Fundamental Studies

Substantial gene flow caused by long-term translocation between natural bank populations of the Peruvian scallop (*Argopecten purpuratus*) is supported by RAD-Seq analyses

Ximena Velez-Zuazo¹,² | Sergio P. Barahona³,⁴ | Omar G. Melo³ | Eric Hanschke⁵ | Ian Hanschke⁵ | Monica C. Santa-Maria³,⁶

¹Center for Conservation and Sustainability, Smithsonian Conservation Biology Institute, Washington, District of Columbia
²ecOceánica, Lima, Perú
³Biotecnología de Alimentos - BIOAL S.A.C, Lima, Perú
⁴Universidad Científica del Sur, Lima, Perú
⁵Seacorp Perú S.A.C., Lima, Perú
⁶Departamento de Ingeniería Ambiental, Universidad de Ingeniería y Tecnología – UTEC, Lima, Peru

Abstract

The Peruvian scallop (*Argopecten purpuratus*, Lamarck 1989) is a marine bivalve of high commercial value in the aquaculture industry, with wild populations distributing from northern Peru to Chile. Growing demand for it in the world aquaculture markets and limited availability of hatchery-based seeds caused long-term seed translocations among wild populations to recover depleted local populations and for production needs. We investigated long-term translocation effects on the genetic diversity and structure of wild populations using next-generation RAD sequencing. We sampled individuals from Sechura, Lobos de Tierra, Samanco, and Bahia Independencia in Peru, and La Rinconada in Northern Chile. We identified 8,345 polymorphic RAD loci and 24,218 SNPs for the five populations. We estimated high observed heterozygosity for all populations and high SNP frequency compared to similar studies on marine bivalves. We detected no spatial divergence among populations in Peru (pairwise FST from 0 to 0.003), but strong differentiation with the population in Chile. Migration rate estimates suggested asymmetric directionality of seed translocation. Overall, our results...
support a remnant effect of an intense historic translocation and ongoing gene flow among wild populations in Peru, challenging the identification of outlier loci and certification of sustainable origin of cultured scallops using genetic markers.

**KEYWORDS**
gene flow, Peruvian scallop, RAD-Seq analysis, seed translocation

## 1 | INTRODUCTION

The Peruvian scallop (*Argopecten purpuratus*) is the most exported marine bivalve from the Peruvian aquaculture industry and an important species in the global bivalve market (Cavero Cerrato & Rodríguez Pinto, 2008). It has a long planktonic larval stage and inhabits sheltered bays along the coasts of Peru and Chile, from Paita (5.1°S) to Valparaiso (33.1°S) (von Brand, Abarca, Merino, & Stotz, 2016; Wolff & Mendo, 2000). This species has been subject to local translocations for decades, fueled by the expansion in Peruvian scallop harvest during the 1980s (Mendo, Wolff, Carbajal, Gonzales, & Badjeck, 2008). In 1982–1983, El Niño Southern Oscillation (ENSO) event caused the overgrowth of *A. purpuratus* populations, resulting in overexploitation, jeopardizing the integrity of natural bank stocks. In response, the Peruvian government promoted the adoption of aquaculture practices to guarantee the sustainability of this resource. But, the limited availability of hatchery-produced seeds led to periodical extractions and translocations of young individuals from healthy natural banks to recover depleted local populations (Mendo et al., 2008). This practice was aimed to promote the recovery of diminished natural stocks and to facilitate scallop production in nearby hatcheries.

Translocation of individuals among wild populations of marine species with commercial value is a widespread practice with consequences that are not yet fully understood (Lemer & Planes, 2012). Translocation is defined as the man-made movement of individuals from a given population to another within the species range of distribution (Beaumont, 2000). This practice serves specific purposes, the main being the recovery of depleted populations for securing harvest (Carlton & Mann, 1996). Evidence indicates that estimates and distribution of genetic diversity and relatedness can change in wild or natural bank populations as a result of translocations. For example, a genetic homogenization in wild populations of previously distinct black-lip pearl oyster (*Pinctada margaritifera*) was found following 10 years of translocations (Arnaud-Haond et al., 2004). A recent study in the same species reported high levels of genetic diversity in wild populations adjacent to farmed populations because of “gene leaking” (Lemer & Planes, 2012).

The intensity and directionality of translocations of Peruvian scallops are poorly documented. An early study investigated the level of genetic structure for three Peruvian scallop populations in Peru (Sechura, Samanco, and Bahia Independencia). They found a significant, albeit weak, population differentiation using nine microsatellite and mitochondrial markers (Marín, Fujimoto, & Arai, 2013). The authors suggested that, despite intense translocations, at least one population retained part of its genetic signature (Bahia Independencia) because of asymmetric translocations and to the effect of geographic distance for realized larval dispersal. A recent study investigated the genetic structure in Peruvian scallop populations spanning its entire geographic distribution using two mitochondrial markers (COI and Cytb genes) and found a slightly phylogeographical structure between Peru and Chile populations (Acosta-Jofré et al., 2020). Their results suggested that Peruvian scallop populations experienced a recent population range expansion, with moderate to low levels of gene flow after the expansion. Both studies attributed the observed patterns to high man-mediated gene flow through seed translocation, and suggested that further studies with higher marker resolution are needed to understand population structuring.

Genome-wide molecular markers could provide the resolution power needed to analyze and detect structure in large populations with high gene flow (Waples, 1998). Genome-wide markers, particularly those generated by
genotyping-by-sequencing approaches (Davey et al., 2011) such as Restriction-Site Associated DNA (RAD) markers (Baird et al., 2008) can detect variation and resolve fine-scale population structure for species with large dispersal capacity and potential high gene flow, for example, mosquitoes (Rašić, Filipović, Weeks, & Hoffmann, 2014) and American lobster (Benestan et al., 2015). We evaluated the genome-wide variation in individuals of the Peruvian scallop from five natural bank populations along their distribution range using thousands of SNP markers obtained from next-generation RAD sequencing. The objective was to investigate the influence of long-term translocations on the current genetic structure of Peruvian scallop wild populations. We (1) estimated levels of genomic diversity among and within wild populations, (2) described the spatial population structure, and (3) assessed the level and directionality of gene flow among them.

Despite its major role in the Peruvian aquaculture industry, the sustainable harvest of Peruvian scallop is not totally guaranteed. To this day, the Peruvian aquaculture industry continues to rely on wild seed extraction, with much of the extracted seed being moved to nearby hatcheries or distant ones adjacent to natural banks (E. Hanschke, personal communication). This hinders the performance of Peruvian producers in international markets (I. Hanschke, personal communication). Traceability tools, such as genome-wide molecular markers, could provide a sustainable origin certification of aquaculture species, demanded by the global bivalve market (Nielsen et al., 2012). In this study, we also discuss perspectives on the use of genetic markers and population-wide genetic descriptors (i.e., heterozygosity and inbreeding coefficients) for traceability.

2 MATERIALS AND METHODS

2.1 Sample collection and DNA purification

Between August 2014 and February 2015, we collected 57 individuals of *A. purpuratus* from natural banks covering a distribution ranging from northern Peru (Sechura-SEC, Lobos de tierra-LDT, Samanco-SAM, and Bahia Independencia-IND) to northern Chile (La Rinconada-RIN) (Figure 1, Table 1). We obtained approximately 200 mg of adductor muscle tissues and stored it in RNAlater (Ambion, Thermo Fisher Scientific, Waltham, MA) at room temperature for the first 24 hr and then stored permanently at –20°C. We isolated whole genomic DNA using the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific, Waltham, MA) following manufacturer’s instructions. DNA concentration (ng/μl) was measured using the Qubit dsDNA BR Assay Kit in a Qubit 2.0 fluorometer (Invitrogen, Thermo Fisher Scientific, Waltham, MA). Genomic DNA integrity was verified by agarose gels electrophoresis.

2.2 Genomic libraries, de novo assembly, and dataset filtering

To construct the Genomic RAD libraries, we followed a recommended method (Etter, Bassham, Hohenlohe, Johnson, & Cresko, 2011) with minor modifications. We digested about 1 μg of high-quality genomic DNA per sample with a low-frequency endonuclease restriction enzyme (“Sbf1”, Thermo Fisher Scientific). We ligated the digested fragments to unique modified Solexa 5 bp-barcode-labeled P1 adapters (see Data S1) using the following reaction parameters: 2 μl P1 adapter, 10 mM rATP, T4 1X DNA ligase buffer (50 mM NaCl, Tris-HCl 10 mM pH 7.5), and 200 CEU T4 DNA Ligase; incubating overnight at 4°C. We pooled the P1-adapter-ligated fragments bearing unique barcode identifiers into four libraries (15 individuals per library) for further fragmentation using a COVARIS S220 Focused-ultrasonicator (Woburn, MA). We recovered fragments ranging from 300 to 600 bp from a low-melting-point agarose gel electrophoresis using the GeneJet Gel Extraction kit (Thermo Scientific). We repaired fragment ends with the Fast DNA End Repair Kit (Thermo Scientific), and added a 3’dA overhang with a Klenow Fragment Exo- Kit (Thermo Scientific). We then ligated a P2 adapter (divergent “Y” adapter; see Data S2) to all libraries using the following reaction parameters: 1 μl P2 adapter, 200 CEU DNA T4 Ligase, and 1X T4 Ligase Buffer;
incubating 1 hr at room temperature. To verify successful ligation of both adapters, we run a PCR pre-enrichment reaction using P1 and P2 specific primers (see Data S3). Reaction setup was as follows: 1X Phusion Master Mix with GC buffer (Thermo Scientific), primer P1 (0.5 μM), primer P2 (0.5 μM), BSA (10 μg), and 5 μl library DNA.

**FIGURE 1** Populations of *Argopecten purpuratus* investigated in this study. The number of individuals per sampling site (n) is shown in parentheses

**TABLE 1** Molecular diversity estimations for five populations of Peruvian scallop

| Population (code)          | Lat–Lon          | n  | Ho       | He       | Π         | Fis            |
|----------------------------|------------------|----|----------|----------|-----------|----------------|
| Sechura (SEC)              | 5°47’S–80°55’W   | 12 | 0.1855 (0.0273) | 0.1814 (0.0183) | 0.1897 (0.0201) | 0.0285 (0.0619) |
| Lobos de Tierra (LDT)      | 6°25’S–80°50’W   | 12 | 0.1924 (0.0322) | 0.1862 (0.0197) | 0.1950 (0.0217) | 0.0311 (0.0707) |
| Samanco (SAM)              | 9°11’S–78°33’W   | 12 | 0.1677 (0.0224) | 0.1756 (0.0174) | 0.1834 (0.0190) | 0.0655 (0.0767) |
| Bahia Independencia (IND)  | 14°14’S–75°27’W  | 9  | 0.2044 (0.0264) | 0.2041 (0.0171) | 0.2164 (0.0193) | 0.0455 (0.0752) |
| La Rinconada (RIN)         | 23°27’S–70°30’W  | 12 | 0.1932 (0.0271) | 0.1970 (0.0193) | 0.2059 (0.0211) | 0.0556 (0.0790) |

*Note: n = number of individuals; Ho, observed heterozygosity; He, expected heterozygosity; Π, nucleotide diversity; Fis, inbreeding coefficient. Standard deviation (SD) is indicated within parenthesis.*
Amplification parameters were as follows: 30 s at 98°C, 25 cycles of 10 s at 98°C, 30 s at 65°C, and 30 s at 72°C, followed by a 5 min extension at 72°C. Amplification products were verified in a 1% agarose gel electrophoresis. Successfully ligated libraries were enriched by PCR, nine PCR replicate reactions per library, using the same PCR pre-enrichment reaction setup and the following amplification parameters: 30 s at 98°C, 18 cycles of 10 s at 98°C, 30 s at 65°C, and 30 s at 72°C, followed by a 5 min extension at 72°C. We recovered amplified fragments ranging from 300 to 600 bp from a low-melting-point agarose gel electrophoresis using the GeneJet Gel Extraction kit (Thermo Scientific) into 30 μl elution buffer. We estimated library concentration in nM using a Qubit 2.0 Fluorometer. RAD-seq libraries were pooled and run in two lanes of an Illumina HiSeq2500 sequencer at the University of Chicago Genomics Facility (USA), with a single end 100 bp kit using fast mode.

We used FastQC (Andrews, 2010) to assess the quality of RAD sequences. Sequences with Phred quality scores greater than 20 were kept. All sequences were pruned to 80 bp to minimize variability in the sequence quality. For sample demultiplexing, de novo assembly, and RAD loci and SNPs identification, we used the Perl modules implemented in Stacks v1.10 (Catchen et al., 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). For the de novo assembly and to generate a catalog of usable loci, we established the following parameters: (1) a minimum depth of six RAD reads per stack or locus (\( m \)), (2) a maximum of two mismatches to add reads (secondary) to a given stack, and (3) a maximum number of three mismatches allowed (\( M \)) between loci within each individual (Benestan et al., 2015). The resulting genotypes were combined using the populations module in Stacks, filtered to include all loci shared for at least 75% of individuals (\( r = 0.75 \) from each population [\( p = 5 \)]) and minimum allele frequency equal or greater than 0.01, to increase the opportunity to identify rare alleles or singletons. To generate final output files for downstream analyses, we used a combination of different output formats such as VCF, Genepop, and Structure (Pritchard, Stephens, & Donnelly, 2000).

### 2.3 Molecular estimates and genetic analyses

We calculated basic molecular estimates at the population level (i.e., observed heterozygosity-Ho, expected heterozygosity-He, and inbreeding coefficient-Fis) using the output provided by the populations module from Stacks.

We analyzed deviations from Hardy–Weinberg equilibrium using 999 iterations with GenoDive software v2.0 (Meirmans & Van Tienderen, 2004). The test was conducted by locus for each population, and a Bonferroni correction (Rice, 1989) for multiple comparisons was applied, starting with the lowest \( p \) value reported. To investigate how genetic diversity is distributed among populations, we conducted an analysis of molecular variance (AMOVA) and tested for population structure estimating the Weir and Cockerham’s \( F_{ST} \) unbiased estimator (Weir & Cockerham, 1984) between all population pairs. We used GenoDive with 999 permutations and a Bonferroni correction (Rice, 1989) to adjust \( p \) values for these calculations. To further test for significant structure among populations, we conducted a principal components analysis (PCA) using the R package adegenet version 2.0.1 (Jombart, 2008; Jombart & Ahmed, 2011), implemented in the software R within Rstudio 1.0.136 (Rstudio Team, 2015). We retained 5-axis components to maintain the original number of putative populations and to observe discrimination.

To investigate the level and directionality of historical translocations between natural bank populations, we estimated the number of migrants (\( N_{m} \)) using a coalescent approach implemented in the program Migrate-n v3.6.4 (Beerli & Felsenstein, 2001). Migrate-n uses Bayesian inference to estimate effective population size (variable \( \theta \)) and migration rates (variable M), which can be used to calculate \( N_{m} \) (\( N_{m} = M \times \theta \)). Migrate-n was run with the full migration Matrix model, adjusting the transition/transversion rate value to 2.0, for nuclear DNA and otherwise using default parameters. The dataset for this analysis included loci that were shared by all individuals from all populations. SNP markers suggested to be under positive selection by BayeScan (Foll & Gaggiotti, 2008) were removed from the dataset for this analysis.
3 | RESULTS

We developed and sequenced four genomic RAD libraries that yielded approximately 260 million sequencing reads, corresponding to single raw reads. On average, <1% of these sequences had ambiguous barcodes, <1% were low quality sequences, and less than 7% had ambiguous RAD loci. These estimates varied among individuals and among populations (see Data S4 and Data S5). After the initial screening and filtering, 225,410,859 sequences in total were retained for further analyses. The number of retained sequences varied among individuals and populations. The number of sequences per individual varied between 498,875 and 9,743,838 (mean value: 3,954,576.47, SD: ± 2,101,529.36) and the mean percentage of retained sequences was 85% (see Data S4). The average Phred score value was 30. All sequences were pruned to 80 bp to minimize variability in the sequence quality. The de novo assembly identified 9,593 RAD loci. After filtering the data to obtain only polymorphic loci for at least 75% of individuals from each population, we recovered 8,345 polymorphic RAD loci (with $\geq 1$ SNP and 246 average depth read) and 24,218 SNPs, which were used for genetic diversity and population structure analysis. The number of loci was reduced to 3,027 (yielding 7,608 SNPs) when including only loci shared by all individuals from all populations (zero missing data) and removing outliers SNPs (outlier SNPs = 4) for gene flow estimates.

3.1 | Genetic diversity

Observed heterozygosity (Ho) varied between 0.17 and 0.20 (Table 1). The population of Bahia Independencia exhibited the highest observed heterozygosity (Ho = 0.20) while Samanco exhibited the lowest value (Ho = 0.17). Similarly, individuals from Bahia Independencia presented the highest nucleotide diversity ($\pi = 0.22$) while Samanco exhibited the lowest diversity ($\pi = 0.1931$). The population from Samanco exhibited the highest estimate for the inbreeding coefficient (Fis = 0.085) while Lobos de Tierra and Sechura had the lowest (Fis = 0.03) (Table 1).

3.2 | Population structure and gene flow analysis

No RAD locus in Hardy–Weinberg disequilibrium was observed. The analysis of molecular variance (AMOVA) indicated that most of the genetic diversity is distributed within individuals (92.6%, $F_{IT} = 0.74$) and to a lesser extent among individuals (6.3%, $F_{IS} = 0.064$, $p = .001$) and among populations (1.1%, $F_{ST} = 0.011$, $p = .001$). Differentiation values among populations ($F_{ST}$) varied between 0 and 0.032 (Table 2). Significant genetic differentiation was observed between the population of La Rinconada (Chile) and the other populations (average $F_{ST} = 0.029$, $p = .001$). The lowest values of genetic differentiation were observed between Samanco compared to Bahia Independencia and Lobos de Tierra. The spatial structure detected between La Rinconada and the other populations was further observed in the PCA (Figure 2).

| Population  | SEC | LDT | SAM | IND | RIN |
|--------------|-----|-----|-----|-----|-----|
| SEC          |     |     |     |     |     |
| LDT          | 0.003 |     |     |     |     |
| SAM          |     | 0.001 |     |     |     |
| IND          | 0.003 | 0.002 | 0   |     |     |
| RIN          | 0.029* | 0.026* | 0.029* | 0.032* |     |

Note: $F_{ST}$ values can be observed under the diagonal. * indicates significant $F_{ST}$ values. SEC, Sechura; LDT, Lobos de Tierra; SAM, Samanco; IND, Bahia Independencia; RIN, La Rinconada.

TABLE 2 Triangular matrix with population differentiation estimations among five bay scallop's natural bank populations
Gene flow analysis implemented in Migrate-n was carried out using 3,027 loci (7,608 SNPs) shared among the five natural bank populations. On average, considering all possible directions, northward gene flow was higher compared with southward gene flow ($N_{m_{\text{Northward}}} = 6,848.17$, $N_{m_{\text{Southward}}} = 2,524.63$). The analysis showed asymmetric gene flow between Chilean and Peruvian populations: a strong gene flow from La Rinconada to the Peruvian populations (mean $N_m = 14,855.4$) and a weak gene flow from the Peruvian populations to La Rinconada (mean $N_m = 1,946.45$). Within the Peruvian populations, high gene flow from Lobos de Tierra to the other populations was observed (mean $N_m = 6,086.25$), while low gene flow was detected among the other populations (Sechura, Samanco, and Bahia Independencia) (Table 3).

4 | DISCUSSION

Despite being the most economically valuable bivalve aquaculture species in Peru, up to date, it is impossible to differentiate batches of exported wild Peruvian scallops from those produced in a sustainable way. This is mostly because of the lack of a reliable regulation system and/or methods for traceability and certification of origin for hatchery produced batches. Genetic tools offer an opportunity for reliable traceability of seafood (Bernatchez
et al., 2017) given detailed information on the genetic structure and distribution of the source populations is available. Two previous studies used mitochondrial markers to assess the genetic diversity and population structure in the Peruvian scallop (Acosta-Jofré et al., 2020; Marín et al., 2013). These markers were useful in determining genetic differences between populations, but they lack the power to detect small, but significant, genetic differences in nearby and large populations (Waples, 1998). In our study, we obtained estimates of genetic diversity and assessed the degree of population structure by exploring the variation in 24,218 SNPs belonging to 8,345 polymorphic RAD loci.

We found high levels of genetic diversity and SNP frequency compared with other bivalve species. The mean estimate of genetic diversity among all natural bank populations in the study, as indicated by $H_0$, was 0.1886, higher when compared with other marine bivalves. For example, a mean $H_0$ of 0.1085 for wild populations was reported for the black-lip pearl oyster *Pinctada margaritifera* (Lal, Southgate, Jerry, & Zenger, 2016). The observed heterozygosity for *A. purpuratus* was nearly equal to the expected heterozygosity ($H_e$, Table 1). This contrasts with other studies reporting heterozygote deficiency as a result of diverse proposed mechanisms (Lal et al., 2016; Miller, Versace, Matthews, Montgomery, & Bowie, 2013; Peñaloza, Bishop, Toro, & Houston, 2014). Similar to the genetic

| Population          | Code  | $\theta$ |
|---------------------|-------|----------|
| Sechura             | SEC   | 1.5      |
| Lobos de Tierra     | LDT   | 4.5      |
| Samanco             | SAM   | 0.5      |
| Bahía Independencia | IND   | 0.5      |
| La Rinconada        | RIN   | 11.5     |

| Migration directionality | $M$     | $N_m$     | Movement |
|-------------------------|---------|-----------|----------|
| SEC → LDT               | 835.8   | 1,253.70  | Southward|
| SEC → SAM               | 1,137.5 | 1,706.25  | Southward|
| SEC → IND               | 1,073.6 | 1,610.40  | Southward|
| SEC → RIN               | 681.8   | 1,022.70  | Southward|
| LDT → SEC               | 1,393.3 | 6,269.85  | Northward|
| LDT → SAM               | 1,378.7 | 6,204.15  | Southward|
| LDT → IND               | 1,337.4 | 6,018.30  | Southward|
| LDT → RIN               | 1,300.6 | 5,852.70  | Southward|
| SAM → LDT               | 1,148.1 | 574.05    | Northward|
| SAM → SEC               | 1,079.5 | 539.75    | Northward|
| SAM → IND               | 1,335.4 | 667.70    | Southward|
| SAM → RIN               | 952.9   | 476.45    | Southward|
| IND → LDT               | 1,100.2 | 550.10    | Northward|
| IND → SEC               | 1,217.1 | 608.55    | Northward|
| IND → SAM               | 1,035.5 | 517.75    | Northward|
| IND → RIN               | 867.9   | 433.95    | Southward|
| RIN → LDT               | 1,320.9 | 15,190.35 | Northward|
| RIN → SEC               | 1,208.1 | 13,893.15 | Northward|
| RIN → SAM               | 1,344.1 | 15,457.15 | Northward|
| RIN → IND               | 1,294   | 14,881.00 | Northward|

Note: $N_m$ was calculated as $M \times \theta$.  

| Table 3 | Mean estimates of mutation-scaled effective population sizes per population ($\theta$), migration rate ($M$) and number of migrants per generation ($N_m$) for each and migration directionality, calculated by Migrate-n |
diversity levels, our estimates of SNP frequency for *A. purpuratus* were also higher compared with other bivalve species. We detected, on average, a SNP every 20 bp, compared with estimates of 1 SNP/30 bp for the Chilean mussel and 1 SNP/40 bp for the Pacific oyster (Sauvage, Bierne, Lapegue, & Boudry, 2007). Our estimates agree with high polymorphic levels previously reported for bivalves.

Overall, the genetic diversity did not differ significantly among populations, albeit the population of Bahía Independencia, which exhibited the highest genetic diversity values (Ho = 0.2044). This agrees with a previous study where the population of Bahía Independencia exhibited higher genetic diversity than populations from Sechura and Samanco (Marín et al., 2013). Sechura and Samanco populations had the lowest genetic diversity estimates, which is in agreement with a previous study (Marín et al., 2013). These two populations are adjacent to hatcheries, so the low levels of diversity could be the result of migrants coming from them. A similar scenario has been observed previously for the black-lip pearl oyster (Lal et al., 2016). Low genetic diversity in hatchery populations is expected because of the reproductive dynamics in captivity, where a few progenitors are responsible for all the individuals produced, thus promoting allele loss. This has been observed in populations of captive invertebrates (Miller et al., 2013).

Samanco (Fis = 0.0655) and La Rinconada (Fis = 0.0556) (Table 1) exhibited the highest inbreeding coefficient while Lobos de Tierra and Sechura exhibited the lowest values. The surplus of homozygotes can be explained by self-fertilization caused by closed circulation in these bays. The Peruvian scallop is a functional hermaphrodite with external cross-fertilization and sperm release before the oocytes, to prevent self-fertilization (Uriarte, Rupp, & Abarca, 2001); the areas in Samanco and La Rinconada present gyres or gyre-like currents (Avendaño, Cantillán, Thouzeau, & Peña, 2007; Tenorio, 2016), which promote larval retention. In contrast, the lower inbreeding values detected in Lobos de Tierra Islands and Sechura Bay (Table 1) may be because of low larval retention in these areas (Flores-Valiente et al., 2019).

Population structure across the distribution range of *A. purpuratus* was detected between La Rinconada (Chile) and the Peruvian populations. Using the loci present in ≥70% of the individuals of all populations, we were able to distinguish La Rinconada population from the rest. This agrees with Acosta-Jofré et al. (2020), who also detected significant genetic differentiation between Antofagasta’s population (very close to La Rinconada) with the Peruvian populations. La Rinconada population is located far from the other populations; the linear distance to the nearest natural bank in Peru is nearly 1,250 km, making it very unlikely to be naturally colonized by larval dispersion from other areas (Acosta-Jofré, Acosta-Jofré et al., 2020). The differentiation was further supported by the ordination analysis, where La Rinconada population clearly distinguishes from the rest (Figure 2). Within Peru, we did not detect significant differentiation among natural banks (Table 2). Pairwise $F_{ST}$ values ranged from 0 (IND vs. SAM, LDT vs. SAM) to 0.003 (LDT vs. SEC, IND vs. SEC). Similarly, the molecular variance analysis indicates most of the variation (92.6%) exists within, rather than among populations, thus providing further evidence for the weak or lack of population differentiation among all Peruvian wild populations. Considering the geographic distance separating the natural banks in Peru, the high level of genetic admixture is remarkable. Exceptions are the populations of Sechura and Lobos de Tierra, which are separated by less than 150 km. If we account for the moderate dispersal capability of the Peruvian scallop and the “weak source (SEC) – strong sink (LDT)” scenario suggested for the populations of Sechura and Lobos de Tierra (Flores-Valiente et al., 2019), high levels of genetic admixture are expected. But, for the other populations, the distance between them is >600 km, right in the upper limit of the dispersal distance estimated for species with high larval dispersion capabilities (Van Wyngaarden et al., 2017). Except for Flores-Valiente et al. (2019), no other study has assessed the potential role that marine currents in Peru play in larval transport and retention (Barahona, Vélez-Zuazo, Santa-Maria, & Pacheco, 2019). In any case, we would expect to detect population structure at least between Bahía Independencia and the rest of the sites. Previous genetic distinction for this population was observed by Marín et al. (2013), who suggested that the geographic and oceanographic characteristics of this bay strongly limited natural gene flow with other populations.

We do not overrule the homogenizing role of larval dispersal facilitated by currents and oceanographic events (Miller et al., 2013), in combination with the large population size and high fecundity of marine bivalves (Cano, Shikano, Kuparinen, & Merilä, 2008). The Peruvian scallop has a moderate pelagic larval duration of 16–25 days in the water column, depending on the water temperature, before settlement (von Brand, Merino, Abarca, & Stotz, 2006). The combination of the northward passive larval displacement caused by the Peruvian Coastal Current
(PCC) and the southward displacement caused by the movement of warm water masses during ENSO events could facilitate the bidirectional population connectivity (Barahona et al., 2019). The recolonization of extinct natural beds (Wolff & Mendo, 2000) and the formation of temporal intermediate aggregations (Ysla, 2009) during and after ENSO could increase the genetic connectivity. But, given the geographical peculiarities and large distances separating Peruvian scallop natural beds, larval dispersal capabilities and marine currents alone are unlikely to explain the level of admixing we observed. Our results thus suggest a significant effect of historical translocations in gene-flow and homogenization of wild populations along the Peruvian coast beyond what would be expected by larval dispersion alone.

In any case, we detected an asymmetrical gene flow among the Peruvian and La Rinconada populations. The strong gene flow observed from the Chilean population agrees with the cold northward Humboldt Current (PCC). We found that mean northward gene flow is stronger than mean southward gene flow, which suggests that the effect of ENSO warm periods is weaker compared with the constant flow of the Peruvian Coastal Current for this species. Among Peruvian natural banks, we found evidence indicating that Lobos de Tierra is a source of migrants, which is consistent with reports of that location being a common source of seeds for translocations (Mendo et al., 2008).

Regardless of the power of using thousands of genome-wide SNPs, we did not detect fine-scale genetic structure among natural banks. Instead, we found a seascape of genetic homogenization for wild populations in Peru and differentiation from the most distant population, in northern Chile, likely the result of long-distance dispersal limitations. The lack of population structure and outlier loci to distinguish wild populations limits the opportunity of using genomic tools to trace Peruvian scallop stocks back to their population of origin. However, the low heterozygosity observed in natural bank nearby hatcheries suggests that genetic population descriptors such as kinship or inbreeding index could be used as an alternative to trace products back to wild stocks or to aquaculture hatcheries (i.e., bred in captivity).

5 | CONCLUSION

We detected a genetic homogeneity among Peruvian populations and a strong differentiation with a population in the north of Chile. Our results suggest that the genetic similarity among Peruvian populations is a footprint of long-term man-made seed translocations, beyond what would be expected by larval dispersion and oceanographic factors. Our study indicates that the traceability of individuals from wild or aquaculture populations using molecular markers would be difficult, but we do not overrule the use of population descriptors, like kinship or inbreeding index, as an alternative.

ACKNOWLEDGMENTS

We are grateful to Dr. Aldo Pacheco, who facilitated access to scallop samples from La Rinconada (Chile), and to Dr. Ronnie Gaviñán and Dr. Omar Cáceres, from the Peruvian National Institute of Health for kindly facilitating the development of genomic libraries. This study was funded by Grant No. 189-FINCyT-FIDECOM-PIPEA-2014 from The Fund for Innovation, Science and Technology (FINCyT/INNOVATE) of the Peruvian Ministry of Production.

CONFLICT OF INTEREST

The authors declare no conflicts of interest of any sort in the production and publication of this study.

AUTHOR CONTRIBUTIONS

Ximena Velez-Zuazo: Conceived and designed the project, wrote the research proposal, coordinated field data collection, analyzed the data, and wrote the study. Sergio P. Barahona: Conducted laboratory analyses including DNA extractions and development of RAD libraries, analyzed the data, and wrote the study. Omar G. Melo: Conducted laboratory analyses including DNA extractions and development of RAD libraries. Eric Hanschke: Contributed to
design of the project, co-wrote the research proposal, and helped collect field data. Ian Hanschke: Contributed to design of the project, co-wrote the research proposal, and helped collect field data. Monica C. Santa-Maria: Conceived and designed the project, wrote the research proposal, coordinated laboratory analyses, and wrote the study.

ORCID
Ximena Velez-Zuazo  https://orcid.org/0000-0002-6556-4101
Sergio P. Barahona  https://orcid.org/0000-0002-0136-7205
Omar G. Melo  https://orcid.org/0000-0003-0852-1559
Monica C. Santa-Maria  https://orcid.org/0000-0003-4789-7444

REFERENCES

Acosta-Jofré, M. S., Sahade, R., Mendo, J., González-Ittig, R. E., Laudien, J., & Chiappero, M. B. (2020). Population genetic structure and demographic history of the scallop Argopecten purpuratus from Peru and northern Chile: Implications for management and conservation of natural beds. *Hydrobiologia*, 847(1), 11–26.

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Babraham Bioinformatics. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arnaud-Haond, S., Vonau, V., Bonhomme, F., Boudry, P., Blanc, F., Prou, J., ... Goyard, E. (2004). Spatio-temporal variation in the genetic composition of wild populations of pearl oyster (*Pinctada margaritifera cumingii*) in French Polynesia following 10 years of juvenile translocation. *Molecular Ecology*, 13(7), 2001–2007.

Avendaño, D., ... Peña, J. B. (2007). Artificial collection and early growth of spat of the scallop Argopecten purpuratus (Lamarck, 1819), in La Rinconada marine reserve, Antofagasta, Chile. *Scientia Marina*, 71(1), 197–205.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 3(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Barahona, S. P., Vélez-Zuazo, X., Santa-Maria, M., & Pacheco, A. S. (2019). Phylogeography of the rocky intertidal periwinkle *Echinolittorina paytensis* through a biogeographic transition zone in the southeastern Pacific. *Marine Ecology*, 40, e12556. https://doi.org/10.1111/maec.12556

Beaumont, A. (2000). Genetic considerations in transfers and introductions of scallops. *Aquaculture International*, 8(6), 493–512.

Beerli, P., & Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, 98(8), 4563–4568.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, 24(13), 3299–3315.

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M., Beacham, T. D., ... Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, 32(9), 665–680.

Cano, J. M., Shikano, T., Kuparininen, A., & Merilä, J. (2008). Genetic differentiation, effective population size and gene flow in marine fishes: Implications for stock management. *Journal of Integrated Field Biology*, 5, 1–10.

Carlton, J. T., & Mann, R. (1996). Transfers and world-wide introductions. In V. S. Kennedy, R. I. E. Newell, & A. F. Eble (Eds.), *The eastern oyster, Crassostrea virginica* (pp. 691–706). Maryland Sea Grant College: College Park.

Catchen, J., Hohenlohe, P. A., Basham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., & De Koning, D. J. (2011). Stacks: Building and genotyping loci De novo from short-read sequences. *G3: Genes|Genomes|Genetics*, 1(3), 171–182.

Cavero Cerrato, P., & Rodríguez Pinto, P. (2008). Producción sostenida de moluscos bivalvos en el Perú: acuicultura y repoblamiento. In A. Lovatelli, A. Farias, & I. Uriarte (Eds.), *Estado actual del cultivo y manejo de moluscos bivalvos y su proyección futura: factores que afectan su sostenibilidad en América Latina* (pp. 209–218). *Taller Técnico Regional de la FAO*. 20–24 de agosto de 2007, Puerto Montt, Chile. Rome: FAO Actas de Pesca y Acuicultura, FAO.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499–510.

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genomics using RAD sequencing (Vol. 772, pp. 157–178). Clifton, NJ: Methods in Molecular Biology.
Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution, 38*(6), 1358–1370.

Wolff, M., & Mendo, J. (2000). Management of the Peruvian bay scallop (*Argopecten purpuratus*) metapopulation with regard to environmental change. *Aquatic Conservation: Marine and Freshwater, 10*, 117–126.

Ysla, L. (2009). Estudio de la variabilidad genética en poblaciones de *Argopecten purpuratus* de Chile y Perú. *Anales Científicos UNALM, 70*(1994), 60–65.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Velez-Zuazo X, Barahona SP, Melo OG, Hanschke E, Hanschke I, Santa-Maria MC. Substantial gene flow caused by long-term translocation between natural bank populations of the Peruvian scallop (*Argopecten purpuratus*) is supported by RAD-Seq analyses. *J World Aquac Soc*. 2021:1–13. [https://doi.org/10.1111/jwas.12795](https://doi.org/10.1111/jwas.12795)