

RESEARCH ARTICLE

High-throughput metabarcoding of SAR11 assemblages from the southwest Atlantic shelf and arid Patagonia: richness and associated rank abundance distributions

Leandro R. Jones^{†,*}, Julieta M. Manrique[†]

Laboratorio de Virología y Genética Molecular, Universidad Nacional de la Patagonia San Juan Bosco, Trelew CP 9100, Argentina

* Correspondence: lrj000@gmail.com

Received December 1, 2022; Revised March 13, 2023; Accepted March 15, 2023

Background: Massively parallel sequencing of environmental DNA allows microbiological studies to be performed in greater detail than was possible with first-generation sequencing. For example, it facilitates the use of approaches hitherto largely applied to flora and fauna, such as rank abundance distribution (RAD) analyses.

Methods: Here, we set out to advance the knowledge on *Ca. Pelagibacterales* (SAR11) communities from southern South America using environmental sequences from the open ocean in the Argentine sea, the uncharted Engaño Bay, as well as a river and an oligohaline shallow lake from the Patagonian Steppe ecoregion. The structures of the SAR11 assemblages present in these ecosystems were dissected by direct and rarefaction-based estimates of species richness, and evaluations of the corresponding abundance distributions (ADs), which was addressed by RAD analyses.

Results: Microbial community composition analyses revealed that the studied SAR11 assemblages coexist with 27 bacterial phyla. SAR11 richness was in general very high, but ADs turned out to be highly uneven. The results were compatible with prior knowledge, and similar to that derived from point estimates of diversity. However, our comprehensive dissection allowed for more detailed quantitative comparisons to be made between the environments surveyed, and revealed differences regarding both richness and the underlying ADs.

Conclusions: Despite SAR11 assemblages being extremely rich, their ADs are very uneven. Richness and ADs can vary, not only between fresh and salt water, but also between oceanic and coastal marine environments. The obtained results provide insights on general topics such as adaptation and the contrast between marine and freshwater radiations.

Keywords: SAR11; richness; species abundance distribution; rank abundance distribution; Patagonia; Argentina

Author summary: The SAR11 clade is likely the most abundant microbial lineage on earth. Here, we provide detailed analyses of the group's richness and AD in poorly studied Southern microbiomes. This requires whole community composition analyses to be performed, which are also provided in conjunction with the SAR11-specific studies. In this way, the work describes unknown aspects of the diversity of an outstanding bacterial lineage and novel data on remote and relatively little-studied microbial communities.

INTRODUCTION

Despite their small size, the great abundance and metabolic versatility of bacteria place them among the

main actors in aquatic ecosystems, where they participate in processes of biogeochemical importance [1]. Among these processes, global carbon cycling, for which dissolved organic matter (DOM) plays a central

[†] Equivalent contribution.

role, supports life on earth [2]. Bacterioplankton contributes significantly to global carbon flux through the microbial carbon pump (MCP) and the microbial loop [3,4]. The MCP, by which the bacterioplankton produces recalcitrant DOM, provides a mechanism that contributes to the known biological carbon pump. On the other hand, heterotrophic bacteria represent a keystone of the microbial loop, since they are capable of assimilating DOM and, ultimately, are grazed by protists, thus acting as a route for introducing carbon into the ecosystem. High-performance sequencing can be considered a disruptive technology with respect to the precision with which it allows studying diversity. Before its development, studies of non-cultivable taxa were based on hundreds of sequences [5,6]; but now hundreds of thousands can easily be used [7,8]. This introduces the possibility of studying complex aspects of specific microbial taxa, such as species richness and abundance distribution (AD), the two pillars of diversity [9], with an unprecedented level of detail.

The order *Ca. Pelagibacterales*, or SAR11, is one of the most prominent bacterial groups on Earth. Members of this taxon are extremely abundant and ubiquitous in aquatic environments, being present in saltwater and freshwater [10]. The goal of this study was to uncover the richness and ADs within SAR11 assemblages from the Southwestern Atlantic Shelf (SAS) and Patagonia. For that, we used previous and novel high-throughput metabarcoding data of the small subunit ribosomal RNA (16S) gene. Besides obtaining point diversity estimates by the widely used Simpson's (*D*), Shannon's (*H*), and Pielou's (*J*) indices, we generated direct and rarefaction-based estimates of species richness, and analyzed the equitability of the corresponding ADs by RAD analyses. Richness and RAD values were consistent with the general trends that could be inferred from the point estimators (*i.e.*, *D*, *H*, and *J*). However, the former provided much more complete depictions of the studied communities, and allowed detailed quantitative comparisons to be made between the environments surveyed. These revealed that, despite richness values being very high, as could be expected from previous knowledge about SAR11 (details in the Discussion), the corresponding ADs were remarkably uneven. We also observed substantial differences between the marine (oceanic and coastal) and freshwater samples, and slighter but significant differences between the oceanic and the coastal samples. The work constitutes one of the first studies of the structure of SAR11 communities from the Argentine sea and the Patagonian Steppe ecoregion, which harbor highly productive and peculiar ecosystems. As discussed later, our results also contribute extra insights regarding the group's biology,

such as adaptation and the difference between marine and freshwater radiations.

RESULTS

We studied twelve samples from remote oceanic (*O1*, *O2*, *O3a*, *O3b*) and coastal (*C1*–*4*) locations in the Argentine sea, and a river (*ChR1*, *ChR2*) and a shallow lake (*LCa*, *LCb*) from the Patagonian Steppe. Samples *C1*–*4* were collected in the same place but at 6-month intervals; therefore, these were treated as separate observations. The *O3a/O3b*, *LCa/LCb*, and *ChR1/ChR2* sample pairs were taken synchronically, so data from each sample pair were pooled. After quality controls of 431,804 spots, we obtained 155,602 high-quality sequences (HQSs) of the V1–V3 region of the 16S gene. Microbial community composition analyses revealed the presence of 27 bacterial phyla (Supplementary Table S1). Overall, *Proteobacteria* were the most abundant, making up almost 80% of the sequences, followed by *Bacteroidetes* (~8%), *Actinobacteria* (~6%) and *Acidobacteria* (~4%) (Fig. 1, Supplementary Fig. S1 and Table S1). The rest of phyla identified had an average abundance of only about 0.14% (0.0005–1.45). These trends showed a relevant dynamism. For example, the actinobacteria and bacteroidetes reached abundances of up to about 10% and as low as 0.006% in some marine samples (Supplementary Table S1). Furthermore, as expected, there were important differences between the sea and continental environments. While in the marine communities the second most abundant phylum (*Bacteroidetes*) accounted for only about 4% of the HQSs, the river and the shallow lake presented relatively large amounts of three other phyla. In the shallow lake, the second and third most abundant taxa, *Bacteroidetes* and *Actinobacteria*, accounted for about 40% and 10% of the sequences, respectively. The samples from the river were the only ones in which the proteobacteria, which presented an abundance of around 24%, were not the most common group. Instead, this environment was dominated by members of the phylum *Acidobacteria* (~38% of the HQSs). The third most abundant phylum in the river was *Actinobacteria*, with an abundance of about 22%. These trends were similar at the class to genus ranks, as can be appreciated from the corresponding distances and the compositional comparisons from Fig. 1. Our group of interest, the SAR11 clade, presented similar abundances in the open ocean and marine coast, but it was relatively scarce in the continental environments, as will be further detailed below.

Sixty-seven thousand five hundred and ninety-four HQSs belonged to SAR11 species (s11HQSs). As anticipated above, the marine samples showed similarly high abundances of s11HQSs (Table 1). However, the

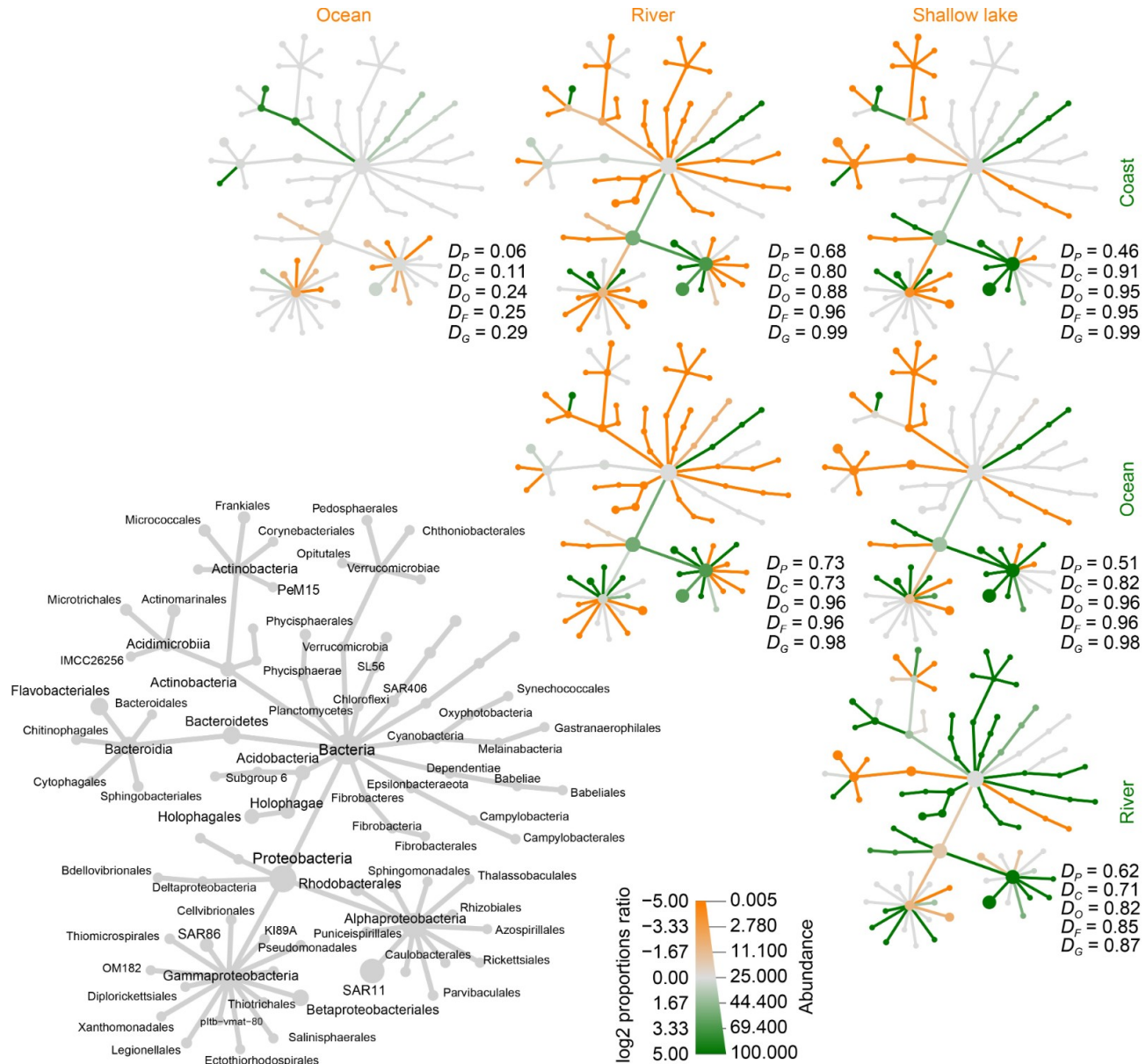


Figure 1. Heat tree matrix comparison of microbial community compositions in the studied environments. The small, colored trees summarize the comparisons between the environments in rows and columns. A group colored green is more abundant in the ecosystem in the rows, and a group colored orange is more abundant in the ecosystem of the column. The numbers close to the trees correspond to Bray-Curtis distances calculated at the phylum, class, order, family and genus ranks (D_p , D_c , D_o , D_f and D_g , respectively). The large tree in the lower left corner is identical to the smaller trees, but its nodes are taxonomically labelled (please see also Supplementary Fig. S1). Unlabeled nodes correspond either to unclassifiable groups or to yet unnamed ones. Node sizes represent abundances in the metacommunity. For improved display, taxonomic information is provided up to the order rank, and data is shown only for operational taxonomic units (OTUs) that were represented by at least 5 HQSs. The full data is provided in Supplementary Table S1.

group was much less represented in the river and virtually absent from the shallow lake, despite the number of HQSs from this last location being almost 40,000.

The large majority of marine sequences belonged to the subgroup Ia (Supplementary Fig. S2 and Table S2),

whereas in the river only species from the IIIb subclade were detected. The only s11HQS detected in the LC samples belonged to the IIIa subclade. The sample coverages (G) were similar to each other when G was calculated from non-equalized data (Table 1). However, when the data were equalized, the G values

Table 1 Diversity of SAR11 assemblages from the Argentine sea and Patagonian Steppe ecoregion

Sample	HQSS	s11HQS	<i>G</i>	<i>eG</i>	<i>D</i>	<i>eD</i>	<i>H</i>	<i>eH</i>	<i>SV</i>	<i>eSV</i>	<i>HR</i>	<i>J</i>	<i>eJ</i>
C1	15,121	9638	0.86	0.72	0.96	0.97	5.17	4.68	1985	424	445(14)	0.68	0.76
C2	9348	7176	0.87	0.75	0.95	0.95	4.87	4.48	1431	394	404(13)	0.67	0.74
C3	11,820	9549	0.91	0.81	0.88	0.88	4.00	3.83	1332	332	319(12)	0.55	0.65
C4	19,112	14,797	0.89	0.74	0.92	0.91	4.89	4.35	2648	456	435(14)	0.62	0.70
O1	14,979	8558	0.84	0.67	0.95	0.95	5.33	4.84	2076	537	513(14)	0.69	0.77
O2	6952	5149	0.83	0.69	0.97	0.97	5.42	5.05	1332	531	503(13)	0.75	0.78
O3a/b	22,147	11,358	0.83	0.60	0.96	0.95	5.99	5.33	3082	674	632(15)	0.74	0.82
ChR1/2	19,069	1368	0.86	–	0.74	–	2.88	–	271	–	261(2)	0.51	–
LCa/b	37,054	1	–	–	–	–	–	–	1	–	–	–	–

HQSS, number of high-quality sequences; s11HQS, number of SAR11 HQSS; *G*, Good's coverage (SAR11 assemblages); *eG*, equalized ($n=1368$) *G*; *D*, Simpson's diversity; *eD*, equalized *D*; *H*, Shannon's diversity; *eH*, equalized *H*; *SV*, number of sequence variants; *eSV*, equalized *SV*; *HR*, Hurlbert's richness (standard deviation); *J*, Pielou's evenness index; *eJ*, equalized *J*.

corresponding to the oceanic samples were lower than those of the coastal ones, suggesting a greater diversity in the open ocean. In agreement with this, Simpson's (*D*) and Shannon's (*H*) diversities were greater among the ocean samples than among the coastal ones (Table 1). *D* and *H* were conspicuously lower in the river than at the sea.

As explained above, this study took advantage of the massiveness of high-throughput sequencing data to obtain a detailed overview of SAR11 species richness and corresponding AD. Richness was first approached by counting the number of sequence variants present in each sample. The detected s11HQSS corresponded to a total of 13,123 such variants (Otu00001–13123; Table 1 and Supplementary S3). The marine samples presented between 1332 and 3082 variants, whereas the freshwater ones presented only 272. Despite the ocean and coastal samples presenting similar proportions of s11HQSS, as detailed above, the former had an average of 2163 sequence variants whereas the latter had, on average, 1849 such variants. In agreement with this, after equalization, the marine samples presented 485.2 variants on average against the only 272 retrieved from the continental samples, and the Engaño Bay samples presented 421 variants on average whereas the oceanic samples had an average of 570 variants. The same ordering was obtained using Hurlbert's rarefaction (Table 1; Supplementary Fig. S3).

Abundance distributions were initially assessed using Pielou's (*J*) evenness index. These analyses revealed that the oceanic samples were more even than the coastal ones, and that evenness was much lower in the river than at either of the marine locations (oceanic and coastal) (Tables 1 and 2; Supplementary Fig. S4). Therefore, to delve into the subject with greater detail, a RAD analysis was performed to compare the relative frequencies of each of our sequence variants in each of the environments surveyed. These were found to be very

Table 2 Comparisons of Pielou's evenness (*J*) through hypothesis tests¹

Comparison	ΔJ^2	$p(\Delta J)^3$	$n\Delta J^4$	$p(n\Delta J)^5$
Coast vs. ocean	0.102	$p = 0.034$	0.113	$p < 0.001$
Coast vs. river	0.245	$p = 0.085$	0.283	$p < 0.001$
Ocean vs. river	0.329	$p = 0.005$	0.373	$p < 0.001$
Sea vs. river	0.307	$p = 0.022$	0.348	$p < 0.001$

¹ $H_0: J_O = J_C = J_R = J$, where J_O , J_C and J_R are the ocean, coast and river evenness, respectively.

² $\Delta J = |J_1 - J_2|$, where J_1 and J_2 are the evenness in the environments compared.

³ Probability that ΔJ is greater than or equal to the observed value.

⁴ $n\Delta J = |nJ_1 - nJ_2|$, where nJ_1 and nJ_2 are evenness values obtained from MaxRank-normalized abundance vectors.

⁵ Probability that $n\Delta J$ is greater than or equal to the observed value.

unequal in general, but were much more uneven in the river than in the sea. While 48% of the riverine sequences corresponded to a single variant (Otu00012), the most abundant variant in the sea was represented by only 11.1% of the marine sequences (Fig. 2A; Supplementary Table S4). Furthermore, the most abundant variant from the river was about 3 times more abundant than the second most abundant variant, and about 19 times more abundant than the third most abundant one. Conversely, in the ocean, the second and third most abundant variants were just 2 and 2.5 times less abundant than the most abundant variant there. Likewise, the rank 4 variant from the river had an abundance of only 1%, whereas, in the sea, the variants with abundances of 1% or less ranked 11th or above. Commonness and rarity patterns within the oceanic and coastal assemblages were also different from each other. The most abundant variant in the coast (Otu00001) corresponded to 16.1% of the coastal sequences (Fig. 2B; Supplementary Table S5). In contrast, only 8.6% of the oceanic sequences belonged to the most abundant oceanic variant (Otu00003). The second most abundant

variant from the coast was about half as abundant as the more abundant one, whereas in the ocean the second-ranking variant was almost as abundant as the first.

Abundance distributions were further assessed and compared by MaxRank normalization, a technique that generates normalized abundance vectors and confidence intervals that can be used in graphic exploratory analyses and also quantitatively compared to each other by multivariate analysis. The normalized RAD (NRAD) of the river was much more right skewed than the marine ones (Fig. 3A). Related to this, the abundances of each of our normalized ranks were very different in

the sea and in the river, with the exception of ranks 50 to 80, which showed relatively small differences. The coastal NRADs presented slightly heavier heads than the oceanic ones (Fig. 3A). In comparison to the coast, the normalized ranks greater than 50 were conspicuously more abundant in the ocean. Furthermore, mid-abundance ranks 3 to 15 were particularly more abundant at the coastal location than at the oceanic sites. A classical multidimensional scaling analysis (MDS), based on Manhattan distances obtained from the standardized abundance vectors in Fig. 3A, separated the samples into two compact groups encompassing

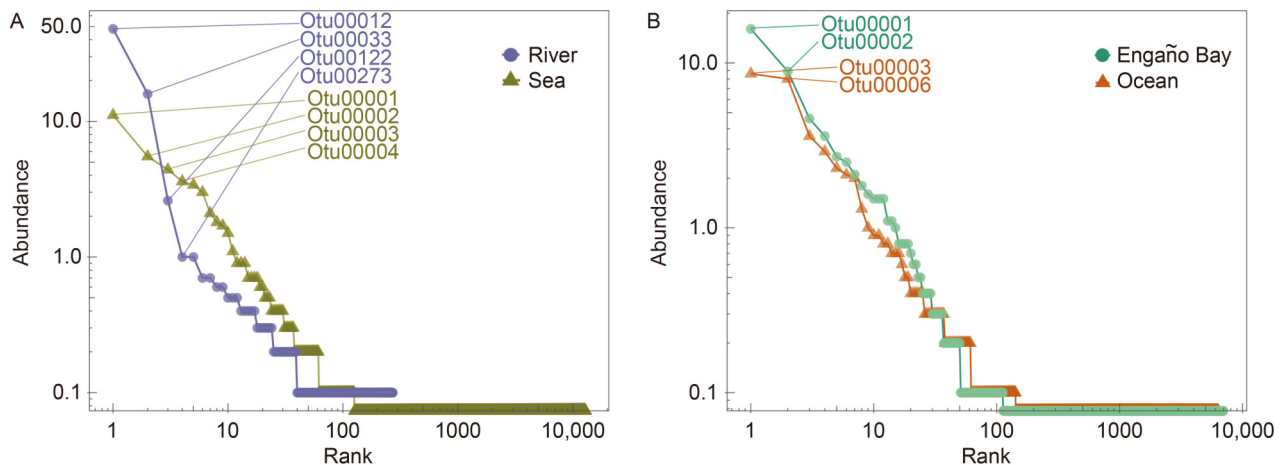


Figure 2. Rank abundance distributions (RAD) of SAR11 assemblages from the Argentine sea and arid Patagonia. Abundances are plotted in decreasing order from the most abundant variant (rank = 1) on the left to the least abundant one on the right. The axes are scaled logarithmically. Panel (A) compares RADs at the sea (Sea) and a river (River). Panel (B) depicts RADs inshore (Engaño Bay) and at the open ocean.

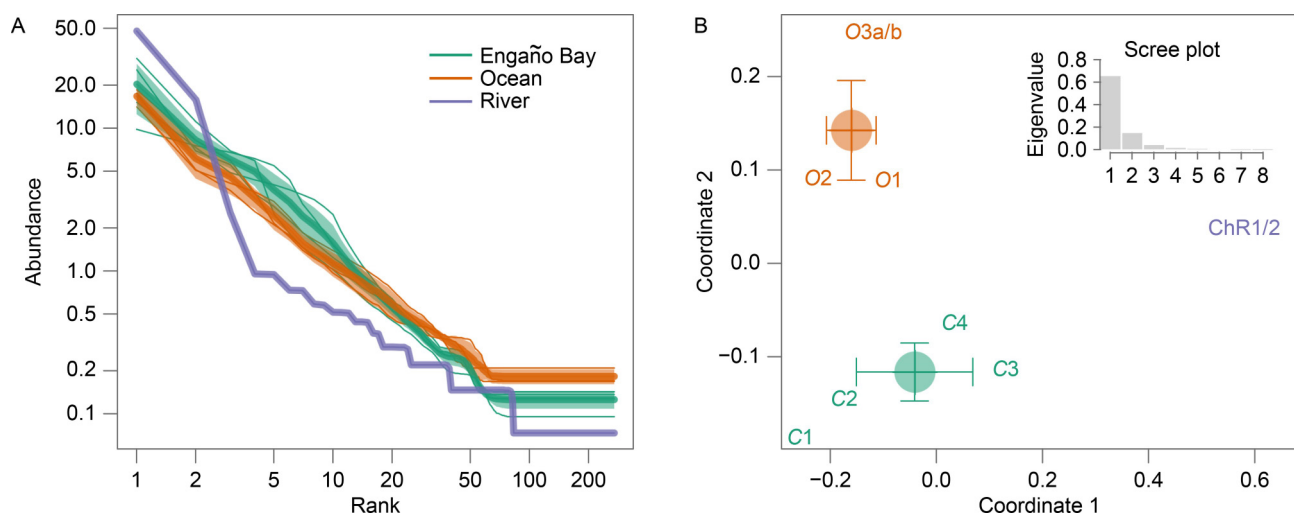


Figure 3. MaxRank normalization analysis. (A) Normalized RADs (NRADs) from the river (River), the open ocean (Ocean), and inshore (Engaño Bay). Shaded areas correspond to 90% confidence intervals. (B) MDS-ordination. The dots and error bars correspond to averages and 90% confidence intervals, respectively. The two coordinates in the MDS plot explain 94.7% of the NRAD distances (inset).

samples O1–3 and C1–4, which also were well discriminated from the river samples in the MDS plot (Fig. 3B).

DISCUSSION

We used high-throughput metabarcoding to investigate species richness and the corresponding ADs within SAR11 assemblages from marine and continental ecosystems from the SAS and arid Patagonia. These territories have unique aquatic environments whose microbiomes have been relatively little studied, as is the case for many other regions in southern South America.

The number of species, or richness, and the equitability of their AD, are a reflection of ecological processes such as niche partitioning, resource distribution, and disturbances [9]. Richness is known to be very high within SAR11 assemblages. High enough, by instance, as to have hampered traditional metagenomic assembly despite the abundance of SAR11 reads in most metagenomes [11]. But studies focused specifically on richness are few compared to those that have dealt with other aspects of the biology of the group. A pioneer metagenomic read recruitment (MRR) study, showed that sequences retrieved from the Sargasso Sea (SS) were between 30% and 80% identical to SAR11 reference strain HTCC1062 at the amino acid level, indicating that the SAR11 sequences from SS are very different to each other [12]. Later MRR studies made elsewhere, and with other reference sequences, produced similar results [13,14]. Likewise, comparisons performed between 140 internal transcribed spacer (ITS) sequences retrieved from 4 metagenomes from the Arctic Ocean, showed that only 29 ITS sequences were identical to at least one of the rest of sequences [15]. On the other side, a metabarcoding study that surveyed ITS molecular clones from the Red Sea, revealed the presence of 69 to 130 SAR11 sequence variants in 7 samples [16]. Our results extend the above-referenced studies, showing that the cells in SAR11 assemblages can present hundreds or thousands of variants of the 16S gene (Table 1). SAR11 species possess a single copy of the 16S gene [17,18], which eases the interpretation of metabarcoding data. Interestingly, small differences in this gene can be associated with important differences at the genome level. For example, *Ca. Pelagibacter ubique* str. HTCC1062 and *Ca. Pelagibacter* sp. str. HTCC7211, two representatives of the *Ia* subclade, are about 99% identical to each other in the 16S gene, but differ in their gene content and are only about 75% identical at the amino-acid level [19,20]. We are not aware of studies that have used high-performance metabarcoding to specifically study richness. So, to the best of our knowledge, this work provides the first data

in that regard. Our results show that high-throughput metabarcoding can provide a very detailed picture of variation at a single locus compared to alternative approaches. For example, the 4 metagenomes from the Arctic Ocean mentioned above provided 140 sequences of the targeted gene, whereas we could generate between 7176 and 14,797 such sequences from our 4 coastal samples (Table 1). MRR has the desirable property of targeting the entire genome, but there is no way to translate the results generated by this method into measures of richness and AD. Furthermore, it requires a reference sequence, whereas metabarcoding is a reference-free technique.

Compared with the marine samples, our riverine samples presented few variants and a much more uneven AD (Tables 1 and 2; Figs. 2 and 3). Furthermore, although we generated almost 40,000 HQSs from the studied shallow lake (LC), we detected no sequences from the *IIIB* freshwater subclade there, which we attribute to the oligohaline nature of this environment (Materials and Methods). Low-frequency sequences from marine lineages have already been observed in freshwater bodies from elsewhere [21]. Furthermore, atypical species from the *IIIA* subclade, which reaches greater abundances in more brackish water and whose abundance is sensitive to salinity [22], have been detected in mesohaline lakes from elsewhere [23,24]. At the time of collection, the salinity at LC was less than typical brackish water salinities, but greater than that observed in the river. Also worth noticing, is that the only variant present at LC was detected in no other environment. Therefore, our overall interpretation is that subclade *IIIB* species cannot thrive at LC, and that the possibility that *IIIA* lineages can thrive there, perhaps at very low frequencies, is plausible in the light of the available evidence. Given that subclades *IIIA* and *IIIB* are closely related to each other, *IIIA* species from unexplored ecosystems may shed light on the marine to freshwater transition of the group, thus their presence at LC deserves further investigation.

Freshwater SAR11 species correspond to a single lineage, whereas nine marine subclades have been described. Two interpretations have been proposed for this difference: (i) that freshwater environments may have been colonized so quickly that the group had no time to diversify, and (ii) that diversification in freshwater environments is limited by strong stabilizing selection [25,26]. The non-equitability observed here at the river (Figs. 2A and 3) supports hypothesis (ii), because stabilizing selection can generate uneven ADs by inducing the supremacy of a narrow set of species. On the other hand, it could be hypothesized that the constant water flow in the river may cause an elevated species turnover. This, together with geographic

isolation, which is known to produce uneven distributions [27], may also explain the extreme concavity of the river RAD, although further analysis will be required to delve into this possibility.

Our C1–4 samples were taken inshore at ~6 months intervals. Other marine samples are from remote oceanic locations; samples *O2* and *O3a/b* are from a region traversed by a cold current, and the *O1* site is close to the confluence of a cold current and a warm one. Therefore, the contrasts observed between our oceanic and coastal samples can be attributed to the location of the samples relative to the coast. This agrees with a previous study in which a sample from the South Pacific Gyre (SPG) presented four times more SAR11 variants than a sample from the Chilean coast. This observation was interpreted as reflecting an advantage of the group in the oligotrophic SPG [28]. However, the ranks of each of our variants varied greatly between the ocean and coastal ecosystems (Fig. 2B; Supplementary Table S5), which suggests that some SAR11 species could be better adapted to the coastal conditions, whereas others might be better adapted to the ocean. This is consistent with evidence that oceanic *Ia* isolates have adaptations that are absent in strains isolated at coastal locations [20]. However, if that were correct, that is if the here-observed ranking differences were a reflection of niche preferences, then there would be no reason for the ocean and coastal assemblages to present different ADs, unless some disturbing factor is intervening. Likewise, additional factors still should be invoked to explain the unequal ADs if ranking differences were to be attributed to genetic or ecological drift [29–31]. Further research will be required to unveil the ultimate cause/s of these interesting patterns of diversity.

MATERIALS AND METHODS

Study region

The SAS is relatively narrow in Brazilian and Uruguayan waters. But below approximately 35 degrees South latitude, roughly coinciding with the Río de La Plata estuary, it begins to widen, reaching extensions of up to almost 900 km in its southernmost region. The Malvinas current, a branch of the Antarctic circumpolar current, sweeps the SAS from south to north, up to a region of mesoscale variability produced by its encounter with the warm Brazil current, approximately at latitude 38 degrees South. The SAS influence region comprises one of largest and richest oceanic ecosystems on Earth [32]. The study of the bacterial communities of this region is one of the fundamental steps in developing productive models applicable to the use and monitoring of the corresponding ecosystem services. Besides studying

SAR11 assemblages from the Argentine sea, we made the first ever study of the SAR11 communities from a river and a shallow lake enclaved in the Patagonian Steppe ecoregion, a semiarid scrub plateau covering most of the southern tip of South America [33–36]. In addition to the absence of previous studies in this unique ecosystem, the sites surveyed represent some of the very few aquatic environments in arid Patagonia. Furthermore, they present a high degree of isolation from similar environments elsewhere, and so they offer special interest regarding the global distribution of the group.

Sampling and high-throughput data generation

The oceanic samples (*O1*, *O2*, *O3a*, and *O3b*) were collected at remote points in the Argentine sea during R/V “Coriolis II” expedition. The *O1* sampling point (39.95°S–55.68°W) is located about 250 km offshore, close to the confluence of the Malvinas and Brazil currents [37]. The *O2* (45.93°S–57.7°W) and *O3a/b* (46°S–59.39°W) sites are located in the Malvinas current, 131 km apart from each other and 685 and 739 km from *O1*, respectively. These sites, which are about 620 and 740 km offshore, respectively, are located close to the Blue Hole region, within one of largest squid fisheries on Earth [38]. The coastal samples (C1–4) were collected at a single station in Engaño Bay (~43.34°S–65.03°W) at 6 months intervals and during high tides. Said station is about 4 km from the Chubut River mouth. However, the estuary front moves 4 to 6 km inland during high tides [33], so our sampling point always presented seawater conditions (Table 3). The collection and processing of samples *O1*–2 and C1–4 were described elsewhere [30,39]. Briefly, 3 L of surface water were picked from a depth of about 1 m, and the picoplankton was physically isolated from the rest of cells in the samples. PCR amplification and high-throughput sequencing of the V1–3 regions of the 16S gene were then carried out. Here-reported data from samples *O3a* and *O3b* were obtained following the same procedures. The continental samples (~3 L each) are from the Chubut River (samples *ChR1* and *ChR2*; 43.446°S–65.945°W) and Chiquichano Chief (in Spanish “Cacique Chiquichano”) shallow lake (samples *LCa* and *LCb*; 43.249°S–65.296°W). Collection was performed using pre-cleaned carboy containers by immersion at about 60 cm depth. Sample processing and 16S data generation were as described for the marine samples. The sequence data used here are available under GenBank BioProject PRJNA312212 (*O1* SRR3180667, *O2* SRR3176906, *O3a* SRR3180683, *O3b* SRR3180684, C1 SRR3180668, C2 SRR3180670, C3 SRR3180669, C4 SRR3180671, *LCa* SRR3180680,

LCb SRR3180681, Chr1 SRR3180678, Chr2 SRR3180679). The samples' origins and main characteristics are summarized in Table 3.

Table 3 Samples' origins and characteristics

Sample	Env.	Lat/Lon	Depth	PSU	Temp	pH
O1	OC	-39,95/-55,68	1.95	33.92	12.39	8.13
O2	OC	-45,93/-57,7	0.53	34.14	12.95	8.24
O3a	OC	-46/-59,39	1.03	34.57	12.83	8.27
O3b	OC	-46/-59,39	1.03	34.58	12.83	8.27
C1	CO	-43,44/-65,11	~1.00	32.96	16.13	8.34
C2	CO	-43,44/-65,11	~1.00	32.88	7.52	8.32
C3	CO	-43,44/-65,11	~1.00	32.84	15.71	8.19
C4	CO	-43,44/-65,11	~1.00	34.86	9.05	8.14
Chr1	Rv	-43,45/-65,92	~0.6	0.13	9.66	8.11
Chr2	Rv	-43,45/-65,92	~0.6	0.14	9.66	8.14
LC1	SL	-43,24/-65,29	~0.6	2.43	12.23	8.28
LC2	SL	-43,24/-65,29	~0.6	2.42	12.23	8.28

Env, environment; OC, ocean; CO, coast; Rv, river; SL, shallow lake; Lat/Lon, latitude/longitude; Depth, collection depth; PSU salinity; Temp, temperature (Celsius degrees).

Generation of HQSs

Quality controls were performed with Mothur [40]. We dismissed flowgrams presenting homopolymeric tracts larger than 8 bases, and less than 360 or more than 720 flows, and the remaining flowgrams were denoised by the PyroNoise algorithm implemented in Mothur. Then, we eliminated the sequences presenting indeterminate positions and those differing in at least 3 bases with respect to the primers and/or 1 base with respect to the barcodes. The sequences that presented 200 bases or less, after trimming primers and barcodes, were dismissed. The remaining data were grouped into groups of sequences displaying no more than 2 substitutions to each other, and potential chimeras were identified by *Chimera.uchime* and eliminated. Finally, the sequences were grouped into operational taxonomic units (OTUs) by the *cluster* function of Mothur setting the *cutoff* parameter to 0.00. This settings clusters the sequences that differ to each other by no more than 0.0049 distance units. Although this procedure may cause some real low-frequency variants to go unnoticed, it has the desirable property of filtering out rare recalcitrant errors, as recommended in the program's documentation. The number of sequences clustered into each OTU can be interpreted as the number of sampled individuals harboring the corresponding sequence variant, or as the number of times each sequence variant was sampled.

Taxonomic profiling and recruitment of SAR11 sequences

Sequences were classified by Mothur's Wang's Naïve Bayesian Classifier (WNBC) and the 132 release of the SILVA database. The obtained data were analyzed by the R packages *metacoder* [41] and *vegan* [42], and by base R functions [43]. After that, non-SAR11 sequences were dismissed, and the remaining sequences were assigned to SAR11 subgroups using WNBC and the reference sequences listed in the Supplementary Material. Besides counting the number of s11HQSS obtained per sample, interpreted as sampling efforts on the target SAR11 communities, sampling efforts were assessed by the Good's coverage index using the R package *QsRutils* [44].

Diversity analyses

Sequence variants distributions and diversity were analyzed by base R functions and the package *vegan*. We used raw and, where indicated, equalized data. Equalization was performed in *vegan* by equating sampling efforts to the sampling effort of the sample that presented the smaller number of s11HQSS ($n=1368$; Table 1). Point estimates of diversity were obtained by the Shannon's (H) and Simpson's (D) indices. The terms diversity and richness are sometimes used interchangeably. But as explained earlier, diversity is made of two components: the number of species, and the equitability in their AD, also known as *evenness*. For example, in Shannon's entropy:

$$H = - \sum_{i=1}^q p_i \log p_i, \quad (1)$$

q accounts for richness, and p for the corresponding AD. Here, we dissected diversity into its two components, an approach that has been little used in microbial ecology. Richness values were estimated by direct counting and by the Hurlbert's rarefaction method [45] in *vegan*, which allows generating confidence intervals of richness values. Evenness values within the studied assemblages were assessed by Pielou's index in *vegan*, and by RAD analyses using *BiodiversityR* [46]. Pielou's index (J) is defined as the ratio of the observed H (Eq. (1)) to its maximum possible value, $H_{\max} = \log q$:

$$J = H/H_{\max}. \quad (2)$$

Thus, J approaches 1 as evenness increases. Statistical contrasts were performed by permutation [9]. We define $\Delta J = |J_1 - J_2|$, where J_1 and J_2 are Pielou's evenness values obtained from environments 1 and 2, respectively. We set H_0 such that evenness is homogeneous across all locations, and, for each contrast,

obtained null distributions by generating 1000 pairs of permuted datasets. Permutations were generated by randomly swapping the elements of the equalized abundance vectors of the compared environments. Furthermore, we also define $n\Delta j = |nJ_1 - nJ_2|$, where nJ_1 and nJ_2 are the evenness values obtained from normalized data obtained by the MaxRank normalization technique [47]. Permutations were obtained as described for Δj , except in the case of the equalization step. These analyses were done with R scripts that are available from the authors upon request. Typically, RADs are visualized as log-scaled, bidimensional plots, where abundances are plotted in decreasing order from greatest (rank = 1, placed on the left in the x -axis) to smallest (placed on the far right of the x -axis). Therefore, in such graphs, the RADs' concavity, or hollowness, increases with increase of unevenness. RAD shapes were quantitatively compared to each other by MaxRank normalization ($R=271$; $N=100$) in *RADanalysis* [48], and MDS on Manhattan distances, as advised by the authors of the method [47]. MDS was performed by the *cmdscale* function ($k=2$, $eig=T$) of R.

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at <https://doi.org/10.15302/J-QB-023-0329>.

ACKNOWLEDGEMENTS

This work was supported by grants PIP 2021–2023 11220200102657CO (CONICET), PICT 2020 series A-03643 (FONCYT), and PI 1657 (UNPSJB). LRJ and JMM are members of CONICET. Continuous support from civil association ArGen (Argentina Genetics) is most appreciated.

COMPLIANCE WITH ETHICS GUIDELINES

Conflicts of interest The authors Leandro R. Jones and Julieta M. Manrique declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

OPEN ACCESS

This article is licensed by the CC By under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted

by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

1. Falkowski, P. G., Fenchel, T. and Delong, E. F. (2008) The microbial engines that drive earth's biogeochemical cycles. *Science*, 320, 1034–1039
2. Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Höglberg, P., Linder, S., *et al.* (2000) The global carbon cycle: a test of our knowledge of earth as a system. *Science*, 290, 291–296
3. Pomeroy, L. Williams, P. leB., Azam, F. and Hobbie, J. (2007) The Microbial Loop. *Oceanography* (Wash. D.C.), 20, 28–33
4. Jiao, N., Herndl, G. J., Hansell, D. A., Benner, R., Kattner, G., Wilhelm, S. W., Kirchman, D. L., Weinbauer, M. G., Luo, T., Chen, F., *et al.* (2010) Microbial production of recalcitrant dissolved organic matter: long-term carbon storage in the global ocean. *Nat. Rev. Microbiol.*, 8, 593–599
5. Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. and Field, K. G. (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, 345, 60–63
6. Amann, R. L., Ludwig, W. and Schleifer, K. H. (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.*, 59, 143–169
7. Cui, H., Li, Y. and Zhang, X. (2016) An overview of major metagenomic studies on human microbiomes in health and disease. *Quant. Biol.*, 4, 192–206
8. Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J. and Crosby, L. (2018) Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems*, 3, e00055–e18
9. Legendre, P. and Legendre, L. (1998) *Numerical Ecology*. Amsterdam: Elsevier
10. Giovannoni, S. J. (2017) SAR11 bacteria: the most abundant plankton in the oceans. *Annu. Rev. Mar. Sci.*, 9, 231–255
11. Haro-Moreno, J. M., Rodriguez-Valera, F., Rosselli, R., Martinez-Hernandez, F., Roda-Garcia, J. J., Gomez, M. L., Fornas, O., Martinez-Garcia, M. and López-Pérez, M. (2019) Ecogenomics of the SAR11 clade. *Environ. Microbiol.*, 22, 1748–1763
12. Wilhelm, L. J., Tripp, H. J., Givan, S. A., Smith, D. P. and Giovannoni, S. J. (2007) Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol. Direct*, 2, 27
13. Delmont, T. O., Kiefl, E., Kilinc, O., Esen, O. C., Uysal, I., Rappé, M. S., Giovannoni, S. and Eren, A. M. (2019) Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. *eLife*, 8, e46497
14. López-Pérez, M., Haro-Moreno, J. M., Coutinho, F. H., Martinez-Garcia, M. and Rodriguez-Valera, F. (2020) The evolutionary success of the marine bacterium SAR11 analyzed through a metagenomic perspective. *mSystems*, 5, e00605–e00620

15. Kraemer, S., Ramachandran, A., Colatriano, D., Lovejoy, C. and Walsh, D. A. (2020) Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *ISME J.*, 14, 79–90
16. Ngugi, D. K. and Stingl, U. (2012) Combined analyses of the ITS loci and the corresponding 16S rRNA genes reveal high micro- and macrodiversity of SAR11 populations in the Red Sea. *PLoS One*, 7, e50274
17. Grote, J., Thrash, J. C., Huggett, M. J., Landry, Z. C., Carini, P., Giovannoni, S. J. and Rappé, M. S. (2012) Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *MBio*, 3, e00252–e12
18. Henson, M. W., Lanclos, V. C., Faircloth, B. C. and Thrash, J. C. (2018) Cultivation and genomics of the first freshwater SAR11 (LD12) isolate. *ISME J.*, 12, 1846–1860
19. Cameron Thrash, J. C., Temperton, B., Swan, B. K., Landry, Z. C., Woyke, T., DeLong, E. F., Stepanauskas, R. and Giovannoni, S. J. (2014) Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J.*, 8, 1440–1451
20. Carini, P., Van Mooy, B. A. S. V., Thrash, J. C., White, A., Zhao, Y., Campbell, E. O., Fredricks, H. F. and Giovannoni, S. J. (2015) SAR11 lipid renovation in response to phosphate starvation. *Proc. Natl. Acad. Sci. USA*, 112, 7767–7772
21. Paver, S. F., Muratore, D., Newton, R. J. and Coleman, M. L. (2018) Reevaluating the salty divide: phylogenetic specificity of transitions between marine and freshwater systems. *mSystems*, 3, e00232–e18
22. Herlemann, D. P., Woelk, J., Labrenz, M. and Jürgens, K. (2014) Diversity and abundance of “Pelagibacterales” (SAR11) in the Baltic Sea salinity gradient. *Syst. Appl. Microbiol.*, 37, 601–604
23. Oh, S., Zhang, R., Wu, Q. L. and Liu, W.-T. (2014) Draft genome sequence of a novel SAR11 clade species abundant in a Tibetan Lake. *Genome Announc.*, 2, e01137–e14
24. Oh, S., Zhang, R., Wu, Q. L. and Liu, W.-T. (2016) Evolution and adaptation of SAR11 and *Cyanobium* in a saline Tibetan lake. *Environ. Microbiol. Rep.*, 8, 595–604
25. Logares, R., Brate, J., Heinrich, F., Shalchian-Tabrizi, K. and Bertilsson, S. (2009) Infrequent transitions between saline and fresh waters in one of the most abundant microbial lineages (SAR11). *Mol. Biol. Evol.*, 27, 347–357
26. Eiler, A., Mondav, R., Sinclair, L., Fernandez-Vidal, L., Scofield, D. G., Schwientek, P., Martinez-Garcia, M., Torrents, D., McMahon, K. D., Andersson, S. G., *et al.* (2016) Tuning fresh: radiation through rewiring of central metabolism in streamlined bacteria. *ISME J.*, 10, 1902–1914
27. Latimer, A. M., Silander, J. A. Jr and Cowling, R. M. (2005) Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science*, 309, 1722–1725
28. West, N. J., Lepère, C., Manes, C.-L. O., Catala, P., Scanlan, D. J. and Lebaron, P. (2016) Distinct spatial patterns of SAR11, SAR86, and actinobacteria diversity along a transect in the ultra-oligotrophic South Pacific Ocean. *Front. Microbiol.*, 7, 234
29. Hellweger, F. L., van Sebille, E. and Fredrick, N. D. (2014) Biogeographic patterns in ocean microbes emerge in a neutral agent-based model. *Science*, 345, 1346–1349
30. Manrique, J. M. and Jones, L. R. (2017) Are ocean currents too slow to counteract SAR11 evolution? A next-generation sequencing, phylogeographic analysis. *Mol. Phylogenet. Evol.*, 107, 324–337
31. Vergin, K., Jhirad, N., Dodge, J., Carlson, C. and Giovannoni, S. (2017) Marine bacterioplankton consortia follow deterministic, non-neutral community assembly rules. *Aquat. Microb. Ecol.*, 79, 165–175
32. Dogliotti, A., Lutz, V. and Segura, V. (2014) Estimation of primary production in the southern Argentine continental shelf and shelf-break regions using field and remote sensing data. *Remote Sens. Environ.*, 140, 497–508
33. Piccolo, M. C. and Perillo, G. M. E. (1999) The Argentina Estuaries: A Review. In: *Estuaries of South America*, Piccolo, M. C. & Perillo, G.M.E. (Ed.), Heidelberg: Springer
34. Carbonell-Sillett, L., Cavallaro, A., Pereyra, D. A., Askenazi, J. O., Goldstein, G., Scholz, F. G. and Bucci, S. J. (2022) Soil respiration and N-mineralization processes in the Patagonian steppe are more responsive to fertilization than to experimental precipitation increase. *Plant Soil*, 479, 405–422
35. Derguy, M. R., Martinuzzi, S. and Arturi, M. (2022) Bioclimatic changes in ecoregions of southern South America: trends and projections based on Holdridge life zones. *Austral Ecol.*, 47, 580–589
36. Mataloni, G. and Quintana, R. D., eds. (2022) *Freshwaters and Wetlands of Patagonia*. Springer International Publishing
37. Matano, R. P., Palma, E. D. and Piola, A. R. (2010) The influence of the Brazil and Malvinas Currents on the Southwestern Atlantic Shelf circulation. *Ocean Sci.*, 6, 983–995
38. Torres Alberto, M. L., Bodnariuk, N., Ivanovic, M., Saraceno, M. and Acha, E. M. (2020) Dynamics of the confluence of Malvinas and Brazil currents, and a southern Patagonian spawning ground, explain recruitment fluctuations of the main stock of *Illex argentinus*. *Fish. Oceanogr.*, 30, 127–141
39. Giaccardi, L. I., Badenas, M. A., Jones, L. R. and Manrique, J. M. (2022) Abundant microbes of surface sea waters of the uncharted Engaño Bay at the Atlantic Patagonian Coast: relevance of bacteria-sized photosynthetic eukaryotes. *Aquat. Ecol.*, 56, 1217–1230
40. Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 75, 7537–7541
41. Foster, Z. S. L., Sharpton, T. J. and Grünwald, N. J. (2017) Metacoder: an R package for visualization and manipulation of community taxonomic diversity data. *PLOS Comput. Biol.*, 13, e1005404
42. Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O’Hara, R. B., Simpson, G. L., Solymos, P., *et al.* (2020) *vegan*: community ecology package, available on the website of cran.r-project
43. R Core Team. (2022) *R: a language and environment for*

- statistical computing, R foundation for statistical computing, Vienna, Austria, available on the website of R-project
44. Quensen, J. (2019) QsRutils: R functions useful for community ecology, available on the website of GitHub
 45. Hurlbert, S. H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, 52, 577–586
 46. Kindt, R. (2005) *Tree Diversity Analysis: A Manual and Software for Common Statistical Methods for Ecological and Biodiversity Studies*. Nairobi: World Agroforestry Centre (ICRAF)
 47. Saeedghalati, M., Farahpour, F., Budeus, B., Lange, A., Westendorf, A. M., Seifert, M., Küppers, R. and Hoffmann, D. (2017) Quantitative comparison of abundance structures of generalized communities: from B-cell receptor repertoires to microbiomes. *PLOS Comput. Biol.*, 13, e1005362
 48. Saeedghalati, M., Farahpour, F. and Hoffmann, D. (2016) RADanalysis: normalization and study of rank abundance distributions. Available on the website of cran.R-project