The native mussel *Mytilus chilensis* genome reveals adaptative molecular signatures facing the marine environment

Cristian Gallardo-Escárate[1*], Valentina Valenzuela-Muñoz[1], Gustavo Nuñez-Acuña[1], Diego Valenzuela-Miranda[1], Fabian Tapia[1], Marco Yévenes[2], Gonzalo Gajardo[2], Jorge E. Toro[3], Pablo A. Oyarzún[4], Gloria Arriagada[5,6], Beatriz Novoa[7], Antonio Figueras[7], Steven Roberts[8], Marco Gerdol[9]

[1]Interdisciplinary Center for Aquaculture Research, University of Concepción, Concepción, Chile.
[2]Laboratorio de Genética, Acuicultura & Biodiversidad, Departamento de Ciencias Biológicas y Biodiversidad, Universidad de Los Lagos, Osorno, Chile.
[3]Facultad de Ciencias, Instituto de Ciencias Marinas y Limnológicas (ICML), Universidad Austral de Chile, Valdivia, Chile.
[4]Centro de Investigación Marina Quintay (CIMARQ), Universidad Andres Bello, Quintay, Chile.
[5]Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad Andrés Bello, Chile.
[6]FONDAP Center for Genome Regulation, Santiago, Chile.
[7]Instituto de Investigaciones Marinas (IIM), Consejo Superior de Investigaciones Científicas (CSIC), Vigo, Spain.
[8]School of Aquatic and Fishery Sciences (SAFS), University of Washington, Seattle, USA.
[9]Department of Life Sciences, University of Trieste, Trieste, Italy.

*Corresponding author: Cristian Gallardo-Escárate (crisgallardo@udec.cl)

ABSTRACT

The blue mussel *Mytilus chilensis* is a key socioeconomic species inhabiting the southern coast of Chile. This endemic marine mussel supports a booming aquaculture industry, which entirely relies on artificially collected seeds from natural beds that are translocated to a diverse physical-chemical ocean conditions´ for farming. Furthermore, mussel production is threatened by a broad range of microorganisms, pollution, and environmental stressors that eventually impact its survival and growth. Herein, understanding the genomic basis of the local adaption is pivotal to developing sustainable shellfish aquaculture. We present a high-quality reference genome of *M. chilensis*, which is the first chromosome-level genome for a Mytilidae member in South America. The assembled genome size was 1.93 Gb, with a contig N50 of 134 Mb. Through Hi-C proximity ligation, 11,868 contigs were clustered, ordered, and assembled into 14 chromosomes in congruence with the karyological evidence. The *M. chilensis* genome comprises 34,530 genes and 4,795 non-coding RNAs. A total of 57% of the genome contains repetitive sequences with predominancy of LTR-retrotransposons and unknown elements. Comparative genome analysis was conducted among *M. chilensis* and *M. coruscus*, revealing genic rearrangements distributed into the whole genome. Notably, Steamer-like elements associated with horizontal transmissible cancer were explored in reference genomes, suggesting putative phylogenetic relationships at the chromosome level in Bivalvia. Genome expression analysis was also conducted, showing putative genomic differences between two ecologically different mussel populations. Collectively, the evidence suggests that local genome adaptation can be analyzed to develop sustainable mussel production. The genome of *M. chilensis* provides pivotal molecular knowledge for the *Mytilus* complex evolution and will help to understand how climate change can impact mussel biology.

Keywords: Chromosome-level, comparative genomics, transcriptomics, *Mytilus chilensis*.

1. INTRODUCTION

The blue mussel *Mytilus chilensis* (Hupé, 1854) is an ecological and socioeconomic key species in Chile, that leads the national shellfish aquaculture. This endemic mussel species constitutes one of the main industries in mussel production worldwide (Gonzalez-Poblete, Hurtado, Rojo, & Norambuena, 2018; Uriarte, 2008). However, the success of mussel aquaculture production in Chile is threatened by a wide range of microorganisms (Detree, Nunez-Acuna, Roberts, & Gallardo-Escarate, 2016; Enriquez, Frosner, Hochsteinmintzel, Riedemann, & Reinhardt, 1992; Gray, Lucas, Seed, & Richardson, 1999), marine pollution (Blanc, Molinet, Subiabre, & Diaz, 2018) and the climate variability that can impact the larval settlement and growth of mussel populations (Harvell, 2002; Hüning, 2013; Vihtakari et al., 2013). To cope with those stressors, mussels and marine invertebrates produce two-component responses, a specific response to the stressor and a more general response involving immune and endocrine pathways (Cooper 1996). Impacts of multi-stressors have been predicted to have additive, synergetic or antagonistic effects on marine organisms' physiology (Crain, Kroeker, & Halpern, 2008a). These different effects are directly linked to the amount of time between the occurrence of stressors, their intensity, and the organism's ability to return to homeostasis before a new stressor occurs (Gunderson, Armstrong, & Stillman, 2016). Despite these predictions, meta-data analyses show that most of the studied multi-stressors had synergetic effects on organisms' physiology (Crain, Kroeker, & Halpern, 2008b). Notably, isolated effects on mussel's immune system of environmental stressors or pathogens infection have been extensively studied (Bibby, Widdicombe, Parry, Spicer, & Pipe, 2008; Lockwood, Sanders, & Somero, 2010; Malagoli, Casarini, Sacchi, & Ottaviani, 2007; Mitta et al., 2000; Pereiro, Moreira, Novoa, & Figueras, 2021; Rey-Campos, Novoa, Pallavicini, Gerdol, & Figueras, 2021; Romero, Novoa, & Figueras, 2022; Sendra et al., 2020). However, mussel's

immune response to the combination of stressors remains unexplored. The interplaying between the immune system and multi-environmental stressors such as ocean acidification, hypoxia, marine heatwaves, HABs, and pathogen infections requires physiological, cellular, and molecular tools that uncover the complexity of mussel biology. Here, high-quality genome assembly at the chromosome level is pivotal to driving the scientific community in this endeavor and mussels represent an outstanding model species. For instance, mussel species display morphologically conserved karyotypes, and there is no evidence of whole-genome duplication events (Perez-Garcia, Moran, & Pasantes, 2014). Compared with other bivalves, the reported mussel genomes share relatively large genome sizes characterized by high heterozygosity and expanded mobile elements (Du et al., 2017; McCartney et al., 2022; Uliano-Silva et al., 2017; S. Wang et al., 2017; Zhang et al., 2012). Unfortunately, these genome features challenge the assembly efforts to avoid genome fragmentation. Up to now, chromosome-level genome assembly for Mytilidae has only been reported for the congeneric species *M. coruscus* (Yang et al., 2021) and for the zebra mussel *Dreissena polymorpha* (McCartney et al., 2022); meanwhile, other members of the Family have been reported as highly contiguous reference assemblies at contig level (Murgarella et al., 2016; Renaut et al., 2018; Xu et al., 2017). Interestingly, the presence-absence variation (PAV) phenomenon has recently been reported for *M. galloprovincialis*, where a pan-genome composed of 20,000 protein-coding genes was observed in conjunction with dispensable genes that are entirely missing in some mussels (Gerdol et al., 2020).

The endemic Chilean blue mussel *M. chilensis*, a close relative of the *M. edulis* species complex of the northern hemisphere (Gaitan-Espitia, Quintero-Galvis, Mesas, & D'Elia, 2016; Larrain, Zbawicka, Araneda, Gardner, & Wenne, 2018), represents an iconic species to explore key questions in ecology (Curelovich, Lovrich, Cueto, & Calcagno, 2018), ecophysiology (Duarte

et al., 2018) and adaptative genomics (Yevenes, Nunez-Acuna, Gallardo-Escarate, & Gajardo, 2021, 2022). It is a keystone taxon in the ecosystem regulating phytoplankton and nutrient flow and contributes to remineralizing organic deposits in the sediment (Hargrave, Doucette, Cranford, Law, & Milligan, 2008; Vinagre, Mendonca, Narciso, & Madeira, 2015). It inhabits rocky substrates in the intertidal and subtidal zones along the southern Pacific Ocean from latitude 38S to 53S (Flores et al., 2015). As a gonochoric species with an annual gametogenic cycle, sexual maturity occurs in spring-summer, where planktonic larvae can drift between 20 and 45 days before settlement (Toro, Ojeda, & Vergara, 2004). Dispersal potential has been estimated to be up to 30 km, allowing different degrees of gene flow among mussel populations (Astorga, Vargas, Valenzuela, Molinet, & Marin, 2018).

In this study, PacBio sequencing, and Hi-C scaffolding technology were jointly used to assemble the first chromosome-level reference genome of *M. chilensis*. Moreover, we conducted a comparative genomics study among reported genome mussel's species and analyzed the molecular signatures in two mussel populations facing distinct physical-chemical ocean conditions. Genomic features revealed putative chromosome rearrangements among mussel species, suggesting phylogenetic relationships for Steamer-like elements in Mytilidae. Candidate genes and single nucleotide polymorphisms were also associated with local adaptation of *M. chilensis*, revealing specific-transcriptome profiles associated with metabolism and immune-related genes. The knowledge gained in this study will provide pivotal information to explore how the marine environment drives phenotypic plasticity, that in turn, reveals genome adaptation signatures in mussel populations.

2. MATERIAL AND METHODS

2.1. Sample collection, NGS libraries, and sequencing

Adult specimens of *M. chilensis* were collected from a natural bed located in Puerto Mont (41º48'S-73º5'W), Chile (Figure 1A). Five mussels were selected for whole-genome sequencing using 1 mL of hemolymph collected from each specimen to reduce the heterozygosity or the number of individuals per pool. The samples were centrifuged at 1200 RPM to isolate the hemocyte cells and preserved by liquid nitrogen. High-quality DNA was isolated using the Qiagen DNA purification kit (QIAGEN, Germantown, MD, USA) following the manufacturer′s instructions and quantified by TapeStation 2200 instrument (Agilent, USA). DNA samples >9.5 in DIN numbers were selected for library preparation. Furthermore, 50 individuals were sampled from Cochamó (41º28'S-72ª18'W) and Yaldad (43º07'S-73º44'W), Southern Chile, to isolate RNA and explore molecular signatures associated with the local genome adaptation. Herein, these mussel populations inhabit contrasting ocean variability characterized by an estuary with continuous input of freshwater and vertical stratification and a bay exposed to open sea influence, respectively. Over the past two decades, the temporal and spatial variability of Sea Surface Temperature (SST) around Puerto Montt, Chiloé island and at sites Yaldad and Cochamó were analyzed using satellite images. Data on sea surface temperature in the region of interest were obtained from MUR-SST (Multi-scale Ultra-High-Resolution SST) distributed by NOAA through its ERDDAP platform. The MUR-SST images have a spatial resolution of 1 km and a temporal resolution of 1 day. *In situ* measurements for temperature (ºC) and salinity (PSU) seawater were obtained for both locations between 2018 and 2019. The raw environmental data were collected from the CHRONOS database, managed by Instituto de Fomento Pesquero, IFOP (Chile).

Samples were prepared according to the SMRTbell guide for sequencing on the PacBio Sequel II System. The genomic DNA isolated from 5 individuals collected from Puerto Montt was sequenced using SMRT Sequencing according to the manufacturer's protocols. SMRT sequencing yielded 882.1 Gigabases and 63 million long reads from 2 HiFi SMRT cells. The subreads N50 and average read lengths were 14,665 and 14,535 bp, respectively. The total HiFi reads yielded 3.7 million with an average quality of Q36 and Q35, respectively. Hi-C libraries were constructed from hemocyte cells using Phase Genomics' Animal Hi-C kit and sequenced on Illumina's Hiseq4000 platform to yield 253 million reads using the same DNA isolated for PacBio sequencing. Short-read sequencing libraries were prepared using an insert size of 150 bp obtained from 1 μg of genomic DNA after fragmentation, end-paired, and ligation to adaptors. The ligated fragments were fractionated on agarose gels and purified by PCR amplification to produce sequencing libraries. The method applied was like that previously published by Lieberman-Aiden et al. (2009). The PacBio and Hi-C Illumina DNA raw data were deposited in the Sequence Read Archive (SRA) repository, accession number SRR20593343 and SRR20966976, respectively.

Moreover, RNA libraries were constructed from hemocytes, digestive gland, gill, and mantle tissues for transcriptome sequencing to obtain whole-transcriptome profiling from the same mussels used for genome DNA sequencing. Additionally, twelve available Sequence Read Archive (SRA) transcriptomic data (GenBank accession number SRP261955), representing gills and mantle tissues collected from individuals of Cochamó and Yaldad mussel populations (Yévenes et al., 2021, 2022), were incorporated to analyze population-specific transcriptome profiles. These twelve transcriptomic SRA data represent to total RNA extracted by the Trizol reagent method (Invitrogen, USA) from 15 individuals. Three biological replicates (5 individual total RNA extractions each one) from each mussel population were analyzed. The quality and integrity of

163 extracted RNAs were measured in a TapeStation 2200 instrument (Agilent, USA), using the R6K
164 Reagent Kit based on the manufacturer's instructions. RNA samples >9 in RIN numbers were
165 selected for high-quality libraries preparation using TrueSeq Stranded mRNA LT Sample Prep Kit
166 and sequenced in HiSeq 4000 (Illumina, USA).
167
168 **2.2. *De novo* genome assembly and Hi-C scaffolding of *M. chilensis***
169 Two HiFi single-molecular real-time cells in the PacBio Sequel platform yielded 53.8 Gb of high-
170 quality DNA genome information. These long reads were assembled with the Hifiasm package
171 using default parameters (Cheng, Concepcion, Feng, Zhang, & Li, 2021). For Hi-C scaffolding,
172 reads were aligned to the primary draft assembly, also following the manufacturer's instructions
173 (Phase-Genomics, 2019). Briefly, reads were aligned using BWA-MEM (H. Li & Durbin, 2010)
174 with the –5SP and –t 8 options specified, and all other options default. The package
175 SAMBLASTER (Faust & Hall, 2014) was used to flag duplicates and then excluded for further
176 analysis. Sequence alignments were filtered with SAMtools (Danecek et al., 2021; H. Li et al.,
177 2009) using the –F 2304 filtering flag to remove non-primary and secondary alignments. This step
178 was conducted to remove alignment errors, low-quality alignments, and other alignment noise due
179 to repetitiveness, heterozygosity, and other ambiguous assembled sequences. Finally, Phase
180 Genomics' Proximo Hi-C genome-scaffolding platform was used to create chromosome-scale
181 scaffolds from the corrected assembly, according to Bickhart et al. (2017).
182
183 **2.3. Karyotype of *M. chilensis***
184 Metaphase plates of 24 hours post-fertilization larvae were used to obtain chromosomes from *M.*
185 *chilensis*, according to Gallardo-Escárate et al. (2004). Briefly, antimitotic treatment with

Colchicine 0.05% solution was applied for 4 hours. Then the larvae were rinsed in clean seawater and immersed in a hypotonic solution (seawater: distilled water, 1:1) for 30 min. Finally, the larvae were fixed in modified Carnoy solution (methanol: acetic acid, 3:1). Chromosome spreads were obtained by dissociating larva tissue in acetic acid (50%), pipetting suspension drops onto slides preheated at 43ºC and air-dried according to Amar et al. (2008). FISH experiment was performed to validate the physical localization of candidate genes. Here, 28S rDNA was labeled following methods previously published (Perez-Garcia et al., 2014). Briefly, metaphase preparations were denatured at 69°C for 2 min and hybridized overnight at 37°C. Signal detection was performed using fluorescein avidin and biotinylated anti-avidin for the biotinylated probes and mouse anti-digoxigenin, goat anti-mouse rhodamine, and rabbit anti-goat rhodamine for the digoxigenin-labeled probes. Fluorescent staining was carried out with 4,6-diamidino-2-phenylindole (DAPI) and mounted with Vectacshield antifading solution. Chromosome spreads were observed using an epifluorescent microscope Nikon Eclipse 80i equipped with a digital camera DS-5Mc.

2.4. Genome annotation of *M. chilensis*

Our repeat annotation pipeline applied a combined homology alignment strategy, and *de novo* search to identify the whole genome repeats. Tandem Repeat was extracted using TRF (http://tandem.bu.edu/trf/trf.html) by *ab initio* prediction. The homolog prediction commonly used Repbase (www.girinst.org/repbase) database employing RepeatMasker (www.repeatmasker.org/) software and its in-house scripts (RepeatProteinMask) with default parameters to extract repeat regions. *Ab initio* prediction was used to build a de novo repetitive elements database by LTR_FINDER (http://tlife.fudan.edu.cn/ltr_finder/), RepeatScout (www.repeatmasker.org/), RepeatModeler (www.repeatmasker.org/RepeatModeler.html) with

default parameters. Then, all repeat sequences with lengths >100bp and gap 'N' less than 5% constituted the raw transposable element (TE) library. A custom library (a combination of Repbase and a custom *de novo* TE library processed by UCLUST to yield a non-redundant library) was supplied to RepeatMasker for DNA-level repeat identification.

The structural annotation approach was applied to incorporate *de novo*, homolog prediction, and RNA-Seq-assisted predictions to annotate gene models. For gene prediction based on *de novo*, Augustus (v3.2.3), Geneid (v1.4), Genescan (v1.0), GlimmerHMM (v3.04) and SNAP (2013-11-29) were used in our automated gene prediction pipeline. For homolog prediction, sequences of homologous proteins were downloaded from Ensembl/NCBI/others. Protein sequences were aligned to the genome using TblastN (v2.2.26; E-value ≤ 1e−5), and then the matching proteins were aligned to the homologous genome sequences for accurate spliced alignments with GeneWise (v2.4.1) software to predict gene structure contained in each protein region. Finally, for RNA-seq data, transcriptome reads assemblies were generated with Trinity (v2.1.1) for the genome annotation. For the genome annotation optimization, the RNA-Seq reads from different tissues were aligned to genome fasta using Hisat (v2.0.4)/TopHat (v2.0.11) with default parameters to identify exons region and splice positions. The alignment results were inputted for Stringtie (v1.3.3)/Cufflinks (v2.2.1) with default parameters for genome-based transcript assembly. The non-redundant reference gene set was generated by merging genes predicted by three methods with EvidenceModeler (EVM v1.1.1) using PASA (Program to Assemble Spliced Alignment) terminal exon support and including masked transposable elements as input into gene prediction. Individual families of interest were selected for further manual curation.

Gene functions were assigned according to the best match by aligning the protein sequences to the Swiss-Prot using Blastp (with a threshold of E-value ≤ 1e−5). The motifs and domains were annotated using InterProScan70 (v5.31) by searching against publicly available databases, including ProDom, PRINTS, Pfam, SMRT, PANTHER, and PROSITE. The Gene Ontology (GO) IDs for each gene were assigned according to the corresponding InterPro entry. We predicted the protein function by transferring annotations from the closest BLAST hit (E-value<1e−5) in the Swissprot database and DIAMOND (v0.8.22) / BLAST hit (E-value<1e−5) in the NR database. We also mapped the gene set to a KEGG pathway and identified the best match for each gene.

Non-coding RNA annotations such as tRNAs were predicted using the program tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/. Since rRNAs are highly conserved, we choose relative species' rRNA sequences as references and predict rRNA sequences using Blast. Other ncRNAs, including miRNAs and snRNAs, were identified by searching against the Rfam database with default parameters using the infernal software (http://infernal.janelia.org/). Additionally, lncRNAs were identified using the pipelines previously proposed (Gallardo-Escarate, Figueras, & Novoa, 2019; Pereiro et al., 2021; Tarifeno-Saldivia, Valenzuela-Miranda, & Gallardo-Escarate, 2017).

2.5. Comparative genomics between *M. chilensis* and *M. coruscus*

Syntenic relationships were carried out among mussel species for which chromosome-level reference genomes are publicly available. Here, the analysis was performed between the two congeneric species *M. chilensis* (this study) and *M. coruscus* (Yang et al., 2021), where gene annotations were explored by MCScanX (Y. P. Wang et al., 2012) implemented in the TBtools

package (Chen et al., 2020). This approach detects groups of orthologous genes and compares their arrangement to identify colinear segments in the compared genomes. MCScanX was used to discover microsynteny relationships, focusing on the local arrangement of genes near the syntenic blocks. The microsynteny arrangement of genes identified by MCScanX was evaluated through GO analysis to identify the primary molecular function and biological processes enrichened for each genomic region where macromutations or chromosome rearrangements were detected.

Disseminated neoplasia is a disease horizontally transmitted by clonal cancer cells, which causes leukemia in mollusk bivalves (Metzger, Reinisch, Sherry, & Goff, 2015; Metzger et al., 2016). The neoplastic cells gradually replace normal hemocytes leading to increased mortality, and it has been detected in 15 species of marine bivalve mollusks worldwide (Elston, Moore, & Brooks, 1992). Notably, disseminated neoplasia has been observed among mussel species with varying epizootic prevalences. For instance, *M. trossulus* has shown high prevalences in some areas, whereas in *Mytilus edulis*, the prevalences are generally lower. Furthermore, *M. galloprovincialis* has been suggested as a resistant species to the disease in Spain and Italy's mussel populations (Ciocan & Sunila, 2005). This observation extends the relevance to exploring mussel species' genetic features associated with disseminated neoplasia. Herein, the molecular characterization of steamer-like elements in *M. chilensis* was conducted by cloning and walking primer method according to Arriagada et al. (2014). The putative *M. chilensis* Steamer-like was scanned through twelve reference genomes assembled at chromosome level for Bivalvia: *Mytilus coruscus* (GCA_017311375.1), *Mytilus edulis* (GCA_019925275.1), *Dreissena polymorpha* (GCA_020536995.1), *Mercenaria mercenaria* (GCA_014805675.2), *Solen Grandis* (GCA_021229015.1), *Ruditapes philippinarum* (GCA_009026015.1), *Pecten maximus* (GCA_902652985.1), *Pinctada imbricata* (GCA_002216045.1), *Crassostrea gigas*

(GCA_902806645.1), *Crassostrea ariakensis* (GCA_020458035.1), and *Crassostrea virginica* (GCA_002022765.4). The putative Long Terminal Repeat (LTR) sequences were identified using BLAST search, where open reading frames (ORFs) between flanking LTRs were detected. The identified Steamer-like elements were aligned using ClustalW and annotated based on NCBI Conserved Domain search. Amino acid sequences for the full-length Gag-Pol polyprotein region were aligned among the studied bivalve species. The Steamer element was reported for *Mya arenaria* (Accession AIE48224.1) and *M. chilensis* (this study). DNA sequence genealogy analysis was conducted to investigate horizontal transmission events among bivalve species. The maximum likelihood (ML) method was conducted on the SLEs loci localized in all the publicly available bivalve genomes assembled at the chromosome level.

2.6. Whole-genome transcript expression analysis in two *M. chilensis* populations

The transcriptome of mussels collected from Yaldad and Cochamó populations were analyzed using an hierarchical clustering approach to detect transcriptional similarities among tissues/populations. The transcripts that were differentially expressed in comparison to normalized expression values were visualized in a clustering heat map and selected according to the identified cluster. For an optimal comparison of the results, k-means clustering was performed to identify candidate genes involved in specific gene expression patterns. The distance metric was calculated with the Manhattan method, where the mean expression level in 5-6 rounds of k-means clustering was subtracted.

Moreover, raw data from mussel tissues collected from Yaldad and Cochamó populations were trimmed and mapped to the *M. chilensis* genome using CLC Genomics Workbench v22 software (Qiagen Bioinformatics, USA). Threshold values for transcripts were calculated from the coverage analysis using the Graph Threshold Areas tool in CLC Genomics Workbench v22

300 software. Here, an index denoted as Chromosome Genome Expression (CGE) was applied to
301 explore the whole-genome transcript expression profiling, according to Valenzuela et al. (2022).
302 The CGE calculates the mean coverage of transcripts mapped into a specific chromosome region,
303 comparing mussel populations and tissues. Specifically, the CGE index represents the percentage
304 of the transcriptional variation between two or more RNA-seq data for the same locus. The
305 transcript coverage values for each dataset were calculated using a threshold of 20,000 to 150,000
306 reads, where a window size of 10 positions was set to calculate and identify chromosome regions
307 differentially transcribed. This approach was used to visualize chromosome regions actively
308 transcribed, identifying genes differentially expressed and observing tissue-specific patterns in the
309 evaluated mussel populations. Finally, the threshold values for each dataset and CGE index were
310 visualized in Circos plots (M. I. Krzywinski et al., 2009).

311 RNA-seq data analyses were carried out using the raw sequencing reads and mapped on
312 the assembled genome by CLC Genomics Workbench v22 software (Qiagen Bioinformatics,
313 USA) for each tissue/population separately. In parallel, *de novo* assembling was performed to
314 evaluate PAVs and dispensable genes affecting the *in-silico* transcription analysis. The assembly
315 was performed with overlap criteria of 70% and a similarity of 0.9 to exclude paralogous sequence
316 variants. The settings used were set as mismatch cost = 2, deletion cost = 3, insert cost = 3,
317 minimum contig length = 200 base pairs, and trimming quality score = 0.05 using CLC Genomics
318 Workbench v22. After assembly, the contigs generated for each data set were mapped on the genes
319 annotated in the reference genome to evaluate de genome coverage and detect PAV features. The
320 analysis did not show bias putatively associated with PAVs between the analyzed mussel
321 populations. Then, mRNA sequences annotated for the *M. chilensis* genome were used to evaluate
322 the transcription level between mussel populations, where differentially expression analysis was

set with a minimum length fraction = 0.6 and a minimum similarity fraction (long reads) = 0.5. The obtained genes from each tissue/population were blasted to CGE regions to enrich the number of transcripts evaluated by RNA-Seq analysis. In addition, sequences were extracted near the threshold areas in a window of 10 kb for each transcriptome. The expression value was set as transcripts per million model (TPM). The distance metric was calculated with the Manhattan method, with the mean expression level in 5-6 rounds of k-means clustering subtracted. Finally, Generalized Linear Model (GLM) available in the CLC software was used for statistical analyses and to compare gene expression levels in terms of the $\log_2$ fold change ($p = 0.005$; FDR corrected).

Moreover, the innate immunity in marine invertebrates may play an important role in speciation and environmental adaptation (Ellis et al., 2011; Rolff & Siva-Jothy, 2003). Herein, we investigate the immune-related genes associated with the TOLL-like receptor (TLR) and Apoptosis signaling pathways due that the functional annotation revealed thar they were mainly enriched between the mussel populations analyzed. In addition, bioinformatic analyses were carried out using the CLC Genomics Workbench software to mine single nucleotide variants (SNV) from the transcriptomes sequenced for Yaldad and Cochamó. Candidate SNVs were called with the following settings: window length = 11, maximum gap and mismatch count = 2, minimum average quality of surrounding bases = 15, minimum quality of central base = 20, maximum coverage = 100, minimum coverage = 8, minimum variant frequency (%) = 35.0, and maximum expected variations (ploidy) = 2. Furthermore, the genotypes of DEGs were also identified for detecting putative genetic variations between mussel populations. Here, singleton, dispersed, tandem, proximal, and whole genome duplication (WGD) gene events were evaluated using MCScanX. The amino acid changes and the zygosity proportions were also estimated in DEGs between Yaldad and Cochamó populations.

346

347 **2.7. GO enrichment analysis**

348 Differentially expressed mRNA were annotated through BlastX analysis using a custom protein
349 database constructed from GeneBank, KEGG, GO, and UniProtKB/Swiss-Prot. The cutoff E-value
350 was set at 1E-10. Transcripts were subjected to gene ontology (GO) analysis using the Blast2GO
351 plugins included in the CLC Genomics Workbench v22 software (Qiagen Bioinformatics, USA).
352 The results were plotted using the Profiler R package (Yu, Wang, Han, & He, 2012). GO
353 enrichment analysis was conducted to identify the most represented biological processes among
354 protein-coding genes proximally located to the CGE regions. The enrichment of biological
355 processes was identified using Fisher's exact test tool of Blast2GO among the different tissues and
356 mussel populations.

357

358 **3. RESULTS AND DISCUSSION**

359 **3.1. Chromosome genome assembly of *M. chilensis* using proximity ligation**

360 With two HiFi single-molecular real-time cells in the PacBio Sequel platform, we generated 53.8
361 Gb of high-quality DNA genome information. This data comprised 63 million reads with a total
362 length of 882 Giga base pairs (Table1). These long reads were assembled with the Hifiasm package
363 using default parameters (Cheng et al., 2021), yielding a primary assembly of 13,762 contigs
364 equivalent to 2.19 Gb, with an N50 of 206 Mb. The size genome assembly made by Hifiasm was
365 comparable with the previous genome size described for closely related species; 1.28 Gb for *M.*
366 *galloprovincialis* (Gerdol et al., 2020), 1.57 Gb for *M. coruscus* (Yang et al., 2021) and 1.79 Gb
367 for *Dreissena polymorpha* (McCartney et al., 2022).

368 *In vivo* Hi-C is a technique that maps physical DNA-DNA proximity across the entire
369 genome (Ghurye et al., 2019; Pal, Forcato, & Ferrari, 2019). The method was introduced as a
370 genome-wide version of its predecessor, 3C (Chromosome Conformation Capture). It has been a
371 powerful tool in chromosome-scale genome assembly of many animals in recent years (Lando,
372 Stevens, Basu, & Laue, 2018; Lin et al., 2018). In this study, Hi-C experiments and data analysis
373 on hemocyte cells were used for the chromosome assembly of the blue mussel *M. chilensis*. Here,
374 two Hi-C libraries were prepared and sequenced by Phase Genomics (Seattle, WA, USA), resulting
375 in ~20x coverage and ~253 million 150-bp paired-end reads (Table 1). The Hi-C analysis
376 evidenced that 44.68% of high-quality reads showed intercontig signals or Cis-close position
377 (<10kbp on the same contig), and an additional 4.09% of sequence reads revealed a Cis-far
378 conformation (>10Kbp on the same contig) (Table 2). Hi-C reads were aligned using Bowtie
379 (Langmead, Trapnell, Pop, & Salzberg, 2009) to order and orient the 13,762 contigs, and
380 scaffolding was performed using Proximo (Phase Genomics, Seattle, WA, USA). We then applied
381 Juicebox for visual inspection and manual correction (Robinson et al., 2018). We also manually
382 removed 1,894 scaffolds that were microbe-sized and disconnected from the rest of the assembly.
383 Herein, 11,868 contigs were used for the first chromosome-level high-quality *M. chilensis*
384 assembly (Table 3). The N50 and total genome length were calculated in 134 Mbp and 1,938 Gbps,
385 respectively. The *M. chilensis* genome provides a useful genomic resource for research in mussel
386 biology and for developing novel sustainable strategies in mussel aquaculture. The Hi-C data
387 generated 14 chromosomes assembled with HiFi consensus long DNA reads (Fig. 1B). The
388 cytogenetic analysis performed for *M. chilensis* revealed a conservative karyotype for the *Mytilus*
389 genus composed by 2n=14 (Perez-Garcia et al., 2014). Physical localization of 28S-rRNA revealed
390 two loci mapped in different submetacentric/subtelocentric chromosome pairs (Fig. 1C),

391 confirming the presence of major rDNA clusters subterminal to the long arms of two chromosome
392 pairs reported in *M. edulis* and *M. galloprovincialis* (Martínez-Lage, González-Tizón, & Méndez,
393 1995). Concerning genome assembly, the largest scaffold was assembled from 998 contigs in a
394 total size of 173.3 Mb. Meanwhile, the smallest scaffold was 117.3 Mb in length, consisting of
395 744 contigs (Table 3). Notably, the number of contigs in scaffolds were 11,868 (100% of all
396 contigs in chromosome clusters, 86.24% of all contigs) and 1.93 Gbps of genome size (100% of
397 all length in chromosome cluster, 88.43% of all sequence length). The completeness of genome
398 assembly was assessed by the single-copy ortholog set (BUSCO, V5.3.2) (Manni, Berkeley,
399 Seppey, Simao, & Zdobnov, 2021). The results showed the following BUSCO scores: i) Eukaryota
400 Odb10; C:94.1% [S:72.9%, D:21.2%], F:3.1%, M:2.8%, n:255. ii) Metazoa Odb10; C:95.1%
401 [S:75.5%, D:19.6%], F:2.5%, M:2.4%, n:954. iii) Mollusca Odb10, C:85% [S:70.1%, D:14.9%],
402 F:3.6%, M:11.4%, n:5295.

403 **3.2. Genome annotation of *M. chilensis***

404 The genome assembly was annotated using *de novo* and protein and transcript-guided methods
405 (Fig. 2A). The first step of the annotation process was to identify the DNA repeats through the *M.*
406 *chilensis* genome. Repetitive elements and non-coding genes in the blue mussel genome were
407 annotated by homologous comparison and *ab initio* prediction. RepeatMasker (Bedell, Korf, &
408 Gish, 2000) was used for homologous comparison by searching against the Repbase database (Bao,
409 Kojima, & Kohany, 2015) and RepeatModeler (Storer, Hubley, Rosen, Wheeler, & Smit, 2021).
410 According to these analyses, about 1.1 Gbps of repeat sequences were annotated, which accounted
411 for 56.73% of the whole genome. Herein, DNA transposons, LINE, and LTR transposable
412 elements were identified (Table 4). Useful genome information for population genetic studies is
413 the identification of simple sequence repeats (SSRs) or microsatellites. The mining of SSRs

414 revealed that the *M. chilensis* genome has 548,360 SSR sequences, where the 9% and 6% of the
415 SSR loci were annotated for each mussel's chromosome (Fig. 1S). The most frequent SSR motif
416 was the tetranucleotide, followed by the dinucleotides, accounting for the 206,103 and 197,700
417 repeats, respectively. The entire SSR sequences accounted for 0.35% of the whole genome. The
418 development of SSR markers offers a shortcut to assessing genetic diversity, which can potentially
419 be applied in food authentication and genetic traceability for mussel species (Ferreira et al., 2020;
420 Larrain, Diaz, Lamas, Uribe, & Araneda, 2014; Vidal, Penaloza, Urzua, & Toro, 2009).
421

422 **3.3. Protein-coding genes prediction and functional annotation in the *M. chilensis* genome**

423 For the identification of protein-coding genes, *de novo*, homolog prediction, and RNA-seq
424 evidence were used as the training set (Fig. 2A). For homologous predictions, the protein
425 sequences from *Crassostrea gigas*, *Mytilus galloprovincialis*, *M. coruscus*, and *Dreissena*
426 *polymorpha* genomes were extracted using the respectively published references and aligned
427 against the blue mussel genome using TBLASTN (e-value<1e-5). The gene sequence structure of
428 each candidate gene and previously mentioned tools were used to predict protein-coding genes.
429 Finally, a non-redundant reference gene set was generated using EvidenceModeler (EVM) and
430 PASA2 tools (Fig. 2A). Taken together, 34,530 protein-coding genes were identified with 6,531
431 bp of average transcript length, 1,377 bp of average CDS length, 4.92 of average of Exons per
432 Gene, and 1,377 and 1,316 of the average length of exons and introns, respectively (Table 6).
433 Additionally, 516 tRNAs were predicted using tRNAscan-SE, and 143 rRNA genes were
434 annotated using RNAmmer. For non-coding RNAs with putative regulatory roles, 1,365 miRNAs
435 and 43,011 long non-coding RNAs were identified and annotated within the *M. chilensis* genome
436 (Table 7). For functional annotation, the predicted proteins within the blue mussel genome were

437 searched by homology against seven databases: Swissprot, Nr, Nt, KEGG, eggnog, GO, and Pfam
438 (Fig. 2A). Overall, 70.45%, 73.01%, 8.98%, 64.94%, 80.57%, 33.61% and 176.33 % of genes
439 matched entries in these databases, respectively. A total of 34,530 genes (100%) were successfully
440 annotated by gene function and conserved protein motifs (Table 8). The genomic features
441 annotated for the native blue mussel *M. chilensis* were displayed using a circus plot (M.
442 Krzywinski et al., 2009). Herein, this graphical representation showed the primary genomic
443 features for the 14 chromosomes. Specifically, gene density, repeat density, GC content, rRNA
444 localization, and ncRNAs were plotted. The transcriptome expression profiles for the mantle, gills,
445 hemocytes, and digestive gland tissues were also displayed in connection with the syntenic blocks
446 (Fig. 2B).
447
448 3.4. **Comparative genomics**
449 Smooth-shelled blue mussels of the genus *Mytilus* represents a model group because of their
450 cosmopolitan distribution, socioecological importance, and their intriguing evolutionary history.
451 This taxon provides new insights into the process of speciation, and how the hybridization and
452 introgression can be one of the biggest threats to global mussel's biodiversity (Gardner, Oyarzun,
453 Toro, Wenne, & Zbawicka, 2021). Survey of single nucleotide polymorphisms (SNPs) on
454 Southern hemisphere blue mussels has provided a new layer for the understanding of their biology,
455 taxonomy and phylogeography (Araneda, Larrain, Hecht, & Narum, 2016; Nunez-Acuna &
456 Gallardo-Escarate, 2013). However, SNP markers cannot be applied as a single tool to evidence
457 chromosome rearrangements events during the *Mytilus* evolution. Here, whole-genome
458 sequencing in smooth-shelled blue mussels and relatives Bivalve species is a priority for global
459 mussel aquaculture, biosecurity and conservation.

With the aim to explore genomic rearrangements in *Mytilus*, the reported reference genome for *M. coruscus* and *M. chilensis* were analyzed. Of the 34,530 predicted genes from the *M. chilensis* genome, 18,758 (54.32%) were found in syntenic collinear blocks after being compared with the *M. coruscus* genome (Fig. 3A). These syntenic blocks consisted of 671 alignments with a minimum of 5 genes per block. The number of alignments per chromosome ranged from 27 on chromosome 13 to 69 on chromosome 3. Chromosomes with the higher number of genes in collinear blocks were chromosomes 1, 4, and 6, with 1,227, 1,091, and 1,088 genes, respectively. Blocks with less than five genes or E-value < 1E-5 were discarded from this analysis. Most collinear blocks were located at the same pair of chromosomes between the two genomes. For example, *M. chilensis* Chr1 had only syntenic blocks with LG01 from *M. coruscus* in the same order. However, chromosomes 6 and 10 from *M. chilensis* had collinearity with chromosomes LG09 and LG02, respectively, in *M. coruscus* but were orientated as two inversed blocks per pair of chromosomes (red lines in Fig. 3A and Fig. 2S). The genes in these alignments from inversed blocks were extracted, blasted, and Gene Ontology terms were identified. Enrichment analyses from GO terms were obtained from Chr10 and LG09 inversed blocks and Chr6 and LG02 pair of chromosomes (Fig. 3B-C). Most of the molecular functions enriched GO terms in the Chr10 and LG09 pair were associated with heat shock protein (HSP) binding. In contrast, in the Chr6 and LG02 pair, most of the enriched GO terms were associated with the mitochondria and biological processes related to autophagy or regulation of the gene expression by epigenetic changes. Notably, chromosome rearrangements have been associated with adaptative genetic traits in marine organisms (Kess et al., 2020), where specific architectural proteins such as HSPs may have distinct roles in establishing 3D genome organization (L. Li et al., 2015).

3.5. Comparative analysis of Steamer-like elements in Bivalvia

To explore the gene expansion of retrotransposon elements among representatives' species from Bivalvia, we primarily characterized the Steamer-like elements (SLEs) in *M. chilensis* using the approach described by Arriagada et al. (2014). The analysis evidenced that the genome of *M. chilensis* contains five copies of SLEs distributed in chromosomes 1, 6, 7, 10, and 11. The alignment showed that all SLEs copies are flanked by two LTRs (5' and 3') containing the Gag-Pol ORFs and the domains annotated to Protease, Reverse Transcriptase, RNAaseH, and Integrase. Notably, an insertion composed of 12 nucleotides at position 933^934 was exclusively found in chromosomes 7 and 11. The translation for the inserted nucleotides suggests four amino acids, K, T, S, and H, in positive orientation. However, the translation evaluated in the reading frame (-1) evidenced a methionine localized before the RNAaseH coding gene (Fig. 4A).

Furthermore, the phylogenetic analysis using publicly available reference genomes assembled at chromosome level for eleven bivalve species using Maximum Likelihood (ML) revealed six chromosomes cluster composed of bivalves belonging to the families Veneridae, Solenidae, Pectinidae, Ostreidae, Pteriidae, and Mytilidae (Fig. 4B). The phylogenetic reconstruction rooted SLEs found in three chromosomes (2, 4 and 17) from *R. philippinarum.* The other Veneridae member, *M. mercenaria* showed a cluster of four chromosomes (10, 12, 13, and 16) and related to two chromosomes of *S. Grandis* (10 and 16). This last species formed a unique cluster composed of three chromosomes (8, 15, and 17), similar to *P. maximus*, with three chromosomes. Concerning the mussel and oyster genomes assembled at the chromosome level, the phylogenic analysis revealed two main clusters composed of species belonging to Ostreida and Mytilidae, where the first taxon was comprised of the Ostreidae and Pteriidae families. Herein, one cluster was rooted with three SLE sequences from *C. virginica*, *C. gigas*, and *C. ariakensis*

located on chromosomes 9, 2, and 5, respectively. The second major cluster was composed of SLEs annotated in chromosomes from Ostreidae and Pteriidae, where *C. virginica* chromosomes were closely related to *P. imbricata*. The third cluster was observed containing three SLE sequences from *C. virginica* and *C. ariakensis*; chromosomes 1, 2, 8, and 1, 2, and 6, respectively (Fig. 4B). The analysis of the Mytilidae family revealed two primary clusters comprised of SLE located in chromosomes from *M. edulis and D. polymorpha*, and *M. coruscus*, *M. chilensis*, respectively (Fig. 4B). This last cluster grouped five chromosomes from *M. coruscus* (Chr. 1, 4, 5, 7 and 11), and two from *M. edulis* (Chr. 4 and 6). The Steamer-like sequenced characterized for *M. chilensis* was also observed in this cluster. Finally, a detailed analysis of the three mussel species reported with genome assemblies at the chromosome level was conducted (Fig. 4C). Notably, a rooted cluster comprising chromosomes 7, 11, 9 for *M. couscous*, and 4 and 6 for *M. edulis* were closely related. Herein, two primary clusters of SLEs located in chromosomes from *M. chilensis*, *M. edulis*, and *M. coruscus* were observed. The analysis suggested that the SLEs identified on *M. chilensis* chromosomes are closely related to the SLEs annotated on chromosomes 9 and 4 in *M. edulis*; meanwhile, the SLEs located in chromosomes 1, 5, and 11 in *M. coruscus* were also identified in the same chromosome cluster. The second main cluster observed was comprised exclusively of SLEs annotated in *D. polymorpha* chromosomes, except by the SLEs copies identified in chromosomes 9 and 10 of *M. edulis*. Interestingly, the SLEs annotated in chromosome 9 from *M. edulis* are shared among the three primary clusters analyzed, suggesting putative translocation gene events in Mytilidae. Overall, the phylogenic relationships of SLEs revealed that the reported bivalve genomes comprise between 3 to 6 loci. A lower number of SLEs was found in Solenidae, Pectinidae, and Veneridae, followed by Mytilidae. A higher number of SLE loci was observed in genomes belonging to the Ostreida order. As far as we know, the

529 evolution of the bivalve chromosomes has mainly been studied using cytogenetical techniques
530 combining molecular probes on candidate genes to detect genome rearrangements that drive the
531 speciation process (Garcia-Souto, Perez-Garcia, Kendall, & Pasantes, 2016; Gonzalez-Tizon,
532 Martinez-Lage, Rego, Ausio, & Mendez, 2000; Y. Wang & Guo, 2004). However, the availability
533 of reference genomes assembled at the chromosome level opens new perspectives to explore the
534 molecular evolution at several taxonomic orders through gene collinearity analysis. The study by
535 Yang et al. (2021) highlighted putative chromosome rearrangements among the king scallop
536 *Pecten maximus*, the blood clam *Scapharca broughtonii*, the hard-shelled mussel *Mytilus*
537 *coruscus*, the pearl oyster *Pinctada martensii*, and the Pacific oyster *Crassostrea gigas* genomes.
538 Notably, the chromosome synteny illustrated that large-scale rearrangements are common events
539 between scallop and oysters but scarce between scallop and mussel genomes. The reported
540 evidence suggested that almost all the chromosome rearrangements between the mussel and the
541 oyster genomes are different, implicating independent chromosome fusion events. The SLEs loci
542 identified in all the genomes analyzed in the current study suggest that SLEs are relatively
543 conserved in chromosome position for some taxa. For instance, the SLEs loci in Veneridae,
544 Pectinidae, and Solenidae appear to be associated with chromosomes 10, 13, 12, and 16. This
545 sharing characteristic can reflect common genetic events during the evolution of these taxonomical
546 groups. Similarly, the Ostreidae and Mytilidae families share SLEs loci annotated to chromosomes
547 1, 2, 8, and 10. The detailed analysis of SLEs in Mytilidae evidence that transposon identified in
548 *M. chilensis* was shared between *M. edulis* and *M. coruscus*, where SLEs in *D. polymorpha* appear
549 to be more phylogenetically distant than *Mytilus* species. Interestingly, the mutation identified on
550 the SLEs localized in the *M. chilensis* genome (insertion of twelve nucleotides), specifically on
551 chromosomes 7 and 11, was shared with the SLE annotated on chromosome 9 in *M. edulis*. This

552 cumulative evidence reveals diverse chromosome rearrangements, reflecting a complex
553 evolutionary history of bivalve chromosomes.
554
555 **3.6. The marine environment of *M. chilensis* populations**
556 Temporal and spatial variability of sea surface temperature (SST) around Chiloé island and at
557 Yaldad and Cochamó were analyzed over the past two decades. The oceanographic variability for
558 the location studied was analyzed from remote sensing data and in situ measurements (Fig. 5A-
559 D). Here, the daily time series of SST extracted from satellite-derived data for both sites evidenced
560 high surface temperature variability between Yaldad and Cochamó, where this last location was
561 constantly higher thought the year (Fig 3S). Notably, the monthly medians computed from the SST
562 time series showed that the main differences were observed during the austral summer (from
563 December to March. During the winter season, the oceanographic variability was less pronounced,
564 showing the temperatures between 13ºc and 10ºC from April to July (Fig. 5C). Furthermore, *in*
565 *situ* data were collected from June 2017 to May 2018 (exact time where the specimens were
566 collected), exhibiting significant differences in both location for temperature and salinity among 0
567 and 20 meters of depth. Interestingly, the oceanographic survey revealed a pronounced vertical
568 stratification with higher temperatures and lower salinity in Cochamó compared with Yaldad (Fig.
569 5D). These observations support the idea that there are two oceanographically different zones in
570 the inner sea of Chiloé Island. In this northern area, mussels from Cochamó and Yaldad were
571 sampled for the current study. Taken together, we can hypothesize that the mussels inhabiting
572 Cochamó are significantly more exposed to environmental stress than the Yaldad mussel
573 population. To date, there are few studies exploring how population genetic variation is related to,
574 or caused by, the marine environmental variation in mussel populations. Notably, a study

575 conducted by Wenne et al. (2022) examined the genetic differentiation of native populations of *M.*
576 *galloprovincialis* throughout its entire geographic range in the Mediterranean Sea, the Black Sea
577 and the Sea of Azov using 53 SNP loci. The results indicated that seven of the 13 environmental
578 variables explained significant variation in population-specific SNP locus allele frequencies.
579 These seven variables explained a total of 75% of the variation in the SNP dataset, suggesting that
580 there is a complex mix of environmental variables that contribute to genetic variation of *M.*
581 *galloprovincialis* populations in the Mediterranean Sea.
582
583 **3.7. Whole-genome transcript expression analysis in two *M. chilensis* populations**
584 The transcriptome profiling among mussels collected from Yaldad and Cochamó evidenced three
585 primary transcriptional clusters, where gene cluster 1 was highly expressed in the gills of mussels
586 exposed to the Yaldad marine conditions; meanwhile, gene clusters 2 and 3 were highly expressed
587 from individuals collected in Cochamó or mussel exposed to estuarine conditions (Fig. 6A). The
588 RNA-seq analysis was performed with the mRNA sequences annotated on the *M. chilensis*
589 genome. Herein, it is essential to note that in mussel species, specifically in *M. galloprovincialis*,
590 the phenomenon of Presence-Absence Variation (PAV) has been described. This fact means that
591 PAVs can bias the analyses of transcriptome profiles in the studied mussel populations. We
592 previously conducted a *de novo* assembling for the RNA-data sets sequenced from Yaldad and
593 Cochamó populations (data not shown). The results evidenced that the completeness of the gene
594 set annotated in the *M. chilensis* genome did not show statistical differences between both mussel
595 populations. The evaluation of differentially expressed genes (DEGs) showed that the main factor
596 of differences in the amount of DEGs was due to the population more than the replicates assessed
597 (Fig. 6B). The proportion of DEGs evaluated among the gene clusters revealed that the cluster 1,

598 highly expressed in Yaldad, accounted the 78.85% of the total DEGs analyzed. Clusters 2 and 3
599 are primarily characterized by high transcription values in the Cochamó population, evidenced by
600 7.32% and 13.82% of DEGs, respectively. The total number of DEGs analyzed was 1,570 (Fig.
601 6C). Notably, the fold-changes values estimated among the replicates and populations revealed
602 high values in the gene transcriptional cluster 1, compared with clusters 2 and 3 where the fold-
603 changes values were significantly lower (Fig. 6D). The functional analysis showed that the cluster
604 1 was enriched by GO terms related with protein modification processes, programmed cell death,
605 immune system processes, defense response, cell differentiation and anatomical structure
606 development (Fig. 6E). The clusters 2 and 3 were less enriched, revealing significant GO terms for
607 transmembrane transport, reproductive processes, protein-containing complex assembly,
608 microtube-based movement, cytoskeleton organization, chromatin organization and metabolic
609 process (Fig. 6E).

610 The cluster gene expression analysis was used to identify genetic polymorphisms annotated
611 in differentially expressed genes (DEGs) between Yaldad and Cochamó mussel populations. The
612 evaluation of DEGs was performed by cluster transcriptome analysis displayed using a Circos plot
613 to visualize specific loci where DEGs were highly transcribed. The fold-change values calculated
614 showed high levels of transcription in clusters 1 and 2 through all chromosomes scanned (see red
615 dots in Fig. 7A). Congruently with the previous RNA-seq results in this study, the highest fold-
616 change values were observed in DGEs annotated in cluster 1 (Yaldad population). On the contrary,
617 cluster 3 showed a small number of DEGs with high fold-changes values. Notably, the DEGs
618 localized on the *M. chilensis* chromosomes evidenced specific-transcriptome patterns where some
619 genes are differentially and spatially expressed through the genome of mussels exposed to the
620 Yaldad and Cochamó environment conditions. The synteny analysis for DEGs and differentially

621 localized in the chromosomes showed a marked pattern of syntenic relationships among
622 chromosomes 5, 7, and 12 for cluster 1; meanwhile, the synteny observed for the DEGs annotated
623 in clusters 2, and 3 revealed a wide distribution along the *M. chilensis* genome (Fig. 7A).
624 Interestingly, the analysis carried out to detect macro-genome mutation in gene families between
625 Yaldad and Cochamó population evidenced a similar number of dispersed genes, suggesting that
626 those might arise from transposition. Tandem genes or repeatedly duplicated were observed with
627 a low proportion in cluster 3 (Cochamó); meanwhile, the proximal genes showed a similar
628 proportion to cluster 1 (Yaldad). These results might suggest small-scale transposition or
629 duplication/insertion events. An interesting finding was observed for whole genome duplication
630 (WGD). The primary proportion was evidenced in cluster 3 (Cochamó), compared with clusters 1
631 and 2 from the Yaldad population (Fig. 7B). Furthermore, the bioinformatic analysis conducted
632 for detecting amino acid changes (AAC) in DEGs showed that 38% of non-synonymous AAC
633 were identified in mussels collected from Yaldad. Contrary, the main proportion of synonymous
634 AAC was detected in mussels exposed to the Cochamó's estuarine conditions (Fig. 7C). Notably,
635 the analysis performed for DEGs annotated in cluster 2 did not show non-synonymous and
636 synonymous AAC in mussels collected from Yaldad. Finally, the evaluation of the zygosity
637 proportion estimated for each mussel population evidenced an inverse pattern between both
638 populations. The Yaldad cluster was higher in the homozygous proportion compared with
639 Cochamó, where AAC heterozygous were detected in a higher proportion (Fig. 7D).

640 To explore the transcriptome signatures between Yaldad and Cochamó mussel populations,
641 we applied the genome chromosome expression (CGE) approach to test differences among tissues
642 and individuals through the *M. chilensis* genome. The CGE analysis revealed high differences
643 among chromosome regions, where the gills tissue was more modulated than the mantle tissue

644 (Fig. 8A). Interestingly, there are some levels of congruence among the CGE annotated for both
645 mussel populations. Using this finding, we conducted a gene ontology enrichment analysis from
646 genes identified by CGE analysis. The results evidenced that gills transcriptomes displayed
647 functional processes associated with transmembrane transport, protein catabolic, nervous system,
648 and metal homeostasis. Notably, immune system process GO terms were highly enriched in gills.
649 Moreover, the chromosome region differentially expressed in mantle tissue revealed that the
650 reproductive process, protein modification process, cell differentiation, anatomical structure
651 development, and gene silencing by RNA were mainly annotated (Fig. 8B). Taken together, the
652 results reported in this study are highly congruent with the previous study conducted by Yévenes
653 et al., (2022) through the transcriptome responses of *M. chilensis* collected in ecologically different
654 farm-impacted seedbeds.

655 The cumulative findings archived in this study strongly suggest that the immune system
656 was primarily modulated between mussels exposed to Yaldad and Cochamó environmental
657 conditions. With the aim to explore the transcription profiling of immune-related genes, we
658 selected two KEGG pathways annotated in the *M. chilensis* genome (Fig. 9). Herein, TOLL-like
659 receptor signaling pathway and apoptosis were analyzed in terms of transcription activity and
660 single nucleotide variation (SNV) between mussel populations. Notably, a non-synonymous SNV
661 was detected on the TLR2 gene (28T>G) in individuals collected from the Yaldad population. The
662 translation evidenced an amino acid change from Phenylalanine to Valine at position 10 in the
663 ORF (Phe10Val) (Fig. 9A). The analysis also evidenced SNV on the genes such as AKT and
664 TAB1, where no amino acid changes were detected. The transcriptome profiling for the TLR
665 pathway evidenced a high modulation of genes such as TLR3, AKT, TRAFF6, FADD, IRAK4,
666 and RAC1 in mussels collected from Yaldad. Interestingly, mitogen-activated protein kinase

667 kinases (MAP2K and MAPK1) and c-Jun N-terminal kinase (JNK) were differentially expressed,
668 suggesting putative roles related to stress signaling pathways (Fig. 9B). Furthermore, the
669 Apoptosis pathway revealed two SNV localized in *eukaryotic initiation factor 2 alpha* (eiF2α) and
670 *inhibitor of apoptosis* (IAP) in mussel sampled from Yaldad and Cochamó population, respectively
671 (Fig. 9C). The 2613delG in *eiF2α* gene produces a frameshift at the Thr872, meanwhile the
672 968_970delCTC localized in the IAP gene produces a deletion of proline at the position 323. The
673 transcriptome profiling of apoptosis-related genes showed a conspicuous differentiation between
674 gills and mantle tissue, where three primary gene expression clusters were identified (Fig. 9D).
675 Notably, in gills tissue, genes such as P53, ERK, TP53, PARP2, and JNK were highly expressed.
676 The gene expression analysis in mantle tissue evidenced high transcriptional activity in genes
677 related to the intrinsic (mitochondria-mediated) pathway, such as *the B-cell lymphoma* (BCL) gene
678 and the second mitochondria-derived activator of caspase (DIABLO) gene. Further functional
679 studies will be conducted to validate the association between single nucleotide polymorphisms and
680 the *fitness* traits observed or how the translocation process associated with the aquaculture activity
681 can evolve in the loss of locally adapted alleles. Interestingly, a recent transplant experiment
682 reported by Jahnsen-Guzmán et al. (2021) demonstrated that *M. chilensis* individuals are adapted
683 to the subtidal environment (4 m depth), as they exhibit significantly higher fitness (growth and
684 calcification rates) than those transferred to the intertidal environment (1 m depth), which showed
685 increased metabolic stress. Herein, mussel lives in an extreme environmental variability, where
686 their ability to cope with perturbations is based on the genome architecture to build plastic and
687 adaptive responses.

688 In summary, we have generated the first chromosome-level assembly of the native blue
689 mussel *Mytilus chilensis* genome. This genomic resource was used to identify genome signatures

690 putatively related to the local adaptation process in the mussel population inhabiting contrasting
691 marine environments. Collectively, the identification of putative mutations associated with
692 immune and metabolic-related genes suggests that mussel populations facing highly variable
693 environments display a genomic adaptation to reduce the number of genes and their transcriptional
694 activity. This evolutionary strategy can suggest that the expression of those genes has evolved a
695 degree of "frontloading" that potentially pre-adapts the mussel populations to frequent heat and
696 salinity stress, contributing to their physiological tolerance and *fitness*. We believe that the
697 generated genomic resource will be instrumental for future research on population genomics
698 informing management and sustainable strategies for the Chilean mussel aquaculture.
699

700 **Author contributions**

701 CGE designed and supervised the study. VV, GNA, DVM, JT, PO, and MY collected and prepared
702 the mussel samples. CGE, CGE, VVM, DVM, GNA, and FT analyzed all sequencing and
703 oceanographic data. CGE, GG, AF, BN, SR and MG wrote the manuscript with the other authors'
704 help. GA and CGE sequenced and analyzed the Steamer-like elements in *M. chilensis*. All authors
705 revised the draft and approved the final manuscript.
706

707 **ACKNOWLEDGMENTS**

708 This study was funded by FONDAP #15110027, FONDECYT #1210852 and #1130852, granted
709 by ANID-Chile.
710

711 **COMPETING INTERESTS**

712 The authors have no conflict of interest to declare.

DATA AVAILABILITY AND BENEFITS-SHARING STATEMENT

Data availability statement:

The *Mytilus chilensis* whole genome sequencing data supporting this study's findings are available in NCBI under BioProject PRJNA861856. The sequencing data supporting this study's findings are available in SRA at SRR20966976, SRR20593343 and SRP261955. Benefits from this research accrue from sharing our data and results on public databases as described above. The assembled genome and the genome annotation results were deposited in the Figshare database (https://doi.org/10.6084/m9.figshare.20995963.v1).

References

Amar, G., Palma, C., von Brand, E., & Jara, P. (2008). Karyotype study in the surf clam *Mesodesma donacium* Lamarck, 1818 (Bivalvia: Veneroida: Mesodesmatidae). *Gayana, 72*(1), 18-22.

Araneda, C., Larrain, M. A., Hecht, B., & Narum, S. (2016). Adaptive genetic variation distinguishes Chilean blue mussels (Mytilus chilensis) from different marine environments. *Ecology and Evolution, 6*(11), 3632-3644. doi:10.1002/ece3.2110

Arriagada, G., Metzger, M. J., Muttray, A. F., Sherry, J., Reinisch, C., Street, C., . . . Goff, S. P. (2014). Activation of transcription and retrotransposition of a novel retroelement, Steamer, in neoplastic hemocytes of the mollusk Mya arenaria. *Proceedings of the National Academy of Sciences of the United States of America, 111*(39), 14175-14180. doi:10.1073/pnas.1409945111

Astorga, M. P., Vargas, J., Valenzuela, A., Molinet, C., & Marin, S. L. (2018). Population genetic structure and differential selection in mussel Mytilus chilensis. *Aquaculture Research, 49*(2), 919-927. doi:10.1111/are.13538

Bao, W. D., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA, 6*. doi:10.1186/s13100-015-0041-9

Bedell, J. A., Korf, I., & Gish, W. (2000). MaskerAid: a performance enhancement to RepeatMasker. *Bioinformatics, 16*(11), 1040-1041. doi:DOI 10.1093/bioinformatics/16.11.1040

Bibby, R., Widdicombe, S., Parry, H., Spicer, J., & Pipe, R. (2008). Effects of ocean acidification on the immune response of the blue mussel Mytilus edulis. *Aquatic Biology, 2*(1), 67-74. doi:10.3354/ab00037

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., . . . Smith, T. P. L. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics, 49*(4), 643-+. doi:10.1038/ng.3802

Blanc, J. M., Molinet, C., Subiabre, R., & Diaz, P. A. (2018). Cadmium determination in Chilean blue mussels Mytilus chilensis: Implications for environmental and agronomic interest. *Marine Pollution Bulletin, 129*(2), 913-917. doi:10.1016/j.marpolbul.2017.10.048

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., & Xia, R. (2020). TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Molecular Plant, 13*(8), 1194-1202. doi:10.1016/j.molp.2020.06.009

751 Cheng, H. Y., Concepcion, G. T., Feng, X. W., Zhang, H. W., & Li, H. (2021). Haplotype-resolved de novo
752 assembly using phased assembly graphs with hifiasm. *Nature Methods, 18*(2), 170-+.
753 doi:10.1038/s41592-020-01056-5
754 Ciocan, C., & Sunila, I. (2005). Disseminated neoplasia in blue mussels, Mytilus galloprovincialis, from
755 the Black Sea, Romania. *Marine Pollution Bulletin, 50*(11), 1335-1339.
756 doi:10.1016/j.marpolbul.2005.04.042
757 Cooper , E. L. a. P., N. (1996). COMPARATIVE IMMUNOLOGIC MODELS CAN ENHANCE
758 ANALYSES OF ENVIRONMENTAL IMMUNOTOXICITY. *Annual Review of Fish Diseases, 6*,
759 179-191.
760 Crain, C. M., Kroeker, K., & Halpern, B. S. (2008a). Interactive and cumulative effects of multiple human
761 stressors in marine systems. *Ecology Letters, 11*(12), 1304-1315. doi:10.1111/j.1461-
762 0248.2008.01253.x
763 Crain, C. M., Kroeker, K., & Halpern, B. S. (2008b). Interactive and cumulative effects of multiple human
764 stressors in marine systems. *Ecology Letters, 11*(12), 1304-1315. doi:10.1111/j.1461-
765 0248.2008.01253.x
766 Curelovich, J. N., Lovrich, G. A., Cueto, G. R., & Calcagno, J. A. (2018). Recruitment and zonation in a
767 sub-Antarctic rocky intertidal community. *Journal of the Marine Biological Association of the*
768 *United Kingdom, 98*(2), 411-422. doi:10.1017/S0025315416001284
769 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., . . . Li, H. (2021). Twelve
770 years of SAMtools and BCFtools. *Gigascience, 10*(2). doi:10.1093/gigascience/giab008
771 Detree, C., Nunez-Acuna, G., Roberts, S., & Gallardo-Escarate, C. (2016). Uncovering the Complex
772 Transcriptome Response of Mytilus chilensis against Saxitoxin: Implications of Harmful Algal
773 Blooms on Mussel Populations. *PLoS One, 11*(10). doi:10.1371/journal.pone.0165231
774 Du, X. D., Fan, G. Y., Jiao, Y., Zhang, H., Guo, X. M., Huang, R. L., . . . Liu, X. (2017). The pearl oyster
775 Pinctada fucata martensii genome and multi-omic analyses provide insights into biomineralization.
776 *Gigascience, 6*(8). doi:10.5524/100240
777 Duarte, C., Navarro, J. M., Quijon, P. A., Loncon, D., Torres, R., Manriquez, P. H., . . . Lagos, N. A. (2018).
778 The energetic physiology of juvenile mussels, Mytilus chilensis (Hupe): The prevalent role of
779 salinity under current and predicted pCO(2) scenarios. *Environmental Pollution, 242*, 156-163.
780 doi:10.1016/j.envpol.2018.06.053
781 Ellis, R. P., Parry, H., Spicer, J. I., Hutchinson, T. H., Pipe, R. K., & Widdicombe, S. (2011). Immunological
782 function in marine invertebrates: Responses to environmental perturbation. *Fish & Shellfish*
783 *Immunology, 30*(6), 1209-1222. doi:10.1016/j.fsi.2011.03.017
784 Elston, R. A., Moore, J. D., & Brooks, K. (1992). Disseminated Neoplasia of Bivalve Mollusks. *Reviews*
785 *in Aquatic Sciences, 6*(5-6), 405-466.
786 Enriquez, R., Frosner, G. G., Hochsteinmintzel, V., Riedemann, S., & Reinhardt, G. (1992). Accumulation
787 and Persistence of Hepatitis-a Virus in Mussels. *Journal of Medical Virology, 37*(3), 174-179.
788 doi:Doi 10.1002/Jmv.1890370305
789 Faust, G. G., & Hall, I. M. (2014). SAMBLASTER: fast duplicate marking and structural variant read
790 extraction. *Bioinformatics, 30*(17), 2503-2505. doi:10.1093/bioinformatics/btu314
791 Ferreira, S., Ashby, R., Jeunen, G. J., Rutherford, K., Collins, C., Todd, E. V., & Gemmell, N. J. (2020).
792 DNA from mollusc shell: a valuable and underutilised substrate for genetic analyses. *PeerJ, 8*,
793 e9420. doi:10.7717/peerj.9420
794 Flores, C. A. M., Gomez, M. A. D., Munoz, C. B. A., Perez, L. E. C., Arribas, S. L. M., Opazo, M. P. A.,
795 & Huaquin, E. J. E. N. (2015). Spatial distribution pattern of Mytilus chilensis beds in the Reloncavi
796 fjord: hypothesis on associated processes. *Revista Chilena de Historia Natural, 88*. doi:
797 10.1186/s40693-015-0041-7
798 Gaitan-Espitia, J. D., Quintero-Galvis, J. F., Mesas, A., & D'Elia, G. (2016). Mitogenomics of southern
799 hemisphere blue mussels (Bivalvia: Pteriomorphia): Insights into the evolutionary characteristics
800 of the Mytilus edulis complex. *Scientific Reports, 6*. doi:10.1038/srep26853

Gallardo-Escárate, C., Álvarez-Borrego, J., Del Río-Portilla, M., & Kober, V. (2004). Karyotype of Pacific red abalone *Haliotis rufescens* (Archaeogastropoda: Haliotidae*)*, using image analysis. *Journal of Shellfish Research, 23*(1), 205-209.

Gallardo-Escarate, C., Figueras, A., & Novoa, B. (2019). The dark matter of the mussel's genome: long non-coding RNAs as key players in Mytilus. *Isj-Invertebrate Survival Journal, 16*, 191-191.

Garcia-Souto, D., Perez-Garcia, C., Kendall, J., & Pasantes, J. J. (2016). Molecular Cytogenetics in Trough Shells (Mactridae, Bivalvia): Divergent GC-Rich Heterochromatin Content. *Genes (Basel), 7*(8). doi:10.3390/genes7080047

Gardner, J. P. A., Oyarzun, P. A., Toro, J. E., Wenne, R., & Zbawicka, M. (2021). Phylogeography of Southern Hemisphere Blue Mussels of the Genus Mytilus: Evolution, Biosecurity, Aquaculture and Food Labelling. *Oceanography and Marine Biology: An Annual Review, Vol 59, 59*, 139-+. doi:10.1201/9781003138846-3

Gerdol, M., Moreira, R., Cruz, F., Gomez-Garrido, J., Vlasova, A., Rosani, U., . . . Figueras, A. (2020). Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biology, 21*(1). doi:10.1186/s13059-020-02180-3

Ghurye, J., Rhie, A., Walenz, B. P., Schmitt, A., Selvaraj, S., Pop, M., . . . Koren, S. (2019). Integrating Hi-C links with assembly graphs for chromosome-scale assembly. *Plos Computational Biology, 15*(8). doi:10.1371/journal.pcbi.1007273

Gonzalez-Poblete, E., Hurtado, C. F., Rojo, C., & Norambuena, R. (2018). Blue mussel aquaculture in Chile: Small or large scale industry? *Aquaculture, 493*, 113-122. doi:10.1016/j.aquaculture.2018.04.026

Gonzalez-Tizon, A. M., Martinez-Lage, A., Rego, I., Ausio, J., & Mendez, J. (2000). DNA content, karyotypes, and chromosomal location of 18S-5.8S-28S ribosomal loci in some species of bivalve molluscs from the Pacific Canadian coast. *Genome, 43*(6), 1065-1072. doi:10.1139/g00-089

Gray, A. P., Lucas, I. A. N., Seed, R., & Richardson, C. A. (1999). Mytilus edulis chilensis infested with Coccomyxa parasitica (Chlorococcales, Coccomyxaceae). *Journal of Molluscan Studies, 65*, 289-294. doi:Doi 10.1093/Mollus/65.3.289

Gunderson, A. R., Armstrong, E. J., & Stillman, J. H. (2016). Multiple Stressors in a Changing World: The Need for an Improved Perspective on Physiological Responses to the Dynamic Marine Environment. *Annual Review of Marine Science, Vol 8, 8*, 357-+. doi:10.1146/annurev-marine-122414-033953

Hargrave, B. T., Doucette, L. I., Cranford, P. J., Law, B. A., & Milligan, T. G. (2008). Influence of mussel aquaculture on sediment organic enrichment in a nutrient-rich coastal embayment. *Marine Ecology Progress Series, 365*, 137-149. doi:10.3354/meps07636

Harvell, C. D. M., C.E. Ward, J.R. Altizer, S. Dobson, A.P. Ostfeld, R.S. Samuel, M.D. . (2002). Climate warming and disease risks for terrestrial and marine biota. *Science, 296*(5576), 2158-2162.

Hüning, A. K. M., F. Thomsen, J. Gutowska, M.A. Krämer, L. Frickenhaus, S. Rosenstiel, P. Pörtner, H.-O. . Philipp, E.E Lucassen M. (2013). mpacts of seawater acidification on mantle gene expression patterns of the Baltic Sea blue mussel: implications for shell formation and energy metabolism. *Marine Biology, 160*(8), 1845-1861.

Jahnsen-Guzman, N., Lagos, N. A., Lardies, M. A., Vargas, C. A., Fernandez, C., San Martin, V. A., . . . Duarte, C. (2021). Environmental refuges increase performance of juvenile mussels Mytilus chilensis: Implications for mussel seedling and farming strategies. *Science of the Total Environment, 751*, 141723. doi:10.1016/j.scitotenv.2020.141723

Kess, T., Bentzen, P., Lehnert, S. J., Sylvester, E. V. A., Lien, S., Kent, M. P., . . . Bradbury, I. R. (2020). Modular chromosome rearrangements reveal parallel and nonparallel adaptation in a marine fish. *Ecol Evol, 10*(2), 638-653. doi:10.1002/ece3.5828

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research, 19*(9), 1639-1645. doi:10.1101/gr.092759.109

Krzywinski, M. I., Schein, J. E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., . . . Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*. doi:10.1101/gr.092759.109

Lando, D., Stevens, T. J., Basu, S., & Laue, E. D. (2018). Calculation of 3D genome structures for comparison of chromosome conformation capture experiments with microscopy: An evaluation of single-cell Hi-C protocols. *Nucleus, 9*(1), 190-201. doi:10.1080/19491034.2018.1438799

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology, 10*(3). doi:10.1186/gb-2009-10-3-r25

Larrain, M. A., Diaz, N. F., Lamas, C., Uribe, C., & Araneda, C. (2014). Traceability of mussel (Mytilus chilensis) in southern Chile using microsatellite molecular markers and assignment algorithms. Exploratory survey. *Food Research International, 62*, 104-110. doi:10.1016/j.foodres.2014.02.016

Larrain, M. A., Zbawicka, M., Araneda, C., Gardner, J. P. A., & Wenne, R. (2018). Native and invasive taxa on the Pacific coast of South America: Impacts on aquaculture, traceability and biodiversity of blue mussels (Mytilus spp.). *Evolutionary Applications, 11*(3), 298-311. doi:10.1111/eva.12553

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*(5), 589-595. doi:10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Proc, G. P. D. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H. Q., Ong, C. T., . . . Corces, V. G. (2015). Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol Cell, 58*(2), 216-231. doi:10.1016/j.molcel.2015.02.023

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., . . . Dekker, J. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science, 326*(5950), 289-293. doi:10.1126/science.1181369

Lin, D., Hong, P., Zhang, S. H., Xu, W. Z., Jamal, M., Yan, K. J., . . . Cao, G. (2018). Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nature Genetics, 50*(5), 754-+. doi:10.1038/s41588-018-0111-2

Lockwood, B. L., Sanders, J. G., & Somero, G. N. (2010). Transcriptomic responses to heat stress in invasive and native blue mussels (genus Mytilus): molecular correlates of invasive success. *Journal of Experimental Biology, 213*(20), 3548-3558. doi:10.1242/jeb.046094

Malagoli, D., Casarini, L., Sacchi, S., & Ottaviani, E. (2007). Stress and immune response in the mussel Mytilus galloprovincialis. *Fish & Shellfish Immunology, 23*(1), 171-177. doi:Doi 10.1016/J.Fsi.2006.10.004

Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution, 38*(10), 4647-4654. doi:10.1093/molbev/msab199

Martínez-Lage, A., González-Tizón, A., & Méndez, J. (1995). Chromosomal markers in three species of the genus *Mytilus* (Mollusca: Bivalvia). *Heredity, 74*, 369 - 375.

McCartney, M. A., Auch, B., Kono, T., Mallez, S., Zhang, Y., Obille, A., . . . Gohl, D. M. (2022). The genome of the zebra mussel, Dreissena polymorpha: a resource for comparative genomics, invasion genetics, and biocontrol. *G3-Genes Genomes Genetics, 12*(2). doi:10.1093/g3journal/jkab423

Metzger, M. J., Reinisch, C., Sherry, J., & Goff, S. P. (2015). Horizontal Transmission of Clonal Cancer Cells Causes Leukemia in Soft-Shell Clams. *Cell, 161*(2), 255-263. doi:10.1016/j.cell.2015.02.042

Metzger, M. J., Villalba, A., Carballal, M. J., Iglesias, D., Sherry, J., Reinisch, C., . . . Goff, S. P. (2016). Widespread transmission of independent cancer lineages within multiple bivalve species. *Nature, 534*(7609), 705-+. doi:10.1038/nature18599

Mitta, G., Vandenbulcke, F., Noel, T., Romestand, B., Beauvillain, J. C., Salzet, M., & Roch, P. (2000). Differential distribution and defence involvement of antimicrobial peptides in mussel. *Journal of Cell Science, 113*(15), 2759-2769.

Murgarella, M., Puiu, D., Novoa, B., Figueras, A., Posada, D., & Canchaya, C. (2016). A First Insight into the Genome of the Filter-Feeder Mussel Mytilus galloprovincialis. *Plos One, 11*(3). doi: e015156110.1371/journal.pone.0151561

Nunez-Acuna, G., & Gallardo-Escarate, C. (2013). Identification of immune-related SNPs in the transcriptome of Mytilus chilensis through high-throughput sequencing. *Fish & Shellfish Immunology, 35*(6), 1899-1905. doi:10.1016/j.fsi.2013.09.028

Pal, K., Forcato, M., & Ferrari, F. (2019). Hi-C analysis: from data generation to integration. *Biophys Rev, 11*(1), 67-78. doi:10.1007/s12551-018-0489-1

Pereiro, P., Moreira, R., Novoa, B., & Figueras, A. (2021). Differential Expression of Long Non-Coding RNA (lncRNA) in Mediterranean Mussel (Mytilus galloprovincialis) Hemocytes under Immune Stimuli. *Genes, 12*(9). doi:10.3390/genes12091393

Perez-Garcia, C., Moran, P., & Pasantes, J. J. (2014). Karyotypic diversification in Mytilus mussels (Bivalvia: Mytilidae) inferred from chromosomal mapping of rRNA and histone gene clusters. *Bmc Genetics, 15*. doi:10.1186/1471-2156-15-84

Phase-Genomics. (2019). Aligning and QCing Phase Genomics Hi-C data. Seattle, WA: Phase Genomics Documentation. *https://phasegenomics.github.io*.

Renaut, S., Guerra, D., Hoeh, W. R., Stewart, D. T., Bogan, A. E., Ghiselli, F., . . . Breton, S. (2018). Genome Survey of the Freshwater Mussel Venustaconcha ellipsiformis (Bivalvia: Unionida) Using a Hybrid De Novo Assembly Approach. *Genome Biology and Evolution, 10*(7), 1637-1646. doi:10.1093/gbe/evy117

Rey-Campos, M., Novoa, B., Pallavicini, A., Gerdol, M., & Figueras, A. (2021). Comparative Genomics Reveals 13 Different Isoforms of Mytimycins (A-M) in Mytilus galloprovincialis. *International Journal of Molecular Sciences, 22*(6). doi:10.3390/ijms22063235

Robinson, J. T., Turner, D., Durand, N. C., Thorvaldsdottir, H., Mesirov, J. P., & Aiden, E. L. (2018). Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Systems, 6*(2), 256-+. doi:10.1016/j.cels.2018.01.001

Rolff, J., & Siva-Jothy, M. T. (2003). Invertebrate ecological immunology. *Science, 301*(5632), 472-475. doi:DOI 10.1126/science.1080623

Romero, A., Novoa, B., & Figueras, A. (2022). Genomic and transcriptomic identification of the cathepsin superfamily in the Mediterranean mussel Mytilus galloprovincialis. *Developmental and Comparative Immunology, 127*. doi:10.1016/j.dci.2021.104286

Sendra, M., Saco, A., Yeste, M. P., Romero, A., Novoa, B., & Figueras, A. (2020). Nanoplastics: From tissue accumulation to cell translocation into Mytilus galloprovincialis hemocytes. resilience of immune cells exposed to nanoplastics and nanoplastics plus Vibrio splendidus combination. *Journal of Hazardous Materials, 388*. doi:10.1016/j.jhazmat.2019.121788

Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mobile DNA, 12*(1). doi:10.1186/s13100-020-00230-y

Tarifeno-Saldivia, E., Valenzuela-Miranda, D., & Gallardo-Escarate, C. (2017). In the shadow: The emerging role of long non-coding RNAs in the immune response of Atlantic salmon. *Developmental and Comparative Immunology, 73*, 193-205. doi:10.1016/j.dci.2017.03.024

Toro, J. E., Ojeda, J. A., & Vergara, A. M. (2004). The genetic structure of Mytilus chilensis (Hupe 1854) populations along the Chilean coast based on RAPDs analysis. *Aquaculture Research, 35*(15), 1466-1471. doi:10.1111/j.1365-2109.2004.01172.x

Uliano-Silva, M., Dondero, F., Otto, T. D., Costa, I., Lima, N. C. B., Americo, J. A., . . . Rebelo, M. D. (2017). A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, Limnoperna fortunei. *Gigascience, 7*(2). doi:10.1093/gigascience/gix128

Uriarte, I. (2008). Estado actual del cultivo de moluscos bivalvos en Chile, FAO Actas de Pesca y Acuicultura.

Valenzuela-Muñoz, V., Gallardo-Escárate, C., Benavente, B. P., Valenzuela-Miranda, D., Núñez-Acuña, G., Escobar-Sepulveda, H., & Váldes, J. A. (2022). Whole-Genome Transcript Expression Profiling Reveals Novel Insights into Transposon Genes and Non-Coding RNAs during Atlantic Salmon Seawater Adaptation. *Biology, 11*(1), 1.

Vidal, R., Penaloza, C., Urzua, R., & Toro, J. E. (2009). Screening of ESTs from Mytilus for the detection of SSR markers in Mytilus californianus. *Molecular Ecology Resources, 9*(5), 1409-1411. doi:10.1111/j.1755-0998.2009.02682.x

Vihtakari, M., Hendriks, I. E., Holding, J., Renaud, P. E., Duarte, C. M., & Havenhand, J. N. (2013). Effects of Ocean Acidification and Warming on Sperm Activity and Early Life Stages of the Mediterranean Mussel (Mytilus galloprovincialis). *Water, 5*(4), 1890-1915. doi:10.3390/w5041890

Vinagre, C., Mendonca, V., Narciso, L., & Madeira, C. (2015). Food web of the intertidal rocky shore of the west Portuguese coast - Determined by stable isotope analysis. *Marine Environmental Research, 110*, 53-60. doi:10.1016/j.marenvres.2015.07.016

Wang, S., Zhang, J. B., Jiao, W. Q., Li, J., Xun, X. G., Sun, Y., . . . Bao, Z. M. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution, 1*(5). doi:10.1038/s41559-017-0120

Wang, Y., & Guo, X. (2004). Chromosomal rearrangement in pectinidae revealed by rRNA loci and implications for bivalve evolution. *Biol Bull, 207*(3), 247-256. doi:10.2307/1543213

Wang, Y. P., Tang, H. B., DeBarry, J. D., Tan, X., Li, J. P., Wang, X. Y., . . . Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research, 40*(7). doi:10.1093/nar/gkr1293

Wenne, R., Zbawicka, M., Pradzinska, A., Kotta, J., Herkul, K., Gardner, J. P. A., . . . Kent, M. P. (2022). Molecular genetic differentiation of native populations of Mediterranean blue mussels, Mytilus galloprovincialis Lamarck, 1819, and the relationship with environmental variables. *European Zoological Journal, 89*(1), 755-784. doi:10.1080/24750263.2022.2086306

Xu, T., Sun, J., Lv, J., Watanabe, H. K., Li, T. Q., Zou, W. W., . . . Qiu, J. W. (2017). Genome-wide discovery of single nucleotide polymorphisms (SNPs) and single nucleotide variants (SNVs) in deep-sea mussels: Potential use in population genomics and cross-species application. *Deep-Sea Research Part Ii-Topical Studies in Oceanography, 137*, 318-326. doi:10.1016/j.dsr2.2016.03.011

Yang, J. L., Feng, D. D., Liu, J., Xu, J. K., Chen, K., Li, Y. F., . . . Lu, Y. (2021). Chromosome-level genome assembly of the hard-shelled mussel Mytilus coruscus, a widely distributed species from the temperate areas of East Asia. *Gigascience, 10*(4). doi:10.1093/gigascience/giab024

Yevenes, M., Nunez-Acuna, G., Gallardo-Escarate, C., & Gajardo, G. (2021). Adaptive Differences in Gene Expression in Farm-Impacted Seedbeds of the Native Blue Mussel Mytilus chilensis. *Frontiers in Genetics, 12*. doi:10.3389/fgene.2021.666539

Yevenes, M., Nunez-Acuna, G., Gallardo-Escarate, C., & Gajardo, G. (2022). Adaptive mitochondrial genome functioning in ecologically different farm-impacted natural seedbeds of the endemic blue mussel Mytilus chilensis. *Comparative Biochemistry and Physiology D-Genomics & Proteomics, 42*. doi:10.1016/j.cbd.2021.100955

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology, 16*(5), 284-287.

Zhang, G. F., Fang, X. D., Guo, X. M., Li, L., Luo, R. B., Xu, F., . . . Wang, J. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature, 490*(7418), 49-54. doi:10.1038/nature11413

List of the Figures

Figure 1. (A) Photograph of the native blue mussel *Mytilus chilensis*. (B) Metaphasic chromosomes from mussel larvae samples and mapping of 28S-rDNA by fluorescence *in situ* hybridization. (C) Heatmap of chromosome interaction intensity in the blue mussel Hi-C assembly. The x-and y-axis represents the length of the chromosomes. The color bar represents the Hi-C contact density.

Figure 2. The native blue mussel *Mytilus chilensis* genome. (A) Workflow of *de novo* whole-genome sequencing project and annotation for *M. chilensis*. The rectangles indicate the steps of the primary data processing, and the arrows indicate output or input data. Pink diamonds indicate the combined strategy based on homolog prediction, de novo and RNA-seq assisted prediction. (B) The circos plot shows the genomic features for the 14 pseudo-chromosomes. From outer to the inner circle: Gene density, Repeat density, GC content, rRNAs localization and ncRNAs. The transcriptome expression for mantle (Ma), gills (Gi), Hemocytes (He), and digestive gland (DG) are shown as blue light profiles. Chromosome syntenies are represented in different colors according to each ideogram. The chromosome size is shown in the Mb scale.

Figure 3. (A) Whole-genome macrosyntenic relationships between *M. chilensis* and *M. coruscus*. Orthologous relationships among mussel chromosomes are highlighted in grey and orthologous relationship between chr6 and LG02, and chr10 and LG09 are highlighted in red. (B-C) Gene Ontology term annotated for the syntenic chromosome regions identified by MCScanX for chr6 vs. LG02, and chr10 vs. LG09. The x-axis indicates the negative $\log_{10}$(q-value)

Figure 4. Molecular characterization of Steamer-like elements (SLE) in the *M. chilensis* genome and phylogenetic analysis using publicly available reference genomes assembled at chromosome-level for bivalve species. (A) Alignment of five SLE copies localized in the chromosomes 1, 6, 7, 10 and 11. All the SLE are flanked by two LTRs (5′ and 3′) containing the Gag-Pol ORFs and the domains annotated for Protease, Reverse Transcriptase, RNAaseH and Integrase. An insertion composed by 12 nucleotides (933^934 position) was found in the chromosomes 7 and 11. The detailed alignment and the translation for the nucleotides inserted are highlighted in the red box. (B) Maximum likelihood (ML) phylogenetic tree of nucleotide sequences from SLEs found in eleven reference genomes for Bivalvia. Colored chromosomes and numbers indicate the SLE genome localization and the bivalve species, respectively. The species analyzed were: Veneridae (pink), *Ruditapes philippinarum* (Rphi) and *Mercenaria mercenaria* (Mmer). Solenidae (grey) *Solen grandis* (Sgra). Pectinidae (blue), *Pecten maximus* (Pmax). Ostreidae (green) *Crassostrea gigas* (Cgig), *C. virginica* (Cvir) and *C. ariakensis* (Caria). Pteriidae (blue light), *Pinctada imbricata* (Pimb). Mytilidae (red), *Mytilus coruscus* (Mcor), *Mytilus edulis* (Medu) and *Dreissena polymorpha* (Dpol). (C) ML analysis of SLEs identified in *M. edulis* (orange), *M. coruscus* (red), *D. polymorpha* (purple) and *M. chilensis* (blue) chromosomes.

Figure 5. Temporal and spatial variability around Chiloé island, and at sites Yaldad and Cochamo, over the past two decades. (A-B) Maps correspond to mean SST computed for December 2017 – March 2018 and the corresponding standard deviations. Stars indicate the location of sampling sites. (C) Monthly medians computed from data in (A), showing the first and third quartiles as error bars; note that the

sequence of months shown on the x-axis begins in August and ends in July. Shaded areas indicate summer (yellow) and fall-winter (blue) periods. (D) *In situ* measurements for temperature (ºC) and salinity (PSU) from 0 to 20 meters of depth. The blue and red lines for (C) and (D) represent the data collected from Yaldad and Cochamo, respectively.

Figure 6. Population-specific transcriptome analysis in the blue mussel *M. chilensis*. (A) Transcriptome patterns of coding genes analyzed in gills from Yaldad and Cochamo population. Three replicates were evaluated from each experimental group. The heatmap was based on Transcripts Per Million (TPM) calculation and hierarchical clustering on Manhattan distances with average linkage. Yellow colors mean upregulated coding genes, black represents downregulated genes. (B) Venn diagram showing shared and unique genes expressed among the locations and replicates. (C) Pie chart showing number of differentially expressed genes (DEGs) annotated for expression cluster 1, 2 and 3 between Yaldad and Cochamo populations. (D) Fold-changes values observed for DEGs identified in each cluster evaluated. (E) GO enrichment of cluster-specific genes (P-value<10–16;|fold-change|>5) annotated for key biological processes differentially expressed. The y-axis indicates the GO term, and the x-axis indicates the negative $\log_{10}$(q-value). The color bar indicates the enriched factor. The bubble size indicates the number of GO terms.

Figure 7. Genetic polymorphisms annotated in differentially expressed genes (DEGs) analyzed between Yaldad and Cochamo mussel populations. The evaluation of DEGs was performed by cluster transcriptome analysis. (A) Circos plot showing DEGs identified in the three analyzed clusters. From outer to the inner circle: Gene density, DEGs cluster 1, DEGs cluster 2, DEG cluster 3 and syntenic relationships between DGEs (each color line represents the cluster analyzed). Red dots represents DEGs with fold-changes values >|100| and purple dots represents fold-changes values <|10|. (B) Genotypes of DEGs identified in *M. chilensis* populations according to the transcription cluster analysis. Singleton means that the gene is single-copy, which should not be the type of members of gene families. Dispersed means that the gene might arise from transposition. Tandem means that genes were repeatedly duplicated. Proximal means that the gene might arise from small-scale transposition or arise from tandem duplication and insertion of some other genes and whole genome duplication (WGD) means that the gene might arise from chromosome duplication region. The analysis was carried out using *MCScanX*. (C) Amino acid changes proportions (%) between Yaldad and Cochamo populations. For each mussel population, the non-synonymous and synonymous were annotated for the DEGs selected according to the cluster analysis. (D) Zygosity proportion (%) estimated for each mussel population. Cluster 1 (Yaldad) is represented by the blue bars and the cluster 2 and 3 (Cochamo) is displayed in brown color bars.

Figure 8. Chromosome-genome transcription in *M. chilensis* tissues collected from Yaldad and Cochamo populations. A) Threshold of gene expression from gills and mantle were mapped and compared on chromosomes regions from the two analyzed mussel populations. Transcriptional differences among locations (Yaldad/Cochamo) and tissues (Gills/Mantle) were used to estimate the CGE index. Heatmap in red shows the variation of gene expression from high to low differences. (B) GO enrichment of tissue-specific genes (P-value<10–16;|fold-change|>5) annotated for key biological processes differentially expressed. The y-axis indicates the GO term, and the x-axis indicates the

negative $\log_{10}$(q-value). The color bar indicates the enriched factor. The bubble size indicates the number of GO terms.

Figure 9. Transcriptome response of the immune system by KEGG pathways analysis and single mutation variant detection between *M. chilensis* populations. (A) TOLL-like receptor (TLR) and (C) Apoptosis signaling pathways comparisons between mussels from Yaldad and Cochamo. Single nucleotide polymorphisms and amino acid changes are shown in blue and brown boxes according to the mussel population. Identified gene families on the KEGG pathways are marked with red in the *M. chilensis* genome. (B) and (D) Transcriptome patterns of coding genes analyzed in gills and mantle from Yaldad and Cochamo population for TLR and Apoptosis-related genes. Three replicates were evaluated from each experimental group. The heatmap was based on Transcripts Per Million (TPM) calculation and hierarchical clustering on Manhattan distances with average linkage. Red colors mean upregulated coding genes, and blue colors represent downregulated genes.

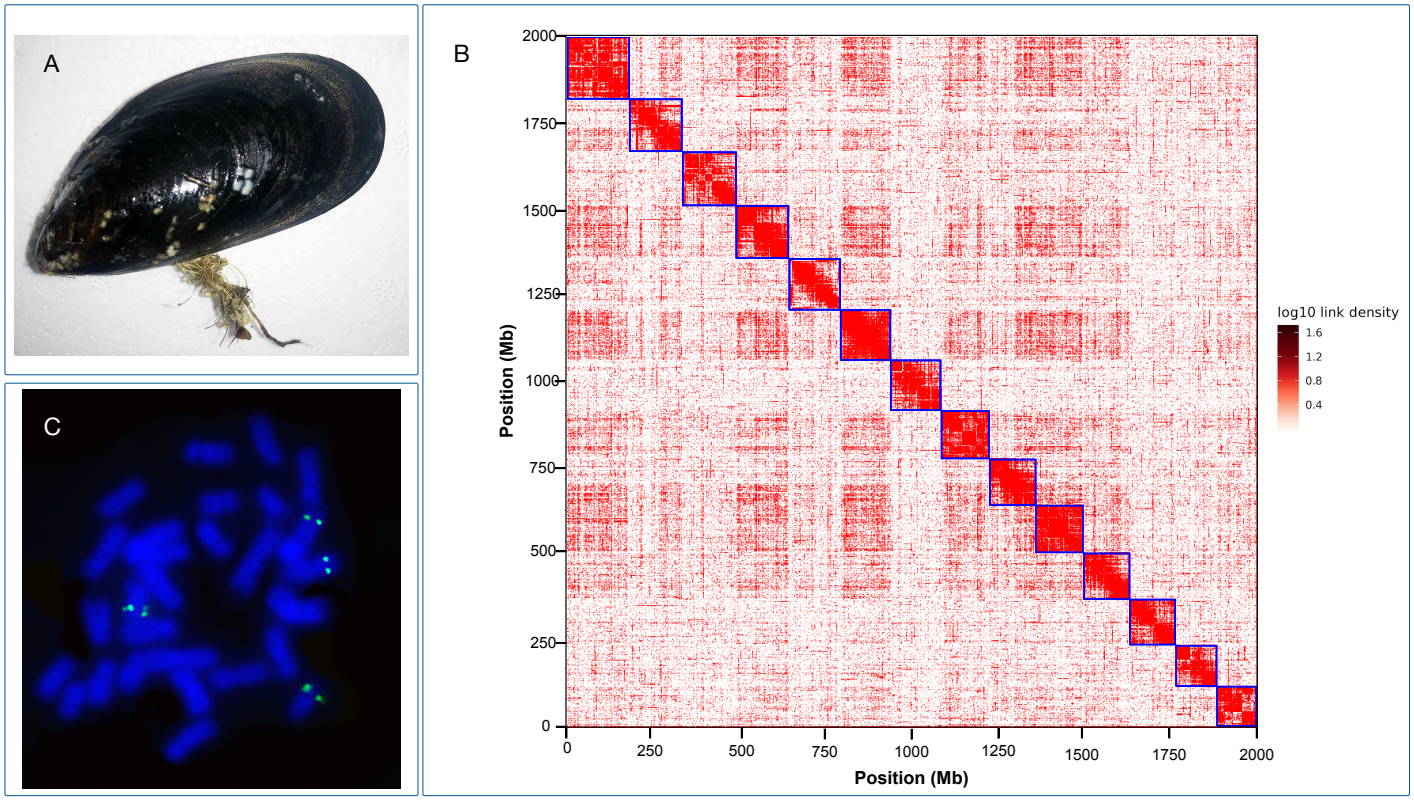

Figure 1

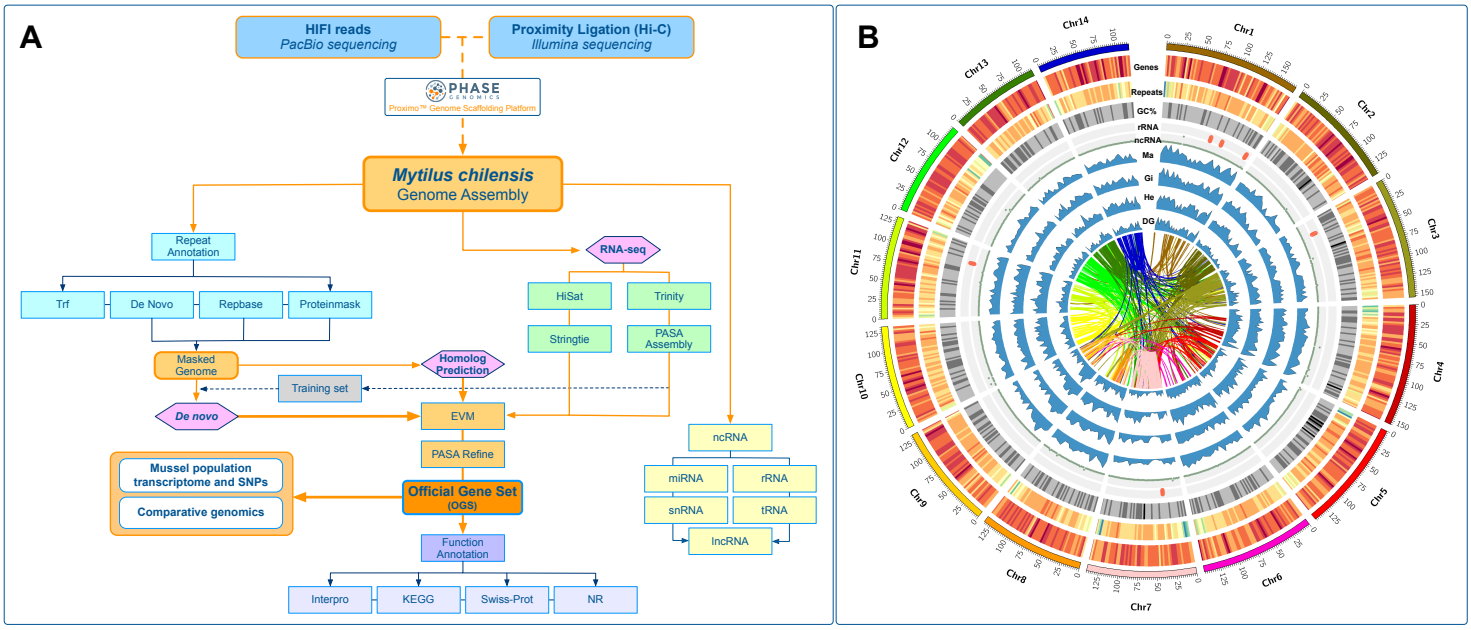

Figure 2

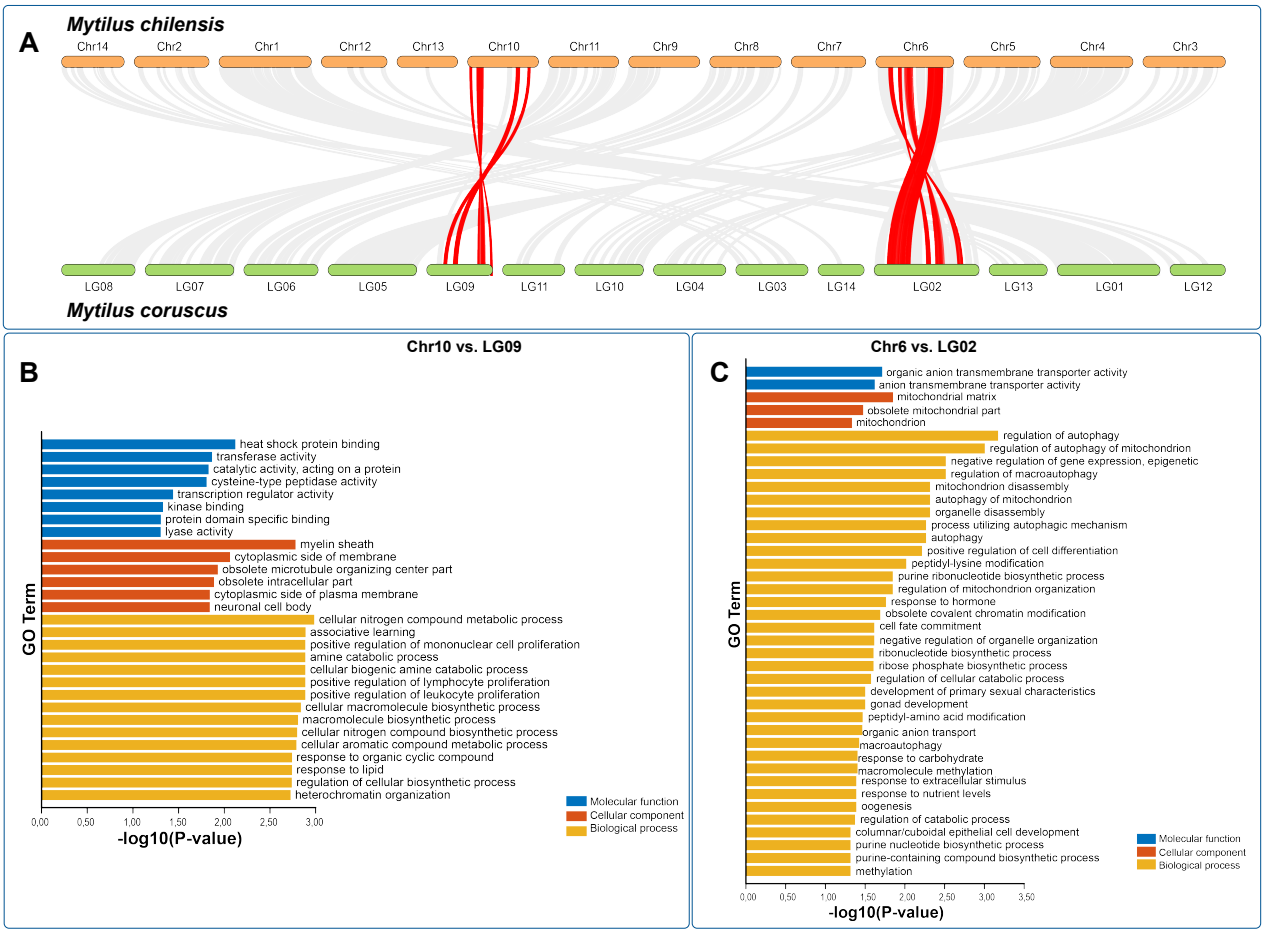

Figure 3

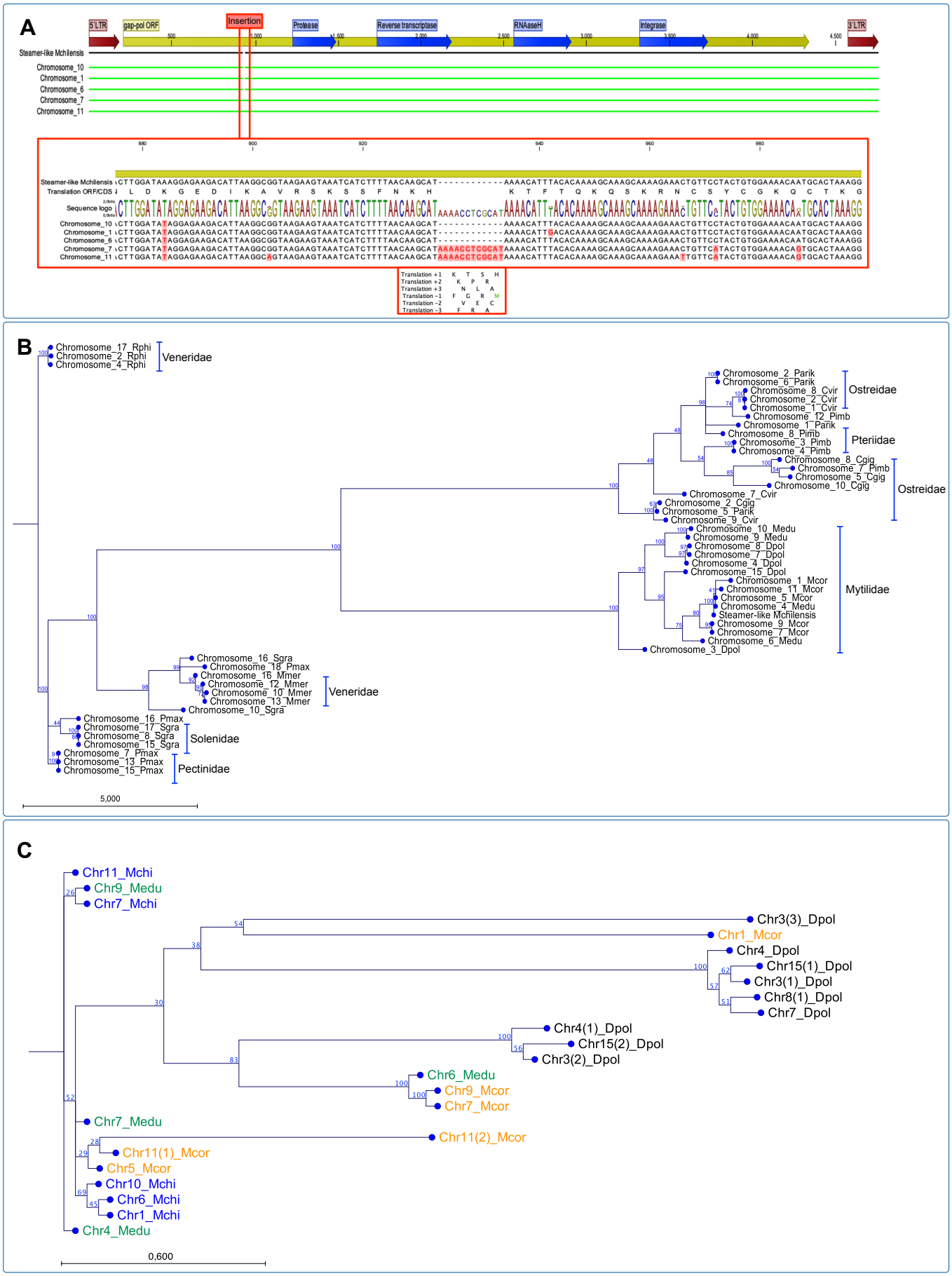

Figure 4

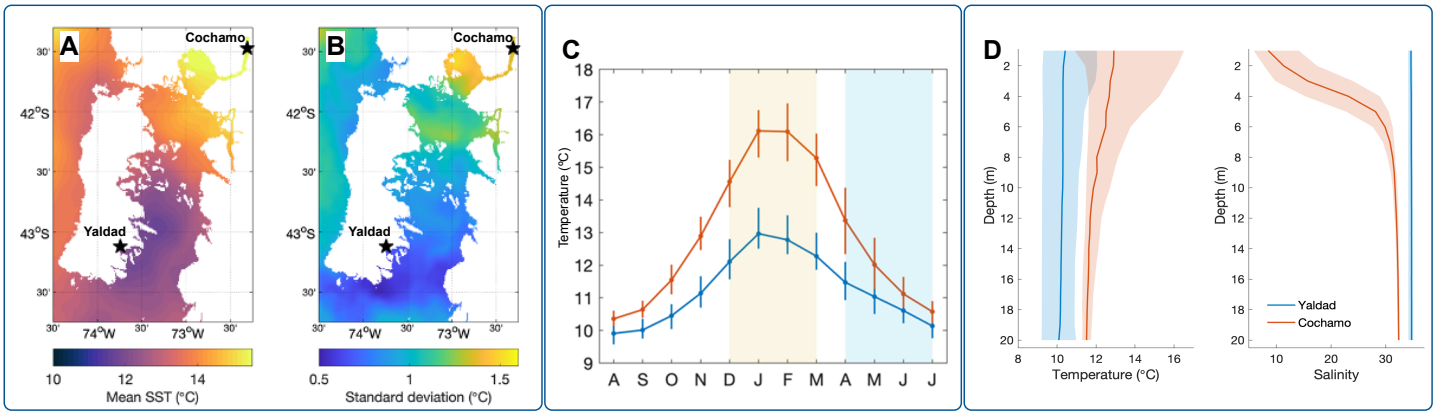

Figure 5

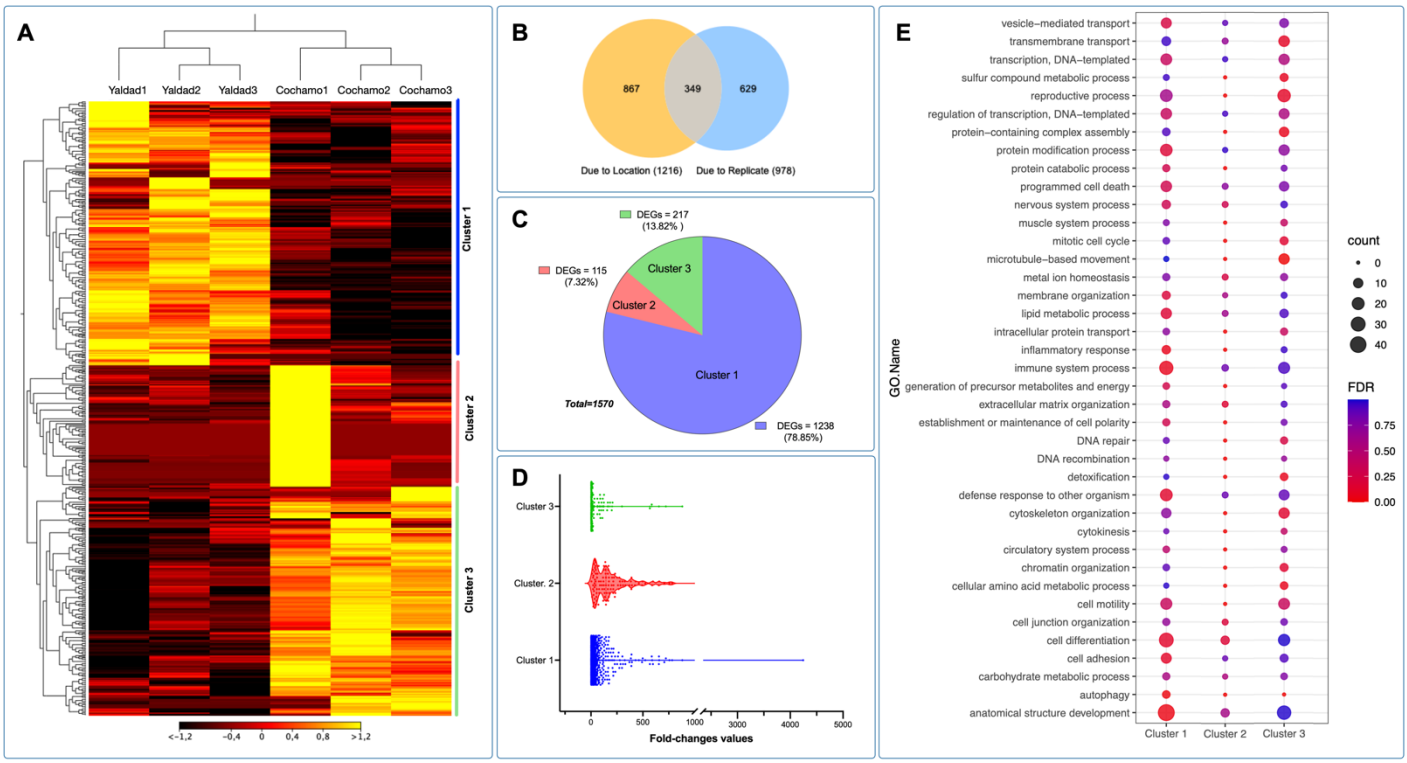

Figure 6

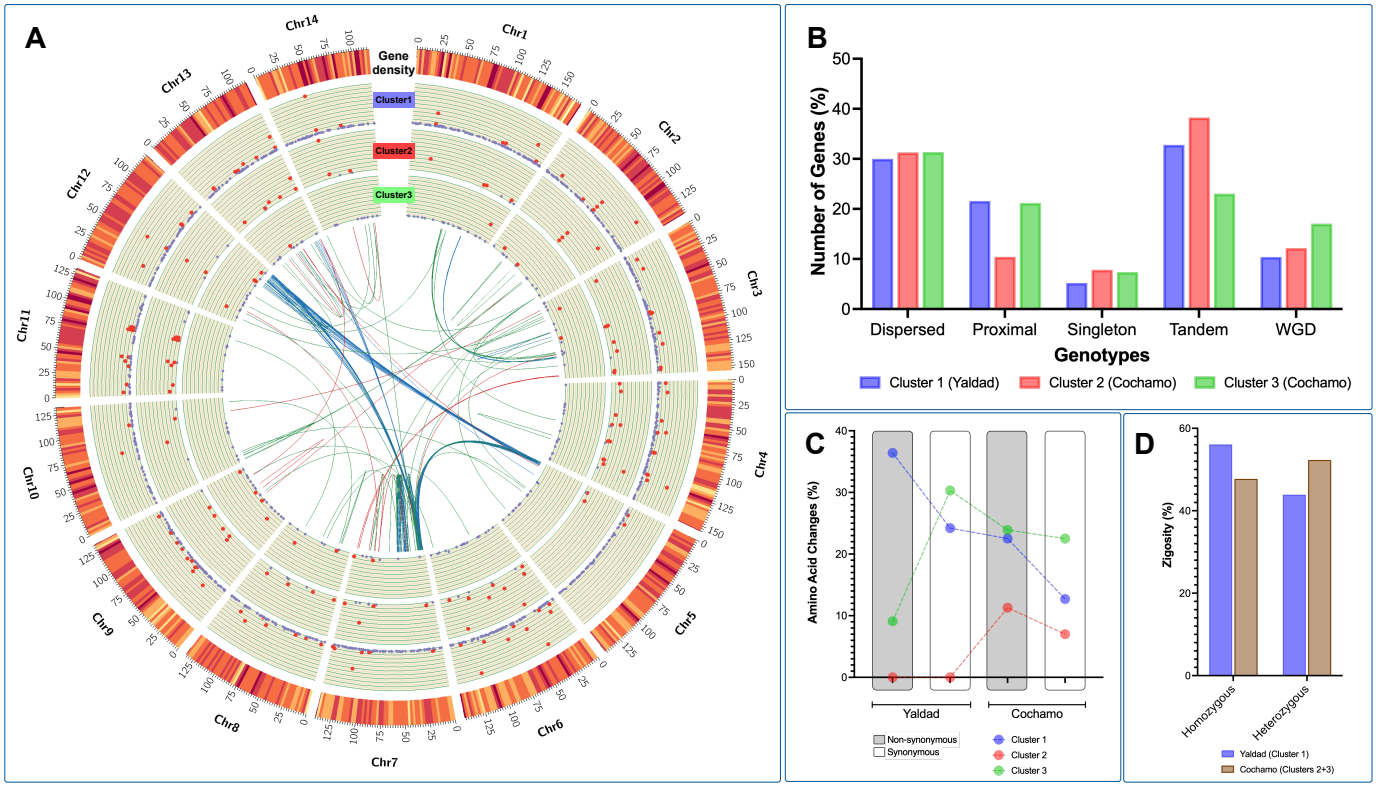

Figure 7

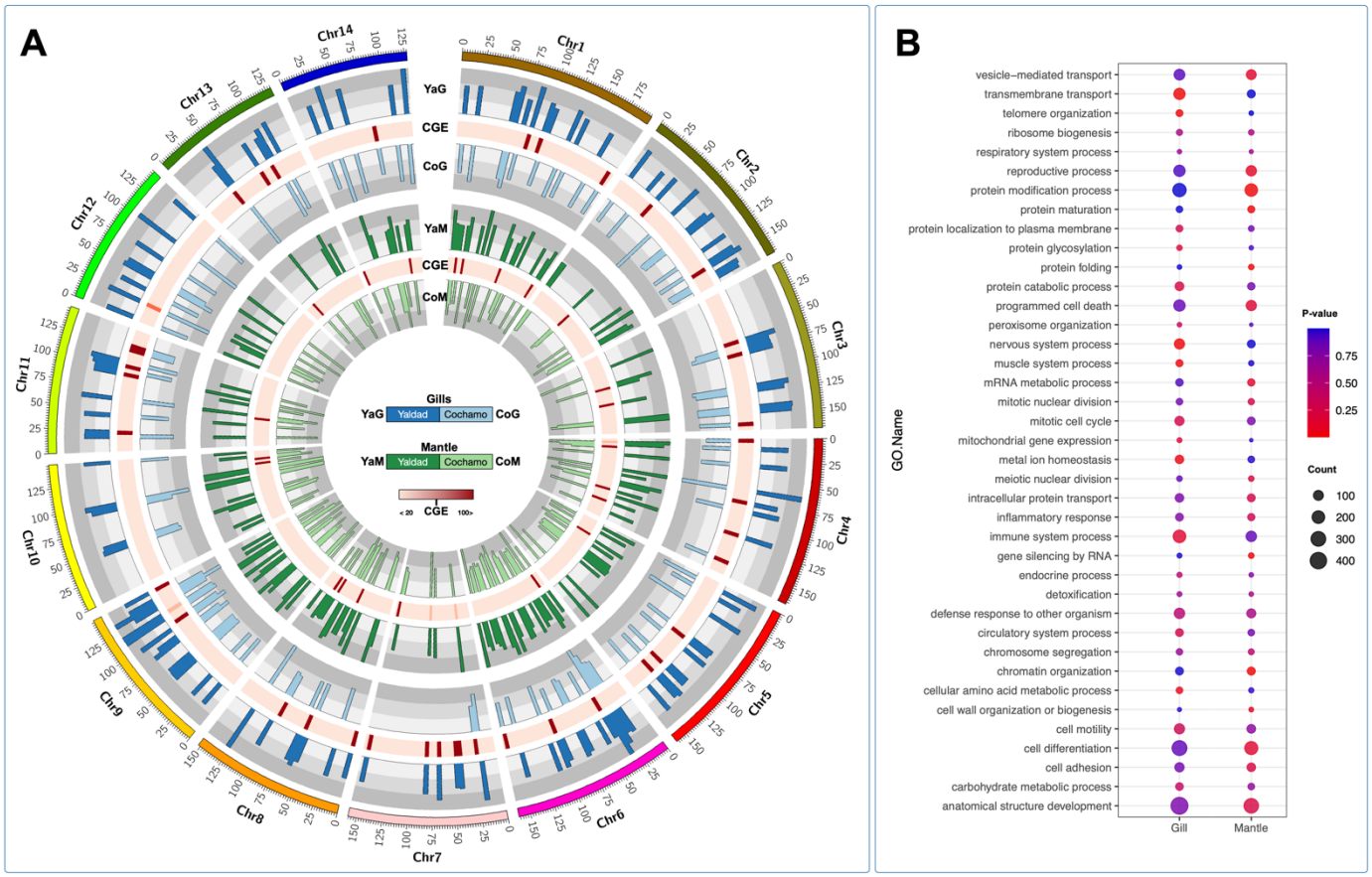

Figure 8

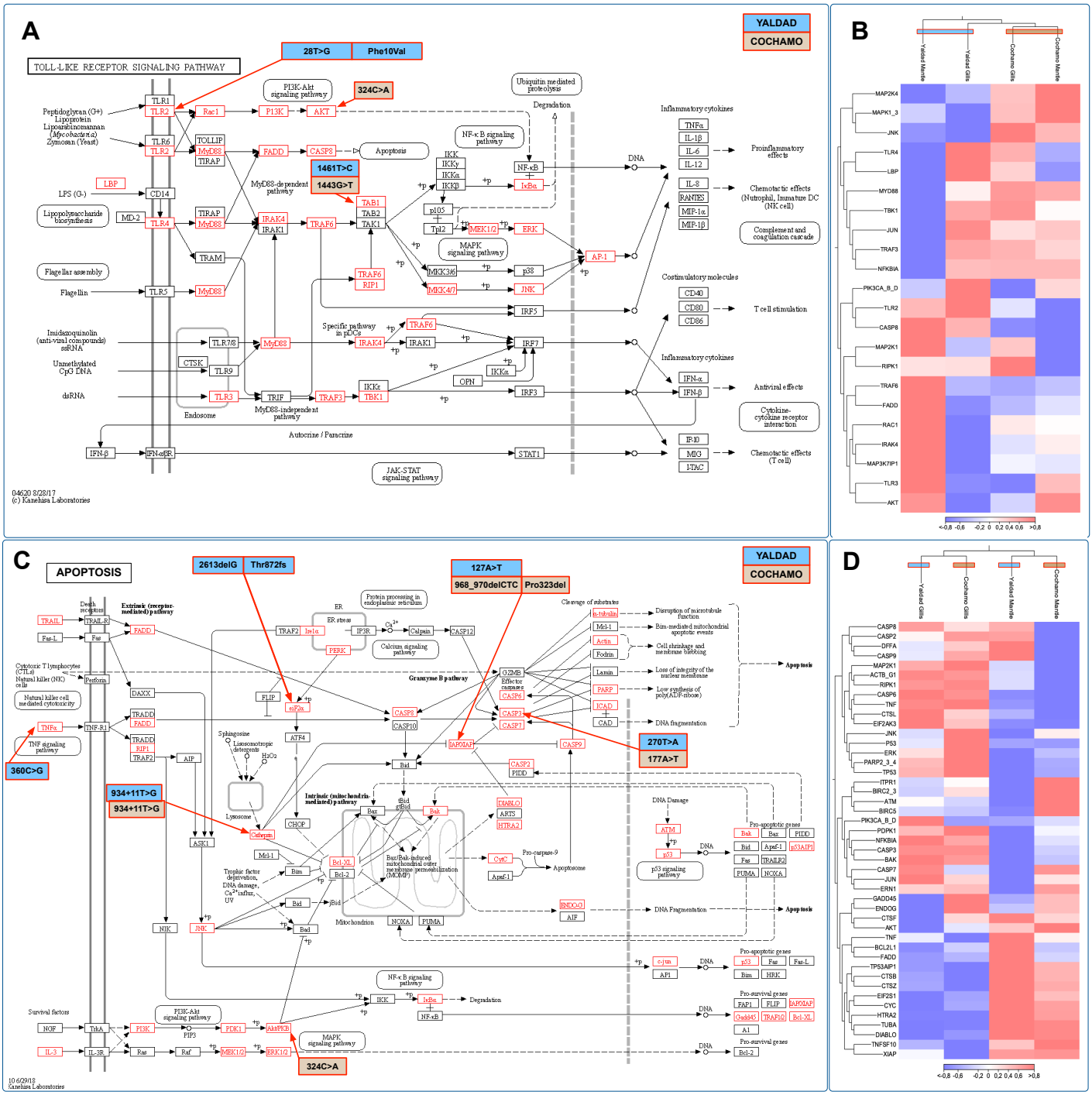

Figure 9

Table 1. Statistics of whole-genome sequence assembly and transcriptome analysis of the blue mussel *Mytilus chilensis* using Illumina, PacBio, and Hi-C.

Types	Method	No. of reads (Millions)	Library size	Length (Giga base pair)
Genome	PacBio SMRT	63M	20kb	882Gbp
	Illumina (Hi-C)	253M	150bp	37Gbp
Transcriptome	Illumina (Hemolymph)	38.3M	100bp	3,838Gbp
	Illumina (Mantle)	37.1M	100bp	3,714Gbp
	Illumina (Gills)	38.6M	100bp	3,865Gbp
	Illumina (Digestive gland)	28.2M	100bp	2,823Gbp

Table 2. Statistics of genome assembly using HiFi reads and proximity ligation analysis for *Mytilus chilensis*.

Label	Statistics
PacBio assembly	
Assembly size	2,191,715,088
Contig (CTG) N50	206,083
CTGs	13,762
Hi-C mapping	
Total read pairs (RPs) analyzed	253,342,981
High quality (HQ)* RPs	14,71%
Clustering usable HQ reads per contig (CTGs >5KB)*	1215.84
RPs >10KB apart (CTGs >10KB)	18.68%
Intercontig HQ RPs	44.68%
Same strand HQ RPs	21.50%
Split reads	37.80%

Table 3. *De novo* assembly of *M. chilensis* genome using proximity ligation (Hi-C).

Scaffold number	Number of contigs	Length (bp)
1	988	173,300,526
2	852	140,500,440
3	957	154,573,458
4	871	155,184,769
5	902	143,621,794
6	845	146,028,403
7	946	139,985,977
8	821	134,004,722
9	794	133,130,516
10	821	132,801,123
11	802	131,378,307
12	790	122,972,572
13	735	113,313,251
14	744	117,342,092
Total	**11,868**	**1,938,137,950**
N50		134,004,722

*Number of contigs in scaffolds: 11,868 (100% of all contigs in chromosome clusters, 86.24% of all contigs)

Table 4. Statistics of the classification results of repeat sequence from *M. chilensis* genome

Type	Length (bp)	% in Genome
DNA:EnSpm	1,578,431	0.08%
DNA:Harbinger	1,793,590	0.09%
DNA:Helitron	18,451,639	0.95%
DNA:MuDR	2,668,407	0.14%
DNA:Other	29,284,120	1.51%
DNA:TcMar	3,828,266	0.20%
DNA:hAT	4,997,238	0.26%
LTR:Copia	1,398,734	0.07%
LTR:Gypsy	12,056,219	0.62%
LTR:Other	12,605,602	0.65%
Low_complexity	30,762	0.00%
NonLTR:LINE	143,701,714	7.41%
NonLTR:SINE	4,939,193	0.25%
Tendem repeat: Satellite & Other	122,642,591	6.32%
Unknown	740,296,986	38.17%
All Repeat	1,100,273,492	56.73%

Table 5. Basic statistical results of gene structure prediction of relative species.

Species	Number of genes	Average Transcript length (bp)	Average CDS length (bp)	Average Exons per gene	Average Exon length (bp)	Average Intron length (bp)
Crassostrea gigas	63,340	17,784	2,008	12.6	3,301	286
Mytilus galloprovincialis	77,414	10,977	1,369	7.8	13,727	260
Mytilus coruscus	37,478	14,735	2,900	5.9	1,290	2,727
Dreissena polymorpha	68,018	13,316	2,603	4.5	1,632	1,194

Table 6. Statistical gene structure prediction for the blue mussel *M. chilensis* genome

	Type	Number	Average Transcript Length (bp)	Average CDS Length (bp)	Average Exons per Gene	Average Exon Length (bp)	Average Intron Length (bp)
De novo	Augustus	83,110	10,628	1,423	5.74	1,423	1,941
	SNAP	70,217	36,619	536	9.46	536	4,265
	GlimmerHMM	39,700	1,973	753	2.85	753	658
	Geneid	49,963	3,111	1,036	2.79	1,036	1,160
	Genscan	83,110	10,628	1,423	5.74	1,423	1,941
	GeneMark	81,256	11,018	1,301	5.82	1,301	270
Homolog	GMAP	250,833	16,107	370	3.10	1,649	6,891
	spaln	136,264	-	-	7.80	179	-
RNAseq	PASA	18,056	-	-	2.05	504	-
EVM		63,943	6,137	1,133	4.54	1,133	1,414
Final set		**34,530**	**6,531**	**1,377**	**4.92**	**1,377**	**1,316**

Table 7. Statistics of Non-coding RNA annotation for the *M. chilensis* genome

Type	Copy number	Average length(bp)	Total length(bp)	% of genome
miRNA	1365	92.8	126678	0.01%
rRNA	143	118.58	16957	0.00%
sRNA	275	83.68	23011	0.00%
snRNA	99	106.71	10564	0.00%
snoRNA	2267	89.38	202622	0.01%
tRNA	516	77.44	39957	0.00%
tRNA pseudogenes	123	72.96	8974	0.00%
All types	4795	89.42	428763	0.02%

Table 8. Statistics of gene function annotation for the *M. chilensis* genome

	Number	**Percentage (%)**
Swissprot	24,325	70.45
Nr	25,212	73.01
Nt	3,102	8.98
KEGG	22,425	64.94
eggNOG	23,181	80.57
GO	11,606	33.61
Pfam	60,887	176.33
Annotated	27,821	80.57
Unannotated	6,709	19.43
Total	34,530	100