A blue mussel chromosome-scale assembly and genomic resources for aquaculture, marine ecology and evolution

Tim Regan[1], Tiago S. Hori[2], Tim P. Bean[1]

[1] The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK.

[2] Atlantic Aqua Farms Ltd., Charlottetown, PE, Canada.

Abstract

The blue mussel is commonly described as the *Mytilus* species complex, encompassing at least three putative species: *M. edulis, M. galloprovincialis* and *M. trossolus*. These three species occur on both sides of the Atlantic and hybridize in nature, and both *M. edulis* and *M. galloprovincialis* are important aquaculture species. They are also invasive species in many parts of the world. Here, we present the first chromosome-level assembly of the *Mytilus* species complex. We used a combination of PacBio sequencing and Dovetail's Omni-C technology to generate an assembly with 14 long scaffolds containing 94% of the predicted length of the *M. edulis* genome (1.6 out of 1.7 Gb). Assembly statistics were total length 1.65 Gb, N50 = 116 Mb, L50 = 7 and, L90 = 13. BUSCO analysis showed 91.3% eukaryote BUSCOs identified. AB-*Initio* annotation using RNA-seq from mantle, gills, muscle and foot predicted 47,128 genes. Using GBS and shotgun sequencing, we also sequenced 3 North American populations of *Mytilus* to characterize single-nucleotide as well as structural variance. This high-quality genome for *M. edulis* provides a platform to develop tools that can be used in breeding, molecular ecology and evolution to address questions of both commercial and environmental perspectives.

Introduction

The blue mussel (*Mytilus edulis*) is common to the North Atlantic from Arctic to Mediterranean regions, with habitat ranging from upper shore to the shallow subtidal (Hayward & Ryland 2017). *M. edulis* is known to hybridise with *M. trossolus* in North America and with *M. trossolus* and *M. galloprovincialis* in Europe. Together, these three species form the *M. edulis* species complex (Fig. 1) (McDonald et al. 1991).

This reef-building bivalve is an ecosystem engineer. Blue mussels dominate fouling communities in shallow and substrata providing important secondary habitat (Norling & Kautsky 2007). Offshore wind energy structure surveys found that they can cover the structures with up to 3.4 kg of biomass m^{-2} (Krone et al. 2013). Through filter-feeding, eutrophication is reduced which can alter ecosystems (Broszeit et al. 2016). This nutrient cycling ability has been harnessed by using *M. edulis* to study the fate of persistent organic and metal pollutants (Chase et al. 2001; McEneff et al. 2014), for the bioremediation of waste (Broszeit *et al.* 2016) and reduce environmental effects from salmon farms (MacDonald et al. 2011).

Mussels are a significant bivalve production species (FAO 2020), however, the industry has faced year on year reductions in wild spat availability for aquaculture sites in the UK and elsewhere (Regan et al. 2021). Rather than a single factor driving these losses, the effect is likely due to a combination of increasing threats and challenges to this species. For instance, the stress of warmer seas has been found to increase *M. edulis* mortalities (Seuront et al. 2019) leading to a poleward contraction in the range of intertidal *M. edulis* populations of the Western Atlantic (Jones et al. 2010). In addition to increased heat stress, warmer oceans also have higher levels of dissolved CO_2 leading to increased Ocean Acidification (OA) which has been shown to affect mussel viability (Asplund et al. 2014). The economic losses to the shellfish industry overall by 2100 due directly to OA have been estimated at over USD$ 1 billion in Europe alone. OA has also been associated with decreased disease resistance in *M. edulis*, both directly and in combinations with increased temperature (Ellis et al. 2015). This makes the mussels more susceptible to bacterial pathogenesis (Ripabelli et al. 1999; Eggermont et al. 2017).

Additionally, at least two relatively recently discovered disseminated neoplasia (DN), both of *M. trossulus* origin, have been associated with reduced fitness in *Mytilus spp.* (Burioli et al. 2021). Furthermore, the effects of hybridisation between the three species of the *Mytilus edulis* complex are yet uncertain with suggested negative effect of *M. trossulus* hybridisation and potential adaptive introgression in the case of *M. galloprovincialis* hybridisation (Fraïsse et al. 2014; Kenchington et al. 2020; Michalek et al. 2021).

In response to these increasing threats and to protect the aquaculture industry, hatchery efforts have been launched in the UK (Regan *et al.* 2021), elsewhere in Europe (Kamermans et al. 2013) and in Canada (Gurney-Smith et al. 2017). However, these efforts have not been without difficulty and a better understanding of the fundamental biology is required. Despite the high commercial significance of these species for aquaculture in addition to the valuable ecosystem services it provides, no chromosomal assembly exists yet for any of these species within the *Mytilus edulis* species complex. Better genomic tools are required to enable further fundamental biological questions to be answered such as inheritance patterns and adaptations.

Like many bivalves, the mussel genome is highly heterozygous (3.5%), and the linear plot of k-mer abundance analysis clearly shows a heterozygous peak in addition to the homozygous peak. The same k-mer analysis estimates a shorter haploid genome length of 1.18 Gb compared to the flow-cytometry data. The estimated repeat content of the *M. edulis* genome is ~43%. These characteristics make assembly of these genomes challenging. However, recently genomes for the American oyster, Pacific Oyster, *M. corruscus* and *M. califoniaus* have been assembled to chromosome-level using short and long-read sequencing technologies as well as Hi-C-based scaffolding (Gomez-Chiarri et al. 2015; Penaloza et al. 2021; Yang et al. 2021; Paggeot et al. 2022). We used a similar approach in this project to produce a highly contiguous assembly. Practical application of this assembly is demonstrated in cross-species synteny analyses and in population structure of *Mytilus* individuals sampled from different regions of the Canadian Atlantic.

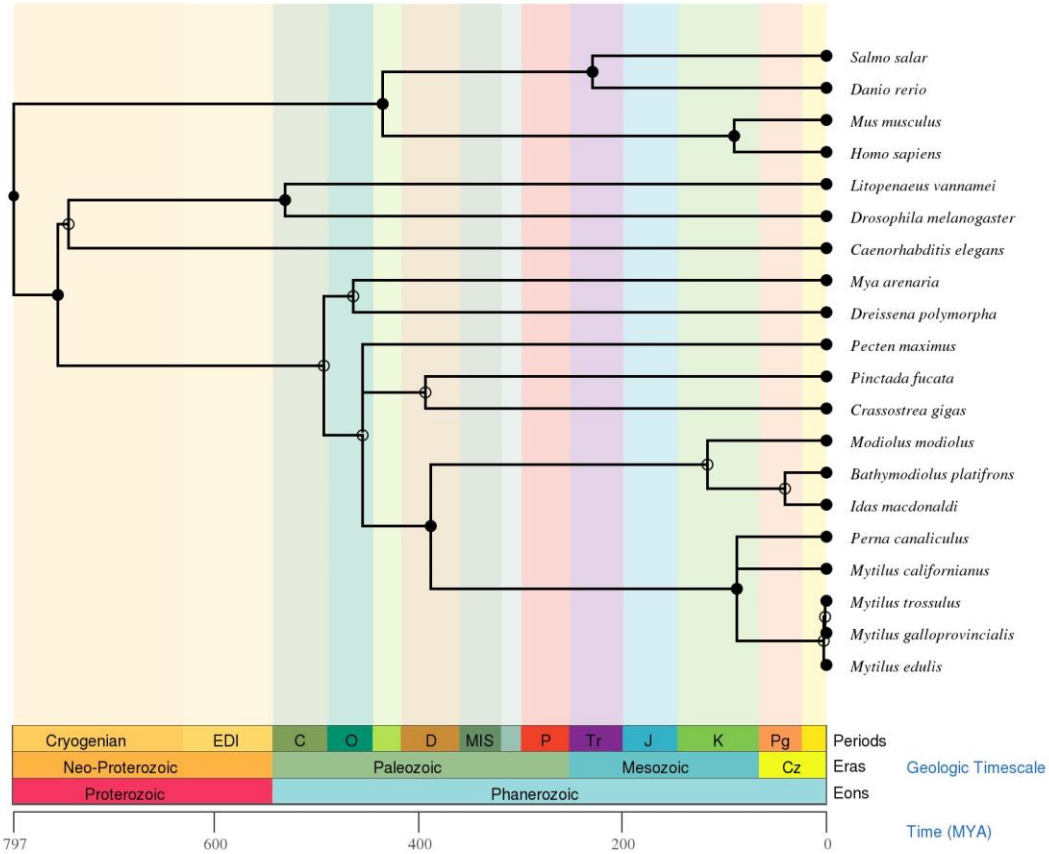

Fig. 1. Mytilidae diverged ~387 MYA. Generated using TimeTree (Kumar et al. 2022).

Methods

DNA extraction, library preparation for genome assembly

One naïve blue mussel sample (Anne) was selected from samples collected by the Provincial Department of Communities and Fisheries in the estuary Foxley river in PE. Foxley River was selected as a sampling site because there is no grow-out aquaculture there (i.e. seed from other bays are not transferred to the area). Due to potential introgression of other species of the *Mytlius* species complex (e.g. *M. trossolus*), this sample was genotyped using 12 SNPs described in (Wilson et al. 2018). We sampled the gill, mantle and muscle of sample "Anne" aseptically, flash froze fresh tissues in liquid nitrogen and preserved them at -80°C. Tissues were shipped to Dovetail Genomics in Scott's Valley, CA, in excess dry ice. Dovetail extracted high molecular weight (HMW) DNA using an in-house modified CTAB method. Dovetail prepared and sequenced PacBIO SMRTbell libraries to a depth of 196X using the Sequel II sequencer. Processed PacBio data (from Dovetail) can be found on SRA under accession number SRX11246493.

Raw contig assembly, scaffold formation and polishing

Dovetail generated a primary contig-level assembly using wtdbg2 (Ruan & Li 2020). The contig-level assembly was filtered of putative duplicated haplotypes and contaminants using Purge Haplotigs (Roach et al. 2018) and Blobtools2 (Challis et al. 2020), respectively. Scaffolding was performed using Omni-C libraries and the HiRise assembler (v1.0). The same DNA sample used by Dovetail was shipped to UWM (University of Wisconsin – Madison) and sequenced to a 100X using the NovaSeq sequencer. These data were trimmed using Trimmomatic (v) and used for polishing with racon (v1.4.3) (Vaser et al. 2017). Completeness was evaluated using BUSCO (v4.0.5 – eukaryota_odb10: 255 BUSCOS, metazoa_odb10: 954 BUSCOS and mollusca_odb10: 5295 BUSCOS) (Manni et al. 2021) and Merqury (Rhie et al. 2020). Merqury analysis was carried out with the same read set used for polishing as the original PacBio CLR reads were not suitable for this analysis. General length metrics were obtained

using QUAST (v5.0.1) (Gurevich et al. 2013; Mikheenko et al. 2018). Synteny mapping between the *M. edulis* and the *M. coruscus* (GCA_017311375.1) (Yang *et al.* 2021) was done using the MCScanX.h function of MCScanX (v2) (Wang et al. 2012). Putative orthologous groups were identified with Orthofinder (v.2.5.4) using predicted gene structures for *M. edulis* (this work) and *M. coruscus* (Yang et al. 2021). Dot plots and circle plots were generated using MCScanX (v.2).

RNA preparation and Isoseq analysis

We shipped flash frozen gill and muscle tissue from sample "Anne" to the Biotechnology Center Core facility at the University of Wisconsin, Madison (UWM). Samples were homogenized using a Qiagen Tissuelyser (2 min @ 20 Hz). RNA was extracted using the RNeasy Mini Kit (Qiagen) with on-column DNAse treatment. UWM performed RNA QC with a nanodrop and bioanalyzer and prepared libraries using the Iso-seq Library SMRTbell express template prep kit. Libraries were sequenced using one Sequel II SMRT cell in CSS mode (i.e. HiFi reads). UWM provided de-multiplexed processed HiFi reads. Putative full-length transcripts were identified using the IsoSeq3 pipeline (https://github.com/ylipacbio/IsoSeq3). Putative open read frames (ORF) were identified using TransDecoder (v5.5.0) (https://github.com/TransDecoder/TransDecoder/wiki).

Annotation

Repeat modeller (4.1.0) (Flynn et al. 2020) was used to predict repeat motifs for *M.* edulis and Repeat Masker (2.0.2a) (Smit et al. 2015) was used to mask the final assembly. Ab-initio annotation was done using Augustus (v3.4.0) (Stanke et al. 2006) trained with genes from *Mytilus galloprovincialis*, *Mytlius coruscus*, *Crassostrea virginica*. Additional hints were generated using short-read RNA-seq (data not shown) and Isoseq data generated herein. The ab-initio annotation was updated using PASA (v2.5) with alignments of a de-novo transcriptome produced using Trinity (v2.8.15) (Grabherr et al. 2011) in Genome-Guided mode. Annotations were updated with PASA twice.

Sample collection, SNP discovery and population structure

We collected gill samples for DNA extractions using standard molecular biology techniques and preserved them in non-denatured 100% ethanol. For population genetics analysis, we sampled mussels in sets of 96 from west to east PEI in Foxley River (FOX) (wild population), Malpeque Bay (MB) (North Shore, Prince County); French River (FR), Stanley Bridge (ST), Whetley River (WR), Tracadie (TRC) (North Shore, Queen county); Orwell (OR) (South Shore, Queen's county; Morell (MRL) (North Shore – King's county) and; Murray River (MR) (South Shore, King county) (Fig. 2). Seed deployed on these sites were originally collected in St. Peter's bay, Brudenell River, and Malpeque bay. We also collected samples in the Bras D'or Lake in Cape Breton (Nova Scotia), the Magdalene Island (Québec) and Notre Dame bay in Newfoundland.

SNP markers, Admixture anlaysis, population structures.

Samples for SNP discovery and population genetics were shipped to LGC genomics in Berlin. Restriction-associated DNA libraries (RAD) libraries were prepared by LGC with using MslI, normalized and sequenced using a NextSeq sequencer. The resulting reads were trimmed and checked for the restriction site by LGC. This final read set was used to identify SNPs and call individual genotypes using Tassel5 (v.5.2.4). SNPs were filtered using MAF (<0.01) and percent of individuals with genotypes. For the SNP discovery, the samples from Cape Breton were excluded from filtering analysis. These samples were shown to be a pure *M. trossolus* population and had missing calls for a significant number of sites. For population genetics analysis, a second set of SNPs that were successfully called across all populations was used. All samples were also genotyped for the 12 species discrimination SNPs from (Wilson *et al.* 2018). All population genetics analyses were performed using the dpca function of the R package Adegenet (v.2.1.7) and STRUCTURE (v.2.3.4) or fastSTRUCTURE (v1.0) (K 3 to K10, 5000 repetitions, 1000 burn in). Analysis of species discrimination SNPs included the genotypes published by (Corrochano-Fraile et al. 2022) as outgroups.

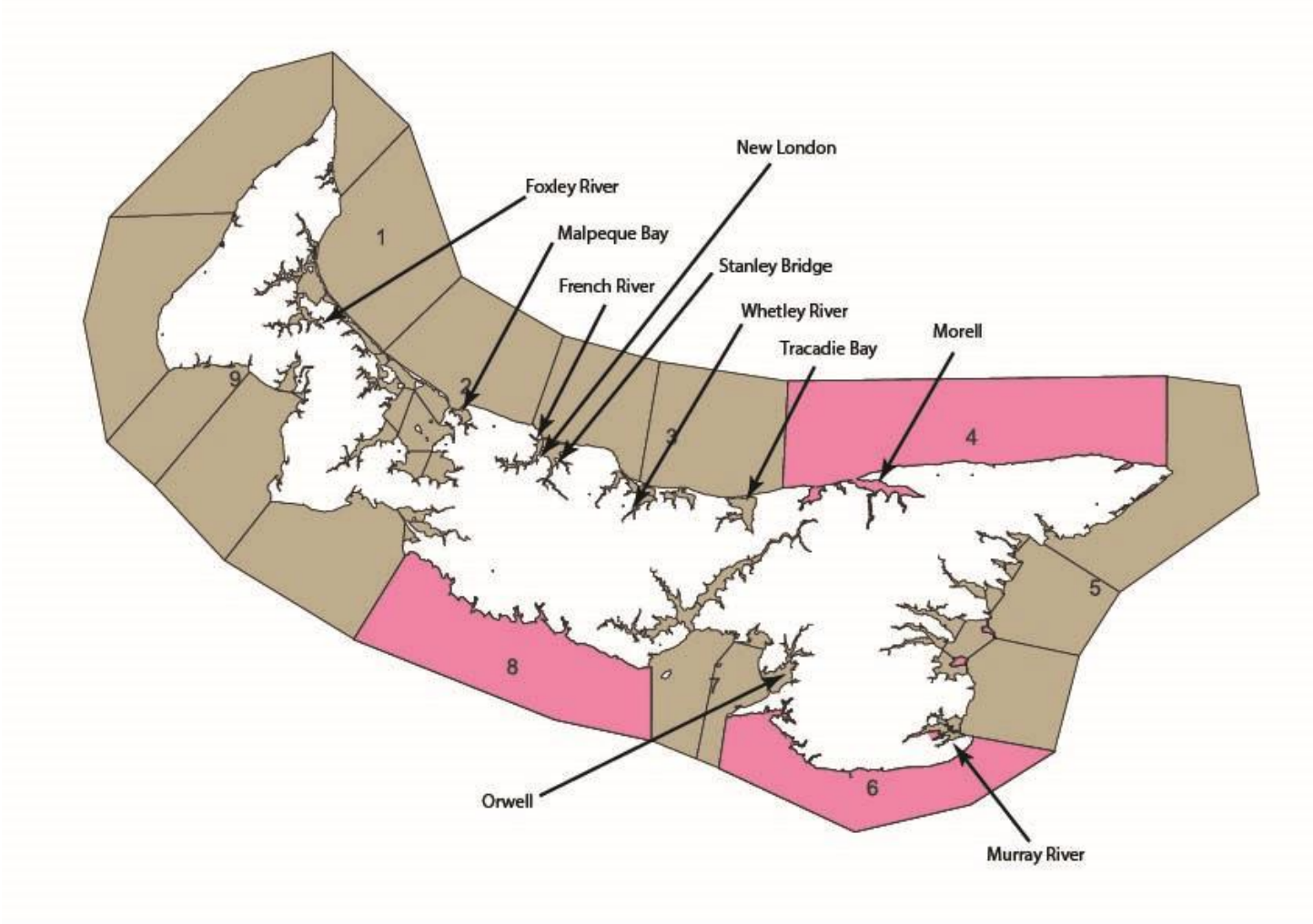

Figure 2: Sampling Locations in PEI

Results

Genome Assembly and Annotation

The chromosome-level assembly presented herein was produced in two stages. First, ~15 million PacBio CLR reads (~340 Gb) were produced, representing coverage of 196X for an estimated genome size of 1.7 Gb (Hinegardner 1974; Rodríguez-Juíz et al. 1996). These reads were assembled into contigs using wtdbg2, which uses uncorrected reads (Ruan & Li 2020). The primary assembly was 1.96 Gb long in 17,825 contigs and a N50 of 443 Kb. After haplotype purging and contaminant removal, the final contig assembly had 10,111 contigs, for a total of 1,65 Gb and a N50 of 518 Kb.

	Anne	**Corrochano-Fraile et al. (2022)**
Length	1,659,567,081 bp	1,827,085,763
# of scaffolds	1,119	3,339
# of contigs	9,866	5,966
N50	116,503,180	1,097,279
Completeness (Merqury)	76.71	NA
QV	32.39	NA

Following scaffolding using Omni-C libraries and the HiRise assembler, we generated a primary chromosome-level assembly made of 2,117 contigs. We removed putative contaminants using Blobtools by eliminating sequences coming from non-molluscan organisms. This assembly was further filtered to contain only sequences > 5,000 bp (998 scaffolds - > 8 Mb – 0.05% of the raw assembly). The resulting draft is deposited on NCBI assembly under accession number GCA_019925275.1. The final assembly is made of 1,119 contigs and has an N50 of 116 Mb. The 14 putative *M. edulis* chromosomes are deposited under accession numbers CM034349.1 to CM34362.1. Detailed statistics for the assemblies can be found in Table 1.

Completeness analysis in Merqury resulted in 76.71% recovery of k-mers present in the polishing Illumina data from the final version of the assembly. We also evaluated the completeness of the assembly when combined with purged haplotypes, which was 83.44%. QV value for the primary assembly was 32.39, while the combined draft had a QV of 30.73. BUSCO analysis showed a recovery of 91.3%, 91.1%, and 79.3% complete BUSCOS against the eukaryote, metazoan and molluscan databases, respectively.

Synteny analysis showed a high degree of collinearity between putative chromosomes of *M. edulis* and *M. coruscus* (Fig. 3). However, putative inversions, transposition and deletion can be observed in almost all chromosomes. Gene order in chromosomes represented by sequences CM034349.1 (*M. edulis* chromosome 1) - CM029595.1 (*M. coruscus* LG1) and CM034343.1 (*M. edulis* chromosome 4) – CM029599.1 (*M. coruscus* LG5) showed the highest conservation. Putative orthologous relationships between *M. edulis* chromosomes and *M. coruscus* linkage groups (LG) are shown in Table 2.

Isoseq3 analysis of CSS data from muscle and gill (2.9 million reads -188,165 Mb) identified 216,343 high-quality putative full-length (FL) transcripts (from 2.8 million reads containing poly-A tails). We identified 196,111 putative open reading frames from the 216,343 FL transcripts. BLASTp Analysis against the uniref90 database returned informative hits for 83% (164,969) of these translated transcripts. *Ab-initio* gene prediction in Augustus detected 46,604 gene models that produced 46,604 transcripts. After two rounds of PASA updates, the final number of gene models was 47,128 and the number of transcripts was 55,138.

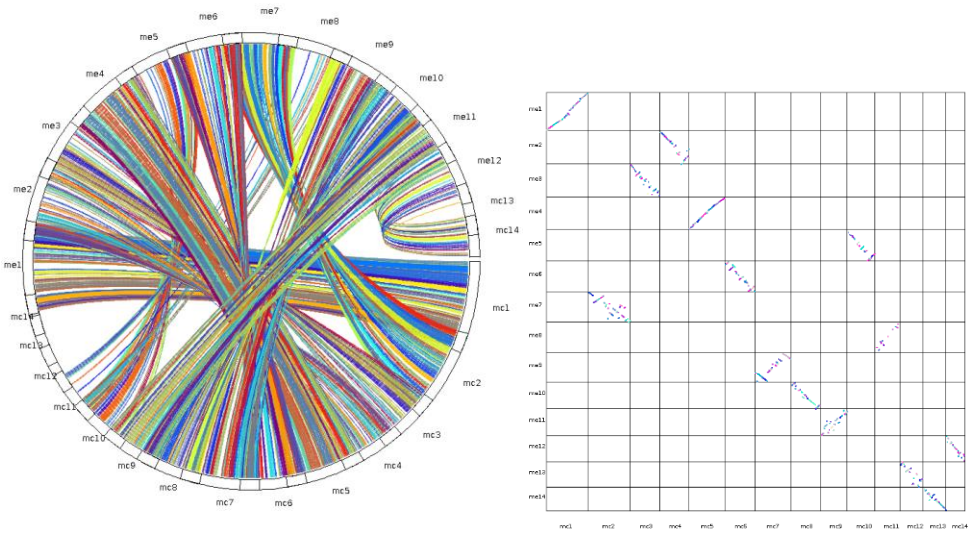

Figure 3. Synteny between the *M. edulis* and *m. coruscus* putative chromosomes

Table 2: Putative synteny between *M. edulis* and *M. coruscus.* Ids are NCBI Assembly database Molecule name and accession number.

M. edulis	*M. coruscus*
Chromosome 1 (CM034349.1)	LG01 (CM029595.1)
Chromosome 2 (CM034350.1)	LG04 (CM029598.1)
Chromosome 3 (CM034351.1)	LG03 (CM029597.1)
Chromosome 4 (CM034352.1)	LG05 (CM029599.1)
Chromosome 5 (CM034353.1)	LG10 (CM029604.1)
Chromosome 6 (CM034354.1)	LG06 (CM029600.1)
Chromosome 7 (CM034355.1)	LG02 (CM029596.1)
Chromosome 8 (CM034356.1)	LG11 (CM029605.1)
Chromosome 9 (CM034357.1)	LG07 (CM029601.1)
Chromosome 10 (CM034358.1)	LG08 (CM029602.1)
Chromosome 11 (CM034359.1)	LG09 (CM029603.1)
Chromosome 12 (CM034360.1)	LG14 (CM029608.1)
Chromosome 13 (CM034361.1)	LG12 (CM029606.1)
Chromosome 14 (CM034362.1)	LG13 (CM029607.1)

SNP discovery and population structure

Population Structure analysis conducted using 12 SNPs (Wilson *et al.* 2018) clearly separated *M. edulis* populations in PEI from other regions of Canada. We generated two sets of SNPs: the first set totalling 71,231 SNPs using only samples from PEI and made a polymorphic collection of SNPs in *M. edulis.* The second set, with ~6,000 markers, is a polymorphic set of SNPs in both *M. edulis* and *M. trossolus.* Population structure and putative admixture are shown in Figure 4. In the DPCA and PCA, untrained clustering clearly separated both CB from PEI/NL/MAG and also the putative populations in the Gulf of St. Lawrence. In green, are the majority of samples from NL and MAG, while blue represents individuals from PEI.

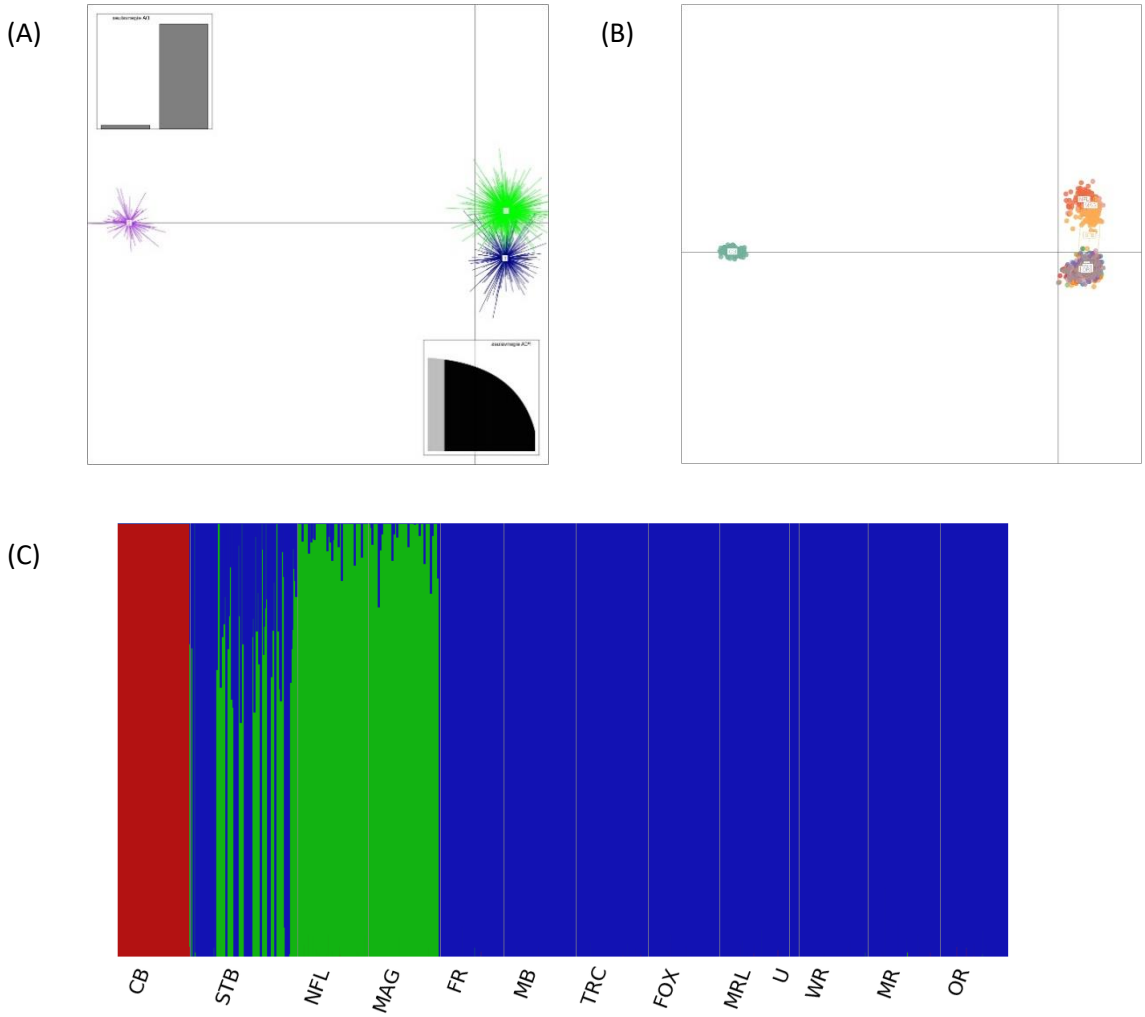

Figure 4 Population structure in the North American Atlantic. (A) DPCA with 6,000 SNPs, (B) PCA with 6,000 SNPs, (C) fastStructure with 6,000 SNPs. Samples from Cape Breton NS (CB), FR, TRC, MB, FOX, MRL, U, WR, MR, OR (PEI), Magdalene Islands QC (MAG), Notre Dame Bay NL (NFL)

Discussion

Here, we describe the first chromosome-level genome assembly for the blue mussel, *M. edulis*. Compared to the assembly published in Corrochano-Fraile et al. (2022), the assembly presented herein has better contiguity (Table 1). Our assembly is also shorter than the assembly from Corrochano-Fraile et al. (2022). The two assemblies' length falls close to the estimated size of the genome based on c-value (1.7) (Rodríguez-Juíz *et al.* 1996) and is significantly longer than what is estimated by k-mer abundance analysis with GenomeScope (1.18 Gb).

Despite the total assembly length being close to that estimated using c-values, k-mer-based completeness analysis recovers only 76% in a set of Illumina reads origination from sample “Anne”. When the putative purged haplotigs were added back to the assembly, recovery was ~83%. This apparent low k-mer recovery is probably a combination of the consensus being different from either haploid genomes, the error rate in the original PacBio data, the fact that the reads used for polishing were not used from the primary assembly, the contigs removed based on length or contaminant status, and the gaps arising from the Omni-C scaffolding. It is also possible that the high heterozygosity affects the accuracy of k-mer abundance analysis, as shown by the large discrepancy between genome size estimates.

The 14 putative chromosomes represent ~96% of the assembly, with lengths varying from 140 Mbp to 90 Mbp. These data and the N50 metric show that this assembly has high contiguity and that this assembly and its annotation will be highly useful for aquaculture, evolution and molecular ecology studies. Herein, we illustrate the possible applications of this assembly by performing population and synteny analyses.

Ab-Initio analysis resulted in 47,128 gene models and 55,138 transcripts after PASA update. The number of transcripts is much lower than that identified with IsoSeq. This is likely due to low sensitivity of PASA to incorporate novel isoforms into the ab-initio annotation. BUSCO analysis using Metazoa,

Eukaryota and Mollusca databases (odb10) shows that the IsoSeq transcriptome recovers more BUSCOs (>90%) than the ab-initio transcripts. Therefore the real number of transcripts for *M.edulis* is likely closer to 126,000 than 55,000. Corrochano-Fraile et al. (Corrochano-Fraile *et al.* 2022) identified larger number of genes (69,246), but did not report number of transcripts arising from these genes. Therefore, the number of genes in *M. edulis* is likely between 47,000 to 69,000.

Due to the close relationship between the members of the *Mytilus* species complex, we wanted to verify that the individual sampled (Anne) was pure *M. edulis*. Using the 12 SNPs published by (Wilson *et al.* 2018) we determined that is the case. We used *M. trossolus*, *M. galloprovincialis* and European *M. edulis* genotypes as outgroups. The later inclusion of 96 individuals from Cape Breton (NS) showed no significant evidence of *M. trossolus* introgression in PEI samples. Given that Cape Breton has long been considered a pure *M. trossolus* population (Wilson et al., 2018), we are confident that the sample Anne represents an *M. edulis* individual. We also genotyped over 500 PEI individuals and 96 samples from Magdalene Island and 96 individuals from Newfoundland with the same 12 SNP panel. As before no significant introgression of *M. trossolus* was detected in PEI. However, we only genotyped 96 samples collected in an area with no grow-out leases (Foxley River). Although unlikely, we cannot rule out the possibility of a sampling bias in aquaculture sites favouring *M. edulis.* DPCA and Structure analysis indicate that there is low population stratification between different regions of PEI, while the Magdalene islands and NL are distinct populations. The populations from NL and the Magdalene Islands are more similar to each other than from the populations from PEI. For animals from one sampling event at S.t Peter's bay, we found evidence of shared genetics between PEI and the population to the northeast of the island.

The taxonomic status of the "species" in the *Mytilus* species complex remains in debate. Chromosome-level assemblies allow the study of macroevolution of the genome by looking at synteny across species. Herein, we present the synteny analysis between the 14 putative chromosomes of *M. edulis* and *M. corruscus* (Yang et al. 2021) to exemplify how chromosome-level assemblies may allow us to

better understand the phylogenetic relationships within the genus *Mytilus*. This analysis shows that some of putative orthologous chromosomes of the 2 species maintain high-levels of collinearity (e.g. chromosomes 1 and 4 from *M. edulis* with LG1 and LG5 from *M. coruscus* respectively) while others present significant levels of re-arrangements (e.g. chromosomes 7 and 11 from *M. edulis* with LG2 and LG9 from *M. coruscus* respectively). An in-depth analysis of chromosome synteny will shed light on the level of collinearity between multiple members of the genus *Mytilus*. Homology between chromosomes is a key element of the viability of hybrids. Reproductive isolation tends to increase during speciation, and these resources will permit further studies on the reproductive compatibility of the species in genus *Mytilus* at the chromosome level.

Acknowledgements

Industry Partners: Mussel King (Scott Dockendorff), PEI Aqua Farms (John Paquet), Atlantic Aqua Farms (Terry Ennis).

PEIMSO: Angela MacDonald, Neil McNair, Aaron Ramsey, Jessica Fry

Genome Atlantic: Kristin Tweel, Charmaine Gaudet, Nil d'Entremont

Service Providers: LGC Berlin and UK, Genome Quebec, Dovetail Genomics

1. Asplund M.E., Baden S.P., Russ S., Ellis R.P., Gong N. & Hernroth B.E. (2014) Ocean acidification and host-pathogen interactions: blue mussels, Mytilus edulis, encountering Vibrio tubiashii. Environ Microbiol 16, 1029-39.
2. Broszeit S., Hattam C. & Beaumont N. (2016) Bioremediation of waste under ocean acidification: Reviewing the role of Mytilus edulis. Marine Pollution Bulletin 103, 5-14.
3. Burioli E.A.V., Hammel M., Bierne N., Thomas F., Houssin M., Destoumieux-Garzón D. & Charrière G.M. (2021) Traits of a mussel transmissible cancer are reminiscent of a parasitic life style. Scientific Reports 11, 24110.
4. Challis R., Richards E., Rajan J., Cochrane G. & Blaxter M. (2020) BlobToolKit – Interactive Quality Assessment of Genome Assemblies. G3 Genes|Genomes|Genetics 10, 1361-74.
5. Chase M., Jones S.H., Hennigar P., Sowles J., Harding G., Freeman K., Wells P., Krahforst C., Coombs K. & Crawford R. (2001) Gulfwatch: Monitoring spatial and temporal patterns of trace metal and organic contaminants in the Gulf of Maine (1991–1997) with the blue mussel, Mytilus edulis L. Marine Pollution Bulletin 42, 490-504.
6. Corrochano-Fraile A., Davie A., Carboni S. & Bekaert M. (2022) Evidence of multiple genome duplication events in Mytilus evolution. BMC Genomics 23, 340.
7. Eggermont M., Bossier P., Pande G.S.J., Delahaut V., Rayhan A.M., Gupta N., Islam S.S., Yumo E., Nevejan N., Sorgeloos P., Gomez-Gil B. & Defoirdt T. (2017) Isolation of Vibrionaceae from wild blue mussel (Mytilus edulis) adults and their impact on blue mussel larviculture. FEMS Microbiology Ecology 93.
8. Ellis R.P., Widdicombe S., Parry H., Hutchinson T.H. & Spicer J.I. (2015) Pathogenic challenge reveals immune trade-off in mussels exposed to reduced seawater pH and increased temperature. Journal of Experimental Marine Biology and Ecology 462, 83-9.
9. FAO (2020) The State of World Fisheries and Aquaculture 2020. Sustainability in action. . In: *The State of World Fisheries and Aquaculture (SOFIA)*, p. #244 p. FAO, Rome, Italy.
10. Flynn J.M., Hubley R., Goubert C., Rosen J., Clark A.G., Feschotte C. & Smit A.F. (2020) RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences 117, 9451-7.
11. Fraïsse C., Roux C., Welch J.J. & Bierne N. (2014) Gene-flow in a mosaic hybrid zone: is local introgression adaptive? Genetics 197, 939-51.
12. Gomez-Chiarri M., Warren W.C., Guo X. & Proestou D. (2015) Developing tools for the study of molluscan immunity: The sequencing of the genome of the eastern oyster, Crassostrea virginica. Fish Shellfish Immunol 46, 2-4.
13. Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N. & Regev A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29, 644-52.
14. Gurevich A., Saveliev V., Vyahhi N. & Tesler G. (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072-5.

15. Gurney-Smith H.J., Wade A.J. & Abbott C.L. (2017) Species composition and genetic diversity of farmed mussels in British Columbia, Canada. Aquaculture 466, 33-40.
16. Hayward P.J. & Ryland J.S. (2017) *Handbook of the marine fauna of North-West Europe*. Oxford University Press.
17. Hinegardner R. (1974) Cellular DNA content of the Mollusca. Comp Biochem Physiol A Comp Physiol 47, 447-60.
18. Jones S.J., Lima F.P. & Wethey D.S. (2010) Rising environmental temperatures and biogeography: poleward range contraction of the blue mussel, Mytilus edulis L., in the western Atlantic. Journal of Biogeography 37, 2243-59.
19. Kamermans P., Galley T., Boudry P., Fuentes J., McCombie H., Batista F.M., Blanco A., Dominguez L., Cornette F., Pincot L. & Beaumont A. (2013) 11 - Blue mussel hatchery technology in Europe. In: Advances in Aquaculture Hatchery Technology (ed. by G. Allan & G. Burnell), pp. 339-73. Woodhead Publishing.
20. Kenchington E.L., MacDonald B.W., Cogswell A., Hamilton L.C. & Diz A.P. (2020) Sex-specific effects of hybridization on reproductive fitness in Mytilus. Journal of Zoological Systematics and Evolutionary Research 58, 581-97.
21. Krone R., Gutow L., Joschko T.J. & Schröder A. (2013) Epifauna dynamics at an offshore foundation – Implications of future wind power farming in the North Sea. Marine Environmental Research 85, 1-12.
22. Kumar S., Suleski M., Craig J.M., Kasprowicz A.E., Sanderford M., Li M., Stecher G. & Hedges S.B. (2022) TimeTree 5: An Expanded Resource for Species Divergence Times. Mol Biol Evol.
23. MacDonald B.A., Robinson S.M.C. & Barrington K.A. (2011) Feeding activity of mussels (Mytilus edulis) held in the field at an integrated multi-trophic aquaculture (IMTA) site (Salmo salar) and exposed to fish food in the laboratory. Aquaculture 314, 244-51.
24. Manni M., Berkeley M.R., Seppey M. & Zdobnov E.M. (2021) BUSCO: Assessing Genomic Data Quality and Beyond. Current Protocols 1, e323.
25. McDonald J., Seed R. & Koehn R. (1991) Allozymes and morphometric characters of three species ofMytilus in the Northern and Southern Hemispheres. Marine Biology 111, 323-33.
26. McEneff G., Barron L., Kelleher B., Paull B. & Quinn B. (2014) A year-long study of the spatial occurrence and relative distribution of pharmaceutical residues in sewage effluent, receiving marine waters and marine bivalves. Science of the Total Environment 476, 317-26.
27. Michalek K., Vendrami D.L.J., Bekaert M., Green D.H., Last K.S., Telesca L., Wilding T.A. & Hoffman J.I. (2021) Mytilus trossulus introgression and consequences for shell traits in longline cultivated mussels. Evolutionary applications 14, 1830-43.
28. Mikheenko A., Prjibelski A., Saveliev V., Antipov D. & Gurevich A. (2018) Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34, i142-i50.
29. Norling P. & Kautsky N. (2007) Structural and functional effects of Mytilus edulis on diversity of associated species and ecosystem functioning. Marine Ecology Progress Series 351, 163-75.
30. Paggeot L.X., DeBiasse M.B., Escalona M., Fairbairn C., Marimuthu M.P.A., Nguyen O., Sahasrabudhe R. & Dawson M.N. (2022) Reference genome for the California ribbed mussel, Mytilus californianus, an ecosystem engineer. J Hered.
31. Penaloza C., Gutierrez A.P., Eory L., Wang S., Guo X., Archibald A.L., Bean T.P. & Houston R.D. (2021) A chromosome-level genome assembly for the Pacific oyster Crassostrea gigas. Gigascience 10.
32. Regan T., Bean T.P., Ellis T., Davie A., Carboni S., Migaud H. & Houston R.D. (2021) Genetic improvement technologies to support the sustainable growth of UK aquaculture. Reviews in Aquaculture 13, 1958-85.
33. Rhie A., Walenz B.P., Koren S. & Phillippy A.M. (2020) Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biology 21, 245.

34. Ripabelli G., Sammarco M.L., Grasso G.M., Fanelli I., Caprioli A. & Luzzi I. (1999) Occurrence of Vibrio and other pathogenic bacteria in Mytilus galloprovincialis (mussels) harvested from Adriatic Sea, Italy. International Journal of Food Microbiology 49, 43-8.
35. Roach M.J., Schmidt S.A. & Borneman A.R. (2018) Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 19, 460.
36. Rodríguez-Juíz A.M., Torrado M. & Méndez J. (1996) Genome-size variation in bivalve molluscs determined by flow cytometry. Marine Biology 126, 489-97.
37. Ruan J. & Li H. (2020) Fast and accurate long-read assembly with wtdbg2. Nature Methods 17, 155-8.
38. Seuront L., Nicastro K.R., Zardi G.I. & Goberville E. (2019) Decreased thermal tolerance under recurrent heat stress conditions explains summer mass mortality of the blue mussel Mytilus edulis. Scientific Reports 9, 17498.
39. Smit A., Hubley R. & Green P. (2015) RepeatMasker Open-4.0. http://www.repeatmasker.org.
40. Stanke M., Keller O., Gunduz I., Hayes A., Waack S. & Morgenstern B. (2006) AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic acids research 34, W435-W9.
41. Vaser R., Sović I., Nagarajan N. & Šikić M. (2017) Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27, 737-46.
42. Wang Y., Tang H., Debarry J.D., Tan X., Li J., Wang X., Lee T.H., Jin H., Marler B., Guo H., Kissinger J.C. & Paterson A.H. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research 40, e49.
43. Wilson J., Matejusova I., McIntosh R.E., Carboni S. & Bekaert M. (2018) New diagnostic SNP molecular markers for the Mytilus species complex. PLoS One 13, e0200654.
44. Yang J.L., Feng D.D., Liu J., Xu J.K., Chen K., Li Y.F., Zhu Y.T., Liang X. & 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.