

# SCIENTIFIC REPORTS



OPEN

## Lineage divergence, local adaptation across a biogeographic break, and artificial transport, shape the genetic structure in the ascidian *Pyura chilensis*

Nicolás I. Segovia<sup>1,2,3</sup>, Cristian Gallardo-Escárate<sup>2</sup>, Elie Poulin<sup>3</sup> & Pilar A. Haye<sup>1,2</sup>

Marine benthic organisms inhabit a heterogeneous environment in which connectivity between populations occurs mainly through dispersive larval stages, while local selective pressures acting on early life history stages lead to non-random mortality, shaping adaptive genetic structure. In order to test the influence of local adaptation and neutral processes in a marine benthic species with low dispersal, in this study we used Genotyping by Sequencing technology to compare the neutral and putatively selected signals (neutral and outlier loci, respectively) in SNPs scattered throughout the genome in six local populations of the commercially exploited ascidian *Pyura chilensis* along the southeast Pacific coast (24°–42°S). This species is sessile as an adult, has a short-lived larval stage, and may also be dispersed by artificial transport as biofouling. We found that the main signal in neutral loci was a highly divergent lineage present at 39°S, and a subjacent signal that indicated a separation at 30°S (north/south), widely reported in the area. North/south separation was the main signal in outlier loci, and the lineage divergence at 39°S was subjacent. We conclude that the geographic structure of the genetic diversity of outlier and neutral loci was established by different strengths of environmental, historical and anthropogenic factors.

Marine benthic species that inhabit patchy and heterogeneous substrates such as rock beds or sandy beaches are likely to be structured as metapopulations<sup>1,2</sup>. These species usually have dispersive larval stages that can act as propagules for connectivity between local populations. However, the amount of connectivity between their populations does not depend only on the presence and duration of dispersive larval stages<sup>3,4</sup>. Local adaptive pressures may differ drastically between local populations and may act as a settlement and survival sieve for early life stages of benthic marine organisms arriving to a local population<sup>5</sup>. Local adaptation can lead to strong differences between larval dispersal potential and realized dispersal<sup>5–7</sup>, by reducing recruitment and establishment of foreign individuals<sup>5,8,9</sup> or by reducing the fitness of juveniles that establish and persist in the receiving population<sup>10</sup>.

The consequences of local adaptation in marine benthic species have been assessed with experimental methods such as reciprocal transplants and common garden experiments. These studies have shown there is differential survival between organisms of different origins in different environmental conditions<sup>5,11</sup>, and that there is non-random, genotype-dependent, selective mortality of recruits and post-colonization mortality<sup>3,12</sup> that lead to high genetic differentiation among adult populations. In spite of the well-recognized effects of local adaptation on the geographic distribution of the genetic diversity of marine populations, there are few examples of phylogeographic studies that compare the spatial genetic signatures of natural selection and neutral evolutionary forces at a broad scale, allowing the evaluation of gene flow, genetic drift, natural selection and other mechanisms that may determine the geographic distribution of the genetic diversity of a species.

<sup>1</sup>Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, 1781421, Coquimbo, Chile. <sup>2</sup>Interdisciplinary Center for Aquaculture Research (INCAR), Departamento de Oceanografía, Universidad de Concepción, 4070386, Concepción, Chile. <sup>3</sup>Instituto de Ecología y Biodiversidad, Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, 7800003, Santiago, Chile. Correspondence and requests for materials should be addressed to P.A.H. (email: phaye@ucn.cl)

Received: 03 November 2016

Accepted: 09 February 2017

Published: 16 March 2017

One experimental approach to analyze neutral and putatively adaptive signals is the use of genome scans to detect thousands of Single Nucleotide Polymorphism (SNP) loci scattered throughout the genome, some of which can be identified as having signals of natural selection by analyzing several individuals from different environments<sup>13</sup>. Spatial analyses of loci putatively affected by natural selection (i.e. outlier loci) provide insights into how adaptive variation could affect phylogeographic structure<sup>14,15</sup> in contrast to the signal of loci that are neutral to natural selection. Studies focused on the assessment of phylogeographic signals of outlier and neutral loci in marine systems have been mostly developed for marine mammals and fishes<sup>16–18</sup>. These studies have generally shown that outlier loci possess a stronger signal of the genetic structure, as is expected by definition<sup>19,20</sup>. However, the main signal is mostly congruent between outlier and neutral loci. This kind of study in marine invertebrates is scarce; only recently a few studies have compared the phylogeographic signal of neutral and outlier loci using genome-wide procedures and have reported a mostly consistent structure between neutral and outlier loci in mussels<sup>19,21</sup> and oysters<sup>22,23</sup>.

In this study we analyze neutral and adaptive spatial genetic structure of the benthic ascidian *Pyura chilensis* Molina 1972 along the Humboldt Current System (HCS) located in the southeast Pacific Ocean. The HCS is characterized by high productivity and regions with strong upwelling<sup>24</sup>. The wide geographic extent of the southeast Pacific coast and the cyclic variations of the HCS promote both temporal and spatial changes in population sizes and distributions of marine benthic species<sup>24</sup>. Despite being an almost linear coastline with apparent absence of geographic barriers between 18°S and 42°S, within the HCS coastal area there is a well-described marine biogeographic boundary located at 30–32°S. This biogeographic break is the limit of the distribution range of several marine species<sup>24</sup>, and phylogeographic studies of benthic marine taxa have shown that species with low dispersal capability have strong genetic discontinuity across the 30°S biogeographic boundary<sup>25,26</sup>. Life history attributes of marine benthic species and the oceanographic conditions seem to be the main factors that contribute to the geographic structure of the genetic diversity and differentiation in this geographic area<sup>24,26</sup>.

*Pyura chilensis* is a conspicuous inhabitant of the HCS widely distributed in the intertidal and shallow sub-tidal from 10°S to 44°S in the southeast Pacific coast<sup>27</sup>, and has been intensively exploited for human consumption<sup>28</sup>. Settlement occurs mainly on rough substrates and on the tunic of conspecifics<sup>29</sup>; it lives either as solitary individuals or as massive clumps of individuals (Supplementary Fig. S1). The species is hermaphroditic, with outcrossing as the main reproductive strategy<sup>28</sup>. *Pyura chilensis* has short-lived (12–24 hr) lecithotrophic tadpole-shaped larvae that provides a very low capacity to disperse between geographic areas along the HCS. In contrast to what has been reported for other sympatric benthic species of the HCS with low larval dispersal capacity<sup>26</sup>, *P. chilensis* did not show a genetic discontinuity at 30°S when analyzed with sequences of the mitochondrial gene Cytochrome Oxidase I (COI) and the nuclear gene Elongation Factor 1 alpha<sup>30</sup>. The lack of a genetic discontinuity at 30°S in the coast of Chile was novel for a benthic invertebrate with short-lived larvae. In fact, some species with planktonic larval durations of more than 2 days display significant genetic divergence across the 30°S biogeographic boundary<sup>26</sup>. By uncovering abundant neutral and adaptive genetic variation in *P. chilensis* it may be possible to detect variation at 30°S, as would be expected in a benthic species with a low dispersal potential.

An alternative means of dispersal for ascidians such as *P. chilensis* is transport in artificial structures such as ship and boat hulls, where the species can grow attached<sup>31–33</sup>. Although this has not been specifically assessed for *P. chilensis*, the species has often been seen on the hulls of boats of artisanal fisheries, on buoys, ropes and on any other marine facility that provides substratum as an important part of the fouling community<sup>34–35</sup>. This has been hypothesized as one of the mechanisms behind the low degree of genetic differentiation of the COI gene along a broad geographic area and across the 30°S biogeographic break<sup>30</sup>; similarly, anthropogenic transport has often been suggested as a main dispersal and invasion mechanism for other ascidians<sup>30–33</sup>.

Another relevant aspect about *P. chilensis* is the genetic structure of the species based on COI sequences. The COI phylogeographic structure of the species was characterized by three mitochondrial haplogroups, two of which were sympatric and widespread, and a third that was highly divergent from the other two and geographically restricted, predominantly found between 39°S and 42°S, being most abundant at 39°S in Los Molinos (LM)<sup>30</sup>. Based on COI sequences, LM was the most differentiated population because it harbored a high proportion of individuals of a unique and divergent haplogroup. Here we evaluate the divergence of the LM population of *P. chilensis* using neutral and putatively adaptive SNP loci, and explore if natural selection follows the same signal as neutral sequence divergence, i.e. if LM is also highly differentiated with putatively adaptive loci. Even though physical transport of a few individuals per generation between local populations should be enough to maintain genetic homogeneity in the whole genome at both individual and population levels<sup>3,36</sup>, non-neutral genetic variation affects fitness, influencing recruitment and survival of individuals in the variable marine environment. Thus, lack of spatial genetic differentiation at neutral or nearly neutral loci does not imply that outlier loci will provide the same signal<sup>36</sup>.

Here we evaluate the spatial genetic structure of neutral and outlier SNP loci in six populations of *P. chilensis* covering from 24°S to 42°S, and across the 30°S biogeographic discontinuity in the southeast Pacific. We aim to assess the influence of contrasting evolutionary forces, including natural selection (local adaptation), isolation and drift (lineage divergence at 39°S), and gene flow (artificial transport), on the spatial genetic structure of the benthic ascidian *P. chilensis* in the HCS.

## Results

**SNP discovery and outlier detection.** Extracted DNA was of high quality, with average 260/280 and 260/230 ratios of  $1.96 \pm 0.01$  and  $2.15 \pm 0.03$ , respectively. Genotyping by Sequencing (GBS) of 90 individuals of *P. chilensis* (15 of each of six local populations; Table 1) gave a total of 202,494,184 reads. Of these, 169,855,853 produced good quality barcode reads. The FastQC results showed a mean per-base quality score (*Phred* score) of 36 and a mean quality score of 33.12. After applying the trimming and demultiplexing approach, 163,692 tags were produced with a mean of 4.38 reads per individual. The UNEAK pipeline detected a total of 81,837 SNP.

Location	Coordinates	Code	N
Pan de Azúcar	26°08'S; 70°39'W	PA	15
Caleta Pajonales	27°44'S; 71°02'W	CP	15
La Herradura	29°58'S; 71°21'W	LH	15
Talcahuano	36°38'S; 71°21'W	TH	15
Los Molinos	39°50'S; 73°23'W	LM	15
Ancud	41°52'S; 73°50'W	AC	15

**Table 1. Sampled local populations of *Pyura chilensis*.** Sampling localities, coordinates, code, number of successfully analyzed individuals for genome scans for *Pyura chilensis* in the Southeast Pacific coast.

Site	Proportion of polymorphic loci		$A_r$		$H_o$		$H_e$		HWE (Outlier)		
	Neutral	Outlier	Neutral	Outlier	Neutral	Outlier	Neutral	Outlier	$\chi^2$	df	P-value
PA	0.2	0.23	1.560 ± 0.43	1.639 ± 0.46	0.271 ± 0.21	0.306 ± 0.20	0.294 ± 0.16	0.338 ± 0.17	89.8	96	0.66
CP	0.59	0.59	1.579 ± 0.42	1.619 ± 0.46	0.304 ± 0.23	0.318 ± 0.20	0.295 ± 0.16	0.358 ± 0.16	118.17	90	0.02
LH	0.48	0.65	1.581 ± 0.42	1.511 ± 0.47	0.305 ± 0.23	0.319 ± 0.26	0.291 ± 0.16	0.299 ± 0.18	89.46	74	0.11
TH	0.37	0.4	1.605 ± 0.40	1.666 ± 0.45	0.285 ± 0.22	0.334 ± 0.19	0.290 ± 0.16	0.336 ± 0.15	69.14	102	0.99
LM	0.46	0.44	1.662 ± 0.39	1.554 ± 0.46	0.169 ± 0.18	0.143 ± 0.15	0.288 ± 0.15	0.250 ± 0.14	197.41	76	<0.001
AC	0.57	0.61	1.604 ± 0.41	1.626 ± 0.46	0.298 ± 0.23	0.345 ± 0.21	0.289 ± 0.16	0.340 ± 0.17	57.08	92	0.998

**Table 2. Population genetic statistics.** Proportion of polymorphic loci and allelic richness ( $A_r$ ) for neutral/outlier loci, observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity and deviations of Hardy-Weinberg Equilibrium (HWE) for outlier loci of *Pyura chilensis* in the Southeast Pacific coast. In bold, significant deviation of HWE after correction with a False Discovery Rate.

	PA	CP	LH	TH	LM	AC
PA		0.003	<b>0.012</b>	<b>0.038</b>	<b>0.414</b>	<b>0.045</b>
CP	<b>0.027</b>		<b>0.015</b>	<b>0.038</b>	<b>0.417</b>	<b>0.047</b>
LH	<b>0.183</b>	<b>0.093</b>		<b>0.045</b>	<b>0.413</b>	<b>0.049</b>
TH	<b>0.524</b>	<b>0.548</b>	<b>0.643</b>		<b>0.399</b>	<b>0.011</b>
LM	<b>0.587</b>	<b>0.609</b>	<b>0.692</b>	<b>0.457</b>		<b>0.393</b>
AC	<b>0.564</b>	<b>0.585</b>	<b>0.672</b>	<b>0.142</b>	<b>0.376</b>	

**Table 3. Population-pairwise  $F_{ST}$  values for neutral and outlier SNP loci in *Pyura chilensis*.** Genetic differentiation measures of pairwise comparisons calculated using 2251 neutral SNP loci (above the diagonal) and 81 outlier SNP loci (below the diagonal) in *Pyura chilensis*. Significant values after 10,000 permutations ( $P < 0.05$ ) are in bold.

After filtering with a minor allele frequency (MAF) of 4.4% and a minimum call rate (mnC) of 90%, the resulting dataset consisted of 2,332 SNPs.

Using BAYESCAN with a False Discovery Rate (FDR) correction of  $q$ -values of 0.05, 81 of the 2,332 loci uncovered showed  $F_{ST}$  values greater than expected under a neutral distribution ( $q$ -values  $< 0.05$ ) (Supplementary Fig. S2). These 81 loci were analyzed separately; thus, there were two datasets, one with the 81 outlier SNP loci and one with 2,251 neutral SNP loci.

**Population genetic diversity and population structure analyses.** Mean expected and observed heterozygosity ( $H_e$  and  $H_o$ , respectively) were mostly similar between neutral and outlier loci.  $H_e$  mean values ranged from 0.290 to 0.294 and 0.250 to 0.358 for neutral and outlier loci, respectively. Mean  $H_o$  values ranged from 0.169 to 0.304 and 0.143 to 0.345 for neutral and outlier loci, respectively (Table 2). The mean allelic richness ( $A_r$ ) within population ranged from 1.560 to 1.662 for neutral loci and 1.511 to 1.626 for outlier loci. Northern populations generally showed lower values of  $A_r$  in neutral loci and similar values between all sampled populations in outlier loci; LH and LM were the sites with the lowest values. The exact test for HWE deviations showed that after FDR correction, only the LM site showed significant deviation (Table 2).

Most of the population-pairwise  $F_{ST}$  values for both neutral and outlier loci were significant, with the exception of PA and CP in neutral loci (Table 3), indicating that *P. chilensis* harbors neutral and adaptive population differentiation.  $F_{ST}$  values ranged between 0.003–0.417 and 0.027–0.692 for neutral and outlier loci, respectively (Table 3). For neutral loci  $F_{ST}$  values were generally low (0.003 to 0.049) with the exception of Los Molinos (LM) that was highly differentiated with values ranging between 0.393 and 0.417, which are an order of magnitude greater than all the rest of the values. In contrast, all  $F_{ST}$  values calculated with outlier loci were, as expected,

Source of variation	Neutral Loci		Outlier Loci	
	N/S	LM	N/S	LM
Among groups	2.91	<b>39.51</b>	<b>48.73</b>	8.74
Among populations within groups	<b>18.4</b>	<b>1.84</b>	<b>12.51</b>	<b>45.47</b>
Within populations	<b>78.69</b>	<b>58.65</b>	<b>38.77</b>	<b>45.79</b>

**Table 4. Analysis of Molecular Variance (AMOVA).** Percentage of variation explained by different hierarchical levels for neutral and outlier loci considering two groupings in *Pyura chilensis*. One separating North and South of 30°S (N/S) and the other one separating Los Molinos (LM) from the rest of the populations. Significant values in bold.

relatively high; remarkably, in neutral loci LM did not show higher levels of differentiation than the other population pairwise comparisons. In the outlier loci population-pairwise differentiation matrix, the values between pairs from the northern (PA, 26°08'S; CP, 27°44'S; LH, 29°58'S) versus southern (TH, 36°38'S; LM, 39°50'S; AC, 41°52'S), *i.e.* north/south of 30°S, showed the highest differentiation values (Table 3).

Using AMOVA analysis the separation of populations north and south of 30°S explained 2.91% and 48.73% of the variation for neutral and outlier loci, being significant only for outlier loci. The analysis of the separation of LM from the rest of the populations explained 39.5% and 8.74% of the genetic variation for neutral and outlier loci, respectively, and was only significant only for neutral loci (Table 4).

Mantel tests did not detect a significant relationship between the linearized genetic distance and geographic distance for either neutral or outlier markers, with nearly significant *p-values* ( $r = 0.034$ ,  $P = 0.089$  and  $r = 0.75$ ,  $P = 0.068$ , respectively) (Fig. 1a,b). Excluding LM from the analysis of neutral loci indicated a marginally significant relationship between genetic and geographic distance ( $r = 0.85$ ,  $P = 0.048$ ) (Fig. 1c). In the case of outlier loci, this led to virtually the same result as when LM was included in the analysis ( $r = 0.81$ ,  $P = 0.064$ ) (Fig. 1d).

In order to detect spatial population genetic structure we used the GENELAND software, which revealed the presence of three groups in the neutral loci dataset. One corresponded to the three northern localities PA, CP and LH; the second included TH and AC and the third was LM (Fig. 2a). Five groups were detected using outlier loci; one group contained PA and CP, and each other local population was assigned to a single genetic group (Fig. 2b).

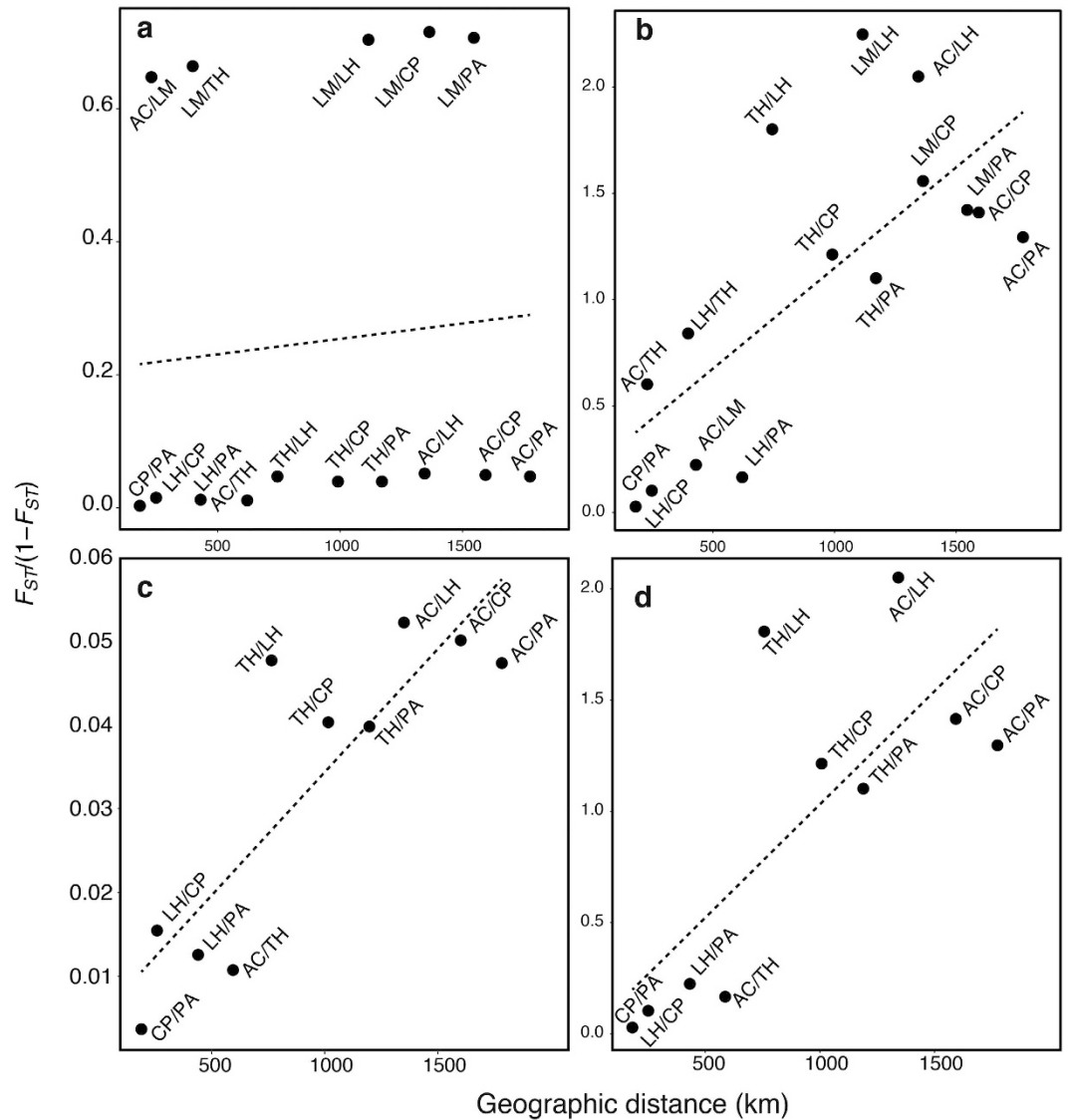
Cluster analyses confirmed the disparate signals of neutral and outlier loci. In the clustering approach of the neutral loci analysis using DAPC (considering the sampling site as a prior), the first principal component axis (vertical) separated LM from all the rest of the populations (Fig. 3a). The second axis (horizontal) separated the north/south of 30°S, with an overlap in the three populations of the north (PA, CP, LH) and an overlap of TH and AC in the south (LM did not show overlap) (Fig. 3a). Thus, the main signal derived from the DAPC analysis of neutral loci was the separation of LM from the rest of the populations and there was a weaker signal associated with the 30°S biogeographic boundary. In outlier loci, north/south of 30°S were well separated in the horizontal axis, with an overlap of PA and CP in the north of TH and LM in the south. The vertical axis separated the PA/CP group from LH, and LM from AC and TH (Fig. 3b).

As with GENELAND analyses, DAPC performed without considering the information of the geographic origin of the samples, and based on the BIC, detected three and five groups as the optimal separation for neutral and outlier loci, respectively (Fig. 3c,d). For neutral loci, the three northern populations appeared in the same cluster, and in the south, TH and AC appeared in the same cluster but LM was clearly separated, with the exception of three individuals that clustered with TH and AC (Fig. 3c). In outlier loci the five groups were mostly consistent with the geographic origin with the exception of PA/CP, which formed a single cluster (Fig. 3d). The third group was composed mainly of individuals of LH (Fig. 3d).

Dendrograms constructed with Identity-by-Site (IBS) without incorporating geographic information detected three and four clusters for neutral and outlier loci, respectively, providing a similar pattern of differentiation as previous analyses (GENELAND, DAPC). With neutral loci LM appeared as a distant separate cluster, while the two other clusters were closer together. These are the three localities north of 30°S and the southern localities TH and AC (Fig. 3e). Again, three individuals of LM appeared together with the individuals sampled in TH and AC. For outlier loci, the north/south of 30°S were well separated (Fig. 3f). The analysis detected two groups in the northern cluster. The first group was composed mostly of individuals of LH and of individuals of PA and CP in lower frequency; the second group was composed exclusively of individuals of PA and CP (Fig. 3f). In the southern cluster, most of the individuals from LM (except three) conformed a well-separated cluster and the rest of the southern individuals including those three individuals from LM formed a single cluster (Fig. 3f).

With Evanno's method,  $K = 2$  was determined as the optimal separation for both data sets in analyses based on STRUCTURE (Fig. 4a,b). For neutral loci one cluster included most of the individuals of LM (12 of 15 individuals), while the second included all other sampling sites plus the three remaining individuals of LM (Fig. 4c). A slight north/south of 30°S separation appeared as a subjacent signal when forcing  $K = 3$  and  $K = 4$  (Fig. 4c).

The optimal clustering separation for outlier loci,  $K = 2$ , corresponded to a separation between north/south of 30°S (Fig. 4d). These results indicate that the optimal  $K$  of neutral and outlier loci provide a different spatial genetic structure of *P. chilensis*. Forcing  $K = 3$  for outlier loci divided the northern cluster in two, one group with PA and CP with strong influence of a third group located mainly in LH. Only when forcing  $K = 4$  an additional cluster appeared in the south corresponding to 12 of the 15 individuals of LM (Fig. 4d). The LM cluster was evident with optimum  $K$  for neutral loci while it appeared with non-optimum  $K = 3-4$  for outlier loci. Exploring from  $K = 5$  to  $K = 7$ , there was no significant difference in the general pattern compared to  $K = 4$  in both datasets

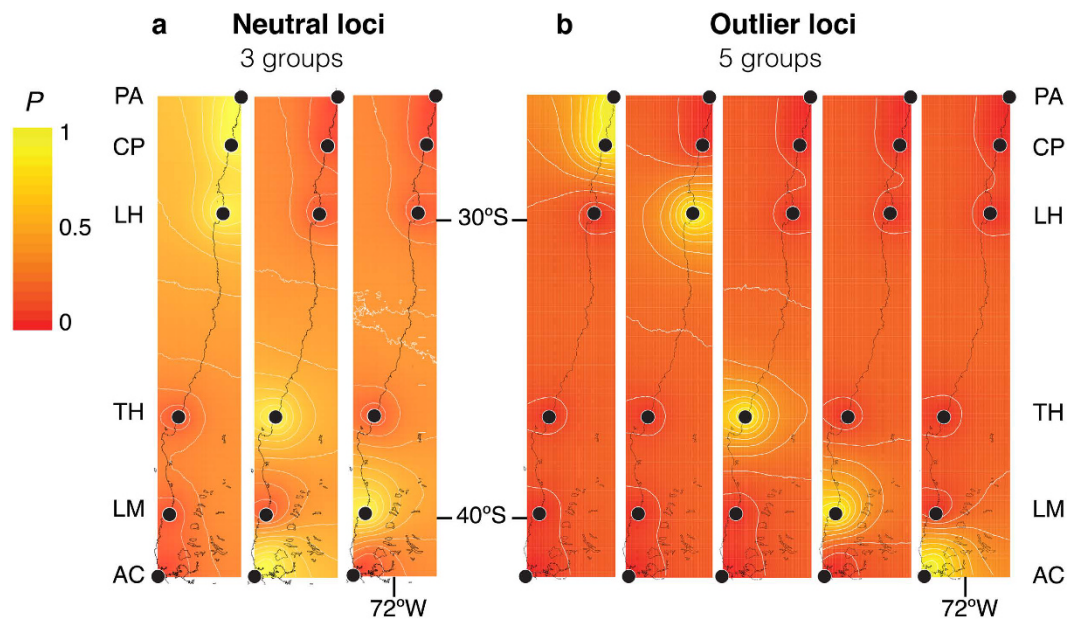


**Figure 1. Relationship between genetic differentiation and geographic distance for neutral and outlier loci of *Pyura chilensis*.** Relationship between linearized genetic differentiation ( $F_{ST}$ ) and geographic distance (Km) along the Humboldt Current System. (a) Relationship using neutral markers, (b) for outlier markers and (c) and (d) show the relationship of neutral and outlier markers, respectively, excluding LM from the analyses. All the pairwise comparisons are marked in the graph.

(results not shown). When LM was excluded from the analyses, the optimal separation also corresponded to  $K=2$  (Fig. 4a,b). For both data sets, the clusters formed corresponded to the north/south of 30°S separation (Fig. 4e,f).

## Discussion

The use of putative selected markers (outlier loci) in marine benthic invertebrates has usually increased the strength of detected phylogeographic structure with respect to neutral loci, as is expected, albeit the general patterns remain the same<sup>19,21–23</sup>. In contrast to this generalization, in the present study neutral and outlier loci showed contrasting spatial genetic structure patterns in the ascidian *Pyura chilensis* along 18° degrees of latitude in the Southeast Pacific coast. Neutral loci had a main signal associated with the lineage history and genetic differentiation of Los Molinos (LM) (39°S), and had a subjacent weaker signal of genetic structure across the 30°S biogeographic boundary. Inversely, outlier loci showed a stronger signal across 30°S and a weaker subjacent signal of lineage divergence at 39°S (LM). The two sets of loci showed the same two phylogeographic patterns, structure across 30°S and 39°S, but their strength was in reverse order. Additionally, the degree of neutral genetic differentiation detected was lower than expected from the short larval duration (2 days) and did not adjust to a pattern of isolation by distance (IBD), suggesting that artificial transport also contributes to the neutral genetic structure of *P. chilensis* along the Humboldt Current System (HCS). SNP sampling and detection of putatively adaptive loci



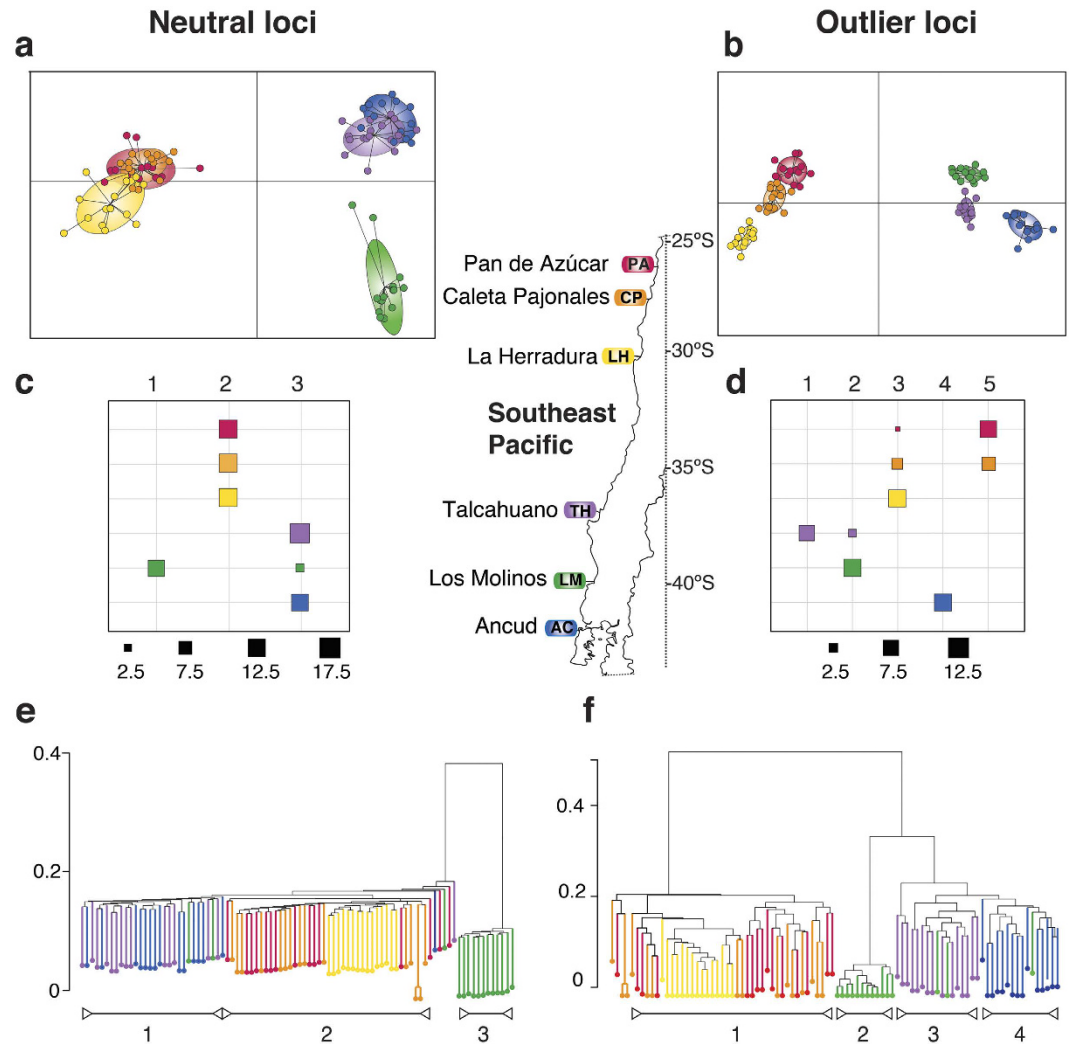
**Figure 2.** Spatial genetic structure of *Pyura chilensis* based on GENELAND analyses. Each map indicates the posterior probability of belonging to one of the three groups found in neutral loci, or the five groups found in outlier loci. Black dots represent the coordinates of the sample sites, and the x and y-axes correspond to geographic coordinates.

allowed the evaluation of the contributions of several evolutionary mechanisms in the phylogeographic structure of *P. chilensis*. We detected a strong adaptive signal of differentiation at the 30°S biogeographic break, a strong neutral signal associated with the divergent lineage present at 39°S, and the influence of artificial anthropogenic transport.

LM had pairwise genetic differentiation values with neutral loci one order of magnitude higher than the rest. This is explained by 12 of the 15 individuals from LM that were highly differentiated from all other individuals of all other sites, independent of whether the analytical approximation used was frequency or distance-based. These differentiated individuals belonged to the divergent mitochondrial DNA haplogroup previously identified in LM<sup>30</sup> (Supplementary Table S1). The extent of the geographic distribution of the divergent lineage that is present in LM needs to be further investigated with an intensive sampling between TH and LM and in the vicinity of LM towards the southern AC locality. It is likely that migration to LM from well-differentiated populations may account for the detected genetic differentiation. The apparent absence of admixture between individuals of LM belonging to the two different clusters as shown by STRUCTURE (Fig. 4e) suggests that these differentiated individuals found in LM may correspond to a reproductively isolated unit. Although these results clearly identified different genetic groups, the existence of a cryptic species in this area should be further investigated, including molecular and morphological systematic studies<sup>37</sup>, before drawing taxonomic conclusions.

As expected, outlier loci showed higher levels of differentiation between all local populations than neutral loci, and the genetic signal provided by outlier loci differed from the neutral signal. The main signal in outlier loci was a strong differentiation north/south of the 30°S biogeographic break, explaining almost half of the total variation in the data. In contrast to the signal detected with neutral markers, outlier loci did not show a main signal associated with the genetic differentiation of LM. Still, the main signal of differentiation north/south of 30°S in outlier loci did not completely expunge the influence of the differentiation of LM; LM appeared as a distinct group when forcing a third suboptimum clustering ( $K = 4$ ) in STRUCTURE (Fig. 4f), and within the southern cluster the IBS dendrograms showed that the same 12 individuals from LM were clearly separated from the other individuals from south of 30°S (Fig. 4f). Additionally, Hardy-Weinberg Equilibrium (HWE) was prevalent; LM was the only population with significant deviations. We explored HWE excluding the 3 individuals of LM that are not highly divergent from the rest of the data set, and after the FDR correction, LM did not deviate from HWE, suggesting that lack of HWE in LM was likely due to a Wahlund effect, in agreement with presence of two divergent lineages in the site<sup>30</sup>. The fact that the 12 divergent individuals from LM did not appear as a separate and external cluster for outlier loci may reflect that natural selection acts globally on these taxa, responding more to environmental constraints than lineage evolution. In other words, the coherence between geographic localities and genetic distribution in outlier loci suggests that selection, independent of the evolutionary history of the lineage, maintains the genetic cohesion of the species mainly based on environmental constraints rather than on lineage history. As genetic variation in neutral markers may reflect historical and contemporary connectivity among populations<sup>38</sup>, selected markers may reflect adaptive processes associated with differences in environment conditions experienced by the species along the HCS.

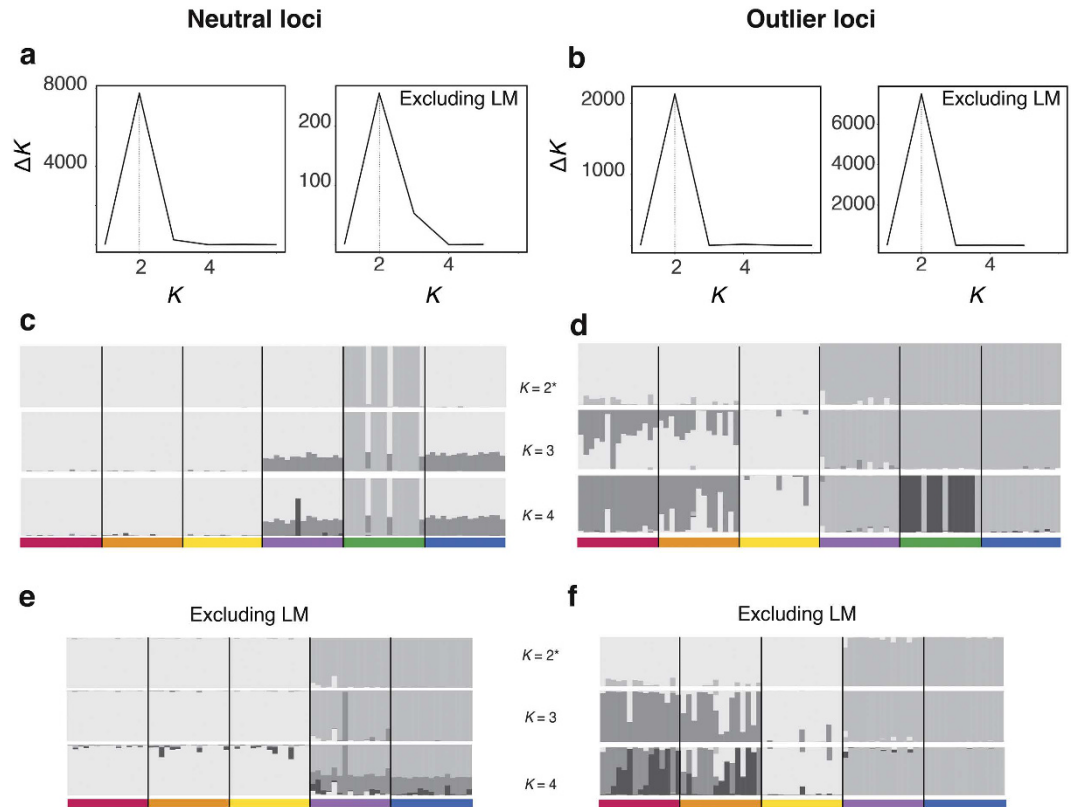
IBD including the all data set was nearly significant and it was marginally significant when LM was excluded from the analysis. However, STRUCTURE analyses showed that when LM was excluded, most of the genetic



**Figure 3. Genetic Clustering of neutral and outlier SNP loci for *Pyura chilensis*.** (a,b) Discriminant Analyses of Principal Components (DAPC) scatterplot showing the first 2 principal components for  $K=3$  and  $K=5$  (neutral and outliers, respectively). Colors represent the sampling location of each individual. (c,d) Proportion of individuals that belong to each cluster based on the DAPC analysis using a BIC criterion with geographic information, (e,f) Identity-by-state (IBS) pairwise relatedness dendrogram for all possible pairs of individuals. The dendrogram shows the three clusters found for neutral markers and the four clusters for outlier markers.

variation was explained by the differentiation to the north and south of 30°S, with optimum  $K=2$ . Therefore, Mantel test results may just reflect a main genetic discontinuity between north/south of 30°S instead of a progressive decrease in migration rate among neighboring sites relative to geographic distance.

Phylogeographic patterns of marine invertebrates along the HCS have shown that the biogeographic boundary at 30°S is a historical discontinuity and that species with low dispersal potential retain the genetic signal and have a concordant phylogeographic break<sup>26</sup>. However, this structure had not been previously detected in *P. chilensis* using sequence data<sup>30</sup>. This is not surprising if we consider that neutral loci revealed that the structure at 30°S is subjacent to a stronger signal of lineage divergence at 39°S, which was the same signal detected with sequence data<sup>30</sup>. The clear differentiation detected by outlier loci between north/south of 30°S is likely linked to the historical and contemporary existence of a marine biogeographic boundary<sup>24</sup>. The contemporary influence of this discontinuity is likely attributed to the heterogeneous environmental conditions that differ markedly north and south of 30°S, mainly, the kinetic energy of the ocean<sup>39</sup>, differences and seasonality in upwelling-favorable winds<sup>40</sup>, and differences in the influence of freshwater on coastal waters<sup>24</sup>. The oceanographic differences between north/south of 30°S could imply that in sessile organisms, such as *P. chilensis*, the effective connectivity (recruitment and survival of juveniles) is restricted, despite the genetic evidence that there is transport of individuals between the two areas<sup>30</sup>. Considering that neutral phylogeographic structure of benthic marine species across 30°S is strongly associated with life history<sup>26</sup>, *P. chilensis* showed phylogeographic structure across 30°S as expected, although it had less differentiation across 30°S than other species with similar dispersal potential based on larval duration.



**Figure 4.** Bayesian clustering inferred with STRUCTURE for *Pyura chilensis*. (a,b) Delta  $K$  results for neutral and outlier loci including and excluding Los Molinos (LM) to infer optimal separations according to Evanno *et al.*<sup>67</sup>, (c–f) Bayesian assignment probabilities of individuals of inferred with STRUCTURE with the entire dataset and excluding LM from the analyses. Each column represents one individual, with the color representing the membership proportion of each of the clusters. The vertical line represents the delimitation of each sampled local population. Graphs show the results for  $K=2-4$ . The optimum  $K$  was denoted by an asterisk.

Differentiation was greater than had been previously reported for *P. chilensis*<sup>30,41</sup>, and also greater than what has been reported for other ascidians that have intensive connectivity driven by anthropogenic transport<sup>31,32,33,42</sup>. The influence of anthropogenic transport was evident from the low, albeit significant differentiation detected between distant local populations, such as the southernmost site analyzed AC and all the northern sites, and general the lack of IBD. The single exception to the significant genetic differentiation detected with neutral loci was the population pairwise genetic differentiation between PA and CP, that was low and non-significant. A possible explanation is that PA and CP are separated by less geographic distance than any other pair of populations, permitting higher gene flow between them than with the rest of the sites. However, short larval life history alone is unlikely to maintain high connectivity between PA and CP, suggesting that artificial transport may be enhancing connectivity between these sites, in agreement, shipping trajectories in the area between PA and CP seem to be geographically continuous and intense (Supplementary Fig. S3).

Physical transport as biofouling may have allowed the latitudinal expansion of *P. chilensis* after periods of isolation during the Pleistocene<sup>30</sup>, but the invasive potential is likely less than in other invasive ascidians. In part this may be explained by the susceptibility of the species to abiotic disturbances<sup>34</sup>, the low and strongly seasonal recruitment rate<sup>43</sup>, and, compared to colonial ascidian growth forms (e.g. *Diplosoma* sp.), a slower growth rate<sup>34</sup>. Higher growth and recruitment rate and rapid reproduction are important life history features in invasive ascidians that allow larvae to colonize open surfaces quickly<sup>44</sup>. Valdivia *et al.*<sup>34</sup> suggested that *P. chilensis* is a good example of a strong competitor with relatively low colonizing ability.

Shipping routes (Supplementary Fig. S3) may also explain the greater differentiation of LM and of LH. LM is the most differentiated population, with a divergent group of individuals, and fittingly, there is very low maritime traffic in the area around LM. There is also low traffic inside La Herradura Bay (LH). The lower differentiation of LH than LM may be explained by the distance from the open coast, which is much smaller in LH, probably allowing slightly greater connectivity by maritime routes. Additionally, differentiation of LH was enhanced by the high proportion of individuals belonging to a distinct mitochondrial haplogroup<sup>30</sup> (Supplementary Table S1). Other sites had from 0 to 4 individuals of this haplogroup, while LH had 7 of the 15. The relatively high genetic differentiation of LH is likely shaped by the lower maritime traffic present in the area, enhancing genetic differentiation.

Although the results are robust and allow inferring the spatial genetic structure of loci affected by contrasting evolutionary forces, it is important to consider alternative explanations associated with the quality of information provided by the data. For example, identified outlier SNP loci may be linked to loci that are the direct target of



natural selection instead of natural selection directly operating on them<sup>45</sup>. Additionally, spatial autocorrelation can cause random associations between the environment and the genetic structure of a species as a result of dispersal following IBD or demographic processes<sup>46</sup>. Some coding regions may have greater genetic-environmental associations than others due to deleterious mutations being selected against in all the study area (purifying selection), and not because advantageous mutations are being selected for in a particular environment (local adaptation<sup>47</sup>). Other factors that may cause loci to behave as outliers are recombination, sampling design, and locus-specific effects<sup>46,48</sup>.

Demonstrating that local adaptation has occurred involves not only the detection of variation at outlier loci, but also the understanding of the functional differences between alleles of individual candidate genes<sup>48</sup>. The knowledge of the physical linkage of SNPs allows the identification of genomic regions with unequal response to natural selection<sup>45</sup>. Since *P. chilensis* is a non-model organism, information on candidate genes is not currently available. In spite of the possible caveats, 3.5% (81) of the total retained SNPs (2,332) showed low and significant *q-values* and higher levels of  $F_{ST}$  for 90 individuals, allowing us to distinguish individuals genetically between sampling locations at a relatively small geographic scale, and suggesting that significant differences could be due to positive selective pressures<sup>49</sup> that differ north/south of 30°S. The percentage of outlier loci detected is proportionally consistent with other studies that show high variation between local populations<sup>19,22,23,50,51</sup>. Further studies are needed to evaluate the genes involved in the putatively adaptive genetic structure detected here, and determine if putatively selected loci are actually under positive selection due to different environmental conditions north and south of the biogeographic boundary at 30°S in the southeast Pacific. Here we showed that even in the presence of a highly divergent lineage, the putatively selected signal could recover a pattern of genetic structure that masks the evolutionary history of the species.

In this study, we obtained several SNPs scattered throughout the genome and analyzed and compared the genetic structure of neutral and putatively adaptive outlier loci in *Pyura chilensis*. The comparison allowed us to conclude that both outlier and neutral loci were diverse and variable in space, and that both have shaped the genetic diversity of *P. chilensis*. The geographic structure of the genetic diversity of outlier and neutral loci was established by different strengths of environmental, historical and anthropogenic factors.

## Methods

**Sample collection and DNA extraction.** Samples of *Pyura chilensis* were obtained by local fisherman from six sites along the Humboldt Current System in the Southeast Pacific Ocean (Table 1). Fisherman sampled individuals from natural substrates in areas with abundant *P. chilensis*. In all sites samples were taken in the shallow subtidal. Even though the degree of relatedness of individuals in a clump has not been assessed, in order to avoid possible effects of relatedness several clumps were obtained, all separated by at least 2 meters. Once in the fishing port, we randomly chose 40–45 clumps or isolated individuals in order to obtain one sample from each, and of those 40–45, we randomly picked 15 per site for this study. Mantle tissue (0.2 g) from each individual was used to extract DNA using the DNeasy Blood® & Tissue Kit (QIAGEN®, USA) according to the manufacturer's instructions. Quantity/purity of DNA was measured in Nanodrop 2,000 (Thermo, USA).

**Genotyping by Sequencing and SNP discovery.** Samples were sequenced in the Biotechnology Center of the University of Wisconsin, USA using Genotyping by Sequencing (GBS)<sup>52</sup>, a method widely used in non-model species and that has proven to be a useful tool to distinguish between local populations at the genome level in marine benthic invertebrates<sup>20</sup>. This is a Reduced Representation Sequencing (RRS) method, which is based on the use of restriction enzymes that reduce the complexity of the genome, with the additional advantage that it can be performed without prior knowledge or reference genomes. This technique allows sequencing a large number of short genomic regions in several individuals<sup>45</sup>, permitting the genotyping of a large number of SNPs randomly located throughout the genome. Previous to sequencing, two additional enzymes were tested (*PstI*, *PstI/TaqI*) and finally *ApeKI*<sup>53</sup> was chosen due to the wide genome distribution of flanking regions described for other species and avoiding repetitive zones<sup>53</sup>.

Libraries were prepared for the DNA of 15 individuals of each population. Following the restriction enzyme digestion, DNA fragments were ligated to a unique barcode adaptor for each individual. The library was prepared in a 96-well plate with 6 wells as blank, using a HiSeq2000 (Illumina, USA) platform, which sequences reads of approximately 100 base pairs (bp). The reads were visualized in FastQC version 0.10.1<sup>54</sup> for quality checks.

All the data were prepared and analyzed using the pipeline Universal Network-Enabled Analysis Kit (UNEAK<sup>55</sup>) using TASSEL v.3<sup>56</sup> that is specially designed for species with no reference genome. TASSEL-UNEAK is a network based SNP detection algorithm that may be less flexible than other pipelines in aspects like read trimming and parameters for *de novo* locus identification<sup>57</sup>, reducing the potential number of total SNPs detected in the dataset. However, TASSEL-UNEAK has proven to be a useful, reliable and reproducible tool for demultiplexing and processing sequence data obtained through GBS for non-model species<sup>55</sup>. With the UNEAK pipeline, the dataset was demultiplexed and the reads were trimmed to 64 bp to remove the barcode sequence and the error-prone tail of the sequences. After the network filtering, identical reads were aligned as tags using an error tolerance rate of 0.03 in order to minimize the chance that real tags were discarded as sequencing errors and to remove potential paralogs before the SNP calling. This filtering is a goodness-of-fit  $\chi^2$  ( $\alpha = 0.05$ ) based on the null hypothesis that the counts of two paired tags of a SNP are equal in all heterozygous individuals<sup>55</sup>.

Additionally, a minor allele frequency (MAF) of 0.044 was used to filter the number of SNP loci whose variability was well represented in the 90 analyzed individuals. This value was determined empirically based on the probability that at least 8 of the 180 possible allele calls (4.4% of the alleles detected in 90 individuals) correspond to the minor allele. Values of minimum call rate (mnC) from 75 to 90% were tested previous to the final SNP calling, finally choosing the most exigent value of mnC (90%) in order to remove all SNPs containing 90% or more "N" values, which designate those individuals for which no allele was assigned for the locus or that were covered

by no more than one tag and because of the amount of missing data detected, preserving a reduced but robust set of SNPs with reliable genotypes without excluding individuals from the analysis. These restrictive filters were performed because all loci were sampled at a mean coverage of  $\geq 10\times$  per individual. Under this condition, in theory each individual had in average 10 or more copies of a locus.

**Outlier detection.** To identify the SNPs putatively under diversifying selection we used the  $F_{ST}$  outlier approach, which is based on the estimation of those loci with greater values of  $F_{ST}$  (outliers) with respect to the expected values under a neutral distribution. We used the software BAYESCAN 2.1<sup>16</sup> for  $F_{ST}$  outlier detection, because it has been reported to have lower rates of false positives with respect to other similar software<sup>58</sup>. BAYESCAN uses a Bayesian framework to calculate the posterior probability that any given locus is under selection<sup>16</sup>. A total of ten separate runs were performed, from 50,000 to 500,000 iterations with a 10% burn-in period to assure the convergence of the MCMC. After the runs, an FDR correction of  $q$ -values of 0.05 was applied in BAYESCAN to avoid the occurrence of false positives. The data set was then separated into two parts, one with the outlier SNPs, which are loci putatively under diversifying selection<sup>16</sup>, and one with neutral SNPs, which are all the rest of the loci that have  $F_{ST}$  values that match the expected under a neutral distribution.

**Population genetic diversity analysis and genetic structure.** From here on, all analyses were performed separately for each of the data sets (neutral and outlier loci).

To evaluate the differences in genetic diversity across the study area in each data set, we calculated the expected heterozygosity, observed heterozygosity and allelic richness with rarefied allele counts, using the HIERFSTAT package version 0.04–22<sup>59</sup> in R v 3.22<sup>60</sup>. Exact tests for Hardy-Weinberg Equilibrium deviations for outlier loci were calculated using GENEPOP version 4.6<sup>61</sup>. Multiple comparisons of HWE were corrected using a FDR.

Pairwise  $F_{ST}$  and their significance were determined using a permutation test (10,000) in ARLEQUIN Version 3.5<sup>62</sup>. ARLEQUIN was also used to perform an Analysis of Molecular Variance (AMOVA) using the north/south of 30°S biogeographic boundary, and the separation of LM as two *a priori* groupings. Additionally, a Mantel test with 10,000 permutations was carried in out in ARLEQUIN 3.5 to test if there is a pattern of isolation by distance, using the relationship between the geographic distance (Km) and Slatkin's linearized genetic distance ( $F_{ST}$ ).

Spatial population genetic structure was assessed using the Bayesian clustering algorithm implemented in GENELAND v 3.1.4<sup>63</sup>. This analysis uses spatial coordinates to separate allele frequencies that can be detected with departure from Hardy-Weinberg and linkage equilibrium into  $K$  clusters. An independent model of allele frequencies was carried out for both data sets, with 10 independent runs with  $1e^6$  iterations and a 10% burn-in period, varying  $K$  from 2 to 6. The optimum  $K$  was inferred from the posterior individual membership probability of each genetic group. Posterior probabilities of membership were plotted with the *shapefiles* of the coastline available in the database of GEOdas (NOAA), filtering in the study area using GEOdas Coastline Extractor v 1.1.3.1 (<https://www.ngdc.noaa.gov/mgg/geodas/geodas.html>).

A clustering approach was used with a discriminant analysis of principal components (DAPC) in the package *adagene*<sup>64</sup> in R v 3.22<sup>60</sup>. In DAPC, a principal components analysis of the multilocus genotypes of the individuals was calculated and then a discriminant analysis was carried out using the PC scores. These analyses are based on the detection of the number of clusters that minimizes between-group variation<sup>64</sup> using the k-means and a Bayesian Information Criterion (BIC) to identify the optimal number of clusters in the data, not assuming underlying structure in population subgroups or panmixia as other similar approaches do. The number of clusters in each data set was determined using  $10e^7$  iterations. To avoid unstable assignments in each cluster, 30 PCs were retained ( $N_{total}/3$ ), using all the 6 discriminant functions. We determined the optimal number of PCs using 1,000 simulations to execute the final DAPC.

To infer SNP-based relatedness structure within populations, an analysis based on pairwise IBS (identity-by-state) was used. This approach uses the information of the genotypes to calculate the probability that two alleles in the same segment have the same ancestor. A cluster analysis of the matrix generated with the IBS pairwise coefficients between all possible pairs of individuals was performed using a permutation score of 10,000 to determine how many groups were present in each data set. This was performed in the package SNPRelate<sup>65</sup> in R environment v 3.22 (R Core Development Team 2015).

To determine the number of genetic groups we used STRUCTURE 2.3.4<sup>66</sup>. Clusters ( $K$ ) varied from one to seven, corresponding to the number of analyzed locations plus one. Ten replicate runs, each with 500,000 MCMC with a 10% burn-in period, were performed under the admixture and correlated frequency model. The most probable  $K$  value was inferred based on the delta  $K$  method proposed by Evanno *et al.*<sup>67</sup>.

**Data Accessibility.** SNP calls for all data set (2,332 SNPs) are available in a VCF file that has been deposited in DataDryad (doi:10.5061/dryad.0s87v).

## References

- Lipcius, R. N. *et al.* Importance of metapopulation connectivity to restocking and restoration of marine species. *Rev. Fisheries. Sci.* **16**, 101–110 (2008).
- Carson, H. S. *et al.* Evaluating the importance of demographic connectivity in a marine metapopulation. *Ecology* **92**, 1972–1984 (2011).
- Hedgecock, D. Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bull. Mar. Sci.* **39**, 550–64 (1986).
- Cowen, R. K. *et al.* Population connectivity in marine systems. *Oceanography* **20**, 14–21 (2007).
- Marshall, D. J. *et al.* Phenotype-environment mismatches reduce connectivity in the sea. *Ecol. Lett.* **13**, 128–140 (2010).
- Pineda, J. *et al.* Transport and dispersal in the coastal ocean and consequences for population connectivity. *Oceanography* **20**, 22–39 (2007).
- Hamilton, S. L. *et al.* Post settlement survival linked to larval life in a marine fish. *Proc. Natl. Acad. Sci. USA* **105**, 1561–1566 (2008).
- Selkoe, K. A. *et al.* Seascape genetics and the spatial ecology of marine populations. *Fish. Fish.* **9**, 363–377 (2008).

9. Cowen, R. K. & Sponaugle, S. Larval dispersal and marine population connectivity. *Ann. Rev. Mar. Sci.* **1**, 443–466 (2009).
10. Pechenick, J. A. Larval experience and latent effects –metamorphosis is not a new beginning. *Integr. Compr. Biol.* **46**, 323–333 (2006).
11. Emlet, R. B. & Sadro, S. S. Linking stages of life history: how larval quality translates into juvenile performance for an intertidal barnacle (*Balanus glandula*). *Integr. Compr. Biol.* **46**, 334–346 (2006).
12. Penney, R. W. *et al.* Genotype-dependent survival, growth, and reproduction in cultured blue mussels, *Mytilus* spp.: results of a reciprocal seed transfer experiment. *J. Shellfish. Res.* **25**, 515–525 (2006).
13. Pritchard, J. K. & Di Rienzo, A. Adaptation-not by sweeps alone. *Nat. Rev. Genet.* **11**, 665–667 (2010).
14. Nosil, P. *et al.* Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**, 375–402 (2009).
15. Schluter, D. & Conte, G. L. Genetics and ecological speciation. *Proc. Natl. Acad. Sci. USA* **106**, 9955–9962 (2009).
16. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993 (2008).
17. Moura, A. E. *et al.* Population genomics of the killer whale indicates ecotype evolution in sympatry involving both selection and drift. *Mol. Ecol.* **23**, 5179–5192 (2014).
18. Candy, J. *et al.* Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Mol. Ecol. Resour.* **15**, 1421–1434 (2015).
19. Aranedo, C. *et al.* Adaptive genetic variation distinguishes Chilean blue mussels (*Mytilus chilensis*) from different marine environments. *Ecol. Evol.* doi: 10.1002/ece3.2110 (2016).
20. Cahill, A. & Levinton, J. S. Genetic differentiation and reduced genetic diversity at the northern range edge of two species with different dispersal modes. *Mol. Ecol.* **25**, 515–526 (2016).
21. Xu, T. *et al.* Genome-wide discovery of single nucleotide polymorphisms (SNPs) and single nucleotide variants (SNVs) in deep-sea mussels: Potential use in population genomics and cross-species application. *Deep-sea. Res. PT II.* doi: 10.1016/j.dsr2.2016.03.011 (2016).
22. Nayfa, M. G. & Zenger, M. R. Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (*Pinctada maxima*). *Mar. Genomics.* doi: 10.1016/j.margen.2016.02.005 (2016).
23. Lal, M. M. *et al.* A Parallel Population Genomic and Hydrodynamic Approach to Fishery Management of Highly-Dispersive Marine Invertebrates: The Case of the Fijian Black-Lip Pearl Oyster *Pinctada margaritifera*. *PLoS ONE* **11**, e0161390 (2016).
24. Thiel, M. *et al.* The Humboldt Current system of northern and central Chile oceanographic processes, ecological interactions and socioeconomic feedback. *Oceanogr. Mar. Biol. Annu. Rev.* **45**, 195–344 (2007).
25. Sánchez, R. *et al.* Spatial pattern of genetic and morphological diversity in the direct developer *Acanthina monodon* (Gastropoda: Mollusca). *Mar. Ecol. Prog. Ser.* **434**, 121–131 (2011).
26. Haye, P. A. *et al.* Phylogeographic Structure in Benthic Marine Invertebrates of the Southeast Pacific Coast of Chile with Differing Dispersal Potential. *PLoS ONE* **9**, e88613 (2014).
27. Lancellotti, D. & Vásquez, J. A. Zoogeografía de macroinvertebrados bentónicos de la costa de Chile: contribución para la conservación marina. *Rev. Chil. Hist. Nat.* **73**, 99–129 (2000).
28. Manríquez P. H. & Castilla J. C. Self-fertilization as an alternative mode of fertilization in the solitary tunicate *Pyura chilensis*. *Mar. Ecol. Prog. Ser.* **305**, 113–125 (2005).
29. Cea, G. Contribución al conocimiento de algunos aspectos de la biología de *Pyura chilensis* Molina 1782 (Chordata, Tunicata, Ascidiacea). *Tesis de Licenciatura en Biología, Universidad de Concepción, Concepción, Chile.* 205 pp (1970).
30. Haye, P. A. & Muñoz-Herrera, N. C. Isolation with differentiation followed by expansion with admixture in the tunicate *Pyura chilensis*. *BMC Evol. Biol.* **13**, 252 (2013).
31. Hudson, J. *et al.* Anthropogenic transport of species across native ranges: unpredictable genetic and evolutionary consequences. *Biol. Lett.* **12**, doi: 10.1098/rsbl.2016.0620 (2016).
32. Ordoñez, V. *et al.* Mixed but not admixed: a spatial analysis of genetic variation of an invasive ascidian on natural and artificial substrates. *Mar. Biol.* **160**, 1645–1660 (2013).
33. Pineda, M. C. *et al.* The whereabouts of an ancient wanderer: global phylogeography of the solitary ascidian *Styela plicata*. *PLoS ONE* **6**, e25495 (2011).
34. Valdivia, N. *et al.* Effects of disturbance on the diversity of hard-bottom macrobenthic communities on the coast of Chile. *Mar. Ecol. Prog. Ser.* **299**, 45–54 (2005).
35. Cifuentes, M. *et al.* Effects of temporal variability of disturbance on the succession in marine fouling communities in northern-central Chile. *J. Exp. Mar. Biol. Ecol.* **352**, 280–294 (2007).
36. Conover, D. O. *et al.* Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *J. Fish. Biol.* **69**, 21–47 (2006).
37. Liu, J. *et al.* Cryptic diversity in the pen shell *Atrina pectinata* (Bivalvia: Pinnidae): high evidence and hybridization revealed by molecular and morphological data. *Mol. Ecol.* **20**, 4332–4345 (2011).
38. Conover, D. O. *et al.* Reversal of evolutionary downsizing caused by selective harvest of large fish. *Proc. R. Soc. Long. Biol. Sci.* **276**, 2015–2020 (2009).
39. Hormazábal, S. *et al.* Coastal transition zone off Chile. *J. Geophys. Res.* **109**, doi: 10.1029/2003JC001956 (2004).
40. Aravena, G. *et al.* Twelve Years of Change in Coastal Upwelling along the Central-Northern Coast of Chile: Spatially Heterogeneous Responses to Climatic Variability. *PLoS ONE* **9**, e90276 (2014).
41. Astorga, M. O. & Ortiz, J. C. Genetic variability and population structure in the tunicate *Pyura chilensis* Molina, 1782, in the coast of Chile. *Rev. Chil. Hist. Nat.* **79**, 423–434 (2006).
42. Pineda, M. C. *et al.* Stochasticity in space, persistence in time: genetic heterogeneity in harbour populations of the introduced ascidian *Styela plicata*. *PeerJ* **4**, e2158 (2016).
43. Ambler, R. & Cañete, J. I. Asentamiento y reclutamiento de *Pyura chilensis* Molina 1792 (Urochordata: Ascidiacea) sobre placas artificiales suspendidas en la Bahía La Herradura, Coquimbo, Chile. *Rev. Biol. Mar.* **26**, 403–413 (1991).
44. Bullard, S. G. & Carman, M. R. Current trends in invasive ascidian research. In: *Invasive Species: Detection, Impact and Control.* Editors C. P., Wilcox & R. B., Turpin. Nova Science Publishers, Inc (2009).
45. Manel, S. *et al.* Genome Assemblies, Genomic Resources and their Influence on the Detection of the Signal of Positive Selection in Genome Scans. *Mol. Ecol.* **25**, 170–184 (2016).
46. Lotterhos, K. E. & Whitlock, M. C. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* **24**, 1031–1046 (2015).
47. Hoban, S. *et al.* Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am. Nat.* **188**, doi: 10.1086/688018 (2016).
48. Hassl, R. J. & Payseur, B. A. Fifteen years of genome-wide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Mol. Ecol.* **25**, 5–23 (2016).
49. Crawford, D. L. & Oleksiak, M. F. Ecological population genomics in the marine environment. *Brief. Funct. Genomics*, doi: 10.1093/bfgp/ew008 (2016).
50. Drury, C. *et al.* Genomic variation among populations of threatened coral: *Acropora cervicornis*. *BMC Genomics*, doi: 10.1186/s12864-016-2583-8 (2016).
51. Ravinet, M. *et al.* Shared and non-shared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol. Ecol.* **25**, 287–305 (2016).

52. De Donato, M. Genotyping-by-Sequencing (GBS): A Novel, Efficient and Cost-Effective Genotyping Method for Cattle Using Next-Generation Sequencing. *PLoS ONE* **8**, e62137 (2013).
53. Elshire, R. J. *et al.* A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **6**, e19379 (2011).
54. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).
55. Lu, F. *et al.* Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* **9**, e1003215 (2013).
56. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
57. Andrews, K. R. *et al.* Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat. Rev. Genet.* **17**, 81–92 (2016).
58. Narum, S. R. & Hess, J. E. Comparison of FST outlier tests for SNP loci under selection. *Mol. Ecol. Resour.* **11**, 184–194 (2011).
59. Goudet, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).
60. R Development Core Team R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. <http://www.R-project.org> (2015).
61. Rousset, F. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106 (2008).
62. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
63. Guillot, G. *et al.* 2005 A spatial statistical model for landscape genetics. *Genetics* **170**, 1261–1280 (2005).
64. Jombart, T. *et al.* Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
65. Zheng, X. *et al.* A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics* **28**, 3326–3328 (2012).
66. Pritchard, J. K. *et al.* Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
67. Evanno, G. *et al.* Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).

## Acknowledgements

Authors thank all fisherman that collected samples along the study area, as well as Natalia Muñoz, Francisca Gálvez and Raúl Vera for help with sample processing. This work was financially supported by Grant FONDECYT 1140862 (PAH & CGE), Grant FONDAP 15110027 (CGE & PAH). NIS was supported by the CONICYT Doctoral Thesis Scholarship 21120350.

## Author Contributions

P.A.H., C.G.E. and N.I.S. conceived the idea, designed the experiments, and performed the experiments. N.I.S. and P.A.H. analyzed the data and wrote the paper. N.I.S., P.A.H. and E.P. contributed intellectually to the interpretation and discussion of results. All authors read and approved the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Segovia, N. I. *et al.* Lineage divergence, local adaptation across a biogeographic break, and artificial transport, shape the genetic structure in the ascidian *Pyura chilensis*. *Sci. Rep.* **7**, 44559; doi: 10.1038/srep44559 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017