Title

Reference genome and population genomics of Mytilus californianus

Author

Paggeot, Lisa Xie

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

Reference genome and population genomics of *Mytilus californianus*

A thesis submitted in partial satisfaction of the requirements for the degree of Master of Science

in

Quantitative and Systems Biology

by

Lisa Xie Paggeot

Committee in charge:
Prof. Michael N Dawson
Prof. Juris Grasis (Chair)
Prof. Maggie Sogin
Prof. Michele Nishiguchi

2023

The thesis of Lisa Xie Paggeot is approved, and is acceptable in quality and form for publication on microfilm and electronically:

Juris Grasis (Chair)

Maggie Sogin

Michele Nishiguchi

Michael N Dawson

University of California, Merced
2023

Acknowledgments

Thank you to my committee members for their patience, guidance, and mentorship throughout the process of putting together this thesis. I appreciate all your support through this unpredictable process that is graduate school.

First, I would like to thank my family for being my strongest support system and for their continuous encouragement and support. Thank you to my mom who shows her support by sending me delicious foods and treats. Thank you to my dad who knows what to say when I am at my lowest point. Thank you to my sister who comes to visit me and encourages me to do my best. I could not be where I am today without them.

Second, I would also like to thank all my friends and colleagues who have helped me here at University of California, Merced. Dr. Melissa DeBiasse, Dr. Lauren Schiebelhut, Dr. Vanessa Guerra, Nattanon Wutthituntisil, Bailey Carlson, Karly Higgins, Sam Fellows, Anabelle Klovrza, Samantha Donohoo – thank you for all the support and guidance you have provided. I would like to give a special thank you to Nattanon Wutthituntisil, who is not only my dear friend but also someone I look up to.

Third, thank you to my funding resources. This thesis is a contribution of the Marine Networks Consortium (PIs Michael N Dawson, Rachael A. Bay) as part of the California Conservation Genomics Project (PI: H. Bradley Shaffer), with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

Lastly, I would like to thank my advisor – Dr. Mike Dawson. Thank you for guiding me through this transformation as I learn new skills and techniques on how to be a better researcher and educator. Thank you for your never-ending patience as I navigate through numerous obstacles while balancing many responsibilities. Thank you for giving me new insight on the high standards we should hold as scientists. And lastly, thank you for all the discussions, field work trips, stories, and laughs we have shared together. I could not have asked for a better advisor.

Table of Contents

List of Figures

taken for the upper and lower limits. (C) A bed of *M. californianus* mussels at Mill Creek, Tehama County, California. Globe image in panel A taken from https://www.clipsafari.com/clips/o29983-globe-showing-north-america. Panel C photo credit: Michael N Dawson.

Figure 4: Graphs of *Mytilus californianus* Fst against position on genome. (A) Fst across SNP loci when samples are grouped by upper and lower limit. (B) Fst across SNP loci when samples are grouped by northern and southern sample sites. Light gray points represent Fst values calculated from SNP loci on odd numbered chromosomes, dark gray points represent Fst values calculated from SNP loci on even numbered chromosomes. Outlier loci Fst were calculated to be significantly different (q-value < 0.05) from the rest of the Fst in the data frame are circled in red.

List of Tables

Reference genome and population genomics of *Mytilus californianus*

Lisa Xie Paggeot

Master of Science

University of California 2023

Committee chair: Dr. Juris Grasis

Abstract

Analyzing the relationship between evolutionary forces in a dynamic environment has been a challenge for marine population genetics. We explored this issue using a reference genome and population genomic analyses of the California ribbed mussel, *Mytilus californianus*—an ecosystem engineer that inhabits an intertidal zone shaped by marine and terrestrial weather and climate. Seventeen samples were collected at the upper and lower limits of intertidal environments from five sites across California. A chromosome-scale reference genome was assembled, and population genomic analyses were conducted to investigate genetic differentiation across latitudinal and intertidal gradients. The reference genome is 1.65 Gb long, with N50 sequence length of 118 Mb, and 86% complete. Low population differentiation aligned with previous studies, indicating very high migration ($F_{ST} \leq 0.0003$). Nonetheless, allele frequencies at a small number (48–103) of outlier loci are differentiated alongside the many environmental factors that change across latitudinal and intertidal gradients.

Chapter 1: Reference genome of *Mytilus californianus*

Abstract

The California ribbed mussel, *Mytilus californianus*, is an ecosystem engineer crucial for the survival of many marine species inhabiting the intertidal zone of California. Here, we describe the first reference genome for *M. californianus* and compare it to previously published genomes from three other *Mytilus* species: *M. edulis, M. coruscus,* and *M. galloprovincialis*. The *M. californianus* reference genome is 1.65 Gb in length, with N50 sequence length of 118Mb, and an estimated 86.0% complete single copy genes. Compared to the other three *Mytilus* species, the *M. californianus* genome assembly is the longest, has the highest N50 value, and the highest percentage complete single copy genes. This high-quality genome assembly provides a foundation for population genetic analyses that will give insight into future conservation work along the coast of California.

Introduction

Ecosystem engineers modify the physical environment in a way that changes available habitats (Jones et al. 1994). Their modifications can lead to alteration, expansion, or formation of novel habitats and promote the success of taxa in their vicinity (Dayton 1972; Bruno et al. 2003). The California ribbed mussel, *Mytilus californianus*, an intertidal species distributed along the west coast of North America from the Aleutian Islands to Isla Socorro, Mexico (Figure 1a; Soot-Ryen 1955), is an ecosystem engineer known for its role in the origin of the Keystone Species concept (Figure 1b; Paine 1966). Ribbed mussels form large compact beds, attached to underlying rock, that generate many protected interstices inhabited by other organisms (Paine 1994; Gutierrez et al. 2003). The mussel-rock attachment is by byssal threads (Waite 2017), a feature that is widespread among marine bivalves though stronger in *M. californianus* than in its congeners (Holten-Andersen et al 2009) and different in mytilids versus other bivalves (Pearce and LaBarbera 2009).

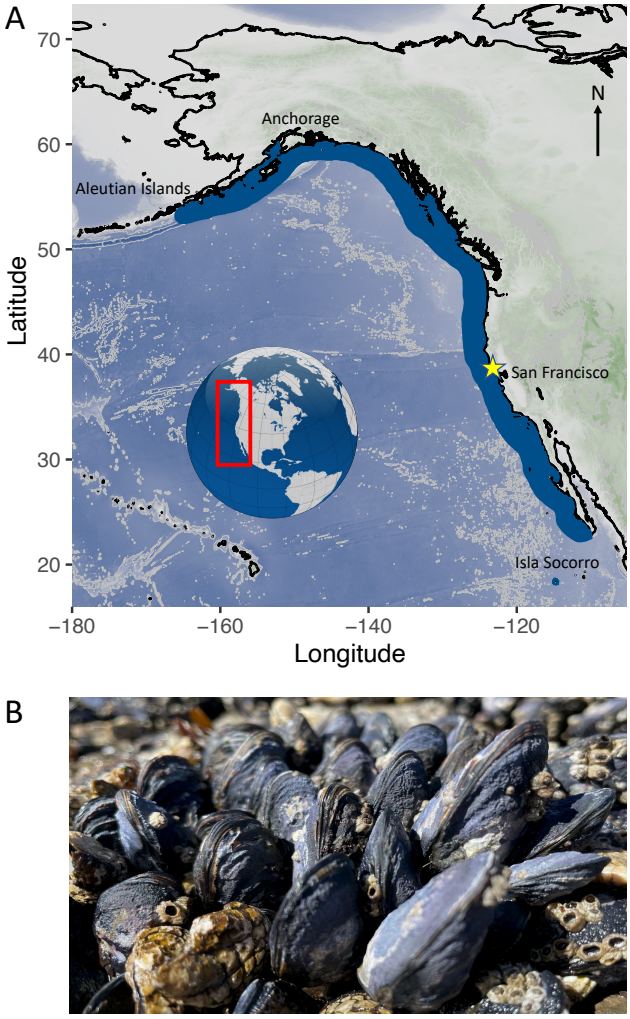

Figure 1: The distribution of the California ribbed mussel, *Mytilus californianus*. (A) World map (inset) and continental map showing the reported geographic range of *M. californianus*, from the Aleutian Islands

to Isla Socorro, Mexico (Soot-Ryen 1955). Other authors have since reported a narrower distribution, such as from Baja California to British Columbia (Sagarin and Somero 2006). Yellow star indicates the geographic location of the sample collected and used to generate the genome. (B) A bed of *M. californianus* mussels at Duxbury Reef, Marin County, California. Note the other species that inhabit the ribbed mussel beds such as the acorn and stalked barnacle. Globe image in panel A taken from https://www.clipsafari.com/clips/o29983-globe-showing-north-america. Panel B photo credit: Michael N Dawson.

In addition to its role in important ecological functions, *Mytilus* has been a rich model system for breaking new ground in the field of marine genetics. For example, analyses of two polymorphic loci (*Lap* and *Pgi*) led Tracey et al. (1975) to suggest that homozygote excess is more apparent in juveniles than in adults due to breeding subpopulation structure within the reproductive population. Analyses of hybrid zones where distributions of *Mytilus* species overlap across the Atlantic and Pacific oceans advanced understanding of evolutionary processes in high dispersal species (Riginos and Cunningham 2004, Springer and Crespi 2006). And, exploiting the bi-parental inheritance of mtDNA in mussels, Śmietanka et al. (2010) sequenced both mitochondrial genomes of *M. trossulus* to identify that the mitogenome, which often has been treated as neutral in phylogeographic studies, likely contains multiple adaptive mutations. However, many of these previous genetic studies in *Mytilus* relied on a small number of loci, therefore limiting inference to small portions of the genome. Extending the genomic resources available would help us better understand the breadth of genomic evolution and its consequences in ecological contexts. Such genomic resources also will illuminate the genetic architecture underlying traits related to physiological tolerance and mechanisms of reproductive compatibility as well as adaptive potential of populations and species in the face of climate change (Place et al. 2008; Savolainen et al. 2013; Kinoshita and Seki 2014).

Here, I present a reference genome for *M. californianus* and compare it with the previously published genomes of three other *Mytilus* species whose ranges cover the Mediterranean Sea (*M. galloprovincialis*), the coasts of China, Korea, Japan (*Mytilus coruscus*), and the western Pacific Ocean (*Mytilus edulis*). The *M. californianus* reference genome will contribute to conservation and management of regional biodiversity in California (Shaffer et al. 2022) and has the potential to provide new insights into responses to anthropogenic events and help identify consequences of local and/or regional genetic variation. In addition, the *M. californianus* reference genome will give us a better understanding of comparative genomics across multiple *Mytilus* species living around the globe.

Methods

Biological materials

One California mussel was collected from McClures Beach, Marin County, California, USA (38.1813, -122.9644) on July 22, 2020 by Michael N Dawson. The specimen was transported live to the University of California, Berkeley, where subsamples of dissected

tissue were flash frozen in liquid nitrogen. A voucher specimen (M0D057914Y) is archived in the Dawson Lab collection at University of California, Merced.

Nucleic acid extraction, library prep, and sequencing

DNA was extracted and libraries were prepared following standard methods established for marine invertebrates by the California Conservation Genomics Project (CCGP) as described in DeBiasse et al. (2022) and available in the supplementary information. In brief, we extracted high molecular weight (HMW) DNA from 40mg of the mantle using the Nanobind Tissue Big DNA kit (Pacific BioSciences [PacBio], CA) with the following minor modifications: we performed an additional wash with the CT buffer for the tissue homogenate and pelleted it by centrifuging at 18000 x *g* (4°C for 5 minutes) to remove any residual buffer before proceeding with the lysis step. We prepared the HiFi SMRTbell library using the SMRTbell Express Template Prep Kit v2.0 (PacBio) and sequenced the 15–20Kb average HiFi SMRTbell library using three SMRT® Cell 8M Trays. We prepared the Omni-C library using the Dovetail™ Omni-C™ Kit (Dovetail Genomics, CA) according to the manufacturer's protocol with the following modifications: we optimized the digest with 2 uL of Nuclease Enzyme Mix input and the proximity ligation reaction with 500ng DNA input. We sequenced the library at the Vincent J. Coates Genomics Sequencing Laboratory at University of California, Berkeley (Berkeley, CA) on an Illumina NovaSeq platform (Illumina, CA) targeting approximately 100 million 150Bp paired end reads per gigabase of genome size.

Nuclear and mitochondrial genome assembly

We assembled the nuclear genome of *M. californianus* following the CCGP assembly protocol Version 3.0 (Table 1, Lin et al. 2022). We assembled the mitochondrial genome from the PacBio HiFi reads using the reference-guided pipeline MitoHiFi (https://github.com/marcelauliano/MitoHiFi) (Allio et al., 2020) and the *Mytilus trossulus* mitochondrial genome (NCBI:GU936626.1) as the starting reference sequence. Full assembly details are available in Lin et al. 2022.

Table 1: Assembly Pipeline and Software Usage. Software citations are listed in the text. § Options detailed for non-default runs.

Assembly	**Software and options** §	**Version**
Filtering PacBio HiFi adapters	HiFiAdapterFilt	Commit 64d1c7b
K-mer counting	Meryl (k=21)	1
Estimation of genome size and heterozygosity	GenomeScope	2
De novo assembly (contiging)	HiFiasm (HiC mode, --primary, p_ctg and a_ctg output)	0.16.1-r375

Remove low-coverage, duplicated contigs	purge_dups	1.2.6
Scaffolding		
Omni-C Scaffolding	SALSA (-DNASE, -i 20, -p yes)	2
Gap closing	YAGCloser (-mins 2 -f 20 -mcc 2 -prt 0.25 -eft 0.2 -pld 0.2)	Commit 20e2769
Omni-C Contact map generation		
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188
SAM/BAM processing	samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	Cooler	0.8.10
Matrix balancing	hicExplorer (hicCorrectmatrix correct --filterThreshold -2 4)	3.6
Contact map visualization	HiGlass	2.1.11
	PretextMap	0.1.4
	PretextView	0.1.5
	PretextSnapshot	0.0.3
Organelle assembly		
Mitogenome assembly	MitoHiFi (-r , -p 50, -o 1)	2 Commit c06ed3e
Genome quality assessment		
Basic assembly metrics	QUAST (--est-ref-size)	5.0.2
Assembly completeness	BUSCO (-m geno, -l mollusca)	5.0.0
	Merqury	1
Contamination screening		
Local alignment tool	BLAST+	2.10
General contamination screening	BlobToolKit	2.3.3
Comparison analysis		
Assembly completeness	BUSCO (https://gvolante.riken.jp/)	5.0.0

Genome size estimation and quality assessment
We estimated genome size, heterozygosity, repeat content, sequencing error, and genome assembly completeness following standard protocols established by the CCGP and described in detail by Lin et al. (2022). We assessed assembly quality of the *M. californianus* primary and alternate genomes using BUSCO (Simão et al. 2015; Seppey et al. 2019) with the Mollusca ortholog database (mollusca_odb10) which contains 5,295 genes. Following data availability and quality metrics established by Rhie et al. (2021), we used the derived genome quality notation x.y.Q, where, x = log10[contig NG50]; y = log10[scaffold NG50]; Q = Phred base accuracy QV (quality value); C = % genome represented by the first 'n' scaffolds, following a known karyotype of 2n = 28 (Ahmed and Sparks 1970). Quality metrics for the notation were calculated on the primary assembly.

Under the assumption that the longest scaffolds contain the majority of the genome sequence and represent the putative chromosomes, we generated a histogram of scaffold lengths for (a) the largest 20 scaffolds and (b) all scaffolds and then performed a k-means clustering in R (R Core Team 2020) to test if a drop-off in scaffold size corresponded to the number of chromosomes predicted for *Mytilus* mussels, including *M. californianus* (Ahmed and Sparks 1970, Pérez-García et al. 2014).

Comparison to previously published Mytilus genomes
We downloaded the complete genome sequences for the mussels *Mytilus edulis* (GCA_019925275.1), *Mytilus galloprovincialis* (GCA_900618805.1), and *Mytilus coruscus* (GCA_017311375.1) from GenBank. We calculated common metrics of assembly completeness across the three published *Mytilus* genomes and the *M. californianus* genome generated here using BUSCO [Version 5.0.0] and the Mollusca ortholog database (mollusca_odb10) as implemented in gVolante (Nishimura et al. 2019).

Results

Nucleic acid extraction, library prep, and sequencing
Extracted HMW DNA had purity 260/280 = 1.83 and 260/230 = 1.91, concentration 169 ng/μl (19.4 μg total), and good integrity with >84% of DNA fragments being 120 Kb or more. Sequencing resulted in 5.1 million PacBio HiFi reads representing ~45 fold coverage (N50 read length 14,054 bp; minimum read length 45 bp; mean read length 13,449 bp; maximum read length of 50,337 bp) based on the Genomescope2.0 genome size estimation of 1.576 Gb. Based on PacBio HiFi reads, we estimated 0.09 % sequencing error rate and 2.73% nucleotide heterozygosity rate. The Illumina sequencing yielded 203.3 million 150 bp paired end Omni-C reads.

Nuclear and mitochondrial genome assembly

We generated a *de novo* nuclear genome assembly of the California mussel (xbMytCali1) for which assembly statistics are reported in Table 2 and Figure 2B. The k-mer spectrum output shows a bimodal distribution with two major peaks, at ~21 and ~42-fold coverage, where peaks correspond to homozygous and heterozygous states respectively of a diploid species. The Omni-C contact map suggests that the primary assembly is highly contiguous (Figure 2C). The alternate assembly, which consists of sequence from heterozygous regions, is less contiguous (Figure S1). We have deposited both the primary and alternate scaffolds to NCBI.

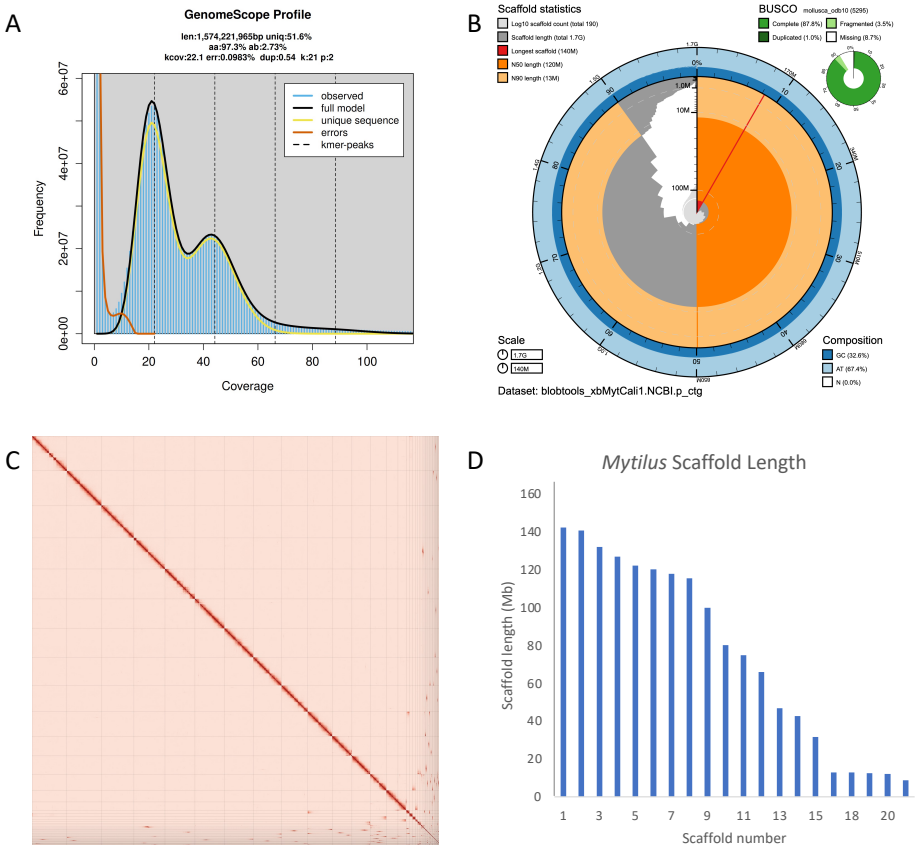

Figure 2. Visual overview of genome assembly metrics. (A) K-mer spectra output generated from PacBio HiFi data without adapters using GenomeScope2.0. The bimodal pattern observed corresponds to a diploid genome. K-mers covered at lower coverage but higher frequency correspond to differences between haplotypes, whereas the higher coverage but lower frequency k-mers correspond to the similarities between haplotypes. The pattern observed corresponds to a k-mer profile for a highly heterozygous species. (B) BlobToolKit Snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *Mytilus californianus* primary assembly (xbMytCali1). The plot circle represents the full size of the assembly. From the inside-out, the central plot covers length-related metrics. The red line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in gray starting from the outside of the central plot. Dark and light orange arcs show the scaffold N50 and scaffold N90 values. The central light gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. White regions in this area reflect the proportion of Ns in the assembly The dark vs. light blue area around it shows mean, maximum and minimum GC vs. AT content at 0.1% intervals (Challis et al. 2020). (C) Omni-C contact maps for the primary genome assembly generated with PretextSnapshot. Omni-C contact maps translate proximity of genomic regions in 3-D space to

contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between two such regions. Scaffolds are separated by black lines, wherein higher density of black lines corresponds to higher levels of fragmentation. (D) Histogram of the 20 longest scaffold reads for *Mytilus californianus*. Scaffold size is given in megabase pairs (Mb).

We generated one mitochondrial genome assembly, 16,730 bp long. The base composition of the final assembly version is A=28.41%, C=13.37%, G= 22.71%, T= 35.48%, and consists of 23 transfer RNAs and 13 protein coding genes.

Table 2: Sequencing and assembly statistics, and accession numbers. * Assembly quality code x.y.Q.C derived notation, from (Rhie et al. 2021). x = log10[contig NG50]; y = log10[scaffold NG50]; Q = Phred base accuracy QV (Quality value); C = % genome represented by the first 'n' scaffolds, following a known karyotype of 2n=28. Quality code for all the assembly denoted by primary assembly (xbMytCali1.0.p). BUSCO Scores. (C)omplete and (S)ingle; (C)omplete and (D)uplicated; (F)ragmented and (M)issing BUSCO genes. n, number of BUSCO genes in the set/database. Bp: base pairs. § Read coverage and NGx statistics have been calculated based on the estimated genome size of 1.576 Gb. ‡ (P)rimary and (A)lternate assembly values

Bio Projects & Vouchers	CCGP NCBI BioProject		PRJNA720569	
	Genera NCBI BioProject		PRJNA765636	
	Species NCBI BioProject		PRJNA777198	
	NCBI BioSample		SAMN24505264	
	Specimen identification		M0D057914Y	
	NCBI Genome accessions		Primary	Alternate
	Assembly accession		GCA_021869535.1	GCA_021869935.1
	Genome sequences		JAKFGE000000000	JAKFGF000000000
Genome Sequence	PacBio HiFi reads	Run	3 PACBIO_SMRT (Sequel II), 5.8 M spots, 81.4 G bases, 48.5 Gb	
		Accession	SRR18000156	
	Omni-C Illumina reads	Run	2 Illumina HiSeq X Ten runs: 203.3 M spots, 61.4 G bases, 20.3 Gb	
		Accession	SRR18000154-55	
Genome Assembly Quality Metrics	Assembly identifier (Quality code *)		xbMytCali1 (7.7.Q60.C86)	
	HiFi Read coverage §		45.05X	
			Primary	Alternate
	Number of contigs		498	38,455
	Contig N50 (bp)		16,323,199	167,345
	Contig NG50 (bp)		17,177,226	251,220
	Longest Contigs (bp)		57,759,394	3,992,896
	Number of scaffolds		176	38,315

	Scaffold N50 (bp)			117,871,512		169,002	
	Scaffold NG50 (bp)			120,330,192		253,008	
	Largest scaffold (bp)			142,435,203		3,992,896	
	Size of final assembly (bp)			1,651,966,901		2,213,012,655	
	Gaps per Gbp (#Gaps)			37 (324)		63 (140)	
	Indel QV (Frame shift)			51.43139759		51.43139759	
	Base pair QV			65.0891		58.9335	
				Full assembly = 60.6658			
	k-mer completeness			67.2429		68.2222	
				Full assembly = 92.3895			
	BUSCO completeness (mollusca) n=5,295		**C**	S	D	F	M
		P ‡‡	86.00%	85.10%	0.90%	3.30%	10.70%
		A ‡‡	85.20%	80.00%	5.20%	4.50%	10.30%
	Organelles		1 Complete mitochondrial sequence			CM038905.1	

Genome size estimation and quality assessment

The primary assembly consists of 176 scaffolds spanning 1.65 Gb with contig N50 of 16.32 Mb, scaffold N50 of 118 Mb, largest contig of 57.8 Mb, and largest scaffold of 142.4 Mb. The final genome size is close to the estimated values from the Genomescope2.0 k-mer spectrum. The primary assembly has a BUSCO completeness score of 86.0% using the Mollusca gene set, a per base quality (QV) of 65, a k-mer completeness of 67.2 and a frameshift indel QV of 51.43. The alternate assembly has a BUSCO completeness score of 85.20% using the Mollusca gene set, a per base quality (QV) of 59, a k-mer completeness of 68.2 and a frameshift indel QV of 51.43. The scaffold length histogram showed that the largest size differences were between scaffolds 9 and 10 (19.6 Mb) and scaffolds 12 and 13 (19.1 Mb) (Figure 2D), the latter corresponding to the split indicated by the k-means clustering, which placed scaffolds 1–12 into a cluster and scaffolds 13–20 into a second cluster.

Comparison to previously published Mytilus genomes

Compared to the most recent assemblies for three other *Mytilus* species (Table 3), the M. californianus genome assembly produced here is the most contiguous (i.e., contained in the smallest number of scaffolds), has the largest N50 value, and is the longest (1,651,966,901 bp), slightly exceeding *M. edulis* (1,651,313,236 bp). The *M. californianus* assembly also has superior BUSCO metrics for core gene completeness (86%), duplication (0.91%), fragmentation (3.3%), and missingness (10.7%). Complete single copy statistics for *M. coruscus*, *M. edulis*, and *M. galloprovincialis* are 81%, 79%, and 71% respectively; duplication statistics are 1.1%, 7.7%, and 5.4%, respectively; fragmentation statistics are 3.4%, 4%, and 4.8%, respectively; missingness statistics are 15.6%, 16.8%, and 24.7%, respectively (Table 3).

Table 3: BUSCO scores for *Mytilus californianus* compared with *M. coruscus, M. galloprovincialis,* and *M. edulis*. *Version of PacBio sequencing chemistry not reported

	Mytilus californianus	*Mytilus coruscus*	*Mytilus edulis*	*Mytilus galloprovincialis*
Citation	This thesis	Yang et al. 2021	Unpublished	Gerdol et al. 2020
GenBank ID	GCA_021869935.1	GCA_017311375.1	GCA_019925275.1	GCA_900618805.1
Sequencing method/technology	PacBio HiFi, Omni-C	Oxford Nanopore, Illumina, Hi-C	PacBio*, Omni-C	PacBio HiFi, Illumina
Assembly length (Gb)	1.65	1.57	1.65	1.28
Sequences	176	4434	1119	10,577
GC-content	32.57	32.45	32.3	32.14
N50 sequence length (Mb)	118	99.5	116.5	0.21
Complete single copies	4553 (86%)	4288 (81%)	4191 (79%)	3735 (71%)
Complete + partial single copies	4730 (89%)	4468 (84%)	4403 (83%)	3988 (75%)
BUSCO duplicated genes	48 (0.91%)	58 (1.1%)	408 (7.7%)	286 (5.4%)
BUSCO fragmented genes	175 (3.3%)	180 (3.4%)	212 (4%)	254 (4.8%)
BUSCO missing genes	567 (10.7%)	827 (15.6%)	892 (16.8%)	1307 (24.7%)

Discussion

The ecological and economic value of *Mytilus* species across the globe has motivated generation of multiple genomic resources for this genus (Murgarella et al. 2016, Yang et al. 2021, BioProject: PRJNA740305). These resources have improved in quality with advances in sequencing and assembly algorithms as can be seen comparing the first mussel genome for *M. galloprovincialis* produced with Illumina short reads (1.74 million scaffolds, N50 = 2651 bp, Murgarella et al. 2016), to the current *M. galloprovincialis* assembly produced with PacBio HiFi long reads and Illumina short reads (10,777 scaffolds, N50 = 32.14 Mb, Gerdol et al 2020). A key advance has been scaffolding assemblies with proximity data from Hi-C or Omni-C libraries, which can greatly increase contiguity. For example, for *M. coruscus*, a genome assembled from Oxford Nanopore Technology (ONT) long reads and Illumina short reads by Li et al. (2020) is contained in 10,484 scaffolds with an N50 of 898 Kb while a genome produced by Yang et al. (2021) using ONT, Illumina, and Hi-C scaffolding reduced scaffold number to 4,434 with a 99 Mb N50. Interestingly, in addition to improvements in contiguity and completeness (as determined by BUSCO metrics), assemblies scaffolded with Omni-C or Hi-C reads have less variation in assembly size (1.57-1.65Gb) than those not scaffolded with proximity data (1.28-1.9Gb), with a possible explanation being that proximity data help resolve highly repetitive areas of the genome, leading to more accurate and precise assembly sizes.

Chromosome-scale reference genomes are powerful tools because their contiguity and completeness provide more power to test important ecological and evolutionary hypotheses than a genome assembly that is fragmented and missing genes or other key genomic features. Previous studies using karyotyping have shown that *Mytilus* mussels, including *M. californianus*, *M. edulis*, *M. galloprovincialis*, and *M. trossulus* have 14 chromosomes (Ahmed and Sparks 1970; Pérez-García et al. 2014). Yang *et al.* (2021) commensurately found 90.9% of *M. coruscus* genome sequence scaffolds in their assembly mapped to 14 chromosomes based on Hi-C proximity data. The Omni-C proximity data we produced here also suggests 14 chromosomes in *M. californianus* (i.e., 14 major bins along the diagonal containing each proximity read and its mate in Figure 2C); k-means clustering suggests a similar number of chromosomes (12; Figure 2d) though the efficacy of this approach may be influenced by the assembly. Mollusc genomes are known to be highly repetitive (Murgarella et al. 2016) and heterozygous (Koehn and Gaffney 1984, Diz et al. 2008), which complicates the assembly process, and likely explains why we recovered more than 14 genomic scaffolds. Regardless, the assembly we produced here is the most complete of the *Mytilus* species available and is a powerful resource for comparative and population genomics.

While mussels have biparental inheritance of the mitochondrial genome (Ladoukakis et al. 2002, Mizi et al. 2005) and we therefore would expect to assemble two scaffolds representing the maternal and paternal contributions (Murgarella et al. 2016), we generated only one. Assembling phased genome assemblies is a bioinformatic challenge

(Chin et al. 2016, Mostovoy et al. 2016), particularly when heterozygosity is high, parental sequences are not available, and an intrinsic part of pipelines is to purge duplicates. Our single mitochondrial genome assembly therefore may represent either only one of the two genomes or a chimera of the two parental contributions (Table 2). Future research efforts should engage with developing tools for better resolving maternal and paternal assemblies individually for mitochondrial, as well as nuclear, genomes. Notwithstanding the preceding caveats, the new reference genome for *M. californianus* will facilitate multiple fields of study. For example, understanding the structure and resilience of the byssal threads may be enhanced by discovering genes encoding the formation of the threads, in a manner paralleling the genome-enabled analyses of structural genes in green mussels, *Perna viridis* (Inoue et al 2021). Additionally, the reference genome can be coupled with comparative ecological studies to give a well-rounded understanding of how functional traits diverged in different environments, e.g. Pearce and LaBarbera (2009) showed that epifaunal species have thicker and more extensive byssal threads than infaunal species, suggesting a correlation with life habits between the two groups of organisms. Furthermore, analyzing the *Mytilus* genome can expand our knowledge of stress response and immune defense in bivalves, for example elucidating the unique bivalve gene families involved in heat shock proteins (Takeuchi et al 2016).

Reference genomes can be useful in multiple applied contexts. First, for example, genomic resources could facilitate husbandry of *M. californianus*, which has not been substantially developed for aquaculture unlike many other *Mytilus* species. *Mytilus galloprovincialis*, the first marine mussel genome to be sequenced (Murgarella et al. 2016), has great value in its native Mediterranean Sea where it constitutes 50% of global EU aquaculture in weight (Robert et al. 2013), but incurs costs as an invasive species in many other parts of the world (Brady and Somero 2006). *Mytilus coruscus* is economically valuable and popular in Asian cuisine due to its high nutritional content (Li et al. 2020; Zhang et al. 2020). The *M. edulis* genome was sequenced for the Prince Edward Island growers to develop tools to implement a breeding program to help with the declining population (https://genomecanada.ca/project/breeding-better-blue-mussels-mytilus-edulis-developing-genomic-tools-implementation-modern-and/). Despite growing larger in length and producing twice as much meat as *M. edulis* in exploratory aquaculture studies (Yamada and Dunham 1989), there is a gap in the literature when it comes to aquaculture studies for *M. californianus* compared to its sister taxa, which this genome may help redress. Second, with biodiversity conservation as a motivator of the CCGP project, the *M. californianus* genome can act as a foundation for future work understanding the genetic diversity and population connectivity for planning marine protected areas (MPAs) and MPA networks (Jeffery et al 2022). This reference genome chapter, coupled with the second chapter regarding population genomics, will deepen our understanding of the evolutionary history of this species and give us a better understanding of how the population is structured, providing a foundation for future genomic studies on ecosystem engineers across the west coast of North America.

Chapter 2: Population Genomics of *Mytilus californianus*

Abstract

The California ribbed mussel, *Mytilus californianus*, is an ecosystem engineer distributed across the heterogeneous intertidal zones of the western coast of North America. Previous studies using allozymes, scDNA, transcriptomics, and DNA barcoding have indicated little population genetic structure in *M. californianus*. This thesis chapter aims to address this question using preliminary whole genome sequencing data. Here we are testing the null hypothesis that *M. californianus* exhibits a panmictic genetic structure across the coast of California. We collected animals from 20 sites along California and sequenced 148 individuals. Preliminary analyses using a subset of 16 individuals sampled along the high and low limits of intertidal sites across four counties (La Jolla, Ventura, Humboldt, and Del Norte). Global Fst was low (Fst = 0.00183) using the Weir and Cockerham F-statistic method. Out of 2,594 Single Nucleotide Polymorphism (SNP) loci, 48 and 103 outlier loci (significant q-value < 0.05) were detected across intertidal and latitudinal gradients, respectively.

Introduction

Marine environments are heterogeneous, with abiotic and biotic factors shaping the distribution of biodiversity among and within geographic locations. Marine intertidal organisms are subject to a particularly large range of abiotic and biotic conditions; examples include steep intertidal (Zardi et al 2011), gradual latitudinal (Connolly et al 2001), continuous (Robles and Desharnais 2002) and discontinuous (Ardisson and Bourget 1992) gradients. The organisms experience these changes in conditions as a function of both averages and variances, spatially and temporally (Sagarin et al 1999, Olabarria and Chapman 2001). With climate change intensifying, both gradual shifts and sudden extremes have been increasing (Helmuth et al 2006). Heatwaves have become a major concern for organisms living in marine intertidal communities (Whalen et al 2023, Ishida et al 2023), as have hypoxia (Marshal and McQuaid 1993, Sussarellu et al 2010), disease (Harvell et al 2002), and their interactions (Dawson et al. accepted).

Whether species living in varying environments can acclimate to the changing conditions, or they must migrate to survive, has become a fundamental question (Reeve and Sherman 1993). Evidence is mixed. Some species appear to have capacity to respond and acclimate to novel conditions induced by climate change, such as the sea snail *Littorina littorea* weighing less and growing slower under low pH and elevated temperature in lab experiments (Melatunan et al 2013). Meanwhile, local thermal adaptation was absent along a latitudinal gradient in porcelain crab *Petrolisthes violaceus* (Gaitán-Espitia et al 2017). Other species exhibit phenotypic plasticity across geographic ranges and abiotic factors. For example, along the Oregon coast, heat shock proteins were highly plastic over small spatial scales in *Mytilus californianus* (Halpin et al 2004). Metabolic rate of the marine star barnacle, *Jehlius cirratus,* was not correlated with latitude but instead with temperature variability in each environment (Broitman et al 2021). These diverse organismal responses to environmental shifts are a means for survival and may be inferred by analyses of genetic diversity.

The distribution of genetic diversity can reveal the relative influences of selection, genetic drift, and migration and how these vary spatially and temporally. Various factors such as distribution range, larval duration, and heat stress tolerance can account for genetic diversity. Genetic structure can be seen across a geographic range due to selection as seen with barnacle *Balanus glandula* showing dramatic variation in frequency of haplotypes between northern and southern populations across the central California coast (Sotka et al 2004). Species that have any planktonic larval duration show a chaotic genetic patchiness (Toonen and Grosberg 2010), however, recent data show low but significant genetic structure at distances below the expected range of larval dispersal (Banks et al 2007, Hogan et al 2012). Selection may lead to divergence within species, such as the intertidal snail *Chlorostoma funebralis* with 34 loci observed to be under divergent selection due to heat stress between the northern and southern populations in California (Gleason and Burton 2016). Synchronously diverging codistributed (SDC) taxa with different dispersal potential patterns show a strong correlation between pelagic duration, relationship with fecundity, and population genetic structure in all but

potentially two species (Dawson et al 2014). Understanding the potential for local adaptation and acclimation in a widely distributed species is important to learn about how populations are changing with various environmental factors.

Mytilus californianus, the California ribbed mussel, is an intertidal ecosystem engineer (Figure 3C)common across much of the northeastern Pacific coastline (Soot-Ryen 1955) and has been used as a species to study population genetic structure due to its role in the Keystone Species concept (Paine 1966). Allozyme homozygosity analyses showed low diversity across the Pacific coastline (Tracey et al 1975, Levinton and Suchanek 1978). Furthermore, analysis of *M. californianus* population genetics structure using allozymes, single-copy nuclear DNA markers, and DNA sequences showed no significant differences among localities and no signal of isolation (Addison et al 2008). No studies to date have used whole genome sequence data to study population genetic structure of *M. californianus* across the intertidal gradient.

Here, I explore *Mytilus californianus* population genetic structure across the coast of California. I test the null hypothesis (H_0) that *M. californianus* is panmictic across the sampling region and consider three alternate hypotheses. First, that over-represented alleles correlate with latitudinal gradients and intertidal zones in parallel (H_1). Second, that there is a difference in population genetic structure across the latitudinal gradient of the California coast (H_2). Finally, genetic structure differentiation is present between the high and low intertidal zones within a sample site (H_3). This study aims to improve our understanding of population genetic structure across the California coast as well as vertically within one site.

Methods

Sample collection

Sixteen samples were collected across four sites: Dike Rock, La Jolla County; Mussel Shoals, Ventura County; Shelter Cove, Humboldt County; and Pyramid Point, Del Norte County (Figure 3A). Mussels were taken from the upper limit of the mussel bed and from the lower limit of the mussel bed (Figure 3B). Sub-sampling of the adductor muscle tissue was done at the University of California, Merced and the tissue was then transported to the University of California, Los Angeles for DNA extraction and library preparation.

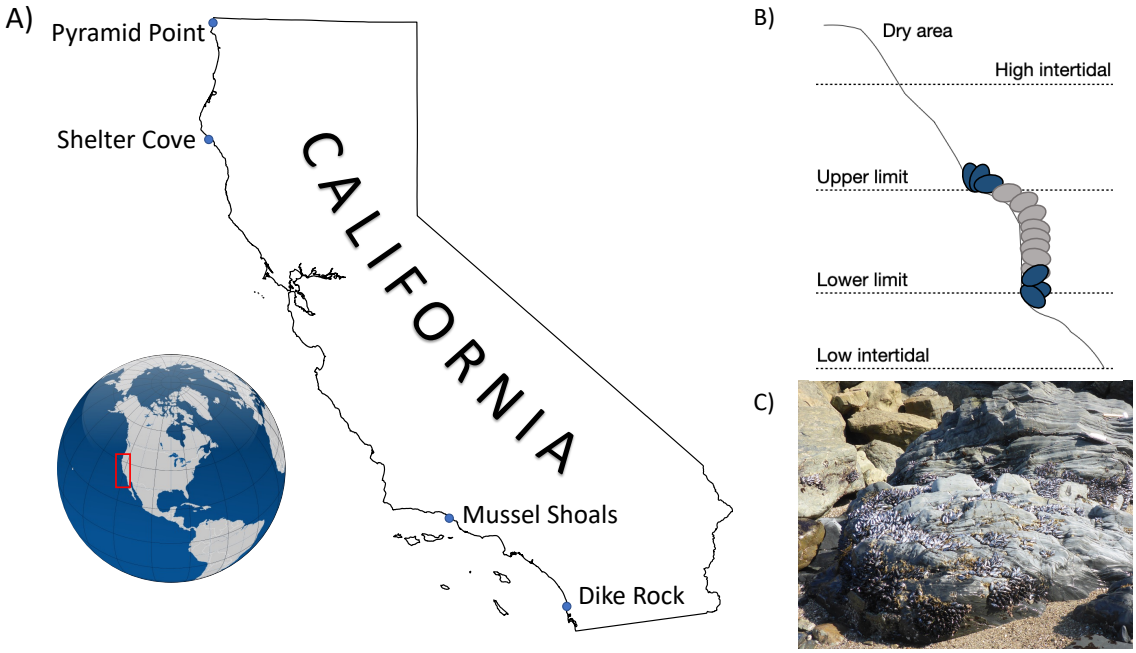

Figure 3: Study design. (A) World map (inset) and California map showing the geographic location of the 4 sites where a subset of samples of *Mytilus californianus* were collected and used for population genetic analyses. (B) A simplified sketch of the rocky intertidal showing the intertidal zones. Blue ovals indicate where samples were taken for the upper and lower limits. (C) A bed of *M. californianus* mussels at Mill Creek, Tehama County, California. Globe image in panel A taken from https://www.clipsafari.com/clips/o29983-globe-showing-north-america. Panel C photo credit: Michael N Dawson.

Omega Bio-tek Mag-Bind Plant DNA DS DNA Isolation

Genomic DNA was extracted from ~10-50mg of the adductor muscle tissue of the mussel. DNA isolations were carried out using a modified version of the Omega Bio-tek Mag-Bind Plant DNA DS kit and was automated for use on an Eppendorf epMotion 5075 TMX liquid handling robot. Samples were digested in CSPL Buffer, Proteinase K, and 0.01M dithiothreitol (DTT) at 56°C for a minimum of 1 hour. Overnight digestion was performed for particularly difficult tissues that did not digest well. DTT was added to help reduce mucopolysaccharides and protein contamination and increase the purity of our samples. Following digestion, samples were centrifuged at 4000xg for 10 minutes and the lysate transferred to a deep well microplate. RNase A was added, and samples incubated for 5 minutes at room temperature before being placed on the epMotion robot. Digested lysate was bound to magnetic Mag-Bind Beads with RBB Buffer following an incubation and then placed on a magnetic rack to remove lysate. Samples then underwent a series of washes using CSPW1 Buffer, CSPW2 Buffer, and SPM Buffer and were dried at 56°C for 10 minutes to remove excess wash buffers. Genomic DNA was eluted with heated Elution Buffer (10mM Tris-HCl, pH 8.5) at 56°C and incubated for a minimum of 5 minutes on a vortexer at 1200rpm.

seqWell plexWell WGS24 Library Preparation

Whole genome libraries for sequencing were prepared using half reactions of the seqWell plexWell WGS24 kit. Using an automated approach on the epMotion robot, 100 ng of sample was fragmented and tagged with a i7 barcode sequence. Following i7 tagmentation, samples were combined into 24-sample library pools, with each sample within a pool containing a unique i7 barcode. Each library pool was then purified using MAGwise Paramagnetic Beads and the concentration assayed using the Qubit 1X Broad Range Assay on a Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Each library pool was then further fragmented and tagged with one of four unique i5 barcode indexes. Each library pool therefore contains 24-samples with 8bp dual indexed reads, allowing for a multiplexing of 96 samples. Following another MAGwise Paramagnetic Bead purification, each library pool was amplified using Kapa HiFi Hot Start ReadyMix (insert citation) according to the following protocol: 10 minutes of fill-in at 72°C, 3 minutes of denaturing at 95°C, 6 cycles of 98°C for 30 seconds, 64°C for 15 seconds, and 72°C for 30 seconds, followed by 72°C for 2 minutes and a 4°C hold. Following amplification, a final 0.8X size selection purification was performed using the MAGwise Paramagnetic Beads to remove fragments less than ~300 bp in size. The purified and amplified libraries were quantified using the Qubit 1X Broad Range Assay and the final

Pooling and Sequencing Procedures

Completed libraries were pooled for sequencing on a NovaSeq S4 6000 150PE at the University of California, Berkeley within the Vincent J. Coates Genomics Sequencing Lab (QB3 Genomics). Sequencing pools were prepped by calculating the sequencing effort need to reach 10X coverage for all individuals within the 24 or 48-sample library pool. Each library pool within a given sequencing pool was adjusted based on genome size of the species contained within a given library pool and sequencing pool.

Bioinformatics

Sequence trimming and SNP calling

For the whole genome sequence fastq files, bbduk from the BBTools package was used for adapter trimming and quality and length filtering (Bushnell, 2021). The whole genome sequences were then mapped to the indexed *Mytilus californianus* genome (NCBI accession number PRJNA796333; Paggeot et al, 2022) with bwa-mem in bwa version 2.2.1 (Vasimuddin et al 2019). The program picard version 2.26.2 and samtools version 1.13 (Danecek et al 2021) were used to mark the sequences in each file that are likely the result of PCR duplication that occurs during the library prep stage. Calling haplotypes was done using GATK version 4.2.6 (McKenna et al 2010) and vcf files were concatenated using bcftools version 1.12. Criteria for haplotype calling are as follows: root mean square quality mapping filter was set to 40, FisherStrand filter was set to greater than 60, QualByDepth filter was set to less than 10, and Depth Quality filter was set to at least 12,500. Single nucleotide polymorphisms (SNPs) were called and filtered using vcftools version 0.1.14. SNPs with genotype quality above 30 and maximum read

depth of 10. After filtering out missing data, linkage disequilibrium was taken into consideration by using vcftools command thin to pick every ten thousandth SNP site.

Population genomics analyses

OutFLANK package (Whitlock and Lotterhos 2015) was used in RStudio to calculate individual SNP loci Fst and parse out outlier loci. Two population datasets were created: (1) comprising upper (n = 8) and lower (n = 8) limit groups, and (2) grouping northern versus southern samples. Pyramid Point and Shelter Cove were grouped together for the northern samples (n = 8); Mussel Shoals and Dike Rock were grouped together for the southern samples (n = 8). Global Fst value was calculated and outliers were parsed out. Significant SNP sites (q-value < 0.05) were parsed out and plotted using the plot() function. RStudio package adegenet version 2.1.10 was used to calculate the Weir and Cockerham Fst value for population structure (Weir and Cockerham 1984) using the wc() function.

Results

SNP calling

67,040,456 SNP loci were called from the 16 samples analyzed. A total of 2,594 SNP loci were retained after filtering and thinning and used in subsequent population genomics analyses.

Population genomics analyses

The global Fst values using all SNP loci are $9.016 * 10^{-5}$ and $2.178 * 10^{-4}$ for intertidal and latitudinal gradients, respectively. 48 and 103 SNP loci were characterized as significant (q-value < 0.05) across intertidal latitudinal gradients, respectively (Figure 4). 14 SNP loci were shared between the two sets of significant SNPs (equivalent to 29.2% of intertidal gradient SNPs, 13.6% of latitudinal gradient SNPs).

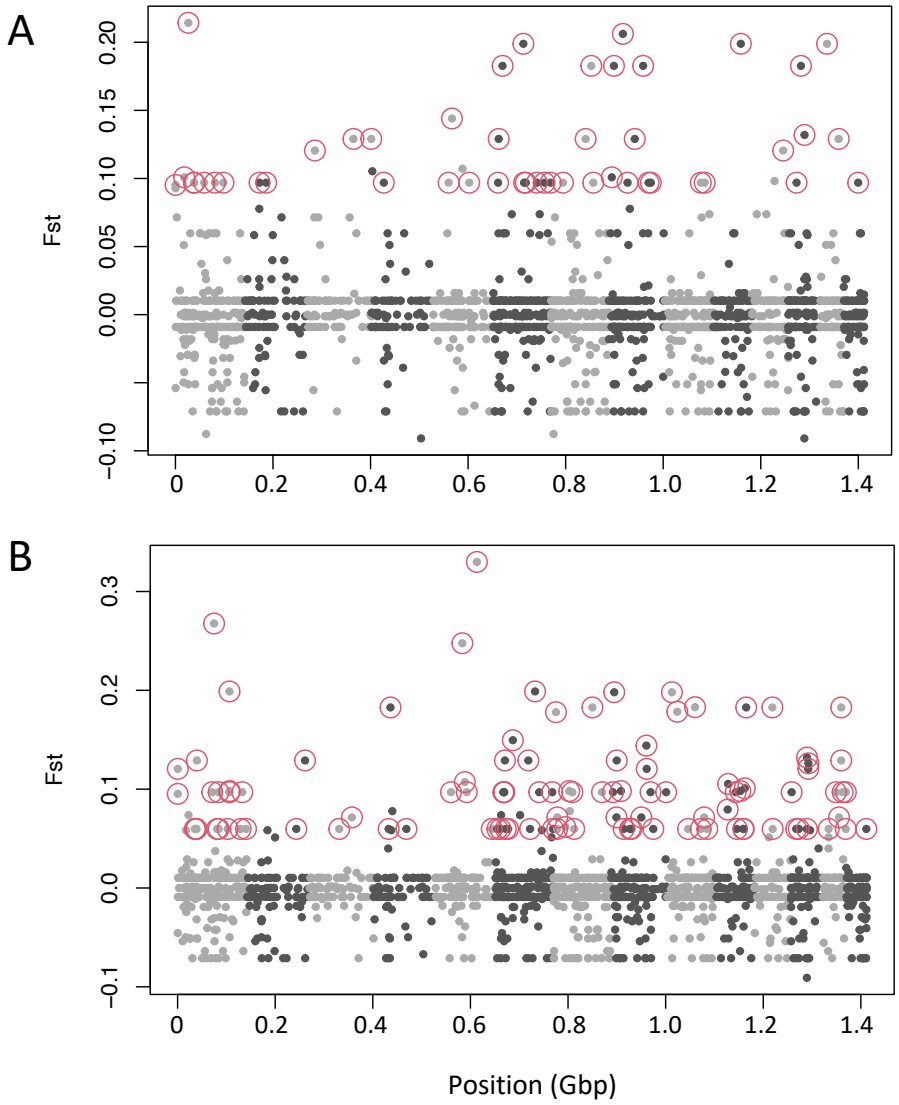

Figure 4: Graphs of *Mytilus californianus* Fst against position on genome. (A) Fst across SNP loci when samples are grouped by upper and lower limit. (B) Fst across SNP loci when samples are grouped by northern and southern sample sites. Light gray points represent Fst values calculated from SNP loci on odd numbered chromosomes, dark gray points represent Fst values calculated from SNP loci on even numbered chromosomes. Outlier loci Fst were calculated to be significantly different (q-value < 0.05) from the rest of the Fst in the data frame are circled in red.

Discussion

The calculated Fst values indicate low genetic differentiation. Genomic analyses thus affirm past studies on *Mytilus* spp genetic population have come to the same conclusion using different methods across a larger geographical range (Skibinski et al 1983, Diz and Presa 2009, Giantsis et al 2014): *Mytilus* spp. are most often panmictic.

The low genetic differentiation observed in *Mytilus californianus* could be attributed to its long pelagic larval duration. Studies have suggested that organisms sharing similar pelagic larval duration, fecundity, and life stages are likely to exhibit comparable phylogeographic structures and gene flow (Addison and Hart 2004, Dawson 2012). An analysis of Fst among eight coral reef fish species revealed a significant difference between species with a pelagic larval duration and those without (Bay et al 2006). The nudibranch species *Goniodoris nodosa*, which possesses a three-month pelagic larval duration, displayed low genetic differentiation when compared to the highly differentiated *Adalaria proxima*, capable of metamorphosis within 1-2 days after release (Todd et al 1998). The pelagic larval duration of *M. californianus* is estimated to range between 10-45 days (Trevelyan & Chang 1983, Strathmann 1987). This duration facilitates extensive larval transport across space and time within the water column, potentially explaining the observed low genetic differentiation.

Despite there being little genetic differentiation across latitudinal and intertidal gradients, *Mytilus californianus* shows high phenotypic variation between the two gradients. Ecological studies of *M. californianus* have shown significant differences between latitudinal gradients including reproductive output (Phillips 2007), acute heat stress survival rate (Logan et al 2012), and growth rates (Blanchette 2007). Terminal sizes of mussels in lower intertidal were significantly larger than higher intertidal (Connor and Robles 2015). Low-zone mussels had significantly higher growth rates than mid-zone and high-zone mussels (Thakar 2017). Researching the relationship between genetic and ecological variation holds the potential to unravel crucial insights into the adaptive capacity of species (Nelson and Cresko 2018), and many other evolutionary processes.

Caveats and future directions

This thesis chapter has low statistical power due to small sample sizes, more samples from more sites are required to properly analyze the population genetic structure. The full dataset contains 148 samples from 20 sites across California. Fst values were calculated on all the SNP loci, outlier or neutral loci will need to be selected for separate analyses for adaptation and migration. Once sequencing is complete, I will have SNP calls for all the samples collected and I will run the pipeline developed in this thesis to analyze all the samples across many sample sites. Further analysis includes looking at the non-neutral sites and referencing the reference genome published to see if it is a coding region. This will give us a deeper understanding of any coding regions with allelic frequencies that are significantly different than the others. Additional analyses will also be performed alluding to linkage and for islands of selection.

Chapter 3: Thesis Summary

I used a chromosome-scale reference genome for *Mytilus californianus* and compared it with other *Mytilus* genomes and concluded that the genome produced is the most complete (86% complete) and contiguous (N50 sequence length of 118 Mb) to date. Trimmed reads were aligned to the reference genome and Single Nucleotide Polymorphisms (SNPs) were called. Afterwards, I used the called SNPs to calculate the fixation index (Fst) values and concluded that the population genetic differentiation across latitudinal and intertidal gradients was low (Fst $\leq$ 0.0003), potentially due to the extensive long pelagic larval duration of *M. californianus*. Leveraging chromosome-scale reference genomes in the analysis of population genetic differentiation improves the reliability of genotype calls and thus enables a deeper understanding of gene flow patterns and genetic structure (Colonna et al 2014, Ragupathi et al 2017).

An understanding of population genetic structure can be applied across various domains, including conservation research, husbandry practices, and evolutionary studies. One application can be to analyze genetic structure across organisms with different pelagic larval dispersal durations. For instance, organisms like *Pollicipes polymerus*, characterized by extended pelagic larval dispersal lasting 1-2 months, exhibit low genetic structure while *Lottia scabra*, with a pelagic larval dispersal lasting less than 2 weeks, exhibit a higher genetic structure (Dawson et al 2012). The workflow developed here holds promise for application to other organisms within the CCGP consortium, such as *Strongylocentrotus purpuratus* (Strathmann 1978), *Tetraclita rubescens* (Dawson et al 2010), and *Anthopleura sola* (McFadden et al 1997). Future work can include evaluating genetic diversity and conducting meta-analyses to investigate genetic variability and adaptation potential. These prospective investigations hold the potential to illuminate evolutionary dynamics and ecological interactions within various ecosystems.

The reference genome serves as a foundation, facilitating an understanding of a species' evolution and genetic diversity across geographic regions and populations (Wong et al 2020). Furthermore, the reference genome offers a robust foundation applicable to aquaculture husbandry practices, enhancing the efficiency and precision of breeding programs (Kim et al 2016). Population differentiation analyses play a role in identifying genes undergoing selection, thus guiding the California Conservation Genomics Project's conservation efforts aimed at preserving genetic diversity. These analyses shed light on a species' adaptive potential in response to environmental changes, particularly vital in predicting and preparing for the impacts of climate change (Clucas et al 2019). Moreover, population differentiation studies inform ecosystem management strategies, enabling the formulation of sustainable practices geared towards maintaining and preserving species diversity and ecological balance (Avise 2010).

References

Addison, JA and Hart, MW. Analysis of population genetic structure of the green sea urchin (*Strongylocentrotus droebachiensis*) using microsatellites. *Marine Biology.* 2004;144, 243–251 https://doi.org/101007/s00227-003-1193-6.

Addison, JA, Ort, BS, Mesa, KA, & Pogson GH. Range-wide genetic homogeneity in the California sea mussel (*Mytilus californianus*): a comparison of allozymes, nuclear DNA markers, and mitochondrial DNA sequences. *Molecular Ecology.* 2008;17: 4222-4232. https://doi.org/10.1111/j.1365-294X.2008.03905.x.

Ahmed M, Sparks AK. Chromosome number, structure and autosomal polymorphism in the marine mussels *Mytilus edulis* and *Mytilus californianus*. *Biol Bull*. 1970;138(1):1–13.

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc, F. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour*. 2020;20(4):892–905.

Ardisson PL and Bourget E. Large-scale ecological patterns: discontinuous distribution of marine benthic epifauna. *Marine Ecology Progress Series*. 1992;83(1), 15-34.

Avise, JC. Perspective: Conservation genetics enters the genomics era. *Conservation Genetics*. 2010; 11, 665–669.

Banks, SC, Piggott, MP, Williamson, JE, Bove, U, Holbrook, NJ, & Beheregaray, LB. Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin *Ecology*. 2007;88, 3055–3064 https://doi.org/101890/07-00911.

Bay LK, Crozier RH, and Caley, MJ. 2006 The relationship between population genetic structure and pelagic larval duration in coral reef fishes on the Great Barrier Reef *Mar Biol*. 2006;149, 1247–1256 https://doi.org/101007/s00227-006-0276-6.

Blanchette, CA, Helmuth, B, Gaines, SD. Spatial patterns of growth in the mussel, *Mytilus californianus*, across a major oceanographic and biogeographic boundary at Point Conception, California, USA. *Journal of Experimental Marine Biology and Ecology*. 2007;340(2), 126-148 https://doi.org/101016/jjembe200609022.

Brady CE, Somero GN. Ecological gradients and relative abundance of native (*Mytilus trossulus*) and invasive (*Mytilus galloprovincialis*) blue mussels in the California hybrid zone. *Mar Biol*. 2006;148(6):1249–1262.

Broitman BR, Lagos NA, Opitz T, Figueroa D, Maldonado K, Ricote N, Lardies MA. Phenotypic plasticity is not a cline: Thermal physiology of an intertidal barnacle

over 20° of latitude. *J Anim Ecol*. 2021;90(8):1961-1972. https://doi.org/101111/1365-265613514. PMID: 33942301.

Bruno JF, Stachowicz JJ, Bertness MD. Inclusion of facilitation into eco- logical theory. *Trends Ecol Evol*. 2003;18(3):119–125.

Bushnell, B. BBTools/BBMap. 2021.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit – interactive quality assessment of genome assemblies. *G3 Genes|Genomes|Genetics*. 2020;10(4):1361–1374. doi:10.1534/ g3.119.400908.

Chin C, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*. 2016;13(12):1050–1054.

Clucas GV, Lou RN, Therkildsen NO, Kovach AI. Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evol Appl*. 2019; 12(10):1971-1987. https://doi.org.10.1111/eva.12861.

Colonna V, Ayub Q, Chen Y, Pagani L, Luisi P, Pybus M, Garrison E, Xue Y, Tyler-Smith C. Human genomic regions with exceptionally high levels of population differentiation identified from 911 whole-genome sequences. *Genome Biol*. 2014; 15, R88. https://doi.org/10.1186/gb-2014-15-6-r88.

Connolly SR, Menge BA and Roughgarden J. A latitudinal gradient in recruitment of intertidal invertebrates in the Northeast Pacific Ocean. *Ecology*. 2001;82: 1799-1813 https://doi.org/101890/0012-9658(2001)082[1799:ALGIRO]20CO;2.

Connor KM and Robles CD. Within-site variation of growth rates and terminal sizes in *Mytilus californianus* along wave exposure and tidal gradients. *Biol Bull*. 2005;228(1):39-51 https://doi.org/101086/BBLv228n1p39 PMID: 25745099.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2), giab008 [33590861].

Dawson MN. Parallel phylogeographic structure in ecologically similar sympatric sister taxa. *Mol Ecol*. 2014;21(4):987-1004. https://doi.org/1101111/j1365-294X201105417x Epub 2012 Jan 9 PMID: 22229665.

Dawson, MN, Hays CG, Grosberg, R and Raimondi, PT. Dispersal potential and population genetic structure in the marine intertidal of the eastern North Pacific. *Ecological Monographs*. 2014;84: 435-456. https://doi.org/101890/13-08711.

Dawson MN, Grosberg RK, Stuart YE, Sanford E. Population genetic analysis of a recent range expansion: mechanisms regulating the poleward range limit in the volcano barnacle *Tetraclita rubescens*. *Mol Ecol*. 2010; 1585-605. https://doi.org/10.1111/j.1365-294X.2010.04588.x.

Dawson, MN, P Giakoumis DM, Schiebelhut LM , Beas-Luna R, Bosley KL, Castilho R, Ewers-Saucedo C, Gavenus KA, Keller A, Konar B, Largier JL, Lorda J, Miner CM, Moritsch MM, Navarrete SA, Raimondi PT, Traiger SB, Turner M, and Wares JP (in review). Almost a decade of death … and other dynamics: deepening perspectives on the diversity and distribution of sea stars and wasting. *Biological Bulletin*.

Dayton P. Toward an understanding of community resilience and the potential effects of enrichment to the benthos at McMurdo Sound, Antarctica. In: Parker B, editor. *Proceedings of the colloquium on conservation problems in Antarctica*. Virginia Polyterchnic Institute and State University; *Allen Press*; 1972.

DeBiasse MB, Schiebelhut LM, Escalona M, Beraut E, Fairbairn C, Marimuthu MPA, Nguyen O, Sahasrabudhe R, Dawson MN. A chromosome-level reference genome for the giant pink sea star, *Pisaster brevispinus*, a species severely impacted by wasting. *J Hered*. 2022;113:689–698.

Diz AP, Presa P. Regional patterns of microsatellite variation in *Mytilus galloprovincialis* from the Iberian Peninsula. *Mar Biol*. 2008;154(2):277–286.

Diz, AP and Presa, P. The genetic diversity pattern of *Mytilus galloprovincialis* in Galician Rías (NW Iberian estuaries). *Aquaculture*. 2009;287(3-5), 278-285. https://doi.org/101016/jaquaculture200810029.

Gaitán-Espitia JD, Bacigalupe LD, Opitz T, Lagos NA, Osores S, Lardies MA. Exploring physiological plasticity and local thermal adaptation in an intertidal crab along a latitudinal cline. *J Therm Biol*. 2017;68(Pt A):14-20. https://doi.org/101016/jjtherbio201702011. PMID: 28689716.

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, Venier P, Naranjo-Ortiz MA, Murgarella M, Greco S, et al. Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biol*. 2020;21(1):275–295.

Giantsis, I, Mucci, N, Randi, E, Abatzopoulos, T, & Apostolidis, A. Microsatellite variation of mussels (*Mytilus galloprovincialis*) in central and eastern Mediterranean: Genetic panmixia in the Aegean and the Ionian Seas. *Journal of the Marine Biological Association of the United Kingdom*. 2014;94(4), 797-809. https://doi.org/101017/S0025315414000174.

Gleason LU and Burton RS. Regional patterns of thermal stress and constitutive gene expression in the marine snail *Chlorostoma funebralis* in northern and southern California. *Marine Ecology Progress Series*. 2016;556, 143-159. https://doi.org/103354/meps11850.

Gutiérrez JL, Jones CG, Strayer DL, Iribarne OO. Mollusks as eco- system engineers: the role of shell production in aquatic habitats. *Oikos*. 2003;101(1):79–90.

Halpin PM, Menge BA, and Hofmann GE. Experimental demonstration of plasticity in the heat shock response of the intertidal mussel *Mytilus californianus. Marine Ecology Progress Series.* 2004;276, 137-145 https://doi.org/103354/meps276137.

Harvell CD, Mitchell CE, Ward JR, Altizer S, Dobson AP, Ostfeld RS, and Samuel MD. Climate Warming and Disease Risks for Terrestrial and Marine Biota. *Science*. 2002;296, 2158-2162. https://doi.org/101126/science1063699.

Helmuth, B, Broitman, B R, Blanchette, C A, Gilman, S, Halpin, P, Harley, C D G, O'Donnell, M J, Hofmann, G E, Menge, B, & Strickland, D. Mosaic Patterns of Thermal Stress in the Rocky Intertidal Zone: Implications for Climate Change. *Ecological Monographs*. 2006;76(4), 461–479. http://www.jstor.org/stable/27646057.

Hogan, JD, Thiessen, RJ, Sale, PF & Heath, DD. Local retention, dispersal and fluctuating connectivity among populations of a coral reef fish, *Oecologia*. 2012;168, 61–71. https://doi.org/101007/s00442-011-2058-1.

Holten-Andersen N, Zhao H, Waite JH. Stiff coatings on compliant biofibers: the cuticle of *Mytilus californianus* byssal threads. *Bio- chemistry*. 2009;48(12):2752–2759.

Inoue K, Yoshioka Y, Tanaka H, Kinjo A, Sassa M, Ueda I, Shinzato C, Toyoda A, Itoh T. Genomics and transcriptomics of the green mussel explain the durability of its byssus. *Sci Rep*. 2021;11(1):5992.

Ishida K, Tachibana M, Yao Y, Wada Y, and Noda T. The impact of marine heatwaves on rocky intertidal communities: evidence of accumulative carryover effects of marine heatwaves. *Frontiers in Marine Science*. 2023;10:1146148. https://doi.org/10.3389/fmars.2023.1146148.

Jeffery NW, Lehnert SJ, Kess T, Layton KKS, Wringe BF, Stanley RRE. Application of omics tools in designing and monitoring marine protected areas for a sustainable blue economy. *Front Genet*. 2022;13(886494).

Jones CG, Lawton JH, Shachak M. Organisms as ecosystem engineers. *Oikos*. 1994;69(3):373–386.

Kartavtsev, YPh, Masalkova, NA, and Katolikova, MV. Genetic and morphometric variability in the settlements of two mussel species (*Mytilus ex gr edulis*), *M trossulus*, and *M galloprovincialis*, in North-West Japan Sea. *J Shellfish Res*. 2018;37(1), 1–17.

Kim C, Guo H, Kong W, Chandnani R, Shuang LS, Paterson AH. Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci*. 2016; 242:14-22. https://doi.org/10.1016/j.plantsci.2015.04.016.

Kinoshita T, Seki M. Epigenetic memory for stress response and adaptation in plants. *Plant Cell Physiol*. 2014;55(11):1859–1863.

Koehn RK, Gaffney PM. Genetic heterozygosity and growth rate in *Mytilus edulis*. *Mar Biol*. 1984;82(1):1–7.

Ladoukakis ED, Saavedra C, Magoulas A, Zouros E. Mitochondrial DNA variation in a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. *Mol Ecol*. 2002;11(4):755–769.

Levinton, JS and Suchanek, TH. Geographic variation, niche breadth and genetic differentiation at different geographic scales in the mussels *Mytilus californianus* and *M edulis. Mar Biol.* 1978;49, 363–375. https://doi.org/101007/BF00455031.

Li R, Zhang W, Lu J, Zhang Z, Mu C, Song W, Migaud H, Wang C, Bekaert M. The whole-genome sequencing and hybrid assembly of *Mytilus coruscus*. *Front Genet*. 2020;11(440):1664–8021.

Lin M, Escalona M, Sahasrabudhe R, Nguyen O, Beraut E, Buchalski MR, Wayne RK. A reference genome assembly of the bobcat, *Lynx rufus*. *J Hered*. 2022;113:515–623.

Logan, CA, Kost, LE, Somero, GN. Latitudinal differences in *Mytilus californianus* thermal physiology. *Mar Ecol Prog Ser*. 2012;450, 93–105. https://doi.org/ 103354/meps09491.

McFadden, C, Grosberg R, Cameron B, Karlton D, Secord D. Genetic relationships within and between clonal and solitary forms of the sea anemone *Anthopleura elegantissima* revisited: evidence for the existence of two species. *Marine Biology*. 1997;128(1): 127-139.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.

Genome Res. 2010;20(9):1297-303. https://doi.org/101101/gr107524110 PMID: 20644199; PMCID: PMC2928508.

Melatunan S, Calosi P, Rundle SD, Widdicombe S, and Moody AJ. Effects of ocean acidification and elevated temperature on shell plasticity and its energetic basis in an intertidal gastropod. *Marine Ecology Progress Series*. 2013;472, 155-168. https://doi.org/103354/meps10046.

Mizi A, Zouros E, Moschonas N, Rodakis GC. The complete maternal and paternal mitochondrial genomes of the Mediterranean mussel *Mytilus galloprovincialis*: implications for the doubly uniparental inheritance mode of mtDNA. *Mol Biol Evol*. 2005;22(4):952–967.

Mostovoy Y, Levy-Sakin M, Lam J, Lam ET, Hastie AR, Marks P, Lee J, Chu C, Lin C, Džakul Z, et al. A hybrid approach for de novo human genome sequence assembly and phasing. *Nat Methods*. 2016;13(7):587–590.

Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C. A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. *PLoS One*. 2016;11(3):e0151561.

Nelson TC and Cresko WA. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters*. 20182(1),9–21. https://doi.org/10.1002/evl3.37.

Nishimura O, Hara Y, Kuraku S. Evaluating genome assemblies and gene models using gVolante. *Methods Mol Biol*. 2019;1962:247– 256.

Olabarria C and Chapman MG. Comparison of patterns of spatial variation of microgastropods between two contrasting intertidal habitats. *Marine Ecology Progress Series*. 2001;220, 201-211. https://doi.org/103354/meps220201.

Paggeot LX, DeBiasse MB, Escalona M, Fairbairn C, Marimuthu MPA, Nguyen O, Sahasrabudhe R, Dawson MN. Reference genome for the California ribbed mussel, *Mytilus californianus*, an ecosystem engineer. *J Hered*. 2022;113(6):681-688. https://doi.org/101093/jhered/esac041.

Paine, RT. Food Web Complexity and Species Diversity. *The American Naturalist*. 1966;100(910), 65–75. http://wwwjstororg/stable/2459379.

Paine RT. Marine rocky shores and community ecology: an experimentalist's perspective. Oldendorf (Luhe, Germany): *Ecology Institute*. 1994.

Parmesan C. Ecological and Evolutionary Responses to Recent Climate Change. *Annual Review of Ecology, Evolution, and Systematics*. 2006;37, 637-669.

Pearce T, Labarbera M. A comparative study of the mechanical properties of Mytilid byssal threads. *J Exp Biol*. 2009;212(10)1442–1448.

Pérez-García C, Morán P, Pasantes JJ. Karyotypic diversification in *Mytilus* mussels (Bivalvia: Mytilidae) inferred from chromosomal mapping of rRNA and histone gene clusters. *BMC Genet*. 2014;15(1):84.

Phillips, NE. A spatial gradient in the potential reproductive output of the sea mussel *Mytilus californianus. Mar Biol*. 2007;151, 1543–1550. https://doi.org/101007/s00227-006-0592-x.

Place SP, O'Donnell MJ, Hofmann GE. Gene expression in the inter- tidal mussel *Mytilus californianus*: physiological response to environmental factors on a biogeographic scale. *Mar Ecol Prog Ser*. 2008;356:1–14.

Ragupathi NK, Sethuvel DP, Inbanathan FY, Veeraraghavan B. Accurate differentiation of *Escherichia coli* and *Shigella serogroups*: challenges and strategies. *New Microbes New Infect.* 2017; 21:58-62. https:/doi.org/10.1016/j.nmni.2017.09.003.

R Core Team. 2020. R: A language and environment for statistical com- puting. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reeve HK and Sherman PW. Adaptation and the Goals of Evolutionary Research. *The Quarterly Review of Biology*. 1993;68(1), 1-32. https://doi.org/101086/417909.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature*. 2021;592(7856):737–746.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference- free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21(1):245–271.

Riginos C, Cunningham CW. Local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilus trossulus*) hybrid zones. *Mol Ecol*. 2004;14(2):381–400.

Robert R, Sanchez JL, Perez-Paralle L. A glimpse on the mollusc industry in Europe. *Aquaculture*. 2013;38(1):5–11.

Robles C and Desharnais R. History and current development of a paradigm of predation in rocky intertidal communities. *Ecology*. 2002;83: 1521-1536. https://doi.org/101890/0012-9658(2002)083[1521:HACDOA]20CO;2.

Sagarin RD, Somero GN. Complex patterns of expression of heat- shock protein 70 across the southern biogeographical ranges of the intertidal mussel *Mytilus californianus* and snail *Nucella ostrina*. *J Biogeogr*. 2006;33(4):622–630.

Sagarin, RD, Barry, JP, Gilman, SE and Baxter, CH. Climate-related change in an intertidal community over short and long time scales. *Ecological Monographs*. 1999;69: 465-490. https://doi.org/101890/0012-9615(1999)069[0465:CRCIAI]20CO;2.

Sarver, SK and Foltz, DW. Genetic population structure of a species' complex of blue mussels (*Mytilus* spp). *Marine Biology*. 1993;117, 105–112 https://doi.org/101007/BF00346431.

Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. *Nat Rev Genet*. 2013;14(11):807–820.

Schiebelhut, LM, Gaylord, B, Grosberg, RK, Jurgens, LJ, & Dawson, MN. Species' attributes predict the relative magnitude of ecological and genetic recovery following mass mortality. *Molecular Ecology*. 2022;31, 5714– 5728. https://doi.org/101111/mec16707.

Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol*. 2019;1962:227–245.

Shaffer HB, Toffelmier Corbett-Detig RB, Escalona M, Erickson B, Fiedler P, Gold M, Harrigan RJ, Hodges S, Luckau TK, Miller C, et al. Landscape genomics to enable conservation actions: the California Conservation Genomics Project. *J Hered*. 2022;113:577–588.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210– 3212.

Skibinski, DOF, Beardnire, JA and Cross, TF. Aspects of the population genetics of *Mytilus* (Mytilidae; Mollusca) in the British Isles. *Biological Journal of the Linnean Society*. 1983;19: 137-183. https://doi.org/101111/j1095-83121983tb00782x.

Śmietanka B, Burzyński A, Wenne R. Comparative genomics of marine mussels (*Mytilus* spp.) gender associated mtDNA: rapidly evolving atp8. *J Mol Evol*. 2010;71:385–400.

Soot-Ryen T. A report on the family Mytilidae (Pelecypoda) Allan Hancock Pacific Expeditions Los Angeles, California. *The University of Southern California Press*. 1955;20(1) p 1–174.

Sotka EE, Wares JP, Barth JA, Grosberg RK, Palumbi SR. Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula. Mol Ecol*. 2004;13(8):2143-56 https://doi.org/101111/j1365-294X200402225x PMID: 15245390.

Springer SA, Crespi BJ. Adaptive gamete-recognition divergence in a hybridizing *Mytilus* population. *Evolution*. 2006;61(4):772–783.

Strathmann MF. Reproduction and development of marine invertebrates of the northern Pacific coast: data and methods for the study of eggs, embryos, and larvae. *University of Washington Press, Seattle, WA*. 1987.

Strathmann R. Length of the pelagic period in echinoderms with feeding larvae from Northeast Pacific. *J Exp Mar Biol Ecol*. 1978; 34:23–27. https://doi.org10.1016/0022-0981(78)90054-0.

Sussarellu R, Fabioux C, Moullac GL, Fleury E, and Moraga D. Transcriptomic response of the Pacific oyster *Crassostrea gigas* to hypoxia. *Marine Genomics*. 2010;3(3–4), 133-143. https://doi.org/101016/jmargen201008005.

Syvänen, AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet*. 2001;2, 930–942. https://doi.org/101038/35103535.

Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, Fujie M, Goto H, Yamasaki S, Nagai K, Morino Y, et al. Bivalve-specific gene ex- pansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zool Lett*. 2016;2(1):3–15.

Thakar, HB, Glasgow, MA, and Blanchette C. Reconsidering evidence of human impacts: Implications of within-site variation of growth rates in *Mytilus californianus* along tidal gradients. *Quaternary International*. 2017;427, 151-159. https://doi.org/101016/jquaint201510018.

Todd CD, Lambert WJ, and Thorpe JP. The genetic structure of intertidal populations of two species of nudibranch molluscs with planktotrophic and pelagic lecithotrophic larval stages: are pelagic larvae “for” dispersal. *Journal of Experimental Marine Biology and Ecology*. 1998;228(1), 1-28. https://doi.org/101016/S0022-0981(98)00005-7.

Toonen RJ and Grosberg R. Testing hypotheses for chaotic genetic structure: patterns of spatial and temporal genetic heterogeneity among recruits, subadults and adults of the intertidal anomuran *Petrolisthes cinctipes*. *Phylogeography and Population Genetics in Crustacea*. S. Koenemann, C. Held, and C. Schubart, Eds., 2010.

Tracey ML, Bellet NF and Gravem, CD. Excess allozyme homozygosity and breeding population structure in the mussel *Mytilus californianus. Mar Biol*. 1975;32, 303–311. https://doi.org/101007/BF00399209.

Tracey ML, Bellet NF, Gravem CD. Excess allozyme homozygosity and breeding population structure in the mussel *Mytilus californianus*. *Mar Biol*. 1975;32(3):303–311.

Trevelyan G and Chang E. Experiments on larval rearing of the California mussel (*Mytilus californianus*). *J World Maric Soc*. 1983;14: 137−148.

Vasimuddin MD, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *IEEE Parallel and Distributed Processing Symposium*. 2019.

Waite JH. Mussel adhesion—essential footwork. *J Exp Biol*. 2017;220(4):517–530.

Weir, BS and Cockerham, CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. 1984;38(6), 1358–1370. https://doi.org/102307/2408641.

Whalen, MA, Starko, S, Lindstrom, SC, and Martone, PT. " Heatwave Restructures Marine Intertidal Communities across a Stress Gradient". *Ecology*. 2023;104(5): e4027 https://doi.org/101002/ecy4027.

Whitlock MC and Lotterhos KE. Reliable Detection of Loci Responsible for Local Adaptation: Inference of a Null Model through Trimming the Distribution of F(ST). *Am Nat*. 2015;186 Suppl 1:S24-36. doi: 101086/682949 PMID: 26656214.

Wong KHY, Ma, W, Wei, CY, Yeh EC, Lin WJ, Wang EHF, Su JP, H FJ, Kao HJ, Chen HH, Chow SK, Young E, Chu C, Poon A, Yang CF, Lin DS, Hu YF, Wu JY, Lee NC, Hwu WL, Boffelli D, Martin D, Xiao M, and Kwok PY. Towards a reference genome that captures global genetic diversity. *Nat Community*. 2020; 11. https://doi.org/10.1038/s41467-020-19311-w.

Yamada SB, Dunham JB. *Mytilus californianus*, a new aquaculture species? *Aquaculture*. 1989;81(3–4):275–284.

Yang JL, Feng DD, Liu J, Xu JK, Chen K, Li YF, Zhu YT, Liang X, Lu Y. Chromosome-level genome assembly of the hard-shelled mussel *Mytilus coruscus*, a widely distributed species from the temperate areas of East Asia. *GigaScience*. 2021;10(4):giab024.

Zardi GI, Nicastro KR, Canovas F, Ferreira Costa J, Serrão EA, Pearson GA. Adaptive Traits Are Maintained on Steep Selective Gradients despite Gene Flow and

Hybridization in the Intertidal Zone. *PLOS ONE*. 2011;6(6): e19402. https://doi.org/101371/journalpone0019402

Zhang W, Li R, Chen X, Wang C, Gu Z, Mu C, Song W, Zhan P, Huang J. Molecular identification reveals hybrids of *Mytilus coruscus* × *Mytilus galloprovincialis* in mussel hatcheries of China. *Aquac Int*. 2020;28(1):85–93.

Supplementary Information

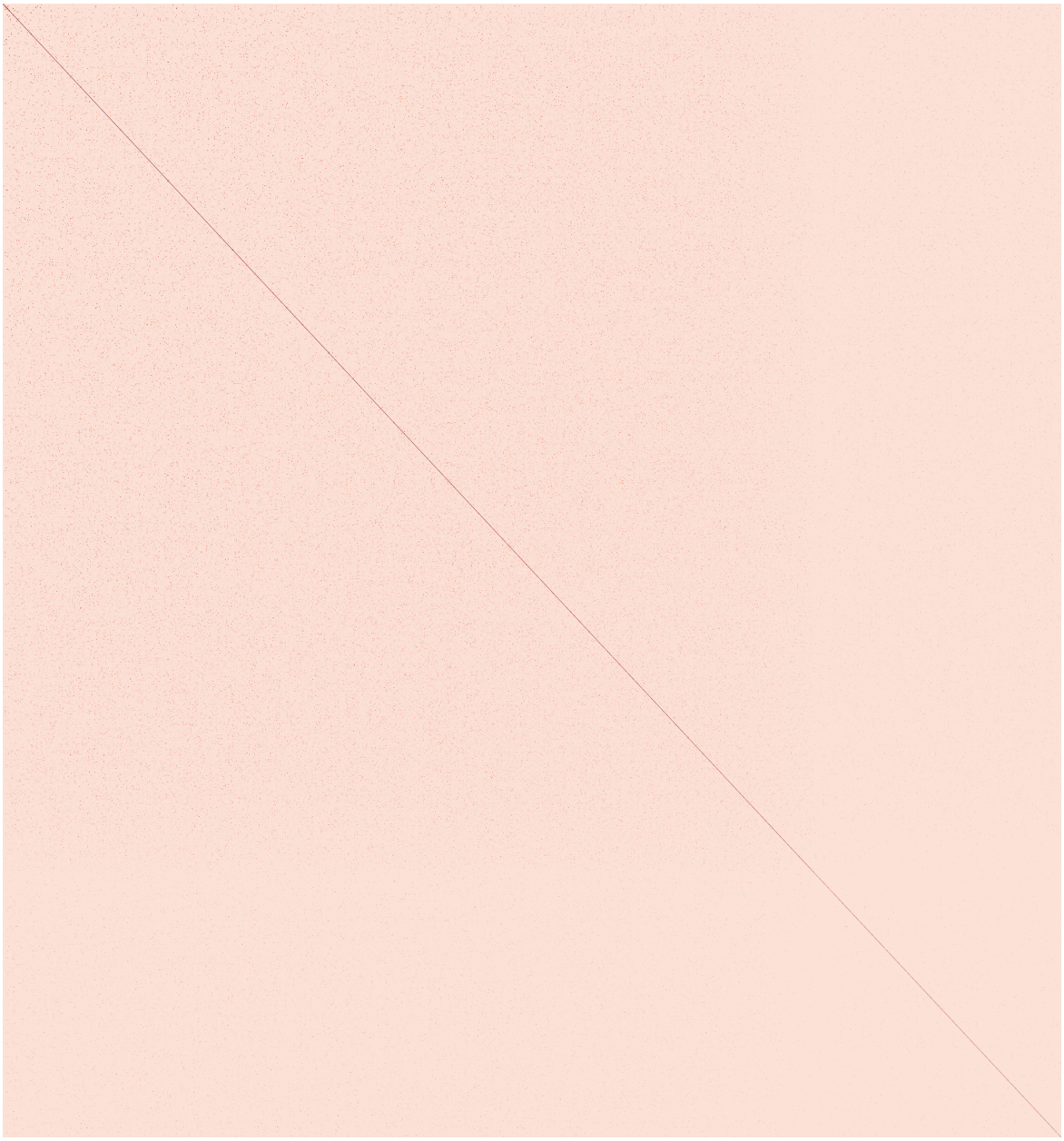

Figure S1: Omni-C contact maps for the alternate genome assembly generated with PretextSnapshot.