Chromosome level reference genome for European flat oyster (Ostrea edulis L.)

Received: 21 April 2022 | Revised: 7 July 2022 | Accepted: 13 July 2022

DOI: 10.1111/eva.13460

ORIGINAL ARTICLE

Evolutionary Applications Open Access WILEY

Chromosome-level reference genome for European flat oyster (*Ostrea edulis* L.)

Manu Kumar Gundappa[1] | Carolina Peñaloza[1] | Tim Regan[1] | Isabelle Boutet[2] | Arnaud Tanguy[2] | Ross D. Houston[1] | Tim P. Bean[1] | Daniel J. Macqueen[1]

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

[2]Station Biologique de Roscoff, Laboratoire Adaptation et Diversité en Milieu Marin (UMR 7144 AD2M CNRS-Sorbonne Université), Roscoff, France

Correspondence

Manu Kumar Gundappa and Daniel J. Macqueen, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Edinburgh, UK.
Email: manu.gundappa@roslin.ed.ac.uk and daniel.macqueen@roslin.ed.ac.uk

Funding information

Biotechnology and Biological Sciences Research Council (BBSRC), Grant/Award Number: BB/S004181/1 and BBS/E/D/10002070

Abstract

The European flat oyster (*Ostrea edulis* L.) is a bivalve naturally distributed across Europe, which was an integral part of human diets for centuries, until anthropogenic activities and disease outbreaks severely reduced wild populations. Despite a growing interest in genetic applications to support population management and aquaculture, a reference genome for this species is lacking to date. Here, we report a chromosome-level assembly and annotation for the European Flat oyster genome, generated using Oxford Nanopore, Illumina, Dovetail OmniC™ proximity ligation and RNA sequencing. A contig assembly (N50: 2.38 Mb) was scaffolded into the expected karyotype of 10 pseudochromosomes. The final assembly is 935.13 Mb, with a scaffold-N50 of 95.56 Mb, with a predicted repeat landscape dominated by unclassified elements specific to *O. edulis*. The assembly was verified for accuracy and completeness using multiple approaches, including a novel linkage map built with ddRAD-Seq technology, comprising 4016 SNPs from four full-sib families (eight parents and 163 F1 offspring). Annotation of the genome integrating multitissue transcriptome data, comparative protein evidence and ab-initio gene prediction identified 35,699 protein-coding genes. Chromosome-level synteny was demonstrated against multiple high-quality bivalve genome assemblies, including an *O. edulis* genome generated independently for a French *O. edulis* individual. Comparative genomics was used to characterize gene family expansions during *Ostrea* evolution that potentially facilitated adaptation. This new reference genome for European flat oyster will enable high-resolution genomics in support of conservation and aquaculture initiatives, and improves our understanding of bivalve genome evolution.

KEYWORDS

aquaculture, bivalves, comparative genomics, conservation and restoration, European flat oyster, genetic map, reference genome

1 | INTRODUCTION

The European flat oyster *Ostrea edulis* (Linnaeus, 1758) is a bivalve mollusc within Ostreidae ('true oysters'). This species is a native of Europe, naturally distributed from 65 degrees North in Norway to 30 degrees North in Morocco, along the North-Eastern Atlantic, and also the entire Mediterranean basin (Thorngren et al., 2019). Introductions of *O. edulis* in the 19th and 20th centuries for aquaculture resulted in the establishment of natural beds in many regions across the world, including North America, New Zealand, Australia and Japan (Bromley et al., 2016). *O. edulis* can reach sizes exceeding 20 cm and has a life span of up to 20 years (Bayne, 2017). This species is a protandrous hermaphrodite that can change sex within a spawning season, and unlike the more widely cultured Pacific oyster *Crassostrea gigas*, brood their larvae in the inhalant chamber for several days before release (Suquet et al., 2018). *O. edulis* exhibits extensive physiological plasticity across its range, for example, the temperature at which spawning occurs (11–25°C degrees) and the duration of the spawning period (from 1 to 2 months to year-round; Bromley, 2015; Bromley et al., 2016).

Ostrea edulis has been an integral part of human diets in Europe for centuries, with evidence for its collection and consumption since at least Roman times. Furthermore, it is thought >700 million oysters were consumed in London alone during 1864 (Pogoda, 2019). However, overfishing and anthropogenic activities have driven the collapse of *O. edulis* stocks throughout its natural range (Merk et al., 2020; Pogoda, 2019). The past 40 years have witnessed a further decline in production, with a peak of 32,995 tonnes in 1961 dropping by >90% to 3120 tonnes by 2016 (FAO, 2020). Human impacts are widely cited as the primary reason for this decline, including habitat destruction, overexploitation, the introduction of non-native species competing for *O. edulis* habitats (Grizel & Héral, 1991; Vera et al., 2019), and the emergence/spread of diseases associated with translocations (Bromley et al., 2016). Key parasites associated with flat oyster population declines include the protist *Marteilia refringens* and the haplosporidian protozoan parasite *Bonamia ostreae*, which causes bonamiosis, for which no effective control methods exist (Sas et al., 2020). Large-scale restoration efforts exemplified by the Native Oyster Restoration Alliance (NORA; https://noraeurope.eu/) are targeting the re-stocking of *O. edulis* at high densities and developing sustainable populations. However, these efforts are strongly hampered by parasitic diseases, especially bonamiosis (Engelsma et al., 2010; Pogoda et al., 2019). While using animals from *Bonamia*-free regions offers a potential short-term solution for restoration and aquaculture efforts, understanding the genetic basis for natural parasite resistance (Sas et al., 2020) will enable selective breeding to enhance *Bonamia* resistance and permanently reduce disease incidence in farmed and wild populations.

Several studies have applied genetic and genomic tools to study *O. edulis* in the absence of a reference genome assembly. Such work has been strongly targeted towards understanding bonamiosis, either by identifying candidate quantitative trait loci (QTL) and genetic outliers linked to *Bonamia* resistance (Harrang et al., 2015; Lallias et al., 2009; Vera et al., 2019) or by studying gene expression responses to *Bonamia* infection (Pardo et al., 2016; Ronza et al., 2018). SNP genotyping arrays with low (Lapègue et al., 2014) and medium (Gutierrez et al., 2017) density have also been developed for genetics applications. The lack of a high-quality reference genome in *O. edulis*, however, contrasts with the situation in the commercially valuable Pacific oyster *C. gigas* (Peñaloza et al., 2021; Qi et al., 2021) and is a current limitation for the research community. An annotated genome for *O. edulis* will enable genetics research in many directions supporting conservation and aquaculture, revealing the physical location of genetic variation with respect to genes and genomic features and offering an essential foundation for functional genomics. A reference genome will also support our understanding of *O. edulis* evolution and environmental adaptation, through comparisons with other bivalve species.

Bivalve genome assembly has classically been hampered by genetic complexities including high heterozygosity and repeat content (Davison & Neiman, 2021), along with the challenge of extracting pure high molecular weight DNA (Adema, 2021). However, recent advances in long-read sequencing technologies have enabled high-quality genome sequences to be generated for multiple bivalves, including *C. gigas* (Peñaloza et al., 2021; Qi et al., 2021), the scallop *Pecten maximus* (Kenny et al., 2020) and hard clam *Mercenaria mercenaria* (Farhat et al., 2022; Song et al., 2021). Here, we integrated multiple sequencing technologies to assemble and annotate a highly contiguous chromosome-level genome assembly for an *O. edulis* individual from the UK, which was confirmed for accuracy by comparison to a novel linkage map for *O. edulis*, and high-quality genome assemblies for several bivalve species. Comparative genomics inclusive of diverse bivalve species allowed us to define gene copy expansions in the *Ostrea* lineage. The high-quality reference genome reported here, and an independent *O. edulis* assembly reported for an individual from a distinct European population in the same issue of this journal by Boutet et al. (2022) (see commentary by Bean et al., 2022), will support ongoing conservation and aquaculture initiatives for the European flat oyster, while improving our comparative understanding of genome evolution and adaptation in the *Ostrea* lineage.

2 | MATERIALS AND METHODS

2.1 | Sampling and sequencing

A single unsexed adult *O. edulis* individual sourced from Whitstable (England, UK) through a commercial supplier (Simply Oysters) was used for all DNA and RNA sequencing performed in this study, as described below. The oyster was depurated in clean seawater for at least 42 h before sampling. Samples of gill, mantle, heart, white muscle, striated muscle, digestive gland, labial palp and gonad were flash frozen using liquid nitrogen and stored at −80°C. High molecular weight DNA was extracted from gill using a cetyltrimethylammonium bromide (CTAB) based extraction method and used to generate short and long-read sequencing libraries. DNA purity was confirmed using a Nanodrop 1000 (Thermo Fisher Scientific). DNA integrity was initially assessed using a Tapestation 4200 (Agilent Technologies). The DNA was purified using Ampure beads (Beckman Coulter™), sheared to a length of ~35 kb using a Megaruptor® (Diagenode) and size selected in the 7–50 kb range on a Bluepippin system (Sage Science) with a 0.75% cassette. The resulting DNA was sequenced on four PromethION flow cells (FLO-PRO002), with base calling performed using Guppy version 3.2.6 + afc8e14. Short-read libraries with an insert size of 350 bp were generated using the same DNA with an Illumina TruSeq DNA library kit, prior to sequencing on an Illumina NovaSeq 6000 by Novogene Ltd (UK) with a paired-end 150 bp configuration. An Omni-C™ library was generated from gill tissue by Dovetail Genomics (Santa Cruz, USA) and sequenced on an Illumina HiSeq X with a paired-end 150 bp configuration.

For RNA-Seq library generation, total RNA was extracted for the eight tissues using a Trizol-based method, before DNAase treatment. RNA integrity was assessed using agarose gel electrophoresis and Bioanalyszer 2100 (Agilent). RNA purity was confirmed via a Nanodrop 1000 system. Illumina TruSeq mRNA libraries were prepared for each sample and sequenced on an Illumina NovaSeq 6000 with a paired-end 150 bp configuration by Novogene Ltd (UK).

2.2 | Genome assembly and scaffolding

Genome size and heterozygosity were estimated using a k-mer approach. The Illumina data were quality assessed using FastQC v0.11.8 (Andrews, 2010), trimmed using TrimGalore 0.4.5 (Krueger, 2015; quality score > 30, minimum length > 40 bp) and processed through Meryl v1.3 (Rhie et al., 2020) to generate a k-mer count database ($k = 20$), which was used to generate a k-mer histogram. The histogram data were used as an input to Genomescope 2.0 (Ranallo-Benavidez et al., 2020) to estimate genome size and heterozygosity.

Contig assembly was performed using the nanopore data with the repeat graph-based assembler Flye 2.7-b1585 (Kolmogorov et al., 2019). Three contig assemblies were generated (*OE_F1*, *OE_F2*, *OE_F3*) setting the *minimum-overlap* parameter to '5000', '10,000' and 'auto', respectively, with all other parameters default. In parallel, the raw nanopore reads were error corrected using the *correct* module within Necat v0.0.1 (Chen, Wang, et al., 2021). The corrected reads were also assembled to contigs using the overlap-based assembler wtdbg2 2.5 (Ruan & Li, 2020) with default parameters, generating the assembly *OE_RB1*. The Flye and wtdbg2 assemblies were passed through pseudohaploid (https://github.com/schatzlab/pseudohaploid) to purge uncollapsed haplotigs. The three purged Flye assemblies (*OE_F1_purged*, *OE_F2_purged*, *OE_F3_purged*) were merged using Quickmerge v0.3 (Chakraborty et al., 2016) setting the parameters *-hco 5.0 -c 1.5 -l n -ml m* to generate a merged assembly (*Flye_Merged*). Finally, the *Flye_Merged* and haplotig purged wtdbg2 (*OE_RB1_purged*) assemblies were merged using Quickmerge v0.3 (as above) to generate a final contig assembly (*OE_contig_v1*), which was polished for two rounds using quality-trimmed Illumina data with Pilon v1.24 (Walker et al., 2014; *OE_contig_pilon_v1*).

The polished contig assembly was scaffolded by Dovetail Genomics using HiRise (Putnam et al., 2016) with the Omni-C™ proximity ligation sequencing data used to orient and link the contigs using 3D contact information. The top 10 super-scaffolds with the HiRise assembly were > 40 Mb and matched the expected *O. edulis* karyotype ($n = 10$; Horváth et al., 2013; Leitao et al., 2002; Thiriot-Quiévreux, 1984; Figure 1a). The next two largest scaffolds (scaffolds 11 and 12, respective sizes: 13.5 and 9.4 Mb) were not assigned to one of the 10 super-scaffolds despite their large size, which led us to hypothesize these regions belonged to the 10 super-scaffolds, yet had not been scaffolded by HiRise. In support of this hypothesis, visualization of the 3D contact information using Juicebox (Durand, Robinson, et al., 2016) revealed 3D contacts between HiRise scaffold 11 and scaffold 6 and between HiRise scaffold 12 and scaffold 1 (Figure S1). To confirm these interactions, we repeated contig scaffolding with the Omni-C™ data using Juicer (default parameters; Durand, Shamim, et al., 2016), and the resultant assembly was aligned and compared with the HiRise assembly using QUAST (Gurevich et al., 2013). Visualization of QUAST alignments in Icarus (Mikheenko et al., 2016) confirmed the locations of scaffolds 11 and 12 within super-scaffolds 6 and 1, respectively (Figure S1). Manual integration of these scaffolds in the HiRise assembly was performed using Scaffolder (Barton & Barton, 2012). Following this work, super-scaffold 6 became the second largest super-scaffold and was therefore renamed to be super-scaffold 2, and this annotation is used hereafter. The resulting scaffolds were polished for one round using Pilon v1.24, leading to the final assembly used in all downstream work (*OE_Roslin_V1*).

2.3 | Genome quality evaluation

OE_Roslin_V1 was screened for the presence of DNA contamination from other taxa using Blobtools v1.1.1 (Laetsch & Blaxter, 2017b) and for misassembly errors using Inspector v1.0.2 (Chen, Zhang, et al., 2021). Structural errors identified in the genome were corrected using the Inspector-correct.py step. The raw nanopore reads were mapped back to the *OE_Roslin_V1* assembly using minimap2 (Li, 2018; parameter *-ax map-ont*) to check for assembly completeness. The genome assembly was compared with a novel linkage map to confirm the accuracy of scaffolding using the chromatin proximity Omni-C™ data (see later section). Assembly quality and efficiency of haplotig purging were evaluated by generating a copy number spectrum plot (tracking the multiplicity of each k-mer in the read set, revealing the number of times it is found in the genome assembly) using Merqury v1.3 (Rhie et al., 2020). Gene completeness was evaluated against a set of 5295 benchmark molluscan orthologous genes (*mollusca_odb10*) using BUSCO v4.1.4 (Simão et al., 2015). We mapped paired-end Ilumina data from the same individual to the finished genome assembly using the minimap2 (Li, 2018; parameter –ax sr). SAMtools (Danecek et al., 2021) was used to extract mean

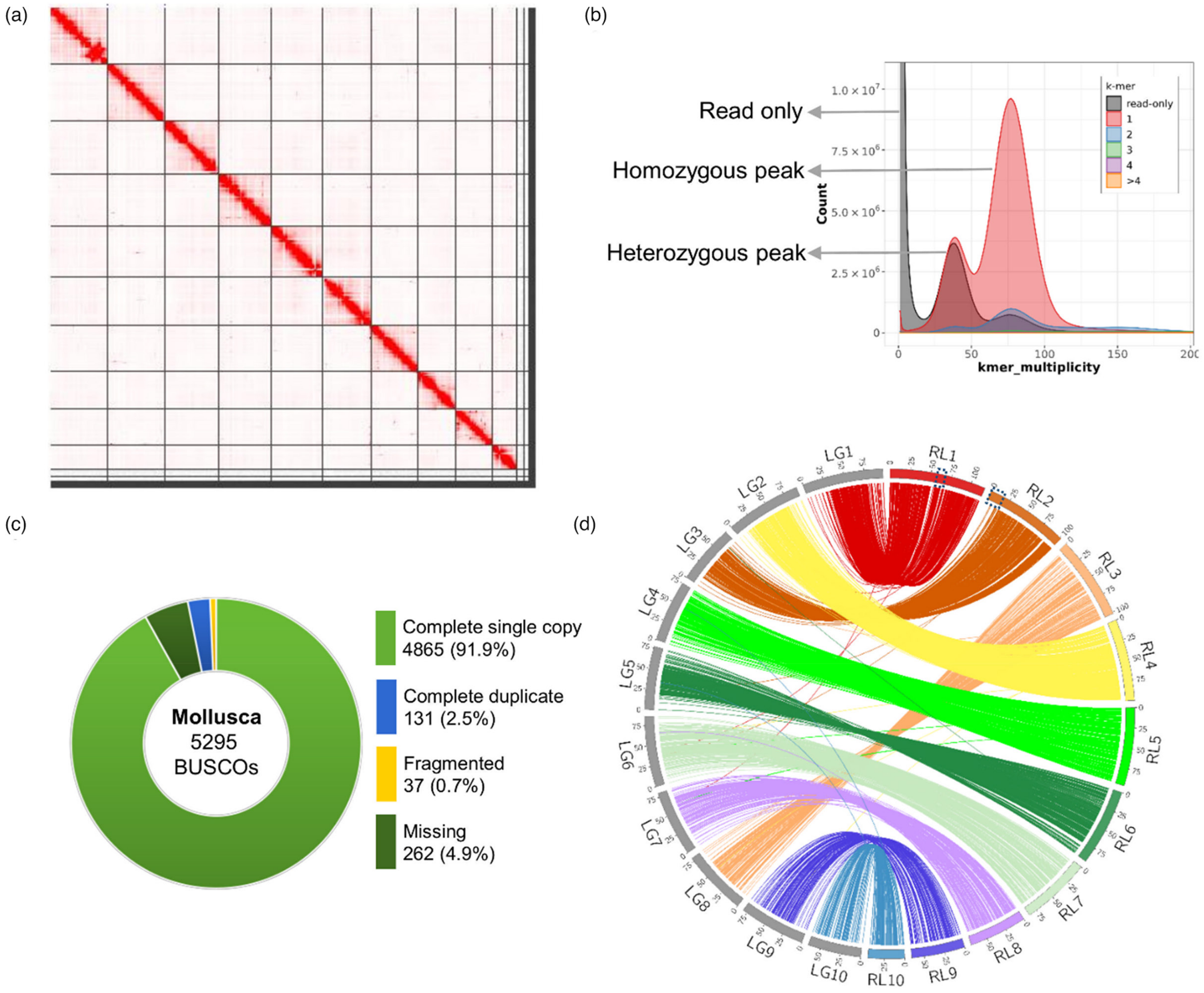


FIGURE 1 *OE_Roslin_V1* assembly quality evaluation. (a) Omni-C contact map highlighting the top 10 super-scaffolds generated by HiRise. The contact map was visualized using Juicebox (Durand, Robinson, et al., 2016). (b) Merqury k-mer copy number spectrum plot for the curated genome assembly. Nearly half of the single-copy k-mers (black region) were missing from the heterozygous peak, indicating efficient purging of haplotigs from the final assembly. K-mers missing from the assembly (black region in the homozygous peak) indicates bases present in the Illumina data missing from the assembly. (c) BUSCO scores for the final scaffolded *OE_Roslin_V1* assembly (mollusca_odb10 database). (d) Circos map highlighting the concordance between the 10 super-scaffolds (RL1 to RL10) and linkage groups (LG1 to LG10). Blue dotted squares within super-scaffolds 1 and 2 highlight the manual scaffolding performed on the basis of 3D contact information in the Omni-C data (Figure S1).

mapping depth values across the entire genome at 100 kb intervals. GC content across the genome was retrieved using BEDTools v2.29.2 (Quinlan & Hall, 2010) at an interval of 500 kb. The mean mapping depth and GC content data were plotted as a circos plot using the package Circlize 0.4.14 (Gu et al., 2014).

2.4 | Genome annotation

De novo repeat prediction was carried out using RepeatModeler v2.0.2 (Flynn et al., 2020). RepeatMasker v4.1.1 (Smit et al., 2015) was used for repeat masking with two databases: (i) RepBase-20,170,127 (Jurka et al., 2005) for Pacific oyster (set using parameters '-s *Crassostrea gigas* –e ncbi') and (ii) the de novo repeat database generated by RepeatModeler. Gene model prediction was carried out on the repeat masked assembly using Funannotate v1.8.7 (Palmer, 2017) after using the Funannotate clean module. Following this, the RNA-seq reads were aligned to the genome using minimap2 v2.21-r1071 (Li, 2018). Protein sequences for *C. gigas* and *C. virginica* from the UniProt database were aligned using Diamond v2.0.9 (Buchfink et al., 2021) and the resultant BAM files utilized for gene model prediction. PASA v2.4.1 (Haas et al., 2003) was then used to predict an initial set of high-quality gene models, which were used to train and run Augustus v3.3.32 (Stanke et al., 2006), SNAP (Korf, 2004) and GlimmerHMM v3.0.4 (Majoros et al., 2004). 40,283 high-quality gene models were automatically extracted from the ab-initio predictions before passing all the data to EVidenceModeler v1.1.1 (Haas et al., 2008) for a final round of gene model prediction. Gene models <50 aa in length (*n* = 2), spanning gaps (*n* = 2) and transposable elements (*n* = 5330) were filtered by Funannotate before the retained gene models underwent UTR prediction using PASA. Functional annotation was performed using the annotate step within Funannotate. Interproscan (Jones et al., 2014) was used to annotate predicted gene products against the following databases: Pfam (El-Gebali et al., 2019), Panther (Mi et al., 2021), PRINTS (Attwood et al., 2012), Superfamily (Pandurangan et al., 2019), Tigrfam (Haft et al., 2013), PrositeProfiles (Sigrist et al., 2013) and Gene Ontology (GO; The Gene Ontology Consortium, 2019). eggNOG-mapper v2.1.2 (Huerta-Cepas et al., 2017) was used to add functional annotation using the fast orthology assignment algorithm. BEDTools v2.29.2 (Quinlan & Hall, 2010) was used to extract data on genic content, gene density, classified repeats across unclassified repeats across the entire genome at a regular interval of 500 kb, all these data were incorporated into a circos plot using the package Circlize 0.4.14 (Gu et al., 2014).

2.5 | Additional validation of manually incorporated scaffolds

As mentioned above, two scaffolds were manually incorporated into the HiRise assembly (also see Results). To confirm the validity of these scaffolds beyond the quality assessments described above, we confirmed the genes present in these regions were: (i) of oyster origin and (ii) showed bioactivity comparable to other regions along the same chromosomes. Firstly, we retrieved the coding sequence of all genes predicted within the manually incorporated and remaining regions of super-scaffolds 1 and 2, which were subjected to BLASTn (Altschul et al., 1997) searches against the Pacific oyster genome (NCBI accession: GCA_902806645.1) and an independent Flat oyster genome (Boutet et al., 2022). The BLASTn cut-off was <1e-20 with the remaining parameters default. Secondly, RNA-Seq data from the heart, striated muscle and gonad were mapped to the genome assembly using STAR (Dobin et al., 2013) with default parameters. Mean RNA-Seq mapping depth for all gene models along super-scaffolds 1 and 2 was retrieved using SAMtools. Graphs comparing statistics between the manually incorporated and remaining regions of super-scaffolds 1 and 2 were generated using ggplot2 (Wickham, 2016).

2.6 | Linkage map construction

Four oyster full-sibling families (*n* = 171 individuals representing 8 parents and 163 F1 offspring) were used to build a novel linkage map for *O. edulis*. The families were produced in the Porscave hatchery (Lampaul-Plouarzel, Brittany, France). DNA was extracted from the parents and the offspring using a standard phenol-chloroform-isoamyl alcohol (PCI; 25:24:1, v/v) protocol. After two washes with PCI, DNA was precipitated overnight with absolute ethanol at −20°C, centrifuged, washed with 70% ethanol, dried and suspended in PCR-grade water. All DNA samples were run in a 1% agarose 1X TBE gel and quantified using a Qubit fluorometer (Thermo Fisher Scientific) with a high-sensitivity dsDNA quantification kit (Invitrogen) according to the manufacturer's instructions. Double-digest RAD-seq (ddRADSeq) libraries were produced for every sample following Brelsford et al. (2016). Briefly, for each individual, 200 ng of genomic DNA was digested using four different enzyme combinations (KasI/AciI, KasI/HpyCH4IV, KasI/MspI and PstI/MseI; New England Biolabs). Barcoded adaptors were ligated to the digested DNA fragments and purified using Nucleo Mag NGS Clean-up and Size Select Kit (Macherel-Nagel). 8 μl of the purified template was used for enrichment and Illumina indexing by PCR using Q5 hot start DNA polymerase (New England Biolabs; PCR conditions: 98°C 30s, 15 cycles 98°C 10s, 60°C 20s, 72°C 30s). A final elongation was done by adding buffer, dNTPs and primers for 15 min at 72°C. PCR products were run in a 1% agarose 1X TBE gel, quantified using a Qubit fluorometer with a high-sensitivity dsDNA quantification kit (Invitrogen) and then pooled in equal proportions into two separate libraries. A 300–800 bp size selection was performed using a 1.5% agarose cassette in a Pippin Prep instrument (Sage Science). Each fraction was run through a DNA chip in a Bioanalyser (Agilent) to determine the mean fragment size. The libraries were pooled at equimolar concentration and sequenced on one lane of a NovaSeq 6000 by Novogene Ltd (USA).

Raw reads were cleaned and demultiplexed with Stacks v2.5.4 (Catchen et al., 2013; Rochette et al., 2019). To avoid reference bias in the quality assessment of the genome assembly, SNP discovery and genotyping were performed using a de novo approach. To identify optimal parameter settings, two Stacks parameters were evaluated: (M) the maximum number of nucleotide mismatches allowed between stacks (or putative alleles) and (m) the minimum number of identical reads used to form a stack. For a subset of 12 samples, values of M were varied from 2 to 9, while parameter m was fixed to either 3 or 5. The final optimal parameter settings (m = 3, M = 4) were chosen as the combination of values that resulted in the highest number of polymorphic loci shared across 80% of the individuals (r80 rule; Paris et al., 2017). Variants were called from the de novo assembled data if the locus was present in more than 80% of the individuals (−r 0.8), after removing sites with observed heterozygosity higher than 0.7 (-max_obs_het 0.7). Genotyping in Stacks resulted in a total of 28,447 assembled loci, with an average depth across polymorphic sites of 79× and 29× in the parental and offspring samples, respectively. The consensus sequences of the catalogued loci were exported and the first 150 bp mapped to *OE_Roslin_V1* using BWA v0.7.8 (Li & Durbin, 2009). Variants within ddRAD loci with a mapping quality (MAPQ) > 4 were retained for subsequent analysis. Among these loci, 98% (24,079 out of 24,522) were uniquely mapped to the *O. edulis* genome and had the same or fewer mismatches than the default value (MAPQ ≥ 25; Menzel et al., 2013).

Further quality control (QC) filters were applied to the genotype data in Plink v1.9 (Chang et al., 2015). Markers and individuals with excess missing data (>10%) were discarded. Principal component analysis revealed that seven individuals were separated from their family cluster (Figure S2). Upon closer inspection, their high levels of Mendelian errors (>100 errors) suggested they had been mislabelled and were therefore removed from the dataset. After QC-filtering, 15,373 SNPs genotyped across 8 parents and 163 offspring were available for the construction of a linkage map using Lep-Map2 and Lep-Map3 (Rastas, 2017; Rastas et al., 2016). Genotype data were converted to genotype likelihoods (posteriors) using the *linkage2post* script in Lep-Map2. Missing or erroneous parental genotypes were imputed using the *ParentCall2* module. SNP markers informative for both parents were assigned to linkage groups (LGs) using the *SeperateChromosomes2* algorithm in Lep-Map3 with lodLimit = 11 and distortionLod = 1. Unassigned SNPs were added to the preliminary map using the *JoinSingles2All* module with lodLimit = 8, lodDifference = 2 and distortionLod = 1. The ordering of markers within LGs was conducted using the *OrderMarkers2* module after filtering markers based on segregation distortion (dataTolerance = 0.01). For each LG, the relative ordering of SNP markers was iterated ten times, and the configuration with the highest likelihood was selected to represent a sex-averaged map for *O. edulis*. One large gap (>10 cm) was identified and manually removed from the distal end of LG 10.

2.7 | Synteny and gene family expansion analyses

Gene level synteny was compared between *OE_Roslin_V1* and genome assemblies for a range of bivalve species using an orthogroup-based approach. A list of putative one-to-one orthologues between *O. edulis* and assemblies for *C. gigas* (NCBI accession: GCF_902806645.1; Peñaloza et al., 2021), *C. virginica* (GCF_002022765.2) and *P. maximus* (GCF_902652985.1; Kenny et al., 2020) were generated using Orthofinder v.2.3.11 (Emms & Kelly, 2019). An independent *O. edulis* genome assembly generated by Boutet et al. (2022; NCBI bioproject: PRJNA772088) was also included. The genomic coordinates of each gene in the one-to-one orthologue list for any two species under comparison were extracted and circos plots generated using the package Circlize 0.4.14 (Gu et al., 2014).

We inferred gene family expansions in *O. edulis* building on a published strategy (Regan et al., 2021). The start-point was all predicted proteins from the genome assemblies of 16 bivalve species, inclusive of protein sequences predicted in the *OE_Roslin_V1* genome (Table S1). The longest isoforms for each protein were retained using AGAT v0.4.4 (Dainat, 2020). These sequences were used to generate orthogroups in Orthofinder v.2.3.11 (Emms & Kelly, 2019). FastTree (Price et al., 2010) was used to infer gene trees per orthogroup, which were compared against the rooted species tree by Orthofinder to infer duplications/losses using a duplication-loss-coalescent model (Emms & Kelly, 2019). Kinfin v1.0 (Laetsch & Blaxter, 2017a) was used to identify orthogroups that showed evidence for gene expansion in *O. edulis* compared with other bivalves (Regan et al., 2021). Orthogroups showing evidence for gene expansions in *O. edulis* were first filtered for a fold change value >2.5 compared with the mean for all other bivalves. Fold change is defined as the number of genes per orthogroup for *O. edulis* divided by the mean number of genes per orthogroup across all other bivalve species. Orthogroups meeting this filter, but with <8/16 species (inclusive of *O. edulis*) represented in the tree, were further removed unless both *C. gigas* and *C. virginica* were present in the tree. Gene expansions in the remaining trees were classified as follows: (i) orthogroups showing >3-fold mean expansion in gene copy number in all Ostreidae species (*O. edulis*, *C. gigas* and *C. virginica*) vs. other bivalves (i.e. potential ancestral Ostreidae expansion), plus a further >3-fold mean expansion in gene copy number comparing *O. edulis* to the mean for *C. gigas* and *C. virginica* (i.e. additional lineage-specific expansion in *Ostrea*), (ii) orthogroups showing >3-fold mean expansion in gene copy number in all Ostreidae species, with no further expansion in gene copy number comparing *O. edulis* to the mean for *C. gigas* and *C. virginica* (i.e. inferred ancestral Ostreidae expansion only), (iii) orthogroups showing >3-fold mean expansion in gene copy number in *O. edulis* vs. other bivalves, with no evidence for expansion in the Ostreidae ancestor (i.e. inferred lineage-specific expansion in *Ostrea* postdivergence from *Crassostrea*), (iv) orthogroups showing >3-fold mean expansion in gene copy number in *O. edulis* compared with the mean for *C. gigas* and *C. virginica* but lacking genes for other bivalve species (i.e. inferred Ostreidae specific genes showing lineage-specific expansion in *Ostrea* post-divergence from *Crassostrea*), (v) orthogroups retaining genes for all three Ostreidae species but lacking any genes for other bivalve species (i.e. inferred Ostreidae specific genes that have not shown further expansion) and (vi) orthogroups showing >3-fold mean expansion in gene copy number in *O. edulis* compared with the mean for other non-Ostreidae bivalve species, absent in both *Crassostrea* species (inferred lineage-specific losses in *Crassostrea* but lineage-specific expansion in *Ostrea*).

Functional annotation of each orthogroup was performed by searching each protein against the eukaryotic SignalP database (Petersen et al., 2011), Gene Ontology database (GO; The Gene Ontology Consortium, 2019) and Pfam database (El-Gebali et al., 2019) using InterProScan v5.47–82.0 (Jones et al., 2014; the top GO/Pfam/InterProScan annotation per orthogroup was recorded) and feeding the results into KinFin (Laetsch & Blaxter, 2017a). Functional annotations were summarized based on their counts across all the expanded orthogroups. Protein sequence alignments from selected orthogroups were retrieved and maximum-likelihood phylogenetic trees were generated using IQTREE v1.6.8 (Nguyen et al., 2015) using the best fitting substitution model (Kalyaanamoorthy et al., 2017) and running the ultrafast bootstrapping (Minh et al., 2013) for 1000 iterations to generate branch support value. The trees were then visualized using iTOL online server (Letunic & Bork, 2021).

3 | RESULTS

3.1 | Contig assembly and quality evaluation

PromethION sequencing yielded 20,061,494 reads summing to 143.42 Gb of basecalled data with N50 length of 9297 bp (Figure S3) and mean length of 7149 bp, which was used for contig assembly. Assuming a haploid genome size of 1.14 Gb following past flow cytometry work involving n = 20 flat oysters sampled from Galicia in Spain (Rodríguez-Juíz et al., 1996), ~120x long-read sequencing depth was achieved, including 26× with reads >15 kb. Around 281 million Illumina short reads (~72× sequencing depth) were used for genome polishing. Around 57.6 million Illumina reads were generated by sequencing the Omni-C™ library, which were used for genome scaffolding. RNA-Seq generated ~50 million Illumina reads per tissue for genome annotation. K-mer-based estimation predicted the *O. edulis* genome to be 881 Mb, with a repeat content of 437 Mb (i.e. 49.8% of genome) and a heterozygosity rate of 1.02% (Figure S4).

The Flye assemblies *OE_F1*, *OE_F2* and *OE_F3* were 976.2, 1027.5 and 964.2 Mb, respectively. Purging for haplotigs resulted in the removal of 2–3% of data across each assembly (Table S2). The purged Flye assemblies had contig N50 values of 0.43, 0.39 and 0.34 Mb, respectively (Table S2). Thus, OE_F1, which used a minimum overlap of 10,000 bp to generate a contig, had the highest contiguity. The wtdbg2 contig assembly *OE-RB1* was 829.1 Mb after purging and had an N50 value of 0.67 Mb (Table S2). All four contig assemblies had a high BUSCO completeness score (~90% complete) compared with the mollusca_odb10 database (Table S2). The final merged and haplotig purged contig assembly *OE_contig_v1* was 934.9 Mb with a contig N50 of 2.38 Mb. Two rounds of genome polishing resulted in minor changes to contiguity but increased BUSCO completeness from 89% to 95.2% (Table S2), indicative of a strong positive effect on sequence accuracy.

3.2 | *O. edulis* chromosome-level genome assembly

Scaffolding using HiRise and Juicer led to assemblies of 935.08 and 936.34 Mb with N50 values of 94.05 and 82.94 Mb, respectively (Table S3). As the HiRise assembly was markedly more contiguous, it was taken forward as the basis for the final reference genome. Based on two lines of 3D contact evidence within the Omni-C data (see Methods), two large scaffolds in the HiRise assembly (scaffolds 11 and 12) were manually inserted into the super-scaffolds of the HiRise assembly. Specifically, scaffold 12 was inserted into super-scaffold 1 (at insertion point 65.4 Mb) and scaffold 11 was inserted at the start of super-scaffold 6 (Figure S1). As noted in the methods, at this stage, super-scaffold 6 was renamed super-scaffold 2 as a product of it becoming the second largest scaffold in the HiRise assembly, maintaining the convention of naming scaffolds according to size (Table S4).

The final assembly including the two manual corrections (*OE_Roslin_V1*) is 935.13 Mb with a scaffold-N50 of 95.56 Mb (Table 1), represented by 10 super-scaffolds comprising 93.65% (875.78 Mb) of the assembly, matching the haploid karyotype of *O. edulis* (i.e. 10 chromosomes; Horváth et al., 2013; Leitao et al., 2002; Thiriot-Quiévreux, 1984). The remaining 59.3 Mb of *OE_Roslin_V1* comprises 1353 unplaced scaffolds. The final assembly size matches closely to the k-mer-based genome-size estimate and is slightly larger than other genome assemblies within Ostreidae, which could be due to lineage-specific repeat expansion (see later section).

TABLE 1 Genome statistics for *O. edulis* (*OE_Roslin_V1* assembly).

Metric	Value
Assembly size (bp)	935,138,052
No. of contigs	2759
Contig N50 (Mb)	2.38
Longest contig (Mb)	16.06
No of scaffolds	1363
Length of top 10 scaffolds (bp)	875,789,595
Longest scaffold (bp)	117,440,623
Assembly N50 (bp)	95,564,955
Gaps (counts)	1534
N's count	153,250
GC content (%)	35.41
Contigs >500 bp	1363
Contigs >1000 bp	1294
Contigs >10,000 bp	846
Contigs >100,000 bp	103
Contigs >1 Mb	18

Detecting and correcting structural errors arising during genome assembly is critical in achieving a high-quality reference genome (Chen, Zhang, et al., 2021). Evaluation of the assembly for structural errors identified 1126 (663 expansions, 387 collapses, 76 inversions) putative structural errors when benchmarked against the raw nanopore reads, which were corrected. Assembly screening revealed little contamination from other taxa (Figure S5). We observed a 97.09% mapping rate of nanopore reads back to the assembly, further demonstrating the accuracy and completeness of the reference genome. A K-mer copy number histogram revealed that haplotig purging was very efficient (Figure 1b). We identified 4865 (91.9%) complete single-copy BUSCO genes and 131 (2.5%) complete duplicated BUSCO genes in the final assembly (Figure 1c).

3.3 | Linkage map and assembly validation

The de novo variant calling pipeline called 24,522 SNPs across the ddRAD-Seq dataset. After stringent filtering (see Methods), the finished genetic map contained 4016 SNPs anchored to the ten expected LGs (Figure S6). We observed an overall high collinearity between these LGs and the *OE_Roslin_V1* genome assembly pseudochromosomes (Figure 1d, Figure S7) confirming the accuracy of the scaffolding performed using the Omni-C data, including at the two manual joins we performed within the scaffold_1 and scaffold_2 of the *OE_Roslin_V1 assembly* (Figure 1d; Figure S7). We observed a potential inversion between LG1 and super-scaffold 1, which was unrelated to the manually scaffolded region (Figure S7). However, on closer inspection, the Hi-C data were ambiguous in this region (Figure 1a), with the opposite orientation of this region within the assembly being impossible to exclude, which would then match LG1.

3.4 | Genome annotation

57.3% (535.9 Mb) of the *OE_Roslin_V1* assembly was identified as repeats (Figure 2a), which falls in a similar range to recently published *C. gigas* genome assemblies (reported as 43% by Peñaloza et al. (2021) and 57.2% by Qi et al. (2021)). A large majority of repeats, comprising 37.65% of the genome, were annotated as unclassified (Figure 2a). A substantial proportion of the genome was annotated as LINE elements (5.98%), DNA transposons (4.37%) and rolling circles repeats (5.47%; Figure 2a). The accompanying sister article to this study provides a more detailed curation of repeat landscape in an independently generated French *O. edulis* genome assembly (Boutet et al., 2022). Note, that this work identified a very similar proportion of repeats (55.1%) using the same bioinformatic pipeline but not all could be confidently annotated.

Gene model prediction identified 35,699 coding genes in the masked genome (Table 2). Genic regions comprised 261.83 Mb (28.42%) of the genome size, with an average gene length of 7411 bp (Figure 2b) and an average coding sequence length of 1224 bp. Functional annotation of the predicted proteins resulted in the annotation of 23,109 gene models with EggNOG hits and provided 17,504 gene models with a GO annotation (Table 2). A range of anotate features are plotted along the genome in Figure 2c.

3.5 | Additional validation of manually incorporated scaffolds

To confirm the validity of the manually scaffolded regions in super-scaffolds 1 and 2, we sought to concretely demonstrate that they belonged to the flat oyster genome. We firstly performed BLASTn (Altschul et al., 1997) searches for all coding genes predicted in these regions against *C. gigas* (Peñaloza et al., 2021) and an independent *O. edulis* assembly (Boutet et al., 2022), and compared the results

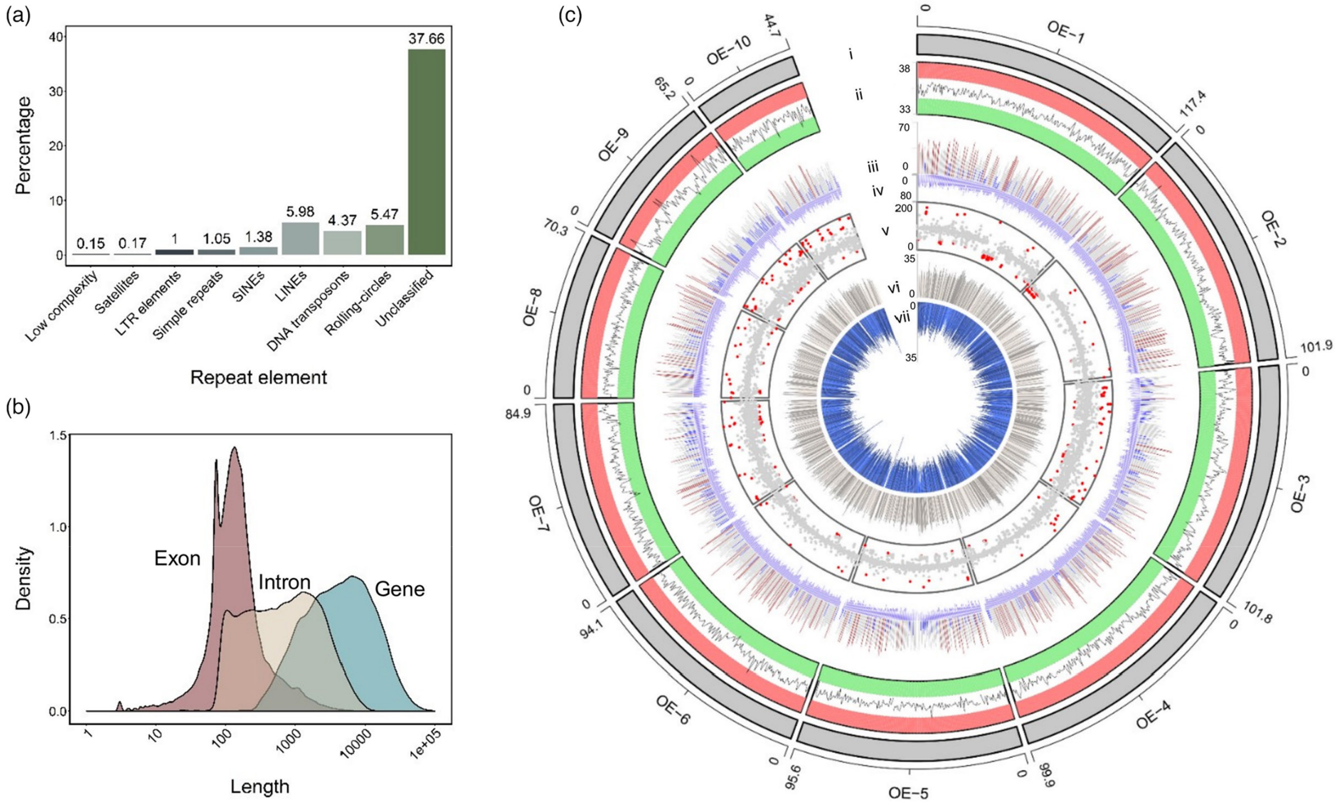


FIGURE 2 Annotation of the *O. edulis OE_Roslin_V1* assembly. (a) Summary of genome repeat classes. (b) Density plot showing gene, exon and intron lengths. (c) Circos plot highlighting annotated features across the ten super-scaffolds (window size 0.5 Mb except track-v, which is 0.1 Mb). Tracks as follows: i: 10 super-scaffolds OE-1 to OE-10; ii: GC percentage (33–38%), with red and green bars indicating GC >36.5% and <34.5%, respectively; iii: genic content (sum of annotated gene models) expressed as percentage of total window size, regions with <20% genic content are coloured blue, while 20 to 40% are coloured grey and >40% are coloured red; iv: gene density (0–80); v: mean Illumina sequencing depth, with values <45 and >150 shown as red points; vi: classified repeats expressed as percentage of total window size (0–35%); vii: novel unclassified repeat elements expressed as percentage of total window size (0–35%).

TABLE 2 Genome annotation statistics for *O. edulis* (*OE_Roslin_V1*).

Metric	Value
Protein-coding genes	35,699
Average gene length (bp)	7411
Average exon length (bp)	241
Single exon transcripts	1631
Multiple exon transcripts	34,068
Total gene length (bp)	265,862,173
Functional annotation (No of proteins)	
GO annotation	17,504
Interproscan hits	19,613
Eggnog hits	23,109
Pfam hits	16,966
Cazyme hits	537
Merops hits	921

with the remaining regions of super-scaffolds 1 and 2 (summarized in Table S5; raw data in Table S6). The proportion and percentage identity of BLAST hits to both oyster genomes were highly comparable for both regions along super-scaffolds 1 and 2. Secondly, RNA-Seq reads (pooled from heart, striated muscle and gonad) mapped with variable depth to approximately 40% of the predicted genes within the manually incorporated regions of super-scaffold 1 and 2 (Figure S8). The RNA-Seq mapping rate and depth were lower in the manually incorporated regions than in the remaining parts of super-scaffolds 1 and 2 (Figure S8).

3.6 | Synteny analysis with other bivalve genomes

Synteny plots of 1-to-1 orthologue gene locations revealed conserved chromosomal-level synteny between *OE_Roslin_V1* and three independently assembled bivalve genomes: *C. gigas* (Figure 3a), *C. virginica* (Figure 3b) and *P. maximus* (Figure 3c). We observed little evidence for major chromosomal rearrangements (i.e. involving megabases of a chromosome undergoing inversion or translocations) between the 10 chromosomes of *O. edulis* and *C. gigas* (Figure 3a), indicating that the ancestral ostreid karyotype has been maintained in both species. Comparison of *OE_Roslin_V1* with *C. virginica* (Figure 3b) provides evidence for possible chromosomal rearrangements in *C. virginica* after its split with *C. gigas*, assuming the chromosome-level synteny between *O. edulis* and *C. gigas* reflects the ancestral state. For instance, super-scaffold 8 in *OE_Roslin_V1*, which shares synteny across the length of *C. gigas* chromosome 4, shares synteny with two major blocks on *C. virginica* chromosomes 5 and 6 (Figure 3b). The synteny relationship between *OE_Roslin_V1* and the extensively rearranged *P. maximus* genome was consistent with that reported between *C. gigas* and *P. maximus* (Yang et al., 2021). We observed genome-wide synteny between *OE_Roslin_V1* and an independently generated assembly for *O. edulis* (Boutet et al., 2022), although there were a small number of chromosomal regions where synteny was broken (Figure 3d).

3.7 | Gene families expanded during Ostrea evolution

Gene duplication is associated with adaptation during evolution (Ohno, 1970), including in bivalves (Phuangphong et al., 2021; Regan et al., 2021). To gain insights into how gene duplication influenced *Ostrea* evolution, we identified gene family expansions in *OE_Roslin_V1* by comparison to 15 additional bivalve genomes. 712 gene families showed evidence of expansion (Table S7; see Methods), categorized into six groups in a phylogenetic framework (Figure 4a). The most common class of putative gene family expansion involved genes distributed among different bivalve families that underwent an expansion in Ostreidae (Figure 4b), with a subset showing evidence of further expansion in *O. edulis* compared with the two *Crassostrea* species (Figure 4c). Similarly, we observed many gene families distributed among several bivalve families, where expansion was specific to *Ostrea* (Figure 4d). We also identified gene families specific to all three Ostreidae members (i.e. absent in other bivalves), among which a large proportion did not show further expansion in *O. edulis* compared with *Crassostrea* (Figure 4e), with a smaller group expanded in *O. edulis* specifically (Figure 4f). Finally, we found a small number of gene families represented by different bivalve families that showed expansion in *O. edulis* but absence in *Crassostrea* species (Figure 4g).

Annotation of protein domains in the expanded gene families may offer clues into biological functions targeted during *Ostrea* evolution (Table S7; summarized in Figure 5a). Among 701 expanded gene families annotated with conserved domains by Interproscan (Jones et al., 2014), 229 were unique to 1 gene family, with the remaining domains present in 2–31 gene families. Thus, many domains were overrepresented among the expanded gene families (Figure 5a), including G protein-coupled receptor, rhodopsin-like (IPR000276; 31 gene families) and secretin-like (IPR000832; 9 gene families). Several domains associated with innate immune function were overrepresented, including C-type lectin (IPR001304; 20 gene families), complement C1q (IPR001073; 15 gene families) and Sushi/SCR/CCP (i.e. complement control protein domain; IPR000436; 9 gene families). There were many overrepresented domains containing zinc finger motifs (including IPR000315; 18 gene families, IPR013087; 9 gene families; and IPR001878; 5 gene families). The highly conserved homeobox domain was annotated in 6 gene families expanded in *O. edulis*. We provide two examples of expanded gene families in Figure 5b,c, both OGs were taken from gene families showing lineage-specific expansion in *Ostrea* after its divergence from Crassostrea.

We further used this dataset to identify extremely expanded gene families in the *O. edulis* genome. For instance, we observed two orthogroups showing the massive tandem expansion of genes encoding proteins with the uncharacterized EB domain

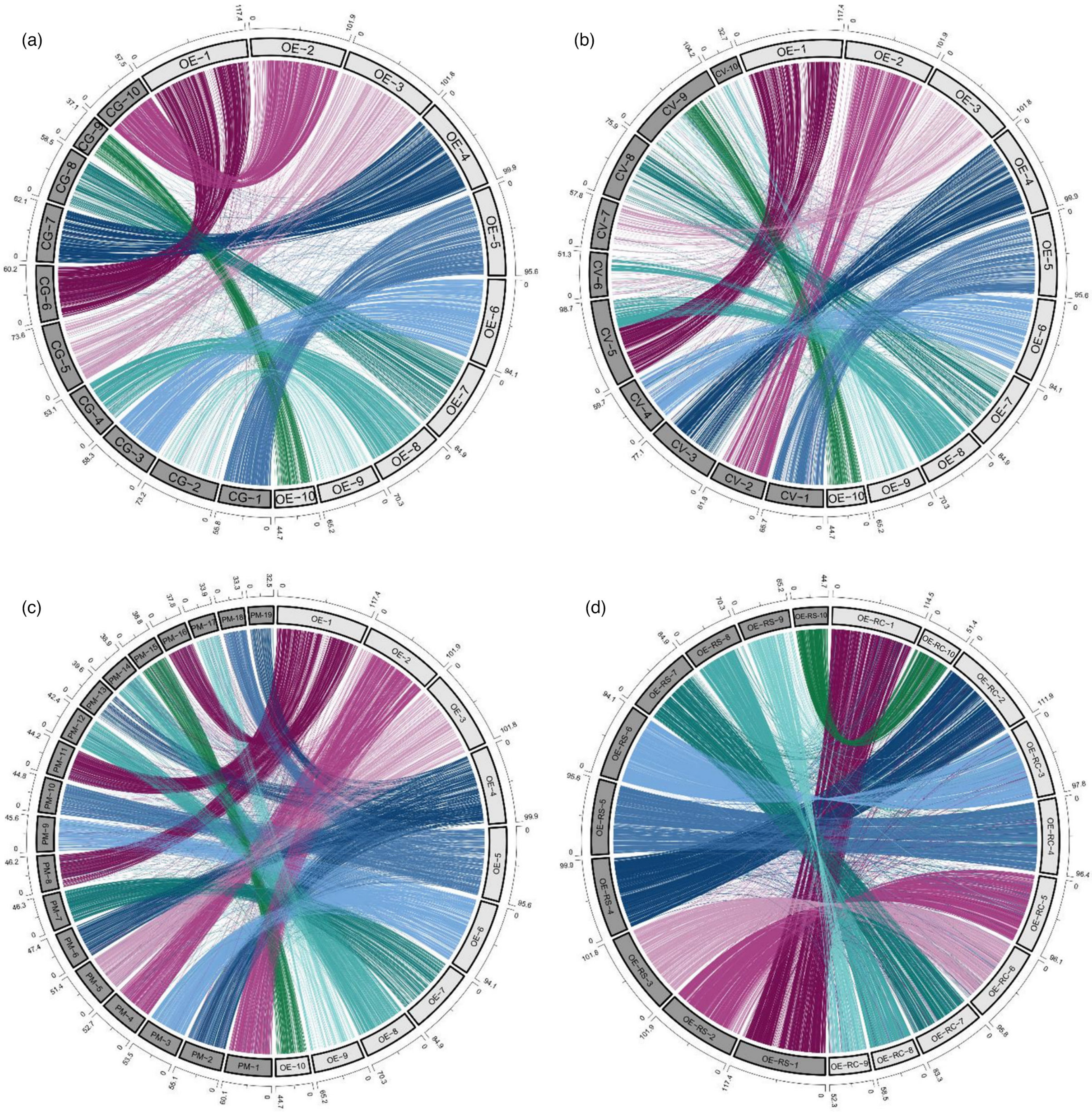


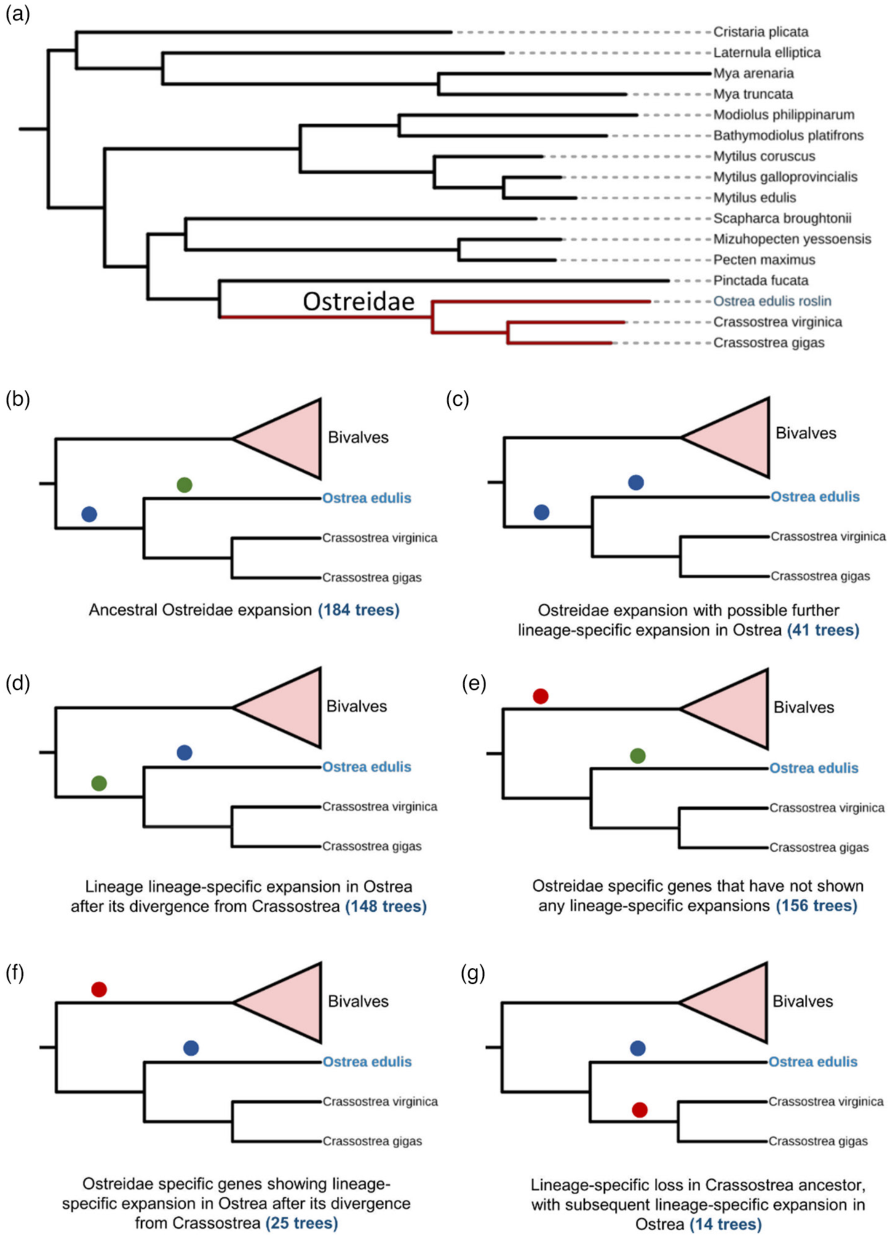
FIGURE 3 Chromosome-level synteny between the *OE_Roslin_V1 O. edulis* assembly and three independent bivalve assemblies. Circos plots are shown comparing the ten super-scaffolds (OE-1 to OE-10) with putative chromosomes of (a) *C. gigas*, (b) *C. virginica*, (c) *P. maximus* chromosomes and (d) an independent *O. edulis* assembly reported in Boutet et al. (2022) ('RC' denotes super-scaffolds from Boutet et al. (2022); 'RS' denotes super-scaffolds from *OE_Roslin_V1*).

(IPR006149). In both cases, these gene families were specific to Ostreidae and present as either 1 or 2 copies in *Crassostrea* species but 31 (orthogroup OG0002210) and 11 copies (orthogroup OG0013280) in *O. edulis* (Table S7). There were many other gene families specifically highly expanded in *O. edulis* (Table S7), including an Ostreidae-specific family (orthogroup OG0001484) encoding proteins containing an SAP domain (41 genes in *O. edulis*, vs. 2 genes each in both *Crassostrea* species), which has been proposed to be involved in the chromosomal organization (Aravind & Koonin, 2000).

4 | DISCUSSION

The high-quality, publicly available genome assembly we have generated and annotated for *O. edulis* serves as a novel reference for

FIGURE 4 Classification of gene family expansion during *O. edulis* evolution. (a) Species tree of bivalve genomes used in the analysis, (b–g) different categories of gene family expansion (classified as described in Methods). Branch annotations: blue circles indicate putative expansion; green circles indicates no expansion; red circle indicates an absence of species along that branch for the affected orthogroups. Full data are provided in Table S7.



(b) Ancestral Ostreidae expansion **(184 trees)**

(c) Ostreidae expansion with possible further lineage-specific expansion in Ostrea **(41 trees)**

(d) Lineage lineage-specific expansion in Ostrea after its divergence from Crassostrea **(148 trees)**

(e) Ostreidae specific genes that have not shown any lineage-specific expansions **(156 trees)**

(f) Ostreidae specific genes showing lineage-specific expansion in Ostrea after its divergence from Crassostrea **(25 trees)**

(g) Lineage-specific loss in Crassostrea ancestor, with subsequent lineage-specific expansion in Ostrea **(14 trees)**

genetics investigations of wild and farmed European flat oyster, in addition to comparative genomic investigations of molluscan taxa. Additional resources of value to the research community have been produced and made publicly available, including multi-organ RNA-Seq data, which we used to support gene model prediction and confirm genome assembly quality but in future can be used to explore patterns of tissue gene expression. In terms of assembly quality, the contig N50 we achieved is among the highest of all bivalve assemblies publicly available. This demonstrates the utility of our choice to merge different contig assemblies using Quickmerge (Chakraborty et al., 2016), which has been shown elsewhere to be effective for generating high-quality assemblies in molluscs (Sun et al., 2021) and other taxa (e.g. Chen, Nie, et al., 2021; Li et al., 2021; Mathers et al., 2021). Genome-wide sequence accuracy was further evidenced by the high mapping rate of nanopore reads back to the assembly, and the limited number of structural errors in the genome, which was lower than reported for the recent *C. gigas* reference genome (Peñaloza et al., 2021). BUSCO scores for our final *O. edulis* assembly are in the range of high-quality molluscan genome assemblies published to date (e.g. Sun et al., 2021), indicating an excellent level of gene representation.

Interestingly, our k-mer-based genome-size estimate (881 Mb), which matched closely with our final assembly length (876 Mb), was only ~77% of the 1.14 Gb genome size previously estimated by flow cytometry in a population of Spanish flat oysters (Rodríguez-Juíz et al., 1996). Similar observations have been made for other bivalve genomes, including *C. gigas* (e.g. Peñaloza et al., 2021). The discrepancy between this past flow cytometry assessment and our own sequencing-based estimates could be partly explained by population differences in genome size, considering the plasticity of genome content within bivalve species (Gerdol et al., 2020). However, this discrepancy cannot be easily explained by an under-representation of

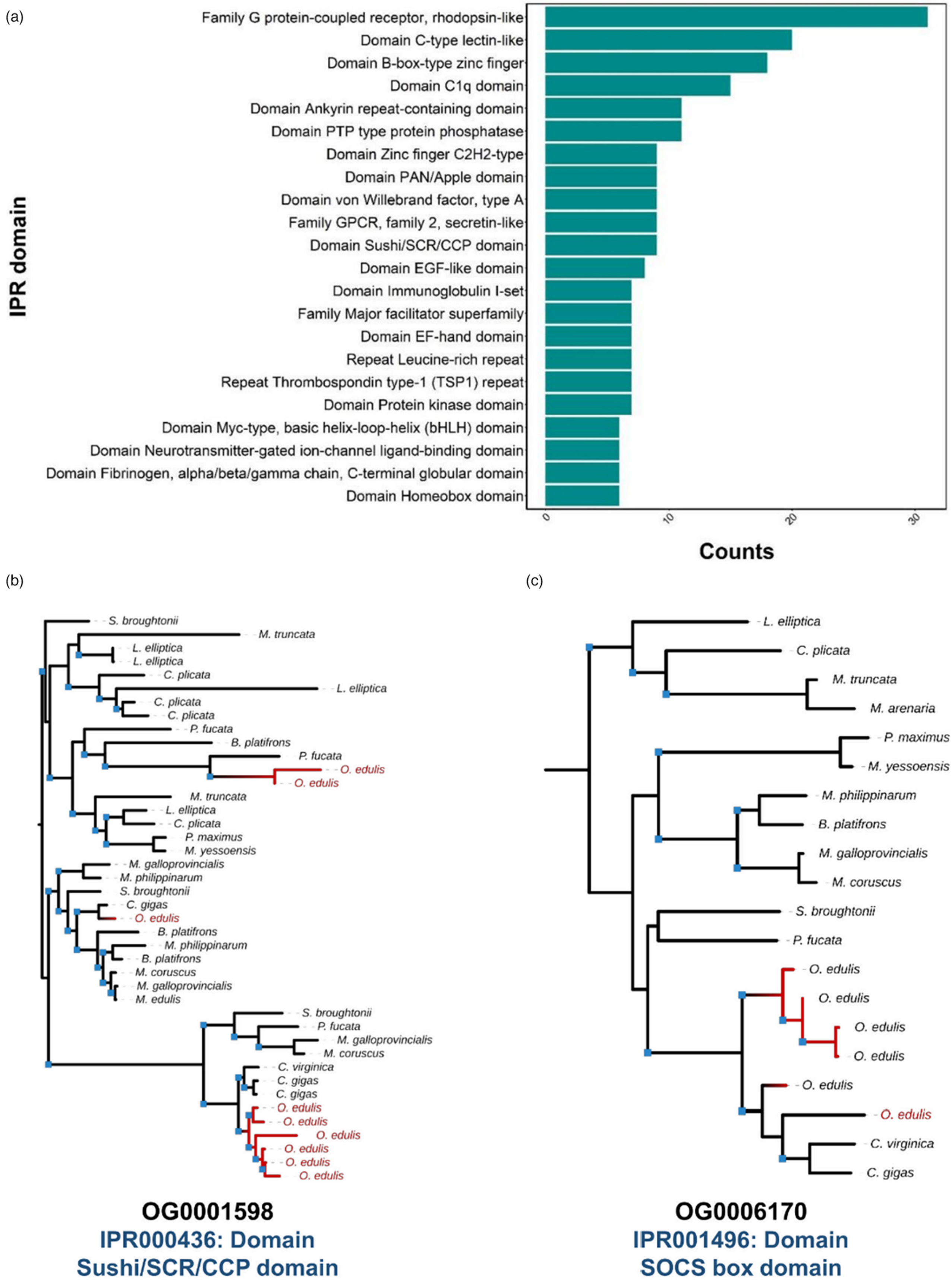


FIGURE 5 Most represented protein domains in expanded *O. edulis* gene families. (a) Top 20 represented IPR domains. (b and c) Example maximum-likelihood phylogenetic trees highlighting gene family expansions in *O. edulis*. Blue squares at nodes indicate bootstrap support value >50%.

repeats in our assembly, considering that >97% of the raw nanopore reads mapped back to the final assembly. Underestimation of genome size can also arise due to high heterozygosity (Liu et al., 2020). Our heterozygosity rate estimate of 1.02% for *O. edulis* was within the range reported for other bivalves, including 1.3% in *C. gigas* (Zhang et al., 2012) and 1.04% in scallop (*Patinopecten yessoensis*; Wang et al., 2017). This is interesting, as these previous estimates were made using individuals selected for reduced heterozygosity via inbreeding (Zhang et al., 2012) or by using a selfing family (Wang et al., 2017), implying a possible loss of genetic diversity in the *O. edulis* population we used for sequencing (e.g. a historic bottleneck). By contrast, an outbred *C. gigas* individual recently sequenced showed a much higher heterozygosity rate estimate of 3.2% (Peñaloza et al., 2021).

With regards to our in-house genome annotation, the average gene length obtained (7411 bp; Figure 2c) is lower than NCBI RefSeq annotations performed for bivalves to date. For example, the NCBI RefSeq annotations for *C. gigas* (PRJNA629593) and *C. virginica* (PRJNA376014) report average gene lengths of 10,990 and 10,828 bp, respectively. When this paper was at an advanced stage of revision, a NCBI RefSeq annotation was completed (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ostrea_edulis/100//), which predicted 38,526 gene models, including 28,480 protein-coding genes and 7875 noncoding genes, with an average gene length of 16,710 bp (average coding sequence length: 1866 bp). These greatly increased gene length statistics demonstrate that our in-house annotation strategy was inefficient in predicting gene models compared with the NCBI RefSeq method, leading to more fragmented and/or partially predicted gene models. However, the annotation we generated still has global utility, considering that we observe extensive 1-to-1 orthologue mapping compared with other genome assemblies (Figure 3), and were able to perform valid comparative genomic analyses both here (i.e. Figures 4 and 5) and in studies that have used our annotation to date (see later paragraph). However, the higher quality Refseq annotation for our assembly supersedes our in-house annotation and will further enhance future genetics and comparative genomic investigations exploiting our genome as a reference. In the longer term, we anticipate that bivalve genomes will benefit from greatly improved functional annotations that extend beyond gene model prediction, incorporating functional assays defined by the FAANG initiative to identify chromatin state modifications, regulatory elements, noncoding RNAs and isoform diversity (Clark et al., 2020).

Our cross-species synteny analysis revealed few major chromosomal reorganizations in the flat oyster genome, consistent with previous reports describing the nearly conserved karyotype across all oysters (Guo et al., 2018). Furthermore, conserved synteny and chromosomal architecture against an independently assembled flat oyster genome assembly (Boutet et al., 2022), coupled with the general high congruency of the assembled super-scaffolds with linkage groups, further confirmed the global quality of our assembly. Expansions to gene families involved in stress responses during bivalve evolution may reflect adaptation to a filter-feeding sessile lifestyle in a hostile environment (Guo et al., 2018; Hu et al., 2022; Regan et al., 2021). Past work has revealed expansions in gene families encoding heat shock proteins, and families involved in apoptosis inhibition and innate immunity, including C-type lectins and C1q complement domain-containing proteins. The gene family expansions reported here mirror these adaptation strategies, with enrichment in functional annotations for pathogen recognition and inflammatory response, e.g. C-type lectins, complement and immunoglobulin domains. The comparative genomic resources provided here can support future evolutionary analyses of gene families and should prove useful when interpreting the fine mapping of genetic variation around flat oyster genes, for instance, those identified in QTL regions.

Future applications of the *O. edulis* reference genome reported here, and for an independent genome assembly described for a French *O. edulis* individual in an accompanying article (Boutet et al., 2022) will address challenges relating to flat oyster conservation and sustainable aquaculture production (see commentary by Bean et al., 2022). These genomes provide researchers with new tools that empower genetic approaches addressing the ubiquitous threat posed by *Bonamia* via a range of technologies (Houston et al., 2020; Potts et al., 2021). In this regard, the genome reported here is proving useful already, with a recent study revealing that SNP markers previously associated with *Bonamia* resistance (Vera et al., 2019) are located in high linkage-disequilibrium across a large region of super-scaffold 8, which contains many candidate immune genes (Sambade et al., 2022). Another recent study has mapped variants genotyped with an existing medium density SNP array (Gutierrez et al., 2017) against our new *O. edulis* genome, identifying QTLs underpinning variation in growth traits on super-scaffold 4 (Peñaloza et al., 2022). Via its public release with all accompanying raw data, we anticipate a rapid uptake of our genome by the research community and envisage the next steps for the field to include broader surveys of genome-wide diversity covering a global representation of populations (Bean et al., 2022). This new phase of genome-enabled biology is like to uncover many secrets on the genetic and functional basis for adaptation and disease resilience in this iconic oyster species.

ACKNOWLEDGMENTS

This study was funded by the Biotechnology and Biological Sciences Research Council (BBSRC) under the AquaLeap consortium (grant code: BB/S004181/1) and received additional support from BBSRC Institute Strategic Programme grant BBS/E/D/10002070. We thank Edinburgh Genomics, especially Marian Thomson, for performing the PromethION sequencing and providing associated advice leading up to the work.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The genome assembly generated in this study along with all raw sequencing data used in assembly and annotation (Oxford Nanopore reads used for contig assembly, Illumina paired-end reads used for contig/scaffold polishing, Dovetail® Omni-C™ paired-end reads used for contig scaffolding, RNA-Seq paired-end reads from 8 tissues used for genome annotation) is available through NCBI under the Bioproject PRJNA772111. The genome annotation and large Supplementary Tables that are not available within the Supplementary Information are available through Figshare (https://doi.org/10.6084/m9.figshare.20050940).

ORCID

Manu Kumar Gundappa https://orcid.org/0000-0003-4328-2178
Arnaud Tanguy https://orcid.org/0000-0002-9670-2693
Tim P. Bean https://orcid.org/0000-0002-2544-9918

REFERENCES

Adema, C. M. (2021). Sticky problems: Extraction of nucleic acids from molluscs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1825), 20200162. https://doi.org/10.1098/rstb.2020.0162

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

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

Aravind, L., & Koonin, E. V. (2000). SAP – A putative DNA-binding motif involved in chromosomal organization. *Trends in Biochemical Sciences*, *25*(3), 112–114. https://doi.org/10.1016/S0968-0004(99)01537-6

Attwood, T. K., Coletta, A., Muirhead, G., Pavlopoulou, A., Philippou, P. B., Popov, I., Romá-Mateo, C., Theodosiou, A., & Mitchell, A. L. (2012). The PRINTS database: A fine-grained protein sequence annotation and analysis resource—Its status in 2012. *Database*, *2012*, bas019. https://doi.org/10.1093/database/bas019

Barton, M. D., & Barton, H. A. (2012). Scaffolder—Software for manual genome scaffolding. *Source Code for Biology and Medicine*, *7*(1), 4. https://doi.org/10.1186/1751-0473-7-4

Bayne, B. L. (2017). *Biology of oysters*. Academic Press.

Bean, T. P., Tanguy, A., Peñaloza, C., Gundappa, M. K., Houston, R. D., Macqueen, D. J., & Boudry, P. (2022). Enhancement of natural populations and revival of aquaculture in the European flat oyster led to two parallel chromosome-level assemblies. *Evolutionary Applications*. https://doi.org/10.1111/eva.13465

Boutet, I., Alves Monteiro, H. J., Baudry, L., Takeuchi, T., Bonnivard, E., Billoud, B., Farhat, S., Gonzales-Araya, R., Salaun, B., Andersen, A., Toullec, J.-Y., Lallier, F., Flot, J. F., Guiglielmoni, N., Guo, X., Allam, B., Pales-Espinoza, E., Hemmer-Hansen, J., Marbouty, M., Koszul, R., & Tanguy, A. (2022). Chromosomal assembly of the flat oyster (*Ostrea edulis* L.) genome as a new genetic resource for aquaculture. *Evolutionary Applications*. Online ahead of print. https://doi.org/10.1111/eva.13462

Brelsford, A., Dufresnes, C., & Perrin, N. (2016). High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, *116*(2), 177–181. https://doi.org/10.1038/hdy.2015.83

Bromley, C., McGonigle, C., Ashton, E. C., & Roberts, D. (2016). Bad moves: Pros and cons of moving oysters – A case study of global translocations of *Ostrea edulis* Linnaeus, 1758 (Mollusca: Bivalvia). *Ocean & Coastal Management*, *122*, 103–115. https://doi.org/10.1016/j.ocecoaman.2015.12.012

Bromley, C. A. (2015). *Science-based management strategies for the commercial and environmental sustainability of the European oyster, Ostrea edulis L.* (Ph.D., Queen's University Belfast). Queen's University Belfast. https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.695264

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354

Chakraborty, M., Baldwin-Brown, J. G., Long, A. D., & Emerson, J. J. (2016). Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Research*, *44*(19), e147. https://doi.org/10.1093/nar/gkw654

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

Chen, S., Wang, Y., Yu, L., Zheng, T., Wang, S., Yue, Z., Jiang, J., Kumari, S., Zheng, C., Tang, H., Li, J., Li, Y., Chen, J., Zhang, W., Kuang, H., Robertson, J. S., Zhao, P. X., Li, H., Shu, S., … Yang, C. (2021). Genome sequence and evolution of *Betula platyphylla*. *Horticulture Research*, *8*, 37. https://doi.org/10.1038/s41438-021-00481-7

Chen, Y., Nie, F., Xie, S.-Q., Zheng, Y.-F., Dai, Q., Bray, T., Wang, Y.-X., Xing, J.-F., Huang, Z.-J., Wang, D.-P., He, L.-J., Luo, F., Wang, J.-X., Liu, Y.-Z., & Xiao, C.-L. (2021). Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nature Communications*, *12*(1), 60. https://doi.org/10.1038/s41467-020-20236-7

Chen, Y., Zhang, Y., Wang, A. Y., Gao, M., & Chong, Z. (2021). Accurate long-read de novo assembly evaluation with inspector. *Genome Biology*, *22*(1), 312. https://doi.org/10.1186/s13059-021-02527-4

Clark, E. L., Archibald, A. L., Daetwyler, H. D., Groenen, M. A. M., Harrison, P. W., Houston, R. D., Kühn, C., Lien, S., Macqueen, D. J., Reecy, J. M., Robledo, D., Watson, M., Tuggle, C. K., & Giuffra, E. (2020). From FAANG to fork: Application of highly annotated genomes to improve farmed animal production. *Genome Biology*, *21*(1), 285. https://doi.org/10.1186/s13059-020-02197-8

Dainat, D. H. J. (2020). AGAT-v0.4.0 (version v0.4.0). https://doi.org/10.5281/zenodo.3877441

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Davison, A., & Neiman, M. (2021). Mobilizing molluscan models and genomes in biology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1825), 20200163. https://doi.org/10.1098/rstb.2020.0163

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. *Cell Systems*, *3*(1), 99–101. https://doi.org/10.1016/j.cels.2015.07.012

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. *Cell Systems*, *3*(1), 95–98. https://doi.org/10.1016/j.cels.2016.07.002

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn,

R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, *47*(D1), D427–D432. https://doi.org/10.1093/nar/gky995

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Engelsma, M., Kerkhoff, S., Roozenburg, I., Haenen, O., van Gool, A., Sistermans, W., Wijnhoven, S., & Hummel, H. (2010). Epidemiology of Bonamia ostreae infecting European flat oysters *Ostrea edulis* from Lake Grevelingen, The Netherlands. *Marine Ecology Progress Series*, *409*, 131–142. https://doi.org/10.3354/meps08594

FAO. (2020). *The state of world fisheries and aquaculture 2020. Sustainability in action*. FAO. https://doi.org/10.4060/ca9229en

Farhat, S., Bonnivard, E., Pales Espinosa, E., Tanguy, A., Boutet, I., Guiglielmoni, N., Flot, J.-F., & Allam, B. (2022). Comparative analysis of the *Mercenaria mercenaria* genome provides insights into the diversity of transposable elements and immune molecules in bivalve mollusks. *BMC Genomics*, *23*(1), 192. https://doi.org/10.1186/s12864-021-08262-1

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 of the United States of America*, *117*(17), 9451–9457. https://doi.org/10.1073/pnas.1921046117

Gerdol, M., Moreira, R., Cruz, F., Gómez-Garrido, J., Vlasova, A., Rosani, U., Venier, P., Naranjo-Ortiz, M. A., Murgarella, M., Greco, S., Balseiro, P., Corvelo, A., Frias, L., Gut, M., Gabaldón, T., Pallavicini, A., Canchaya, C., Novoa, B., Alioto, T. S., … Figueras, A. (2020). Massive gene presence-absence variation shapes an open pangenome in the Mediterranean mussel. *Genome Biology*, *21*(1), 275. https://doi.org10.1186/s13059-020-02180-3

Grizel, H., & Héral, M. (1991). Introduction into France of the Japanese oyster (*Crassostrea gigas*). *ICES Journal of Marine Science*, *47*(3), 399–403. https://doi.org/10.1093/icesjms/47.3.399

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*, *30*(19), 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Guo, X., Li, C., Wang, H., & Xu, Z. (2018). Diversity and evolution of living oysters. *Journal of Shellfish Research*, *37*(4), 755–771. https://doi.org/10.2983/035.037.0407

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gutierrez, A. P., Turner, F., Gharbi, K., Talbot, R., Lowe, N. R., Peñaloza, C., McCullough, M., Prodöhl, P. A., Bean, T. P., & Houston, R. D. (2017). Development of a medium density combined-species SNP Array for Pacific and European oysters (*Crassostrea gigas* and *Ostrea edulis*). *G3 Genes Genomes Genetics*, *7*(7), 2209–2218. https://doi.org/10.1534/g3.117.041780

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Jr., Hannick, L. I., Maiti, R., Ronning, C. M., Rusch, D. B., Town, C. D., Salzberg, S. L., & White, O. (2003). Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research*, *31*(19), 5654–5666. https://doi.org/10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., & Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Research*, *41*(D1), D387–D395. https://doi.org/10.1093/nar/gks1234

Harrang, E., Heurtebise, S., Faury, N., Robert, M., Arzul, I., & Lapègue, S. (2015). Can survival of European flat oysters following experimental infection with *Bonamia ostreae* be predicted using QTLs? *Aquaculture*, *448*, 521–530. https://doi.org/10.1016/j.aquaculture.2015.06.019

Horváth, Á., Kuzman, A., Bubalo, A., Bartulović, V., Várkonyi, E. P., Urbányi, B., & Glamuzina, B. (2013). Karyological study reveals a putatively distinctive population of the European flat oyster (*Ostrea edulis*) in Mali Ston Bay, Croatia. *Acta Adriatica*, *54*(1), 111–116.

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., Selly, S. L. C., Martin, S. A. M., Stevens, J. R., Santos, E. M., Davie, A., & Robledo, D. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, *21*(7), 389–409. https://doi.org/10.1038/s41576-020-0227-y

Hu, B., Tian, Y., Li, Q., & Liu, S. (2022). Genomic signatures of artificial selection in the Pacific oyster, *Crassostrea gigas*. *Evolutionary Applications*, *15*(4), 618–630. https://doi.org/10.1111/eva.13286

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, *110*(1–4), 462–467. https://doi.org/10.1159/000084979

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kenny, N. J., McCarthy, S. A., Dudchenko, O., James, K., Betteridge, E., Corton, C., Dolucan, J., Mead, D., Oliver, K., Omer, A. D., Pelan, S., Ryan, Y., Sims, Y., Skelton, J., Smith, M., Torrance, J., Weisz, D., Wipat, A., Aiden, E. L., … Williams, S. T. (2020). The gene-rich genome of the scallop Pecten maximus. *GigaScience*, *9*(5), giaa037. https://doi.org/10.1093/gigascience/giaa037

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, *5*(1), 59. https://doi.org/10.1186/1471-2105-5-59

Krueger, F. (2015). *Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files*. https://github.com/FelixKrueger/TrimGalore

Laetsch, D. R., & Blaxter, M. L. (2017a). KinFin: Software for taxon-aware analysis of clustered protein sequences. *G3: Genes - Genomes - Genetics*, *7*(10), 3349–3357. https://doi.org/10.1534/g3.117.300233

Laetsch, D. R., & Blaxter, M. L. (2017b). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*(1287), 1287. https://doi.org/10.12688/f1000research.12232.1

Lallias, D., Stockdale, R., Boudry, P., Beaumont, A. R., & Lapègue, S. (2009). Characterization of 27 microsatellite loci in the European flat oyster *Ostrea edulis*. *Molecular Ecology Resources*, *9*(3), 960–963. https://doi.org/10.1111/j.1755-0998.2009.02515.x

Lapègue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadieu, C., Gayral, P., Ballenghien, M., Genestout, L., Barbotte, L., Mahla, R., Haffray, P., & Klopp, C. (2014). Development of SNP-genotyping arrays in two shellfish species. *Molecular Ecology Resources*, *14*(4), 820–830. https://doi.org/10.1111/1755-0998.12230

Leitao, A., Chaves, R., Santos, S., Boudry, P., Guedes Pinto, H., & Thiriot Quievreux, C. (2002). Cytogenetic study of *Ostrea conchaphila* (Mollusca: Bivalvia) and comparative karyological analysis within Ostreinae. *Journal of Shellfish Research*, *21*(2), 685–690.

Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, J.-T., Wang, Q., Huang Yang, M.-D., Li, Q.-S., Cui, M.-S., Dong, Z.-J., Ballenghien, M., Genestout, L., Barbotte, L., Mahla, R., Haffray, P., & Wang, X.-Y. (2021). Parallel subgenome structure and divergent expression evolution of Allo-tetraploid common carp and goldfish. *Nature Genetics*, *53*(10), 1493–1503. https://doi.org/10.1038/s41588-021-00933-9

Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., & Fan, W. (2020). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *ArXiv*. 1308.2012 [q-Bio]. http://arxiv.org/abs/1308.2012

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics*, *20*(16), 2878–2879. https://doi.org/10.1093/bioinformatics/bth315

Mathers, T. C., Wouters, R. H. M., Mugford, S. T., Swarbreck, D., van Oosterhout, C., & Hogenhout, S. A. (2021). Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and Long-term conservation of the X chromosome. *Molecular Biology and Evolution*, *38*(3), 856–875. https://doi.org/10.1093/molbev/msaa246

Menzel, P., Frellsen, J., Plass, M., Rasmussen, S. H., & Krogh, A. (2013). On the accuracy of short read mapping. In N. Shomron (Ed.), *Deep sequencing data analysis* (pp. 39–59). Humana Press. https://doi.org/10.1007/978-1-62703-514-9_3

Merk, V., Colsoul, B., & Pogoda, B. (2020). Return of the native: Survival, growth and condition of European oysters reintroduced to German offshore waters. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *30*(11), 2180–2190. https://doi.org/10.1002/aqc.3426

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T., & Thomas, P. D. (2021). PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research*, *49*(D1), D394–D403. https://doi.org/10.1093/nar/gkaa1106

Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., & Gurevich, A. (2016). Icarus: Visualizer for de novo assembly evaluation. *Bioinformatics*, *32*(21), 3321–3323. https://doi.org/10.1093/bioinformatics/btw379

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Ohno, S. (1970). *Evolution by gene duplication*. Springer-Verlag.

Palmer, J. (2017). *Funannotate: Fungal genome annotation scripts*. https://github.com/nextgenusfs/funannotate

Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B., & Gough, J. (2019). The SUPERFAMILY 2.0 database: A significant proteome update and a new webserver. *Nucleic Acids Research*, *47*(D1), D490–D494. https://doi.org/10.1093/nar/gky1130

Pardo, B. G., Álvarez-Dios, J. A., Cao, A., Ramilo, A., Gómez-Tato, A., Planas, J. V., Villalba, A., & Martínez, P. (2016). Construction of an Ostrea edulis database from genomic and expressed sequence tags (ESTs) obtained from Bonamia ostreae infected haemocytes: Development of an immune-enriched oligo-microarray. *Fish & Shellfish Immunology*, *59*, 331–344. https://doi.org/10.1016/j.fsi.2016.10.047

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, *8*(10), 1360–1373. https://doi.org/10.1111/2041-210X.12775

Peñaloza, C., Barria, A., Papadopoulou, A., Hooper, C., Preston, J., Green, M., Helmer, L., Hammerson, J. K., Schulze, J. N., Minardi, D., Gundappa, M. K., Macqueen, D., Hamilton, J., Houston, R. D., & Bean, T. P. (2022). Genome-wide association and genomic prediction of growth traits in the European flat oyster (*Ostrea edulis*). *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.926638

Peñaloza, C., Gutierrez, A. P., Eöry, 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*(3), giab020. https://doi.org/10.1093/gigascience/giab020

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

Phuangphong, S., Tsunoda, J., Wada, H., & Morino, Y. (2021). Duplication of spiralian-specific TALE genes and evolution of the blastomere specification mechanism in the bivalve lineage. *EvoDevo*, *12*(1), 11. https://doi.org/10.1186/s13227-021-00181-2

Pogoda, B. (2019). Current status of European oyster decline and restoration in Germany. *Humanities*, *8*(1), 9. https://doi.org/10.3390/h8010009

Pogoda, B., Brown, J., Hancock, B., Preston, J., Pouvreau, S., Kamermans, P., Sanderson, W., & von Nordheim, H. (2019). The native oyster restoration Alliance (NORA) and the Berlin oyster recommendation: Bringing back a key ecosystem engineer by developing and supporting best practice in Europe. *Aquatic Living Resources*, *32*, 13. https://doi.org/10.1051/alr/2019012

Potts, R. W. A., Gutierrez, A. P., Peñaloza, C. S., Regan, T., Bean, T. P., & Houston, R. D. (2021). Potential of genomic technologies to improve disease resistance in molluscan aquaculture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1825), 20200168. https://doi.org/10.1098/rstb.2020.0168

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLoS One*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., Troll, C. J., Fields, A., Hartley, P. D., Sugnet, C. W., Haussler, D., Rokhsar, D. S., & Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*, *26*(3), 342–350. https://doi.org/10.1101/gr.193474.115

Qi, H., Li, L., & Zhang, G. (2021). Construction of a chromosome-level genome and variation map for the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Resources*, *21*(5), 1670–1685. https://doi.org/10.1111/1755-0998.13368

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, *11*(1), 1432. https://doi.org/10.1038/s41467-020-14998-3

Rastas, P. (2017). Lep-MAP3: Robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics*, *33*(23), 3726–3732. https://doi.org/10.1093/bioinformatics/btx494

Rastas, P., Calboli, F. C. F., Guo, B., Shikano, T., & Merilä, J. (2016). Construction of Ultradense linkage maps with Lep-MAP2: Stickleback F 2 recombinant crosses as an example. *Genome Biology and Evolution*, *8*(1), 78–93. https://doi.org/10.1093/gbe/evv250

Regan, T., Stevens, L., Peñaloza, C., Houston, R. D., Robledo, D., & Bean, T. P. (2021). Ancestral physical stress and later immune gene family expansions shaped bivalve Mollusc evolution. *Genome Biology and Evolution*, *13*(8), evab177. https://doi.org/10.1093/gbe/evab177

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*(1), 245. https://doi.org/10.1186/s13059-020-02134-9

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Rodríguez-Juíz, A. M., Torrado, M., & Méndez, J. (1996). Genome-size variation in bivalve molluscs determined by flow cytometry. *Marine Biology*, *126*(3), 489–497. https://doi.org/10.1007/BF00354631

Ronza, P., Cao, A., Robledo, D., Gómez-Tato, A., Álvarez-Dios, J. A., Hasanuzzaman, A. F. M., Quiroga, M. I., Villalba, A., Pardo, B. G., & Martínez, P. (2018). Long-term affected flat oyster (*Ostrea edulis*) haemocytes show differential gene expression profiles from naïve oysters in response to *Bonamia ostreae*. *Genomics*, *110*(6), 390–398. https://doi.org/10.1016/j.ygeno.2018.04.002

Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature Methods*, *17*(2), 155–158. https://doi.org/10.1038/s41592-019-0669-3

Sambade, I. M., Chiclana, A., Blanco, A., Gundappa, M. K., Bean, T. P., Macqueen, D. J., Houston, R. D., Villalba, A., Vera, M., Kamermans, P., & Martínez, P. (2022). A single genomic region involving a putative chromosome rearrangement in flat oyster (*Ostrea edulis*) is associated with divergent selection to the parasite *Bonamia ostreae*. *Evolutionary Applications*. https://doi.org/10.1111/eva.13446

Sas, H., Deden, B., Kamermans, P., Zu Ermgassen, P. S. E., Pogoda, B., Preston, J., Helmer, L., Holbrook, Z., Arzul, I., van der Have, T., Villalba, A., Colsoul, B., Lown, A., Merk, V., Zwerschke, N., & Reuchlin, E. (2020). Bonamia infection in native oysters (*Ostrea edulis*) in relation to European restoration projects. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *30*(11), 2150–2162. https://doi.org/10.1002/aqc.3430

Sigrist, C. J. A., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., & Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Research*, *41*(D1), D344–D347. https://doi.org/10.1093/nar/gks1067

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit, A. F. A., Hubley, R., & Green, P. (2015). *RepeatMasker Open-4.0. 2013–2015*. http://www.repeatmasker.org

Song, H., Guo, X., Sun, L., Wang, Q., Han, F., Wang, H., Wray, G. A., Davidson, P., Wang, Q., Hu, Z., Zhou, C., Yu, Z., Yang, M., Feng, J., Shi, P., Zhou, Y., Zhang, L., & Zhang, T. (2021). The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in Bivalvia. *BMC Biology*, *19*(1), 15. https://doi.org/10.1186/s12915-020-00943-9

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006). AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, *34*(suppl_2), W435–W439. https://doi.org/10.1093/nar/gkl200

Sun, J., Li, R., Chen, C., Sigwart, J. D., & Kocot, K. M. (2021). Benchmarking Oxford Nanopore read assemblers for high-quality molluscan genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1825), 20200160. https://doi.org/10.1098/rstb.2020.0160

Suquet, M., Pouvreau, S., Queau, I., Boulais, M., Le Grand, J., Ratiskol, D., & Cosson, J. (2018). Biological characteristics of sperm in European flat oyster (*Ostrea edulis*). *Aquatic Living Resources*, *31*, 20. https://doi.org/10.1051/alr/2018008

The Gene Ontology Consortium. (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338. https://doi.org/10.1093/nar/gky1055

Thiriot-Quiévreux, C. (1984). Analyse comparée des caryotypes d'Ostreidae (Bivalvia). *Cahiers de Biologie Marine*, *25*, 407–418.

Thorngren, L., Bergström, P., Dunér Holthuis, T., & Lindegarth, M. (2019). Assessment of the population of *Ostrea edulis* in Sweden: A marginal population of significance? *Ecology and Evolution*, *9*(24), 13877–13888. https://doi.org/10.1002/ece3.5824

Vera, M., Pardo, B. G., Cao, A., Vilas, R., Fernández, C., Blanco, A., Gutierrez, A. P., Bean, T. P., Houston, R. D., Villalba, A., & Martínez, P. (2019). Signatures of selection for bonamiosis resistance in European flat oyster (*Ostrea edulis*): New genomic tools for breeding programs and management of natural resources. *Evolutionary Applications*, *12*(9), 1781–1796. https://doi.org/10.1111/eva.12832

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*(11), e112963. https://doi.org/10.1371/journal.pone.0112963

Wang, S., Zhang, J., Jiao, W., Li, J., Xun, X., Sun, Y., Guo, X., Huan, P., Dong, B., Zhang, L., Hu, X., Sun, X., Wang, J., Zhao, C., Wang, Y., Wang, D., Huang, X., Wang, R., Lv, J., … Bao, Z. (2017). Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution*, *1*(5), 1–12. https://doi.org/10.1038/s41559-017-0120

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.

Yang, J.-L., Feng, D.-D., Liu, J., Xu, J.-K., Chen, K., Li, Y.-F., 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*(4), giab024. https://doi.org/10.1093/gigascience/giab024

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P. W. H., Paps, J., Zhu, Y., Wu, F., Chen, Y., Wang, J., … Wang, J. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, *490*(7418), 49–54. https://doi.org/10.1038/nature11413

SUPPORTING INFORMATION

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

How to cite this article: Gundappa, M. K., Peñaloza, C., Regan, T., Boutet, I., Tanguy, A., Houston, R. D., Bean, T. P., & Macqueen, D. J. (2022). Chromosome-level reference genome for European flat oyster (*Ostrea edulis* L.). *Evolutionary Applications*, *00*, 1–17. https://doi.org/10.1111/eva.13460