



OPEN

Genetic diversity of *Anadara tuberculosa* in two localities of the Colombian Pacific Coast

Luis Fuentes^{1,2}, Marcela Guevara-Suarez^{1,2}, María Mercedes Zambrano³, Pedro Jiménez⁴, Jorge Duitama⁵ & Silvia Restrepo^{1,6}✉

Piangua, *Anadara tuberculosa*, is an economically important mollusk for the human population living on the Colombian Pacific Coast. In the last years, the demand and exploitation of this mollusk have increased, putting it at risk to the point of being endangered. This research aimed to identify the genetic diversity and population structure of piangua in two localities on the Pacific Coast of Colombia. We assembled a chromosome-level genome using PacBio-Hifi and Arima sequencing. We obtained 274 scaffolds with an N50 of 45.42 Mbp, a total size of 953 Mbp, and a completeness of 91% based on BUSCO scores. The transposable elements accounted for 30.29% of the genome, and 24,317 genes were annotated. Genome-guided variant calling for 89 samples using DArT sequencing data delivered 4,825 bi-allelic SNPs, which supported genetic diversity and population structure analyses. Data showed that the piangua populations in the two localities were under expansion events more than 100k years ago. However, results also showed a reduction in genetic diversity, as evidenced by the loss of heterozygosity, which may be caused by high levels of inbreeding, probably due to a recent overexploitation. Furthermore, although we evidenced gene flow between the two localities, there is also a subtle geographical population structure between the two localities and among mangroves in one of the localities. This is the first study in Colombia that provides relevant genetic information on piangua to lay the foundations for conservation strategies.

Colombia is considered one of the most biodiverse countries in the world, with approximately 10% of the diversity of the Earth¹, and one of the Colombian most diverse regions is the Pacific coast. This region belongs to the Chocó biogeographic region, considered the ninth most biodiverse hotspot in the world^{2–4}. However, social indicators on the Colombian Pacific coast are among the worst in the country^{2,5}. Consequently, the communities have seen the need to exploit natural resources as economic and nutritional sources through mining, hunting, and fishing².

On the Colombian Pacific coast, mollusk fishing is carried on in the mangroves, an ecosystem rich in biodiversity⁶. In the last years, shellfish extraction has increased as it has become an important economic activity for local communities, mainly in regions such as Bahía Málaga, Bahía Solano, Iscuandé, and Tumaco⁷, where the piangua (*Anadara tuberculosa*) is one of the mollusks with significant importance. Demand for the piangua has led to high capture levels in Colombia, where its export increased from 100 tons/year to 3,283 tons/year between 1980 and 2004⁸. This led to a reduction of piangua of up to 60% in Bahía Málaga⁹, while in Tumaco, the per capita fishing rate has decreased considerably since 1975⁵. This effect is also reported for Ecuador and Perú¹.

Deforestation, overfishing, and overharvesting are the most important threats to biodiversity of marine species in Colombia. These phenomena generate important genetic losses between and within species, evidenced by a decrease in effective population size, which induces inbreeding and accelerates the reduction of diversity through genetic drift. This process leads to a reduction in species fitness, which could lead to extinction.

The decrease in the piangua population has placed this mollusk among the endangered species¹⁰. For all these reasons, the Colombian government is committed to adopting bioeconomy (Global Bioeconomy Summit Communiqué, 2018) as a sustainable approach to exploiting its natural resources. Within this framework, scientific knowledge of the diversity and population genomics of piangua is crucial because it will provide information for the pursuit of conservation strategies and sustainable exploitation.

¹Laboratory of Mycology and Phytopathology (LAMFU), Department of Biological Sciences, Department of Food and Chemical Engineering, Universidad de Los Andes, Bogotá, Colombia. ²Applied genomics research group, Vice president of Research, Universidad de Los Andes, Bogotá, Colombia. ³Corpogen Research Center, Bogotá, Colombia. ⁴Faculty of Basic and Applied Sciences, Universidad Militar Nueva Granada, Cajicá, Colombia. ⁵Department of System and Computing Engineering, Universidad de Los Andes, Bogotá, Colombia. ⁶Boyce Thompson Institute, Ithaca, NY, USA. ✉email: srestrep@uniandes.edu.co

Population genetics studies in different species use reduced representation DNA sequencing (RRS) approaches, such as DArTseq¹¹. Although the analysis of RRS data can be performed *de novo*¹², using a reference genome can lead to more robust information to infer population parameters and diversity trends¹³. The *de novo* assembly of eucaryote genomes remains challenging due to the quality and length of DNA sequencing reads needed to assemble genomes with a high percentage of repetitive sequences and high heterozygosity¹⁴. High-quality chromosome-level assemblies for some complex genomes have been achieved using a combination of long-read sequencing, such as PacBio or NanoPore, and scaffolding techniques, such as the high-throughput chromatin conformation capture (Hi-C) approach^{15,16}. The hybrid methodology has allowed the chromosome-level assembly of *Anadara brouthonii* and *Anadara kagoshimensis*, two species related to *A. tuberculosa*^{17,18}.

Therefore, in this study, we report a chromosome-level genome assembly of *A. tuberculosa*, and its use as a reference in the analysis of its genetic diversity and population structure in two regions of the Colombian Pacific coast. Despite our initial goal to sample along the complete Pacific coast of Colombia, our sampling was limited by socioeconomic and political limitations, making it difficult to fully cover the species' range. Although we limited our analysis and conclusions to two localities (Fig. 1), we gathered important information to unravel the genetic situation of the species in this understudied region. The results are relevant for discussing conservation strategies with the appropriate stakeholders.

Results

The genome of *Anadara tuberculosa*

We assembled a genome combining Pac-Bio-Hifi and Hi-C sequencing data. A total of 33 Gb of Pac-Bio Hifi data, containing 3,698,171 raw reads with a mean length of 12,751 base pairs and a quality > 34, were used for the assembly. The primary assembly obtained running the Hifiasm software had a size of 953 Mbp, fragmented into 366 contigs with an N50 of 9.96 Mbp and a 28X median read depth. Scaffolding of these contigs was performed using 77 Gb of Hi-C pair-end sequencing data (Table 1). A total of 525,668,541 reads were aligned to the primary assembly of Hifiasm with a 78.31% overall alignment rate. Scaffolding was performed using this alignment, obtaining 274 scaffolds with a scaffold N50 of 45.419 Mbp. The mitochondrial genome was identified as scaffold 212 of assembly with a size of 38 Kbp. It showed 80.3% identity with the mitochondrial genome of *Anadara kagoshimensis* and 71.5% identity with that of *Anadara brouthonii*.

Based on the alignment of conserved genes in the Mollusca phylum, this primary assembly had a predicted completeness of 91% [Single-Copy: 87.6%, Duplicated: 3.4%], with 2.8% of fragmented genes and 6.2% of missing genes. According to the Hi-C heat map (Fig. 2A), the intense red diagonal indicates that the scaffolds showed a well-organized intra-chromosomal interaction. However, the intense red dots away from the diagonal indicate many interactions between scaffolds, mainly between small scaffolds.

We annotated transposable elements (TEs), accounting for 30.29% of the genome. The DNA transposons spanned 20.66%, and the LTR elements spanned 8.86%. After masking regions with TE annotations, a total of 24,305 genes were annotated with an average CDS and protein length of 1,264.79 bp and 420.45 bp, respectively. A synteny analysis of orthologous genes between the annotated genome assemblies of *A. tuberculosa* and *A. brouthonii* revealed 511 blocks, from which 52 *A. tuberculosa* scaffolds shared homologous genes with the 19 *A. brouthonii* chromosomes (Fig. 2B). From these, the largest scaffold (scaffold one) was paired with the chromosomes 12, 14, and 17 of *A. brouthonii* (crossed lines in Fig. 2A).

For the functional analysis (Fig. S1), only 18,760 of the 24,305 predicted proteins were classified, representing 92.4%. These proteins were characterized in Clusters of Orthologous Groups (COGs), which, in turn, were classified into four general categories. Thus, 37.9% of proteins were classified in the cellular processes and signaling category, 18.6% in metabolism, 17% in information storage and processing, and 26.4% in poorly characterized. The proteins present in the last category were proteins of unknown function.

Identification of the SNPs of the *A. tuberculosa* populations

Genotyping was performed by mapping the DArTseq reads to the chromosome-level genome presented above. An average of 2,000,000 trimmed and filtered reads for each sample were mapped with an overall alignment rate of 85%. A total of 78,889 SNPs were initially identified. After applying filters, the number of bi-allelic SNPs was reduced to 4,825, with a missing data of 21.5%. Out of all these SNPs, 1,381 rejected the assumptions of Hardy-Weinberg Equilibrium (HWE) with a corrected p-value < 0.01. The minor allele frequency (MAF) distribution (Fig. S2) showed that 3,293 SNPs had frequencies lower than (0.1). A total of 356 of these SNPs belonged to the 1,381 HWE-rejected SNPs. Moreover, 49% of HWE-rejected SNPs had frequencies lower than (0.2). To compare diversity information among localities, the samples were classified by localities, and population statistics were recalculated keeping all 4,825 SNPs within each locality.

Subtle genetic structure of piangua populations in two localities on the Pacific coast

Genetic diversity in both sampling points was similar, as assessed by the similarity of MAF distributions (Fig. S3). More than half of the SNPs had minor allele frequencies below < 0.1. Specifically, the minor alleles of 826 SNPs were segregating only in Buenaventura, while minor alleles of 788 SNPs were segregating only in Iscuandé (Fig. S8). These private SNPs represent approximately 16.6% of all SNPs. In 97% of the cases, the minor allele frequency of these SNPs within their population was lower than 0.1. Consequently, only 16 SNPs from Buenaventura and 27 SNPs from Iscuandé rejected the HWE assumption.

A loss of heterozygosity was evident due to a high difference between expected (H_e) and observed (H_o) heterozygosity in both regions. In Buenaventura, the average H_e was 0.154 while the average H_o was 0.048, indicating a loss of heterozygosity of about 68.8%. Similarly, in Iscuandé, the average H_e was 0.166 while the

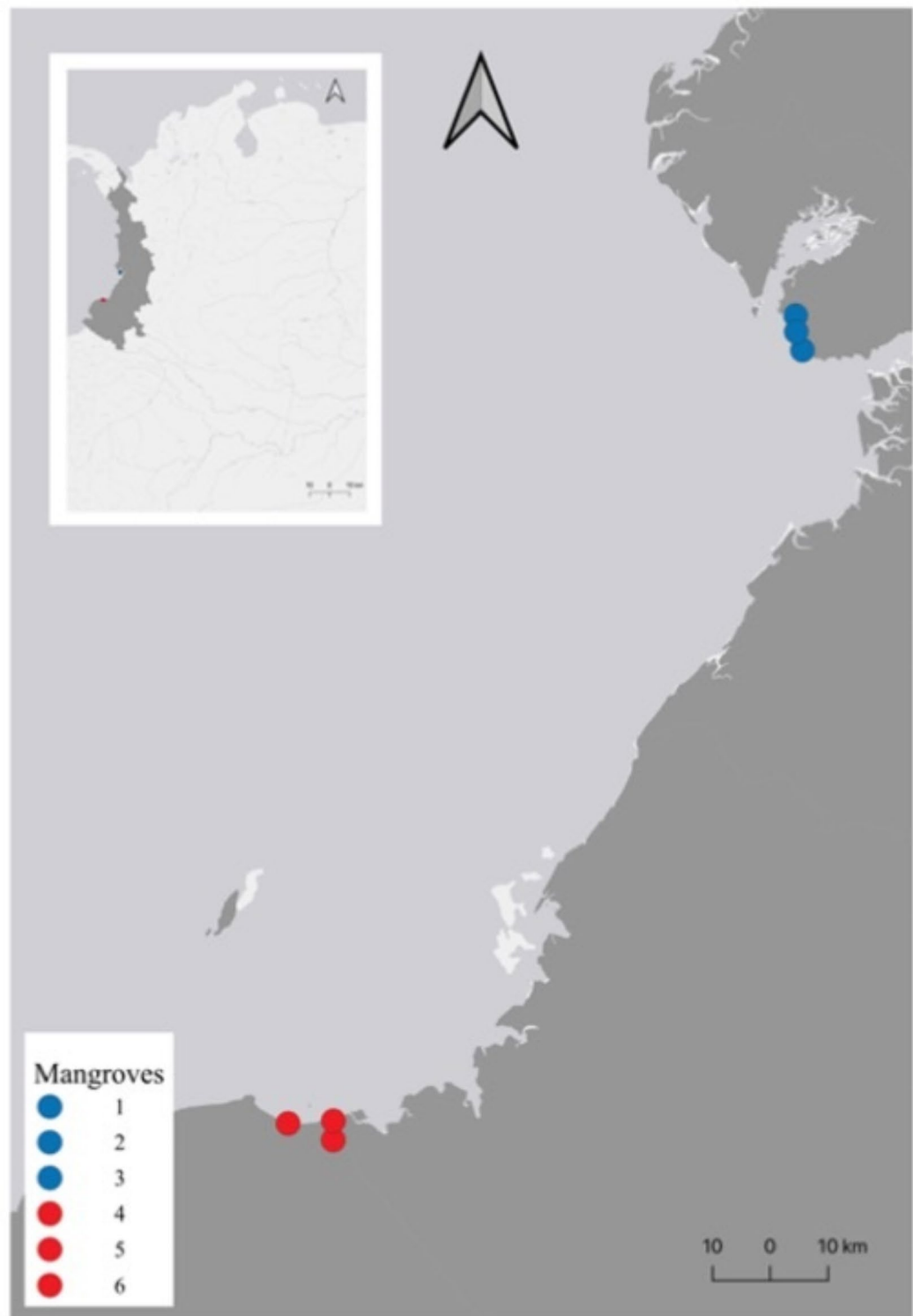


Fig. 1. Map of sampling points. The Pacific region is represented in dark-gray color. The sampling points (mangroves) from Buenaventura are in blue, and the sampling points from Iscuandé are in red. The map was made using QGIS (<https://www.qgis.org/es/site/>).

average H_o was 0.045, indicating a loss of heterozygosity of about 72.9%. The distribution of heterozygosity was similar in both regions, where ~86% of SNPs had frequencies below 0.1 (Fig. 3).

Nucleotide diversity was compared between both localities, 5.57×10^{-8} in Buenaventura and 6.03×10^{-8} in Iscuandé. This showed no difference in nucleotide diversity between them and corroborated the MAF distribution results. The fixation index, F_{st} , showed low differentiation in the allele frequencies among the two localities

PacBio - HiFi	Reads	3,698,171
	Contigs	366
	N50	9.93 Mbp
	L50	30
	GC	32.39%
	Length	953.1 Mbp
	Depth	28X
Arima	Reads	525,668,541
	Scaffolds	274
	N50	45.42 Mbp
	L50	9
	GC	32.39%
	Length	953.1 Mbp
	Depth	28X
BUSCO	Single-copy	87.6%
	Duplicated	3.4%
	Fragmented	2.8%
	Missing	6.2%

Table 1. Summary statistic for the Chromosome-level genome.

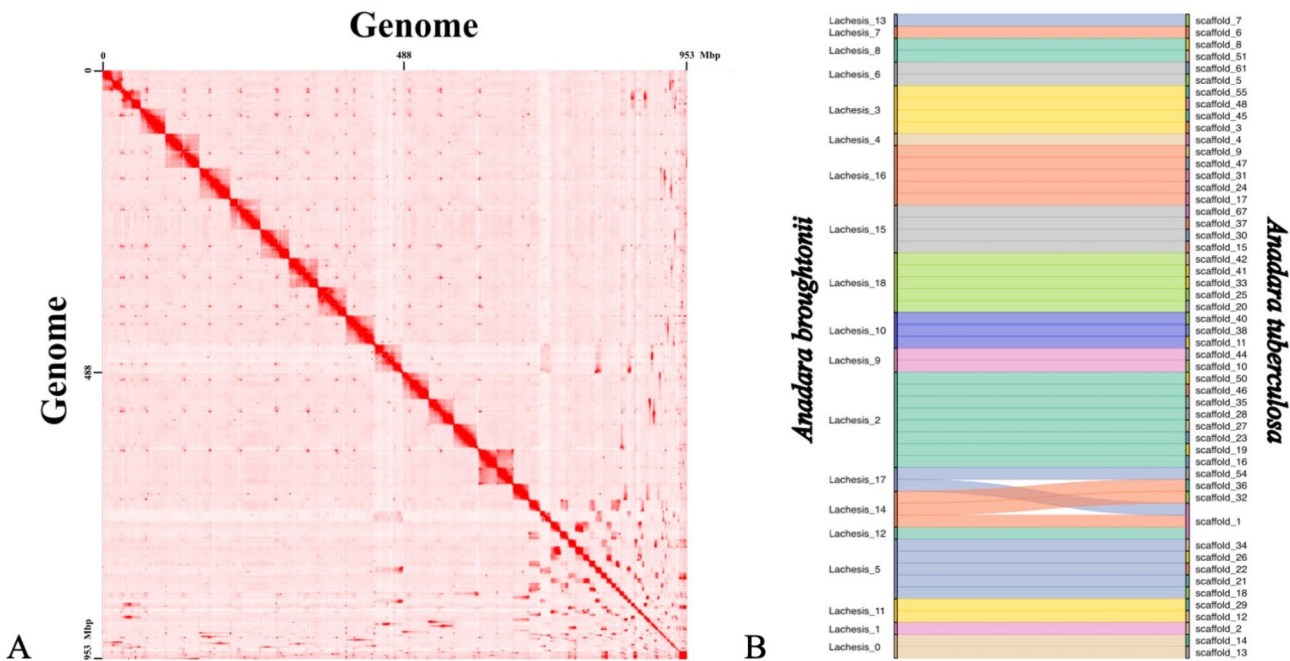


Fig. 2. Genomic analysis. **(A)** Hi-C heatmap. The x-axis and y-axis represent the 274 scaffolds. Intense-red dots show the interactions between regions. **(B)** Synteny diagram between the chromosome-level genome of *A. tuberculosa* and the chromosome-level genome of *A. broughtonii*.

(0.004) (Table 2). The low differentiation was constant within each locality because the pairwise F_{st} between mangrove swamps did not exceed 0.0104, and the F_{st} values between Iscuandé sampling points were negative (Table S1). Although the highest pairwise F_{st} was 0.01, some F_{st} values were significantly higher than zero, suggesting an incipient population differentiation. Two of these cases were observed between sampling points within Buenaventura. The inbreeding coefficient (F_{is}) and the total inbreeding coefficient (F_{it}) were high when the samples were not filtered by locality, with F_{is} of 0.713 and F_{it} of 0.714. However, these nonrandom mating indexes were also high within localities, 0.655 and 0.656 in Buenaventura and 0.684 in Iscuandé for both indexes (Table 2). This suggests that the variation within localities can explain most of the total variation.

According to the population genetic analysis (Fig. 4), there is a subtle genetic differentiation based on the geographical origin of the specimens. The STRUCTURE analysis (Fig. 4A) showed that neither locality was represented by a single cluster (yellow or black); however, admixture was evident between both localities.

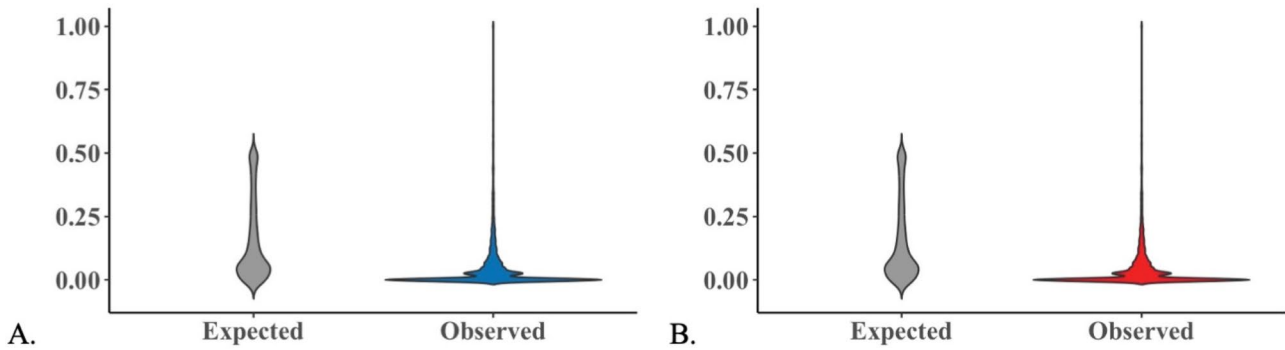


Fig. 3. Violin plot of the distribution of the heterozygosity in both localities: **(A)** Heterozygosity density distribution for Buenaventura, the gray violin representing the expected heterozygosity density, and the blue violin representing the observed heterozygosity density. **(B)** Heterozygosity density distribution for Iscuandé. The gray violin representing the expected heterozygosity density, and the red violin representing the observed heterozygosity density. The x-axis is the frequency of heterozygosity.

Locality	Buenaventura	Iscuandé
Samples	49	40
F _{st}	0.004	
F _{is}	0.713*	
	0.655	0.684
F _{it}	0.714*	
	0.656	0.684
H _e	0.155	0.166
H _o	0.048	0.045
π	5.57e ⁻⁰⁸	6.03e ⁻⁰⁸
Tajima' D	-0.11	-0.10

Table 2. Population genetics values of piangua populations.

Although the Evanno method suggested that the best clustering for the data was two (Fig. S4A), the raw likelihood does not increase with the number of possible populations (Fig. S4B). The result of the LEA (Fig. S4C) and the NetView (Fig. S5) tests corroborated the admixture shown by STRUCTURE. However, the DAPC analysis (Fig. 4B) resolved partially overlapping clusters showing a subtle, limited genetic structure between the two sampled localities.

The admixture effect found in STRUCTURE was also evident when the specimens were sorted by mangrove (sampling point) to assess if there was a structure within localities. The gene flow analysis showed comparable gene exchange at all sampling points, with a minimum flow of 6% from mangrove five to mangrove three (Fig. S6), and with a mean of 93% between locations (Fig. 4C).

Demography history

The extended Bayesian skyline plot (EBSP) revealed two significant population expansion events, the first around 130K years ago and the second around 100K years ago followed by a period of population stability (Fig. S7). This finding was supported by the neutrality test, where the Tajima' D index yielded negative results for both localities, with -0.11 for Buenaventura and -0.10 for Iscuandé (Table 2). However, recent trends indicate a rapid population decline, as evidenced by the confidence intervals (CIs) for effective population size (Ne) where the values for Buenaventura, Iscuandé and the combined population fall below one thousand (Table S2).

Discussion

Human impact is one of the main reasons for the loss of biodiversity^{19,20}. Hence, it has been necessary to propose conservation strategies that consider information on genetic variability between and within species²¹. In recent years, genetic information has increased with the development of high-throughput sequencing technologies that have improved genetic knowledge in marine species^{22–24}. Conservation strategies have also been proposed based on monitoring epigenetic markers and performing transcriptome assays²⁵. In *A. tuberculosa*, the conservation strategies have been deficient, making it necessary to improve genetic knowledge of this mollusk. In this manuscript, we present the results of our efforts to use high throughput sequencing (HTS) through reduced representation libraries to assess the genetic diversity and population structure of *A. tuberculosa* in two localities on the Pacific Coast of Colombia. The samplings analyzed in this study represent only a subset of the full distribution of Piangua along the Colombia Pacific coast. It is important to acknowledge that the inference

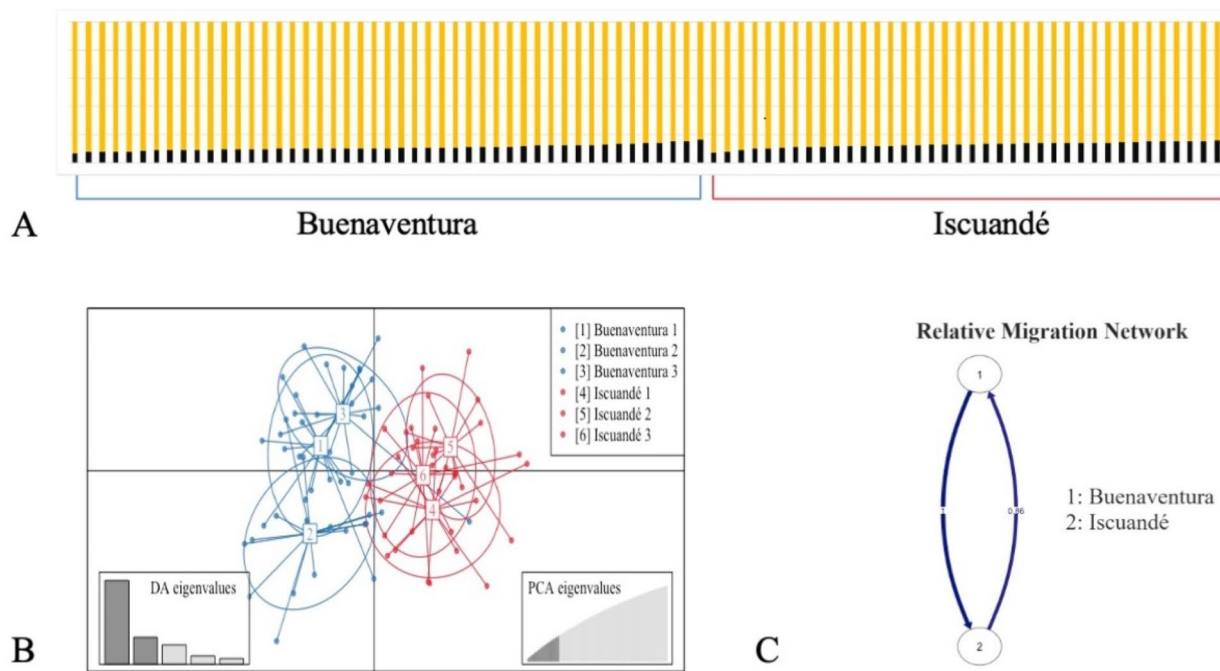


Fig. 4. Population genetic analysis. (A) Cluster distribution by sample, according to Evanno et al. method. Thus, each cluster was represented by one color (black or yellow), and each bar corresponds to a sample. A total of 49 samples from Buenaventura (left) and 40 samples from Iscuandé (right) were sorted by the proportion of one cluster. Each bar represents an individual and colors are the genetic cluster of each individual. (B) The DAPC cross-validation. The clusters 1–3 represented the Buenaventura sampling points and the clusters 4–6 represented the Iscuandé sampling points. (C) Guided graph representing gene flow between the Buenaventura (1) and Iscuandé (2) samples. Blue arrows indicate genetic migration between localities.

of the genetic values and population numbers inferred from this study are geographically limited due to the incomplete samples caused by the constraints mentioned in the introduction^{26,27}. However, this is the first study using a technique based on HTS to obtain information on genetic diversity in Piangua, and the analysis will allow us to plan suitable conservation strategies with the local communities. We also offer the first chromosome-level genome for this species, an important resource for future genetic studies.

Piangua genome

The lack of genetic information on Piangua could limit understanding of its biological and ecological processes. A genome of high quality and contiguity plays a crucial role in obtaining the best representation of the genetic diversity of species and understanding SNP distribution throughout the genome²⁸. The chromosome-level genome reported in this study is the first genome of Piangua species and is of great relevance in subsequent studies.

We reported a genome of 953.1 Mbp, with an N50 scaffold of 45.42 Mbp and 274 scaffolds. These statistics were close to the genomes reported in *S. kagoshimensis*¹⁸ and *A. broughtonii*¹⁷, two species of the same genus. Using synteny analysis, we observed that the genome was at the chromosomal level, as our scaffolds reached completeness for all chromosomes reported in the *A. broughtonii* genome (Fig. 2). Furthermore, our genome at the chromosomal level had a completeness comparable to *Crassostrea gigas* (95.6%), which is the most investigated molluscan genome²⁹, and with related species.

In terms of annotation, we found that 30.29% of the genome is represented by TEs. Recent studies have shown the importance of TEs as drivers of phenotypic variation in eukaryotes^{30–32}. Within these regions, 8.86% of the genome corresponds to long terminal repeats (LTRs). LTRs are spread out throughout the Mollusca phylum, and they provide information about the genome evolution processes in this phylum³¹. Furthermore, the number of TEs found in the Piangua genome were slightly different from the TEs reported in *A. broughtonii*, and *A. kagoshimensis*^{17,18}. Therefore, it is important to correlate this result with the evolutionary process and transposable element analysis of *A. tuberculosa*. The number of predicted genes was almost identical to those reported in related species, although this number differed to *C. gigas*.

Genetic diversity of *A. tuberculosa* on the Pacific Coast

Despite some small conservation efforts made by the local communities, the piangua population has suffered a considerable decline in both localities. This phenomenon could explain the large difference between H_e and H_o (Table 2), where it is possible to evidence a loss of heterozygosity of ~71% (Buenaventura = 68.8% and Iscuandé = 72.9%). It could imply a possible loss of genetic diversity³³ in both piangua populations. The MAF distribution (Fig. S3) was similar in both localities, suggesting that the allele frequency in both Piangua populations was

fairly stable. The nucleotide diversity (Table 2) showed that Buenaventura and Iscuandé shared similar levels of genetic variation.

A demographic analysis of the genotypic data suggests that the Piangua population in both localities has steadily increased after two population expansions around one hundred thousand years ago (Supplementary Fig. 7, Table 2). This result is similar to those observed for other bivalve species^{34–36}. However, this analysis is not able to capture recent events (within the last 50 years) in which the populations can be experiencing rapid declines. In this case, the decline is supported by the observed loss of heterozygotes and the consequent decrease in genetic diversity in the piangua population in Buenaventura and Iscuandé. This heterozygote deficiency could be supported by the influence of null alleles and the Wahlund effect^{37,38} and high inbreeding coefficients may be related to the constant process of overexploitation of the resource, considering that Piangua species has been overfished for the last 30 years in Colombia⁹. Interpreting the effective population size was challenging because there is no previous information about Piangua N_e . However, in marine species with high fishing pressure, the N_e values might be smaller and the populations suffer a loss of genetic diversity³⁹. Our analysis predicted a N_e value of 895 individuals in the combined population, which is lower than that reported for other bivalves^{40–42}.

The expression of important survival traits may be reduced due to inbreeding depression^{21,43,44}. This process is common in mollusks due to mortality processes in early life stages⁴⁵ and population reduction. A similar result has been reported in the bivalve *Crassostrea gigas*, where the inbreeding depression caused a reduction in survival rate and growth⁴⁶. Moreover, studies on other marine species, such as *Pagrus auratus*⁴⁷, and *Pleuronectes platessa* *L.*³⁹, have also shown that inbreeding resulted in the disappearance of multiple traits. Therefore, recognizing whether the Piangua population is undergoing a process of high inbreeding is important to propose strategies that favor a population expansion process. By increasing the effective population size²¹ it is possible to improve heterozygosity within populations.

Population structure of piangua populations in two localities of the Colombian Pacific coast

Establishing the population structure in bivalves is highly challenging, due to the high capacity of migrations by the high flow gene. In the case of Piangua, the STRUCTURE analysis did not show a clear genetic differentiation at the population level between Buenaventura and Iscuandé, even though they are geographically distant, approximately 180 km. Similar to other mollusks, the high genetic similarity between both localities may be attributed to the high capacity of dispersion of Piangua at the larva stage. Additionally, STRUCTURE analysis often struggles to detect weak or subtle genetic structures²⁷. In contrast, the DAPC cross-validation analysis (Fig. 4B) and pairwise F_{st} values suggest a subtle genetic structure between two localities, with a possible sub-structure within Buenaventura. This phenomenon could be explained by a very recent overexploitation of this mollusk in both localities. However, the result can also be attributed to temporal chaotic genetic patchiness (CGP)⁴⁸. Thus, this situation would require close and continuous genetic monitoring of the populations in the two localities to determine if human intervention is indeed causing the structuring of a panmictic population.

The low value of F_{st} (0.004) and the admixture observed within samples in both locations imply that there is a high gene flow between the two populations (see Fig. 4C and Fig. S6). This high gene flow is the consequence of frequent migration among individuals across the localities which is driven by the patterns of oceanic gene flow across the Colombian Pacific coast⁴⁹ and the capacity of the species to disperse over large distances in the early life stages⁵⁰. However, the p -values among the sampling points in Buenaventura showed a significant difference between points 1 and 3 compared to point 2 (Table S1), a result corroborated by DAPC analysis (Fig. 4B). This finding suggests the development of a recent sub-structure in Buenaventura, likely driven by the loss of heterozygosity or CGP⁴⁸.

The results obtained in this study agree with the analysis of Diringer et al., which showed no differentiation in population structure between two *A. tuberculosa* populations that are also located north of the equator⁵⁰. They analyzed the COI mitochondrial marker in 48 specimens of two localities above the equator, 24 for Esmeralda, Ecuador, and 24 for Tumaco, Colombia⁵⁰. They showed that the F_{st} between these localities was 0.011 and argued that the main reason for this result is that the trochophore larvae can disperse for a few hundred kilometers⁵⁰. Also, some studies have demonstrated that oceanic currents can mark patterns in the genetic variations of mollusks such as *A. broughtonii* and *Pinctada maxima*^{37,50–52}.

Conservation implications

One of the most critical concerns for the community in these regions is the decline of the piangua population over the years. Marine conservation strategies must be improved, and research efforts must catch up. One main reason is the lack of knowledge of marine species^{53,54}, mainly in the genetic field. Providing genetic insight allows an understanding of the evolutionary process, population dynamics, and genetic flow, contributing to the design of conservation strategies^{54–57}. In this study, we found that piangua populations in two localities of the Colombian Pacific coast have remained stable for several thousand years. However, we observed a recent and considerable loss of genetic diversity, and a very subtle population structure with a probable sub-structure in Buenaventura likely influenced by overfishing.

Piangua has been subjected to constant overexploitation in recent years, which could lead to a reduction of population size and might be the consequence of the inbreeding process and the loss of diversity^{39,58}. Recognizing the genetic information makes it possible to establish strategies such as restoration interventions and environmental monitoring^{54,59}. Thus, it is crucial to intensify efforts to study the broad dispersal range of Piangua to gain a comprehensive understanding of its current population status. It is essential to establish correlations between this status and the effects of the fishing pressure and its impact on the weak population structure. Additionally, the potential role of private alleles of each locality should be assessed since some of them could be involved in early adaptation processes to each environment⁶⁰.

Materials and methods

Sample collection

Specimens were collected from two localities on the Colombian Pacific coast, Buenaventura, in the Valle del Cauca department, and Iscuandé, in the Nariño department (Fig. 1). The collection points were selected based on information provided by local communities. All specimens had an average size of 5 cm in diameter and were not discriminated between females and males. For Buenaventura, 51 specimens were collected from three mangrove swamps: 17 from mangrove 1 (3.846030, -77.283847), 16 from mangrove 2 (3.900195, -77.294537), and 18 from mangrove 3 (3.8747614, -77.2935579). For Iscuandé, 43 specimens were collected from three mangrove swamps: 15 from mangrove 4 (2.61278, -78.01833), 14 from mangrove 5 (2.64306, -78.01806), and 14 from mangrove 6 (2.63889, -78.08917). The specimens were dissected, and a piece of the adductor muscle and mantle of each specimen was extracted and conserved separately in 1 mL of salt-saturated DMSO buffer (20% DMSO, 250 mM EDTA pH 8, and NaCl to saturate the solution)⁶¹. The remaining tissue was conserved in Falcon tubes with approx. 10 mL of salt saturated DMSO buffer. All specimens were brought to the Mycology and Phytopathology Laboratory (LAMFU), where each specimen's genomic DNA was extracted.

Genome sequencing and assembly

Tissues from two specimens collected at Buenaventura were sent to the Vertebrate Genomes Laboratory (VGL) of Rockefeller University, where whole-genome sequencing was done using PacBio-Hifi and Arima technologies. According to their protocols, VGL extracted the genomic DNA (gDNA) for sequencing. The data from PacBio-Hifi was analyzed using NanoPlot⁶². The genome assembly was performed through *de novo* methodology using the Hifiasm assembler⁶³ with default parameters, a ploidy of two, and the Hi-C reads integration with the tags --h1 and --h2. The scaffolding was performed through Salsa⁶⁴ according to its indications, and then it was evaluated using Juicer tools (<https://github.com/aidenlab/juicer>). The assembly was assessed using Quast⁶⁵ and BUSCO⁶⁶ with the mollusca_odb10 database. The mitochondrial genome was identified through local alignment via blastn command, with our chromosome-level genome as the query and the mitochondrial genome of *Anadara broughtonii* (OM807134.1) as the target. The identified scaffold corresponding to the mitochondrial genome was then isolated and analyzed on the BLAST website (<https://blast.ncbi.nlm.nih.gov/>) against *Anadara* taxon (taxid:6554).

The transposable elements were identified through the EDTA pipeline⁶⁷; the assembly was then masked using the NGSEP GenomeAssemblyMask command. The genes were annotated using MAKER⁶⁸, which received the masked assembly, and the *A. broughtonii* proteome, downloaded from MolluscoDB (<http://mgbase.qnlm.ac/home>). The annotation was evaluated using the NGSEP TranscriptomeAnalyze command, and the synteny was performed through the JCVI MCscan pipeline⁶⁹. Finally, a functional analysis was performed through eggNOG-mapper v2⁷⁰ using as input the *A. tuberculosa* proteome obtained from the annotation evaluation.

Genomic DNA extraction

The genomic DNA was extracted from each specimen using the CTAB method⁷¹. Briefly, approx. 30 mg of adductor muscle was incubated with 600 µL of CTAB buffer (100 mM Tris-HCl pH 8, 1.4 M NaCl, 20 mM EDTA, and 2% CTAB) supplemented with 0.2% Mercaptoethanol and 0.1 mg/mL proteinase K (New England Biolabs) at 60 °C for 1 h, then the genomic DNA was extracted using 500 µL of chloroform: isoamyl alcohol (24:1). The mix was centrifuged at 10,000 x g for 5 min, the supernatant was collected in a new tube, and 300 µL of isopropanol were added and incubated at -20 °C overnight. The DNA was harvested by centrifugation at 10,000 x g for 5 min and then washed once with 1 mL 70% of ice-cold ethanol. The pellet was air-dried at room temperature. Finally, the DNA was resuspended in 70 µL of DNase and RNase-free water. The RNA contamination was eliminated through digestion using 1 mg/mL RNase A (New England Biolabs) at 37 °C for 1 h. The concentration was estimated using Qubit™ 4 (Thermo Scientific) by the Core Facility - GenCore (Universidad de Los Andes), and the quality was evaluated on a 1% agarose gel.

Genotyping analysis

A total of 89 specimens of piangua were genotyped through DArTseq technology¹¹. Approximately 50 to 100 ng of genomic DNA for each individual was used for this analysis. The DArT libraries were generated by the digestion of genomic DNA with *Pst*I and *Taq*I restriction enzymes, then the fragments were ligated with Illumina adaptors and sequenced on Illumina (Illumina Technologies). DArT sequencing data were obtained for 49 Buenaventura and 40 Iscuandé specimens. The barcode sequences were removed from the raw reads using Trimmomatic⁷², and the 3'-end was cleaned up from the adaptor sequence using an awk command. The raw reads with sizes greater than 40 bases were kept and stored in a new fastq file.

SNP discovery and genotyping were performed following the reference-guided pipeline implemented in NGSEP v 4.3.1⁷³. The raw reads were aligned to the piangua genome using the NGSEP ReadsAligner command. Then, the bam files were sorted by Picard software (<https://github.com/broadinstitute/picard>). The SNPs were detected using the NGSEP MultisampleVariantsDetector command with the sorted bam files and the tag -minMQ with a value of 30. Other parameters were used by default.

The dataset of genotyped SNPs was stored in variant call format, which is the standard for storing DNA polymorphism data⁷⁴. The VCF was filtered to obtain only bi-allelic SNPs and SNPs with a minor allele frequency greater than 0.01. The NGSEP VCFFilter command was used with tags -s, -minMAF with a value of 0.01, -q of 40, and -m with a value of 50 to retain only SNPs genotyped in at least 50 individuals. Multiallelic SNVs were not taken into account because they were only 592 (10.17% of the total), and the allele frequency of the third allele was on average lower than 0.04. Diversity statistics were obtained from the filtered VCF using the NGSEP commands VCFSummaryStats and VCFDiversityStats. Also, for each locality, a VCF file was obtained from the filtered VCF using the NGSEP VCFFilter command and the -saf tag, which receives a file with the IDs of the

samples for each locality. These VCFs kept the bi-allelic SNPs, and the MAF filter was not applied to keep the private allele for each locality.

Population genetics and clustering

The SNPs were analyzed for the Hardy-Weinberg Equilibrium with the test proposed by Wigginton et al.,⁷⁵ implemented in the vcftools package⁷⁴. The raw p-values were corrected for multiple testing using the Bonferroni method. So, the expected heterozygosity (H_e), observed heterozygosity (H_o), and the Weir and Cockerham⁷⁶ fixation indexes F_{st} , F_{is} , and F_{it} were determined using the Adegnet⁷⁷, the Hierfstat⁷⁸, and Pegas packages. The p-values for F_{st} pairwise were estimated using the gl.fst.pop function of the dartR package. For this, we implemented a bootstrap of 10000 and a confidence interval of 95%. The nucleotide diversity (π) was estimated using vcftools with a window of 10,000,000 bp, and the Tajima' D index was calculated with the SAMBAR_v1.10 package^{79,80}. The discriminant analysis principal component (DAPC) was carried out using the DAPC cross-validation function from Adegnet R packages where 30 principal components (PCs) were retained to achieve the lowest mean squared error (MSE), and the plot was performed using the scatter.dapc function. The gene flow analysis was estimated using the divMigrate function of diveRsity R packages⁸¹. All plots were performed using the ggplot2 package⁸².

The population structure analysis between localities was analyzed using STRUCTURE with the admixture model⁸³. The VCF file was transformed into a STRUCTURE format by the NGSEP VCFConverter command with the -structure tag. In the mainparams file, the burn-in parameter was modified at 20,000 and the numreps parameter at 50,000 sampling iterations. Finally, STRUCTURE was run to determine the optimal number of populations or genetic clusters (K) with K values ranging from 1 to 6, with 20 replicates for each K. The best K was estimated using the Evanno et al. method⁸⁴, and the snmf function of the LEA R package⁸⁵. Additionally, population structure was estimated using the superparamagnetic clustering method which was integrated into the NetView R package⁸⁶.

Finally, the effective population size (N_e) was estimated using the linkage disequilibrium model (LD) implemented in NeEstimator⁸⁷. N_e was analyzed for the entire population (combining Buenaventura and Iscuandé), as well as for each locality. Additionally, the population history was inferred using the extended Bayesian skyline plot, following the tutorial from Trucchi et al., (2014)⁸⁸. Briefly, the VCF file was converted to NEXUS format using the script vcf2phyip.py⁸⁹. The extended Bayesian skyline plot was performed through BEAST2⁹⁰ using a substitution rate of 1.0, the clock rate of 1.0, MCMC chain length of 1,000,000, and the remaining parameters by default. Finally, the plot was done using ggplot2.

Data availability

The genome of this article is available in NCBI with the BioProject ID PRJNA997345.

Received: 26 July 2023; Accepted: 4 November 2024

Published online: 18 November 2024

References

- Arbeláez-Cortés, E. et al. Colombian frozen biodiversity: 16 years of the tissue collection of the Humboldt institute. *Acta Biológica Colombiana*. **20**, 163–173 (2015).
- Gutiérrez, M. P., Correa García, D., Zárate Ospina, M. F., Noriega Gómez, M. P. & Acosta (2020). M. P. Informe región pacífica.
- Moreno-Cavazos, M. P., Soto-Medina, E. A., Torres-González, A. M. & Llano-Almario, M. Patrones altitudinales de las comunidades vegetales en el Chocó biogeográfico del Valle del Cauca, Colombia. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* **42**, 269–279 (2018).
- Pérez-Escobar, O. A. et al. The origin and diversification of the hyperdiverse flora in the Chocó biogeographic region. *Front. Plant. Sci.* **10**, 1328 (2019).
- Cruz, R. & Borda, C. A. *Estado De explotación y pronóstico de la pesquería de Anadara tuberculosa* (Sowerby, 1833) en *El Pacífico Colombiano* (Revista de Investigaciones Marinas, 2003).
- Pérez, J. U. & Giraldo, L. E. U. Gestión Ambiental De Los ecosistemas de manglar. Aproximación Al Caso Colombiano. *Gestión Y Ambiente* **12**, (2009).
- Lucero, C. H., Kintz, C., Gil-Agudelo, D. L. & J. R. & Hermafroditismo en Los bivalvos *Anadara tuberculosa* y *Anadara similis* Sowerby 1883 (Arcidae) en Los manglares del Pacífico colombiano. *Boletín De Investigaciones Marinas Y Costeras - INVEMAR*. **50**, 163–170 (2021).
- Espinosa, S., Delgado, M. F., Orobio, B., Mejía, L. M. & Gil, D. L. Estado de población y valoración de lagunas estrategias de conservación del recurso piangua *Anadara tuberculosa* (sowerby) en sectores de Bazán y Nerete, costa pacífica Nariñense de Colombia. *BIM* **39**, (2016).
- Lucero, C., Cantera, J. & Neira, R. Pesquería Y Crecimiento De La Piangua (Arcoida: Arcidae) *Anadara tuberculosa* en la Bahía de Málaga del pacífico colombiano, 2005–2007. *Revista De Biología Trop.* **60**, 203–217 (2012).
- Ardila, N., Navas, G. & Reyes, J. *Libro Rojo de Invertebrados Marino De Colombia*. (INVEMAR. Ministerio Del Medio Ambiente (La serie Libros rojos de especies amenazadas de Colombia, 2002).
- Kilian, A. et al. Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol. Biol.* **888**, 67–89 (2012).
- Parra-Salazar, A., Gomez, J., Lozano-Arce, D., Reyes-Herrera, P. & Duitama, J. Robust and efficient software for reference-free genomic diversity analysis of genotyping-by-sequencing data on diploid and polyploid species. (2021).
- Shafer, A. B. A. et al. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol. Evol.* **8**, 907–917 (2017).
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
- Liao, X. et al. Current challenges and solutions of de novo assembly. *Quant. Biol.* **7**, 90–109 (2019).
- Kong, W., Wang, Y., Zhang, S., Yu, J. & Zhang, X. Recent advances in assembly of plant complex genomes. *Genom. Proteom. Bioinform.* <https://doi.org/10.1016/j.gpb.2023.04.004> (2023).
- Bai, C. M. et al. Chromosomal-level assembly of the blood clam, *Scapharca* (*Anadara*) *broughtonii*, using long sequence reads and Hi-C. *GigaScience* **8**, giz067 (2019).

18. Teng, W. et al. Chromosome-level genome assembly of *Scapharca kagoshimensis* reveals the expanded molecular basis of heme biosynthesis in ark shells. *Mol. Ecol. Resour.* **22**, 295–306 (2022).
19. Cardinale, B. J. et al. Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
20. Schratzberger, M. & Somerfield, P. J. Effects of widespread human disturbances in the marine environment suggest a new agenda for meiofauna research is needed. *Sci. Total Environ.* **728**, 138435 (2020).
21. Willi, Y. et al. Conservation genetics as a management tool: the five best-supported paradigms to assist the management of threatened species. *Proc. Natl. Acad. Sci. U S A.* **119**, e2105076119 (2022).
22. Bertolotti, A. C. et al. The structural variation landscape in 492 Atlantic salmon genomes. *Nat. Commun.* **11**, 5176 (2020).
23. Hu, B., Tian, Y., Li, Q. & Liu, S. Genomic signatures of artificial selection in the Pacific oyster, *Crassostrea gigas*. *Evol. Appl.* **15**, 618–630 (2022).
24. Kon, T. et al. Whole-genome resequencing of large yellow croaker (*Larimichthys crocea*) reveals the population structure and signatures of environmental adaptation. *Sci. Rep.* **11**, 11235 (2021).
25. Shafer, A. B. A. et al. Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* **30**, 78–87 (2015).
26. Meirmans, P. G. Subsampling reveals that unbalanced sampling affects structure results in a multi-species dataset. *Heredity* **122**, 276 (2019).
27. Puechmille, S. J. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol. Ecol. Resour.* **16**, 608–627 (2016).
28. Totikov, A. et al. Chromosome-level genome assemblies expand capabilities of genomics for conservation biology. *Genes (Basel)* **12**, 1336 (2021).
29. Peñaloza, C. et al. A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *Gigascience* **10**, giab020 (2021).
30. Martellosi, J. et al. Multiple and diversified transposon lineages contribute to early and recent bivalve genome evolution. *BMC Biol.* **21**, 145 (2023).
31. Thomas-Bulle, C. et al. Mollusc genomes reveal variability in patterns of LTR-retrotransposons dynamics. *BMC Genom.* **19**, 821 (2018).
32. Gozashti, L. et al. Transposable elements drive intron gain in diverse eukaryotes. *Proceedings of the national academy of sciences* **119**, e2209766119 (2022).
33. Wright, L. I., Tregenza, T. & Hosken, D. J. Inbreeding, inbreeding depression and extinction. *Conserv. Genet.* **9**, 833–843 (2008).
34. Oosting, T. et al. Mitochondrial genomes reveal mid-pleistocene population divergence, and post-glacial expansion, in Australasian snapper (*Chrysophrys auratus*). *Heredity* **130**, 30–39 (2023).
35. Othmen, A. B. et al. Lack of mitochondrial genetic structure in the endangered giant clam populations of *Tridacna maxima* (Bivalvia: Cardiidae: Tridacninae) across the Saudi Arabian coast. *Acta Oceanol. Sin.* **39**, 28–37 (2020).
36. Boissin, E. et al. Chaotic genetic structure and past demographic expansion of the invasive gastropod *Tritia Neritea* in its native range, the Mediterranean Sea. *Sci. Rep.* **10**, 21624 (2020).
37. Thomas, L. & Miller, K. J. High gene flow in the silverlip pearl oyster *Pinctada maxima* between inshore and offshore sites near Eighty Mile Beach in Western Australia. *PeerJ* **10**, e13323 (2022).
38. Wright, A. D., Garrison, N. L., Williams, A. S., Johnson, P. D. & Whelan, N. V. Range reduction of Oblong Rocksnail, *Leptoxis compacta*, shapes riverscape genetic patterns. *PeerJ* **8**, e9789 (2020).
39. Hoarau, G. et al. Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (*Pleuronectes platessa* L.). *Proc. Royal Soc. B: Biol. Sci.* **272**, 497–503 (2005).
40. O'Hare, J. A., Momigliano, P., Raftos, D. A. & Stow, A. J. Genetic structure and effective population size of Sydney rock oysters in eastern Australia. *Conserv. Genet.* **22**, 427–442 (2021).
41. Haltiner, L., Spaak, P., Dennis, S. R. & Feulner, P. G. D. Population genetic insights into establishment, adaptation, and dispersal of the invasive quagga mussel across perialpine lakes. *Evol. Appl.* **17**, e13620 (2023).
42. Smith, C. H., Johnson, N. A., Robertson, C. R., Doyle, R. D. & Randklev, C. R. Establishing conservation units to promote recovery of two threatened freshwater mussel species (Bivalvia: Unionida: Potamilus). *Ecol. Evol.* **11**, 11102–11122 (2021).
43. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**, 783–796 (2009).
44. Kardos, M. et al. The crucial role of genome-wide genetic variation in conservation. *PNAS Direct Submiss.* <https://doi.org/10.1073/pnas.2104642118> (2021).
45. Hollenbeck, C. M. & Johnston, I. A. Genomic tools and selective breeding in molluscs. *Front. Genet.* **9**, 253 (2018).
46. Fang, J., Xu, C. & Li, Q. Transcriptome analysis of inbreeding depression in the Pacific oyster *Crassostrea gigas*. *Aquaculture* **557**, 738314 (2022).
47. Hauser, L., Adcock, G. J., Smith, P. J., Bernal Ramírez, J. H. & Carvalho, G. R. Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proc. Natl. Acad. Sci.* **99**, 11742–11747 (2002).
48. Vendrami, D. L. J. et al. Sweepstakes reproductive success and collective dispersal produce chaotic genetic patchiness in a broadcast spawner. *Sci. Adv.* **7**, eabj4713 (2021).
49. Ye, Y. Y., Wu, C. W. & Li, J. J. Genetic population structure of *Macridiscus multifarius* (Mollusca: Bivalvia) on the basis of mitochondrial markers: strong population structure in a species with a short planktonic larval stage. *PLoS One* **10**, e0146260 (2015).
50. Diringer, B. et al. Genetic structure, phylogeography, and demography of *Anadara tuberculosa* (Bivalvia) from East Pacific as revealed by mtDNA: implications to conservation. *Ecol. Evol.* **9**, 4392–4402 (2019).
51. Selkoe, K. & Toonen, R. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Mar. Ecol. Prog. Ser.* **436**, 291–305 (2011).
52. Yu, H., Gao, S., Chen, A., Kong, L. & Li, Q. Genetic diversity and population structure of the ark shell *Scapharca broughtonii* along the coast of China based on microsatellites. *Biochem. Syst. Ecol.* **58**, 235–241 (2015).
53. Kerry, R. G. et al. An overview of remote monitoring methods in biodiversity conservation. *Environ. Sci. Pollut. Res. Int.* **29**, 80179–80221 (2022).
54. Lotze, H. K. Marine biodiversity conservation. *Curr. Biol.* **31**, R1190–R1195 (2021).
55. Johnson, R. N. et al. Adaptation and conservation insights from the koala genome. *Nat. Genet.* **50**, 1102–1111 (2018).
56. Katsanevakis, S. et al. Twelve recommendations for advancing marine conservation in European and contiguous seas. *Front. Mar. Sci.* **7**, (2020).
57. Locke, H. et al. Three global conditions for biodiversity conservation and sustainable use: an implementation framework. *Natl. Sci. Rev.* **6**, 1080–1082 (2019).
58. van Oppen, M. J. H. & Coleman, M. A. Advancing the protection of marine life through genomics. *PLoS Biol.* **20**, e3001801 (2022).
59. Duarte, C. M. et al. *Rebuilding Mar. life Nat.* **580**, 39–51 (2020).
60. Wu, Q., Zang, F., Ma, Y., Zheng, Y. & Zang, D. Analysis of genetic diversity and population structure in endangered *Populus wulianensis* based on 18 newly developed EST-SSR markers. *Global Ecol. Conserv.* **24**, e01329 (2020).
61. Gaither, M. R., Szabó, Z., Crepeau, M. W., Bird, C. E. & Toonen, R. J. Preservation of corals in salt-saturated DMSO buffer is superior to ethanol for PCR experiments. *Coral Reefs* **30**, 329–333 (2011).
62. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).
63. Chen, Y. et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat. Commun.* **12**, 60 (2021).

64. Ghurye, J., Pop, M., Koren, S., Bickhart, D. & Chin, C. S. Scaffolding of long read assemblies using long range contact information. *BMC Genom.* **18**, 527 (2017).
65. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
66. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
67. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
68. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**, 491 (2011).
69. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
70. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
71. Chakraborty, S., Saha, A. & Neelavar Ananthram, A. Comparison of DNA extraction methods for non-marine molluscs: is modified CTAB DNA extraction method more efficient than DNA extraction kits? *3 Biotech* **10**, 69 (2020).
72. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
73. Tello, D. et al. NGSEP3: accurate variant calling across species and sequencing protocols. *Bioinformatics* **35**, 4716–4723 (2019).
74. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
75. Wigginton, J. E., Cutler, D. J. & Abecasis, G. R. A note on exact tests of Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* **76**, 887–893 (2005).
76. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the analysis of Population structure. *Evolution* **38**, 1358–1370 (1984).
77. Jombart, T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
78. Goudet, J. Hierfstat, a package for r to compute and test hierarchical F-statistics. (2004).
79. de Jong, M. J., de Jong, J. F., Hoelzel, A. R. & Janke, A. SambaR: an R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. *Mol. Ecol. Resour.* **21**, 1369–1379 (2021).
80. Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C. & Georges, A. dartR v2: an accessible genetic analysis platform for conservation, ecology and agriculture. *Methods Ecol. Evol.* **13**, 2150–2158 (2022).
81. Keenan, K. et al. An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* **4**, 782–788 (2013).
82. Wickham, H. *Ggplot2* (Springer International Publishing, 2016).
83. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
84. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
85. Frichot, E. & François, O. L. E. A. An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**, 925–929 (2015).
86. Steinig, E. J., Neuditschko, M., Khatkar, M. S., Raadsma, H. W. & Zenger, K. R. Netview p: a network visualization tool to unravel complex population structure using genome-wide SNPs. *Mol. Ecol. Resour.* **16**, 216–227 (2016).
87. Do, C. et al. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (ne) from genetic data. *Mol. Ecol. Resour.* **14**, 209–214 (2014).
88. Trucchi, E. et al. King penguin demography since the last glaciation inferred from genome-wide data. *Proc. Biol. Sci.* **281**, 20140528 (2014).
89. Ortiz, E. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. (2019).
90. Bouckaert, R. et al. BEAST 2: a Software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).

Acknowledgements

We thank the Ministerio de Ciencia, Tecnología e Innovación (MinCiencias) of Colombia, which funded this project under the registration code 75729, and also the Autoridad Nacional de Licencias Ambientales (ANLA) for allowing us to collect the specimens according to the resolution number IDB0359. Finally, the communities of the Colombian Pacific, mainly Buenaventura, and Iscuandé, for their cooperation.

Author contributions

LF, MGS, MMZ, and SR conceived and designed the study. LF, MGS, and JD performed experiments and data analysis. LF, JD, and SR contributed to data interpretation. LF drafted the manuscript. MGS, MMZ, PJ, JD, and SR critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-78869-3>.

Correspondence and requests for materials should be addressed to S.R.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024