Cary SC, Lacap DC, Lee CK, Papke RT, Warren-Rhodes KA, Wong FK, et al: **Ancient origins determine global biogeography of hot and cold desert cyanobacteria.** *Nat Commun* 2011, **2:**163.
6. Stegen JC, Lin X, Konopka AE, Fredrickson JK: **Stochastic and deterministic assembly processes in subsurface microbial communities.** *ISME J* 2012, **6:**1653-1664.
7. Van der Gast CJ: **Microbial biogeography: the end of the ubiquitous dispersal hypothesis?** *Environ Microbiol* 2015, **17:**544-546.
8. Gilbert JA, Steele JA, Caporaso JG, Steinbruck L, Reeder J, Temperton B, Huse S, McHardy AC, Knight R, Joint I, et al: **Defining seasonal marine microbial community dynamics.** *ISME J* 2012, **6:**298-308.
9. Fuhrman JA, Hewson I, Schwalbach MS, Steele JA, Brown MV, Naeem S: **Annually reoccurring bacterial communities are predictable from ocean conditions.** *Proc Natl Acad Sci U S A* 2006, **103:**13104-13109.

10. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB: **Beyond biogeographic patterns: processes shaping the microbial landscape.** *Nat Rev Microbiol* 2012, **10:**497-506.
11. Hanson CA, Müller AL, Loy A, Dona C, Appel R, Jørgensen BB, Hubert CR: **Historical factors associated with past environments influence the biogeography of thermophilic endospores in Arctic marine sediments.** *Frontiers in microbiology* 2019, **10**.
12. Papke RT, Ramsing NB, Bateson MM, Ward DM: **Geographical isolation in hot spring cyanobacteria.** *Environmental Microbiology* 2003, **5:**650-659.
13. Whitaker RJ, Grogan DW, Taylor JW: **Geographic barriers isolate endemic populations of hyperthermophilic archaea.** *Science* 2003, **301:**976-978.
14. Almasia R, Carú M, Handford M, Orlando J: **Environmental conditions shape soil bacterial community structure in a fragmented landscape.** *Soil Biology and Biochemistry* 2016, **103:**39-45.
15. Wu L, Ning D, Zhang B, Li Y, Zhang P, Shan X, Zhang Q, Brown M, Li Z, Van Nostrand JD, et al: **Global diversity and biogeography of bacterial communities in wastewater treatment plants.** *Nature Microbiology* 2019.
16. Bay SK, McGeoch MA, Gillor O, Wieler N, Palmer DJ, Baker DJ, Chown SL, Greening C, Gilbert JA: **Soil bacterial communities exhibit strong biogeographic patterns at fine taxonomic resolution.** *mSystems* 2020, **5**.
17. Chase AB, Martiny JB: **The importance of resolving biogeographic patterns of microbial microdiversity.** *Microbiology Australia* 2018, **39:**5-8.
18. Larkin AA, Martiny AC: **Microdiversity shapes the traits, niche space, and biogeography of microbial taxa.** *Environmental microbiology reports* 2017, **9:**55-70.
19. Fenchel T, Finlay BJ: **The ubiquity of small species: patterns of local and global diversity.** *Bioscience* 2004, **54:**777-784.
20. Flaviani F, Schroeder DC, Lebret K, Balestreri C, Highfield AC, Schroeder JL, Thorpe SE, Moore K, Pasckiewicz K, Pfaff MC, Rybicki EP: **Distinct oceanic microbiomes from viruses to protists located near the Antarctic Circumpolar Current.** *Front Microbiol* 2018, **9:**1474.
21. Gruber N, Landschützer P, Lovenduski NS: **The variable Southern Ocean carbon sink.** *Annual review of marine science* 2019, **11:**159-186.
22. Griffiths HJ: **Antarctic marine biodiversity–what do we know about the distribution of life in the Southern Ocean?** *PloS one* 2010, **5:**e11683.
23. Halanych KM, Mahon AR: **Challenging dogma concerning biogeographic patterns of Antarctica and the Southern Ocean.** *Annual Review of Ecology, Evolution, and Systematics* 2018, **49:**355-378.
24. Díaz A, Féral J-P, David B, Saucède T, Poulin E: **Evolutionary pathways among shallow and deep-sea echinoids of the genus *Sterechinus* in the Southern Ocean.** *Deep Sea Research Part II: Topical Studies in Oceanography* 2011, **58:**205-211.
25. Hüne M, González-Wevar C, Poulin E, Mansilla A, Fernández DA, Barrera-Oro E: **Low level of genetic divergence between *Harpagifer* fish species (Perciformes: Notothenioidei) suggests a Quaternary colonization of Patagonia from the Antarctic Peninsula.** *Polar Biology* 2015, **38:**607-617.
26. González-Wevar CA, Hüne M, Rosenfeld S, Nakano T, Saucède T, Spencer H, Poulin E: **Systematic revision of *Nacella* (Patellogastropoda: Nacellidae) based on a complete**

phylogeny of the genus, with the description of a new species from the southern tip of South America. *Zoological Journal of the Linnean Society* 2019, **186:**303-336.

27. Poulin E, González-Wevar C, Díaz A, Gérard K, Hüne M: **Divergence between Antarctic and South American marine invertebrates: What molecular biology tells us about Scotia Arc geodynamics and the intensification of the Antarctic Circumpolar Current.** *Global and Planetary Change* 2014, **123:**392-399.
28. Strugnell JM, Rogers AD, Prodöhl PA, Collins MA, Allcock AL: **The thermohaline expressway: the Southern Ocean as a centre of origin for deep-sea octopuses.** *Cladistics* 2008, **24:**853-860.
29. González-Wevar CA, Hüne M, Segovia NI, Nakano T, Spencer HG, Chown SL, Saucède T, Johnstone G, Mansilla A, Poulin E: **Following the Antarctic Circumpolar Current: patterns and processes in the biogeography of the limpet Nacella (Mollusca: Patellogastropoda) across the Southern Ocean.** *Journal of biogeography* 2017, **44:**861-874.
30. Koubbi P, De Broyer C, Griffiths H, Raymond B, D'Udekem d'Acoz C, Van de Putte A, Danis B, David B, Grant S, Gutt J: **Conclusions: Present and future of Southern Ocean biogeography.** 2014.
31. Waters JM: **Driven by the West Wind Drift? A synthesis of southern temperate marine biogeography, with new directions for dispersalism.** *Journal of Biogeography* 2008, **35:**417-427.
32. González-Wevar CA, Segovia NI, Rosenfeld S, Ojeda J, Hüne M, Naretto J, Saucède T, Brickle P, Morley S, Féral JP: **Unexpected absence of island endemics: long-distance dispersal in higher latitude sub-Antarctic *Siphonaria* (Gastropoda: Euthyneura) species.** *Journal of biogeography* 2018, **45:**874-884.
33. Moon KL, Chown SL, Fraser CI: **Reconsidering connectivity in the sub-Antarctic.** *Biological Reviews* 2017, **92:**2164-2181.
34. Nikula R, Fraser C, Spencer H, Waters J: **Circumpolar dispersal by rafting in two subantarctic kelp-dwelling crustaceans.** *Marine Ecology Progress Series* 2010, **405:**221-230.
35. Cumming R, Nikula R, Spencer H, Waters J: **Transoceanic genetic similarities of kelp-associated sea slug populations: long-distance dispersal via rafting?** *Journal of Biogeography* 2014, **41:**2357-2370.
36. Gérard K, Bierne N, Borsa P, Chenuil A, Féral J-P: **Pleistocene separation of mitochondrial lineages of *Mytilus* spp. mussels from Northern and Southern Hemispheres and strong genetic differentiation among southern populations.** *Molecular Phylogenetics and Evolution* 2008, **49:**84-91.
37. Frugone MJ, López ME, Segovia NI, Cole TL, Lowther A, Pistorius P, Dantas GP, Petry MV, Bonadonna F, Trathan P: **More than the eye can see: Genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean.** *Molecular phylogenetics and evolution* 2019, **139:**106563.
38. Fraser CI, Nikula R, Spencer HG, Waters JM: **Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum.** *Proceedings of the National Academy of Sciences* 2009, **106:**3249-3253.
39. De Broyer C, Koubbi P, Griffiths H, Raymond B, d' U, Van de Putte A, Danis B, David B, Grant S, Gutt J, et al: *Biogeographic Atlas of the Southern Ocean.* 2014.

40. Shade A, Dunn RR, Blowes SA, Keil P, Bohannan BJ, Herrmann M, Küsel K, Lennon JT, Sanders NJ, Storch D: **Macroecology to unite all life, large and small.** *Trends in ecology & evolution* 2018, **33:**731-744.
41. Murray AE, Grzymski JJ: **Diversity and genomics of Antarctic marine micro-organisms.** *Philos Trans R Soc Lond B Biol Sci* 2007, **362:**2259-2271.
42. Raes EJ, Bodrossy L, Van De Kamp J, Bissett A, Ostrowski M, Brown MV, Sow SL, Sloyan B, Waite AM: **Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean.** *Proceedings of the National Academy of Sciences* 2018, **115:**E8266-E8275.
43. Edgar RC, Flyvbjerg H: **Error filtering, pair assembly and error correction for next-generation sequencing reads.** *Bioinformatics* 2015, **31:**3476-3482.
44. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML: **Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences.** *ISME J* 2015, **9:**968-979.
45. Buttigieg PL, Ramette A: **A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses.** *FEMS Microbiol Ecol* 2014, **90:**543-550.
46. Lindström ES, Langenheder S: **Local and regional factors influencing bacterial community assembly.** *Environmental Microbiology Reports* 2012, **4:**1-9.
47. Szekely AJ, Berga M, Langenheder S: **Mechanisms determining the fate of dispersed bacterial communities in new environments.** *ISME J* 2013, **7:**61-71.
48. Langenheder S, Lindström ES: **Factors influencing aquatic and terrestrial bacterial community assembly.** *Environmental microbiology reports* 2019, **11:**306-315.
49. Astorga A, Oksanen J, Luoto M, Soininen J, Virtanen R, Muotka T: **Distance decay of similarity in freshwater communities: do macro-and microorganisms follow the same rules?** *Global Ecology and Biogeography* 2012, **21:**365-375.
50. Guillaumot C, Fabri-Ruiz S, Martin A, Eléaume M, Danis B, Féral JP, Saucède T: **Benthic species of the Kerguelen Plateau show contrasting distribution shifts in response to environmental changes.** *Ecol Evol* 2018, **8:**6210-6225.
51. Poulin E, Feral J-P: **Pattern of spatial distribution of a brood-protecting schizasterid echinoid, *Abatus cordatus*, endemic to the Kerguelen Islands.** *Mar Ecol Prog Ser* 1995, **118:**179-186.
52. David B, Choné T, Mooi R, De Ridder C: **Antarctic echinoidea. Synopses of the Antarctic benthos.** *Lichtenstein: ARG Gantner* 2005.
53. Díaz A, González-Wevar CA, Maturana CS, Palma AT, Poulin E, Gerard K: **Restricted geographic distribution and low genetic diversity of the brooding sea urchin *Abatus agassizii* (Spatangoidea: Schizasteridae) in the South Shetland Islands: A bridgehead population before the spread to the northern Antarctic Peninsula?** *Rev Chil Hist Nat* 2012, **85:**457-468.
54. Schwob G, Cabrol L, Poulin E, Orlando J: **Characterization of the gut microbiota of the Antarctic heart urchin (Spatangoida) *Abatus agassizii*.** *Frontiers in Microbiology* 2020, **11**.
55. Bowman JP, McCuaig RD: **Biodiversity, community structural shifts, and biogeography of prokaryotes within Antarctic continental shelf sediment.** *Appl Environ Microbiol* 2003, **69:**2463-2483.

56. Beiruti ZA, Bhatnagar S, Tegetmeyer HE, Geelhoed JS, Strous M, Ruff SE: **Transient exposure to oxygen or nitrate reveals ecophysiology of fermentative and sulfate-reducing benthic microbial populations.** *bioRxiv* 2017**:**146886.
57. Parada AE, Needham DM, Fuhrman JA: **Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples.** *Environ Microbiol* 2016, **18:**1403-1414.
58. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA, Caporaso JG: **Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing.** *Nat Methods* 2013, **10:**57-59.
59. Chessel D, Dufour A-B, Dray S: **Package 'ade4': Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences.** *R-package Version: 15-1* 2009.
60. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H: **Vegan: Community ecology package, 2011.** *R package version* 2011**:**1.17-18.
61. Martinez Arbizu P: **pairwiseAdonis: Pairwise multilevel comparison using adonis.** *R package version* 2017, **1**.
62. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML: **Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data.** *Methods Ecol Evol* 2013, **4**.
63. Hsieh T, Ma K, Chao A: **iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers).** *Methods in Ecology and Evolution* 2016, **7:**1451-1456.
64. Team RC: **R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.** ISBN 3-900051-07-0; 2012.
65. Paradis E: **pegas: an R package for population genetics with an integrated–modular approach.** *Bioinformatics* 2010, **26:**419-420.
66. Paradis E, Blomberg S, Bolker B, Brown J, Claude J, Cuong HS, Desper R, Didier G: **Package 'ape'.** *Analyses of phylogenetics and evolution, version* 2015, **2:**4-1.
67. Schneider S, Roessli D, Excoffier L: **Arlequin: a software for population genetics data analysis.** *User manual ver* 2000, **2:**2496-2497.
68. Hudson RR: **A new statistic for detecting genetic differentiation.** *Genetics* 2000, **155:**2011-2014.
69. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25:**1451-1452.
70. Leigh JW, Bryant D: **POPART: full-feature software for haplotype network construction.** *Methods in Ecology and Evolution* 2015, **6:**1110-1116.
71. Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, Rockhold ML, Konopka A: **Quantifying community assembly processes and identifying features that impose them.** *ISME J* 2013, **7:**2069-2079.
72. Feng Y, Chen R, Stegen JC, Guo Z, Zhang J, Li Z, Lin X: **Two key features influencing community assembly processes at regional scale: Initial state and degree of change in environmental conditions.** *Molecular ecology* 2018, **27:**5238-5251.
73. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.** *Systematic biology* 2010, **59:**307-321.
74. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic acids research* 2004, **32:**1792-1797.

75. Stegen JC, Lin X, Fredrickson JK, Konopka AE: **Estimating and mapping ecological processes influencing microbial community assembly.** *Front Microbiol* 2015, **6:**370.
76. Chase JM, Kraft NJ, Smith KG, Vellend M, Inouye BD: **Using null models to disentangle variation in community dissimilarity from variation in α-diversity.** *Ecosphere* 2011, **2:**1-11.
77. Richter-Heitmann T, Hofner B, Krah FS, Sikorski J, Wust PK, Bunk B, Huang S, Regan KM, Berner D, Boeddinghaus RS, et al: **Stochastic dispersal rather than deterministic selection explains the spatio-temporal distribution of soil bacteria in a temperate grassland.** *Front Microbiol* 2020, **11:**1391.
78. Ramoneda J, Le Roux JJ, Frossard E, Frey B, Gamper HA: **Experimental assembly reveals ecological drift as a major driver of root nodule bacterial diversity in a woody legume crop.** *FEMS Microbiology Ecology* 2020, **96:**fiaa083.
79. Wang IJ: **Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation.** *Evolution* 2013, **67:**3403-3411.
80. Assis J, Tyberghein L, Bosch S, Verbruggen H, Serrão EA, De Clerck O: **Bio-ORACLE v2. 0: Extending marine data layers for bioclimatic modelling.** *Global Ecology and Biogeography* 2018, **27:**277-284.
81. Vavrek MJ: **Fossil: palaeoecological and palaeogeographical analysis tools.** *Palaeontologia Electronica* 2011, **14:**16.
82. Adamack AT, Gruber B: **PopGenReport: simplifying basic population genetic analyses in R.** *Methods in Ecology and Evolution* 2014, **5:**384-387.
83. Kuhner MK: **LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters.** *Bioinformatics* 2006, **22:**768-770.
84. Halsey SJ, Bell TJ, McEachern K, Pavlovic NB: **Comparison of reintroduction and enhancement effects on metapopulation viability.** *Restoration Ecology* 2015, **23:**375-384.
85. Ochman H, Elwyn S, Moran NA: **Calibrating bacterial evolution.** *Proceedings of the National Academy of Sciences* 1999, **96:**12638-12643.
86. Turon X, Antich A, Palacín C, Præbel K, Wangensteen OS: **From metabarcoding to metaphylogeography: separating the wheat from the chaff.** *Ecological Applications* 2020, **30:**e02036.
87. Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, Iudicone D, Karsenti E, Speich S, Troublé R: **Open science resources for the discovery and analysis of Tara Oceans data.** *Scientific data* 2015, **2:**1-16.
88. Ceballos SG, Lessa EP, Licandeo R, Fernandez DA: **Genetic relationships between Atlantic and Pacific populations of the notothenioid fish Eleginops maclovinus: the footprints of Quaternary glaciations in Patagonia.** *Heredity (Edinb)* 2016, **116:**372-377.
89. Trovant B, Orensanz JM, Ruzzante DE, Stotz W, Basso NG: **Scorched mussels (BIVALVIA: MYTILIDAE: BRACHIDONTINAE) from the temperate coasts of South America: phylogenetic relationships, trans-Pacific connections and the footprints of Quaternary glaciations.** *Mol Phylogenet Evol* 2015, **82 Pt A:**60-74.
90. Iriarte PF, González-Wevar C, Segovia N, Rosenfeld S, Hüne M, Fainburg L, Nuñez JD, Haye P, Poulin E: **Quaternary ice sheets and sea level regression drove divergence in a marine gastropod along Eastern and Western coasts of South America.** *Scientific Reports* 2020, **10:**1-15.

91. Phillpot H: **Physical geography–climate.** *Antarctica* 1985**:**23.
92. Arkhipkin AI, Middleton D, Sirota A, Grzebielec R: **The effect of Falkland Current inflows on offshore ontogenetic migrations of the squid Loligo gahi on the southern shelf of the Falkland Islands.** *Estuarine, Coastal and Shelf Science* 2004, **60:**11-22.
93. Nagylaki T: **The strong-migration limit in geographically structured populations.** *Journal of mathematical biology* 1980, **9:**101-114.
94. Charlesworth D: **Effects of inbreeding on the genetic diversity of populations.** *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* 2003, **358:**1051-1070.
95. Thornhill DJ, Mahon AR, Norenburg JL, Halanych KM: **Open-ocean barriers to dispersal: a test case with the Antarctic Polar Front and the ribbon worm Parborlasia corrugatus (Nemertea: Lineidae).** *Mol Ecol* 2008, **17:**5104-5117.
96. Hunter RL, Halanych KM: **Evaluating connectivity in the brooding brittle star Astrotoma agassizii across the Drake Passage in the Southern Ocean.** *Journal of Heredity* 2008, **99:**137-148.
97. Bargelloni L, Marcato S, Zane L, Patarnello T: **Mitochondrial phylogeny of notothenioids: a molecular approach to Antarctic fish evolution and biogeography.** *Systematic Biology* 2000, **49:**114-129.
98. González-Wevar CA, Hüne M, Cañete JI, Mansilla A, Nakano T, Poulin E: **Towards a model of postglacial biogeography in shallow marine species along the Patagonian Province: lessons from the limpet *Nacella magellanica* (Gmelin, 1791).** *BMC Evolutionary Biology* 2012, **12:**139.
99. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L: **Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization.** *Molecular ecology* 2013, **22:**5983-5999.
100. Güller M, Puccinelli E, Zelaya DG: **The Antarctic Circumpolar Current as a dispersive agent in the Southern Ocean: evidence from bivalves.** *Marine Biology* 2020, **167**.
101. Macaya EC, Zuccarello GC: **DNA Barcoding and genetic divergence in the Giant Kelp Macrocystis (Laminariales) 1.** *Journal of Phycology* 2010, **46:**736-742.
102. Leese F, Agrawal S, Held C: **Long-distance island hopping without dispersal stages: transportation across major zoogeographic barriers in a Southern Ocean isopod.** *Naturwissenschaften* 2010, **97:**583-594.
103. Fraser CI, Morrison A, Rojas PO: **Biogeographic Processes Influencing Antarctic and sub-Antarctic Seaweeds.** In *Antarctic Seaweeds.* Springer; 2020: 43-57
104. Sul WJ, Oliver TA, Ducklow HW, Amaral-Zettler LA, Sogin ML: **Marine bacteria exhibit a bipolar distribution.** *Proc Natl Acad Sci U S A* 2013, **110:**2342-2347.
105. Logares R, Deutschmann IM, Junger PC, Giner CR, Krabberod AK, Schmidt TSB, Rubinat-Ripoll L, Mestre M, Salazar G, Ruiz-Gonzalez C, et al: **Disentangling the mechanisms shaping the surface ocean microbiota.** *Microbiome* 2020, **8:**55.
106. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A: **Structure and function of the global ocean microbiome.** *Science* 2015, **348**.
107. Milici M, Vital M, Tomasch J, Badewien TH, Giebel HA, Plumeier I, Wang H, Pieper DH, Wagner-Döbler I, Simon M: **Diversity and community composition of particle-associated and free-living bacteria in mesopelagic and bathypelagic Southern Ocean water masses: Evidence of dispersal limitation in the Bransfield Strait.** *Limnology and Oceanography* 2017, **62:**1080-1095.

108. Mestre M, Hofer J: **The Microbial Conveyor Belt: Connecting the Globe through Dispersion and Dormancy.** *Trends Microbiol* 2020.
109. Locey KJ: **Synthesizing traditional biogeography with microbial ecology: the importance of dormancy.** *Journal of Biogeography* 2010, **37:**1835-1841.
110. Miyazaki M, Sakai S, Yamanaka Y, Saito Y, Takai K, Imachi H: ***Spirochaeta psychrophila* sp. nov., a psychrophilic spirochaete isolated from subseafloor sediment, and emended description of the genus *Spirochaeta*.** *Int J Syst Evol Microbiol* 2014, **64:**2798-2804.
111. Hakim JA: *Comparison of gut microbiomes in laboratory cultured sea urchins revealing selective attributes of microbial composition based upon their feed and surroundings.* The University of Alabama at Birmingham; 2015.
112. Petro C, Starnawski P, Schramm A, Kjeldsen KU: **Microbial community assembly in marine sediments.** *Aquatic Microbial Ecology* 2017, **79:**177-195.
113. Meysman FJ, Middelburg JJ, Heip CH: **Bioturbation: a fresh look at Darwin's last idea.** *Trends in Ecology & Evolution* 2006, **21:**688-695.
114. Grossart H-P, Dziallas C, Leunert F, Tang KW: **Bacteria dispersal by hitchhiking on zooplankton.** *Proceedings of the National Academy of Sciences* 2010, **107:**11959-11964.
115. Serebryakova A, Aires T, Viard F, Serrao EA, Engelen AH: **Summer shifts of bacterial communities associated with the invasive brown seaweed Sargassum muticum are location and tissue dependent.** *PLoS One* 2018, **13:**e0206734.
116. Bowley J, Baker-Austin C, Porter A, Hartnell R, Lewis C: **Oceanic hitchhikers–assessing pathogen risks from marine microplastic.** *Trends in Microbiology* 2020.
117. Louca S, Parfrey LW, Doebeli M: **Decoupling function and taxonomy in the global ocean microbiome.** *Science* 2016, **353:**1272-1277.
118. Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC: **Drivers of bacterial β-diversity depend on spatial scale.** *Proceedings of the National Academy of Sciences* 2011, **108:**7850-7854.
119. Sintes E, De Corte D, Ouillon N, Herndl GJ: **Macroecological patterns of archaeal ammonia oxidizers in the Atlantic Ocean.** *Molecular ecology* 2015, **24:**4931-4942.
120. Zhou J, Ning D: **Stochastic community assembly: does it matter in microbial ecology?** *Microbiology and Molecular Biology Reviews* 2017, **81**.
121. Vellend M: *The theory of ecological communities.* Princeton: Princeton University Press; 2016.
122. Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, Knelman JE, Darcy JL, Lynch RC, Wickey P: **Patterns and processes of microbial community assembly.** *Microbiology and Molecular Biology Reviews* 2013, **77:**342-356.
123. Langenheder S, Lindstrom ES: **Factors influencing aquatic and terrestrial bacterial community assembly.** *Environ Microbiol Rep* 2019, **11:**306-315.
124. Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt DE, Benoit J, Sarma-Rupavtarm R, Distel DL, Polz MF: **Genotypic diversity within a natural coastal bacterioplankton population.** *Science* 2005, **307:**1311-1313.
125. Chust G, Villarino E, Chenuil A, Irigoien X, Bizsel N, Bode A, Broms C, Claus S, Fernandez de Puelles ML, Fonda-Umani S, et al: **Dispersal similarly shapes both population genetics and community patterns in the marine realm.** *Sci Rep* 2016, **6:**28730.

126. Zure M, Fernandez-Guerra A, Munn CB, Harder J: **Geographic distribution at subspecies resolution level: closely related Rhodopirellula species in European coastal sediments.** *ISME J* 2017, **11:**478-489.
127. White TA, Stamford J, Rus Hoelzel A: **Local selection and population structure in a deep-sea fish, the roundnose grenadier (Coryphaenoides rupestris).** *Molecular Ecology* 2010, **19:**216-226.
128. Everroad RC, Wood AM: **Phycoerythrin evolution and diversification of spectral phenotype in marine Synechococcus and related picocyanobacteria.** *Molecular Phylogenetics and Evolution* 2012, **64:**381-392.
129. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, Polz MF, Alm EJ: **Population genomics of early events in the ecological differentiation of bacteria.** *science* 2012, **336:**48-51.
130. Heino J, Melo AS, Siqueira T, Soininen J, Valanko S, Bini LM: **Metacommunity organisation, spatial extent and dispersal in aquatic systems: patterns, processes and prospects.** *Freshwater Biology* 2015, **60:**845-869.
131. Logares R, Tesson SVM, Canback B, Pontarp M, Hedlund K, Rengefors K: **Contrasting prevalence of selection and drift in the community structuring of bacteria and microbial eukaryotes.** *Environ Microbiol* 2018, **20:**2231-2240.
132. Nosil P, Egan SP, Funk DJ: **Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection.** *Evolution* 2008, **62:**316-336.

Figure and table titles:

Figure 1. Sampling localities across the Southern Ocean, encompassing Possession Bay and Puerto Deseado in Atlantic Patagonia (PAT1 and PAT2, respectively), King George Island in Maritime Antarctica (KGI), and Port-aux-Français in Kerguelen Islands (KER).

Figure 2. NMDS scatter diagram of the *Spirochaeta* OTUs composition in gut tissue samples across the localities.

Figure 3. Clustering of *Spirochaeta* OTUs based on their relative abundances in each site.

Figure 4. Median-joining oligotype networks of the three selected OTUs of *Spirochaeta*.

Figure 5. Gene flows summary and proposed dispersal routes across the Southern Ocean for each *Spirochaeta* taxa.

Table 1. Experimental design and sequencing data.

Table 2. Summary of oligotypes number and genetic indices per OTU and per site for the three most abundant *Spirochaeta* OTUs found in all sampling localities.

Table 3. Genetic (F_{ST}) and phylogeographic structure (ϕ_{ST}) of the *Spirochaeta* populations among localities.

Table 4. Effective numbers of migrants per generation (Nm) among *Spirochaeta* populations of the OTU6, OTU7 and OTU40.

Table 5. Quantitative parsing of ecological processes driving populations turnover within *Spirochaeta* OTUs.

Table 6. Multiple Matrix Regression with Randomization (MMRR) to quantify the relative effects of isolation by distance (IBD) and isolation by environment (IBE) on oligotypes assemblage within *Spirochaeta* OTUs.

Figure and table legends:

Figure 1. The Antarctic Polar Front (APF) and the Southern Antarctic Circumpolar Front (SACCF) are represented.

Figure 2. Presence/absence matrix converted in Bray-Curtis distances (left panel) and Unweighted Unifrac distance (right panel). Colors are assigned to the locality.

Figure 3. Clustering of Bray-Curtis distances matrix using the complete linkage method. The size of the circles indicates the repartition of a given OTU sequences among the 4 localities. The percentages indicate the OTUs' proportions in the *Spirochaeta* dataset. Framed labels and black arrows indicate the selected OTUs that were selected to be process through the Minimum Decomposition Entropy pipeline (MED).

Figure 4. Each circle represents a distinct oligotype. Colors indicate the locality of provenance. Circle size is scaled on the oligotype frequency normalized by the number of sequences in each locality, within the corresponding OTU dataset. Raw abundances are provided in Table S3.

Figure 5. Only the gene flows with Nm values > 0.25 are graphically represented. Discontinuous arrows represent Nm values > 0.25 and < 1, continuous arrow represent Nm values > 1. Continuous arrows' width is proportional to Nm values.

Table 1. N: number of samples, Nseq.: total number of cleaned sequences obtained, Relat. Abund.: (relative abundance in the global dataset).

Table 2. *N*: number of sequences, *k*: number of oligotypes, *S*: number of polymorphic sites, *H*: genetic diversity, *Π*: mean number of pairwise diversity. The mean and standard deviation were calculated from a total of 1,000 bootstraps, performed by randomly subsampling per site a number of sequences equal to the minimum number of sequences obtained among sites for a given OTU. *In the case of OTU40, the number of sequences in the PAT2 site was used to perform the resampling.

Table 3. Structure values are beneath each diagonal and *p*-values are above them. *p*-values were obtained through 1,000 permutations and the significance level was set < 0.05. *p*-values of 0 indicate value < 0.00001.

Table 4. Only gene flows with Nm values > 0.25 are considered as significant. Mean and standard deviation values were calculated from the 10 runs performed for each OTU.

Table 5. According to the Stegen et al. (2013) approach, percentage refers to the percentage of pairs of communities that appear to be driven by either homogeneous selection, homogenizing dispersal, ecological drift, dispersal limitation or variable selection.

1053 **Table 6.** The first statistical test (t) individually estimates the effect of the environmental distance
1054 and the geographic distance matrices, whereas the second one (F) evaluates the global fit of the
1055 model considering both distance matrices. *p*-values are considered as significant < 0.05.

1056 **Supplementary material:**

1057 **Additional file 1.** Pairwise PERMANOVA on *Spirochaeta* OTUs composition dissimilarities
1058 among localities. *p*-values are adjusted using the default Bonferroni method implemented in
1059 the *pairwiseAdonis* R package and are considered as significant < 0.05.

1060 **Additional file 2.** Abundance and closest sequence retrieved from Blast analysis for each of the
1061 three OTUs analysed through the MED pipeline.

1062 (1) Bowman JP, McCuaig RD: **Biodiversity, community structural shifts, and biogeography of**
1063 **prokaryotes within Antarctic continental shelf sediment.** *Appl Environ Microbiol* 2003, **69:**2463-
1064 2483.

1065 (2) Acosta-González A, Rosselló-Móra R, Marqués S: **Characterization of the anaerobic**
1066 **microbial community in oil-polluted subtidal sediments: aromatic biodegradation potential**
1067 **after the Prestige oil spill.** *Environmental microbiology* 2013, **15:**77-92.

1068 **Additional file 3.** Relative contribution of each locality in the total abundance of OTU6, OTU7
1069 and OTU40 sequences. Colors are assigned to the different localities.

1070 **Additional file 4.** Summary of number of oligotypes in each locality and per OTU of *Spirochaeta*.

1071 **Additional file 5.** Accumulation curves of OTU6 (A), OTU7 (B) and OTU40 (C) oligotypes
1072 richness. Colors are assigned to each locality. Extrapolation is calculated from Hill numbers of
1073 richness (q=0).