**Title**

Molluscan genomics: implications for biology and aquaculture

**Author**

Takeshi Takeuchi*

Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan

*Correspondence: t.takeuchi@oist.jp

**Keywords**

molluscan genome; genotyping; aquaculture

## Abstract

## Purpose of review

As a result of advances in DNA sequencing technology, molluscan genome research, which initially lagged behind that of many other animal groups, has recently seen a rapid succession of decoded genomes. Since molluscs are highly divergent, the subjects of genome projects have been highly variable, including evolution, neuroscience, and ecology. In this review, recent findings of molluscan genome projects are summarized, and their applications to aquaculture are discussed.

## Recent findings

Recently 14 molluscan genomes have been published. All bivalve genomes show high heterozygosity rates, making genome assembly difficult. Unique gene expansions were evident in each species, corresponding to their specialized features, including shell formation, adaptation to the environment, and complex neural systems. To construct genetic maps and to explore quantitative trait loci (QTL) and genes of economic importance, genome-wide genotyping using massively parallel, targeted sequencing of cultured molluscs was employed.

## Summary

Molluscan genomics provides information fundamental to both biology and industry. Modern genomic studies facilitate molluscan biology, genetics, and aquaculture.

## Introduction

The Mollusca is one of the most speciose animal phyla, including at least 70,000 described species [1]. They account for about one-quarter of all marine animal species, and their habitats include brackish water, freshwater, and land, as well as extreme environments such as deep-sea hydrothermal vents. Their abilities to adapt to various environments are of great interest in ecology and evolution.

In the realm of aquaculture, molluscs are the second-largest resource after finfish, constituting 22% of total global aquaculture production [2]. The production volume of molluscs reached 16.1 million metric tons ($19 billion US) in 2014, roughly a 20% increase from 2004 [3]. Despite their immense diversity in nature, aquaculture development focuses on a limited number of species. According to FAO data, 104 molluscan species or species groups have been farmed, but 5 bivalve species comprise about 40% of all molluscan aquaculture production [2, 4]. Since the majority of bivalves are filter-feeders, they can be cultured without feeding, so mollusc aquaculture is less costly and environmentally benign.

Mollusc aquaculture has a long history. For example, in his book, "*Naturalis Historia*," Pliny the Elder recorded that the ancient Roman merchant, Caius Sergius Orata, established artificial oyster beds in Lucrine Lake in 95 B.C. Scientific bivalve aquaculture has been investigated since the 1960's and breeding programs have been conducted with the aim of genetically improving the strains (e.g. literature cited by [5]). However, most cultured molluscs still remain in a wild state, and they are not genetically improved, compared to domesticated vertebrates and plants. In other words, productivity and quality of molluscan aquaculture products could be considerably improved by selective breeding. In traditional breeding programs, prospective broodstocks are chosen based on their phenotypes and pedigrees, while recent breeding strategies in livestock production are transitioning to genomic selection, which uses genome-wide genetic markers to estimate breeding value [6, 7]. To this end, whole genome information is desired for mollusc species.

Since the mid-2000s, revolutionary advances in DNA sequencing technology have decreased the cost and time required for whole genome sequencing. For example, massive parallel platforms produce 10 to 900 Giga bases (Gb) of data per run (single flow cell), costing tens of US$ per Gb [8]. This provides researchers with an unprecedented opportunity to decode

mollusc genomes, which have fallen far behind those of model organisms, livestock, and crop species. In 2012, draft genomes of the pearl oyster, *Pinctada fucata*, and the Pacific oyster, *Crassostrea gigas*, were the first molluscan genomes published [9, 10]. Since then, genomes of 13 mollusc species in 3 classes (Bivalvia, Gastropoda, and Cephalopoda) have been published [11-22] (Figure 1). In addition, some molluscan genome assemblies, such as that of *Aplysia californica*, are publicly available, although I will not discuss them since the research results are not yet published. Molluscan genome research tends to focus on basic biology including animal evolution, environmental adaptation, neuroscience, and biomineralization. On the other hand, it is clear that genome information could contribute to development of effective breeding and sustainable mollusc aquaculture.

In this review, I first discuss general aspects of molluscan genomes demonstrated by various sequencing projects. In particular, the issue of heterozygosity in bivalve genome assembly is addressed. Next, two bivalve genome projects, the pearl oyster, *P. fucata*, and the Pacific oyster, *C. gigas*, are discussed, having received much attention from the aquaculture industry. Other molluscan genome projects, including two major phyla, the Gastropoda and Cephalopoda, are also summarized, examining various aspects of molluscan biology. Finally, potential contributions of genome data to the aquaculture industry are discussed.

## Heterozygosity in bivalve genomes

Although sequencing technology has drastically improved, constructing a high-quality *de novo* genome assembly is a major challenge for bivalves because the bivalve genome is very heterozygotic (i.e. there are many loci at which individuals have more than one allele). To date nine bivalve nuclear genome assemblies have been published (Figure 1), and all of them display high heterozygosity rates [9, 10, 13, 14, 17-21]. For instance, polymorphism percentages, including single-nucleotide polymorphisms (SNP) and short insertions/deletions (indels), in *Patinopecten yessoensis* and *Crassostrea gigas* genomes are 1.04% and 1.30% per individual, respectively. These rates are 7- to 9-fold higher than in humans (0.14%) [10, 23, 24]. On the other hand, the *Octopus* (cephalopod) genome has a much lower rate (0.08%) [12]. The high heterozygosity rate in bivalves may reflect their large population sizes and their expansive habitats in the open sea, or their enormous fecundity [25], which requires high rates of germline mitosis, causing high mutation rates [26]. In the case of cultured species, artificial admixtures between populations, with expected heterosis or hybrid vigor, may contribute to their high

heterozygosity rates.

High heterozygosity is an obstacle to generating continuous genome assemblies. In contrast to conventional Sanger sequencing, recent high-throughput sequencers generate huge numbers of short-read sequences, typically ranging from 50 to 300 bases. In order to re-construct the original genomic DNA sequence, a computational process or assembly based on the de Bruijn graph framework with a short substring (k-mer) is generally performed [27-29]. This strategy is suitable for dealing with massive numbers of short reads, and this reduces the calculation cost. In general, however, it is difficult to assemble highly heterozygotic genomes. When a heterozygotic diploid genome is sequenced, two unique k-mers are generated from a polymorphic locus. This results in contigs that bifurcate at the variant nucleotide. Consequently, the assembly becomes fragmented, resulting in a considerable number of redundant sequences and mis-assembled duplications [30, 31].

A fundamental solution is to generate an inbred line with reduced heterozygosity. For genome sequencing of *C. gigas*, four generations of full-sibling matings resulted in removal of about half the polymorphism [10]. In the scallop genome project, self-fertilizing progeny were generated from a single hermaphroditic parent, leading to a 50% reduction of polymorphism [18]. The inbreeding strategy reduces heterozygotic loci to some extent, although it seems unrealistic to establish a nearly homozygotic line, because of inbreeding depression [32].

The choice of sequencing and assembly strategy is critical to construct better assemblies. A fosmid-pooling strategy combined with whole-genome, shotgun sequencing was used for the Pacific oyster genome sequencing [10]. By this method, fosmid pools were sequenced separately and assembled, resulted in longer contigs and scaffolds, since each pool covers only 0.57% of the genome, thereby reducing the possibility of co-occurrence of heterozygotes and repetitive sequences in each pool. For the pearl oyster genome assembly, redundant contigs caused by heterozygosity were removed *in silico* [13]. When raw reads are mapped to the assembly, sequence coverage depth of contigs derived from heterozygotic regions is one-half of that of homozygotic regions. Thus, if two contigs show high sequence similarity and low coverage depth, they may be haplotype copies so that one of them can be discarded so as to develop a non-redundant, haploid assembly. This strategy dramatically improved the subsequent scaffolding and final assembly of the pearl oyster genome [13]. Incorporating long-read

sequences, such as those from PacBio or Nanopore may be a more effective strategy to overcome the obstacles of heterozygosity.

## Genome size and repetitive elements

Based on records deposited in the Animal Genome Size Database (http://www.genomesize.com), genome sizes of molluscs range from 290 Mb (Aplacophora, *Neomenia permagna* [33]) to 7.6 Gb (Gastropoda, *Diplommatina kiiensis kiiensis* [34]). Cephalopods have larger genomes (3.8 Gb on average) than those of bivalves (1.6 Gb) and gastropods (2.2 Gb). Since the number of chromosomes is significantly increased, whole genome duplication at the base of cephalopod lineage was inferred [35, 36]. However, this hypothesis was not supported by the whole genome survey of *Octopus bimaculoides* [12].

Varied genome sizes among molluscs reflect, in part, the number of repetitive sequences. In the *O. bimaculoides* genome, which is the largest molluscan genome decoded to date (2.68 Gb), repeat elements account for at least 45% of the genome [12]. SINE retrotransposons are one of the major components of repetitive elements (3.6%) in the octopus genome. Among bivalves, the proportion of repetitive elements varies from 62% in *Modiolus philippinarum* to 36% in *C. gigas* [10, 17]. A large proportion of the repetitive elements in molluscan genomes are dissimilar to those deposited in public databases such as Repbase [37]. For example, 27% of the repetitive elements in *M. philippinarum* were assigned as "unknown" [17]. This suggests that a considerable number of unidentified repetitive elements are present in mollusc genomes.

## The pearl oyster: a model for the study of biomineralization

The pearl oyster, *Pinctada fucata*, has been cultured in eastern and southeastern Asia since pearl farming was established there at the end of 19th century [38]. Molecular mechanisms of pearl formation are substantially the same as those of calcareous shell formation. Epithelial cells in mantle tissue secrete an organic matrix and the matrix regulates construction of microstructure and crystallization of the shell or pearl. Therefore, identification and functional analysis of components in the organic matrix is a topic of major research interest, with the aim of improving pearl quality using genetic information and molecular biology techniques. The draft genome of *P. fucata* was decoded in 2012 [9], followed by an improved version of the genome assembly (version 2.0) in 2016 [13], providing substantial information for identifying various

biological mechanisms, including those involved in development [39-42], physiology [43], reproduction [44], and biomineralization [45]. The genome assembly of another strain of *P. fucata martensii* was published in 2017 [20]. Genes responsible for pearl and shell formation were thoroughly investigated in *Pinctada* species by transcriptomic and genomic approaches [46, 45]. Proteins in the shell called shell matrix proteins are considered key factors of shell formation. Their localization in the shell means that they can interact directly with the crystal phase and can control shell formation. In order to identify shell matrix proteins, organic fractions extracted from shells are analyzed by mass spectrometry, and retrieved peptide sequences are searched against the transcriptome or genome sequence. This proteomic analysis can identify tens or hundreds of shell matrix proteins [47, 48]. It should be emphasized that functional analysis with gene knockdown by RNA interference (RNAi) is applicable for *P. fucata* [20, 49, 50]. Genome-wide surveys of shell-forming genes combined with gene knockdown experiments will eventually reveal the entire shell or pearl formation process at the molecular level. In addition, comparative genomics and proteomics may reveal the evolutionary course of mollusc shell formation. *Pinctada*, *Crassostrea*, and *Lottia*, from which both the genome and shell proteome have been analyzed, have different gene repertoires of shell matrix proteins, while some conserved functional domains such as chitin-binding, VWA, and EGF domains are commonly utilized for mollusc shell formation [10, 48, 51-54]. The *P. fucata* genome revealed tandem duplications and rapid molecular evolution of shell-forming genes [13, 45, 55]. These findings about the molecular basis of shell and pearl formation will be useful for selective breeding for high-quality pearl farming.

## The Pacific oyster: a cosmopolitan bivalve with remarkable adaptability

The Pacific oyster, *Crassostrea gigas*, occurs naturally in the Northwest Pacific, and has become even more widespread after being introduced in many countries for commercial production [56-60]. It is now the second most widely produced mollusc species, behind the Japanese clam, *Ruditapes philippinarum* [61]. The sedentary lifestyle of oysters in the intertidal zone and estuaries, where they are exposed to dynamic environmental stresses including high temperatures, low salinity, and desiccation, necessitates great tolerance to fluctuating conditions. Oysters are suspension feeders, meaning that they have excellent innate immune systems in order to defend themselves against aquatic microbes. These adaptive capabilities enable *C. gigas* to colonize habitats worldwide. *C. gigas* is one of the most studied molluscs, and its

molecular mechanisms, especially gene expression responses to biotic and abiotic challenges, have been heavily investigated [24]. The *C. gigas* genome, decoded in 2012, showed expanded gene families, such as molecular chaperone heat shock proteins (HSPs), inhibitor of apoptosis proteins (IAPs), and superoxide dismutases (SODs). Their up-regulated gene expression represents a response to environmental stresses [10]. Gene families responsible for innate immunity, such as C1q and Toll-like receptors (TLR), are also expanded [62-64]. Notably, some genes in these families respond to abiotic changes (temperature, salinity, and air exposure), indicating that some of the duplicated "immune" genes have been co-opted to accommodate environmental stresses [63]. Understanding the physiology of oysters is essential to improve production and maintain food security of this important mollusc.

## Molluscan genomics for various biological issues

Apart from their importance for the aquaculture industry, mollusc genomes have been studied to address diverse range of biological questions. The phylum Mollusca belongs to the Lophotrochozoa, which comprises one of major clades within the Bilateria. Since genomic information for lophotrochozoans is scarce, mollusc genomes are of particular value to study animal genome evolution. The genome of the owl limpet, *Lottia gigantea*, and two annelid genomes have been sequenced, allowing reconstruction of 17 bilaterian ancestral linkage groups (ALGs) [11]. The genome of the scallop, *Patinopecten yessoensis*, showed remarkable preservation of bilaterian ALGs, as well as intact Hox and ParaHox clusters, which together may represent the ancestral state of lophotrochozoans [18]. Expression of Hox and ParaHox genes showed subcluster-level temporal co-linearity, and this could be an ancestral pattern in bilaterians [18]. The genome of the deep sea mussel, *Bathymodiolus platifrons*, was compared with that of the shallow water mussel, *Modiolus philippinarum*, in order to study the genetic basis for adaptation to extreme environments [17]. In the *B. platifrons* genome, HSP70 and ABC transporter gene families are expanded and highly expressed in gill tissue, suggesting a role in resistance to physical stresses and toxic chemicals in the deep-sea environment. A molecular mechanism for acquiring methane oxidizing symbionts is also hypothesized from expanded gene families, such as Toll-like receptors, adhesion genes (syndecan and protocadherin), and apoptosis-related genes [17]. The freshwater snail, *Biomphalaria glabrata*, is an intermediate host of the blood fluke, *Schistosoma mansoni*, therefore it may be possible to interrupt snail-mediated parasite transmission. Genome analysis of *B. glabrata* provides basic information about its biological process such as interactions between the snail and the parasite

[15]. Cephalopods command special interest because of their specialized body plans and complex neural systems. The genome of the octopus, *Octopus vulgaris*, demonstrated a large number of protocadherin genes, which are responsible for neuronal development [12]. The C2H2 zinc finger transcription factor gene family is also expanded, and mRNAs of tandemly arranged C2H2 genes are expressed in adult brain, optic lobe, axial nerve cord, and in embryonic tissues. Extensive RNA editing in neural tissue is also evident, enabling complex neural excitability [12, 65].

In addition to molluscan genome studies mentioned above, genomes of the mussel (*Mytilus galloprovincialis*), the clam (*Ruditapes philippinarum*), the scallop (*Argopecten irradians*), the freshwater snail (*Radix auricularia*), and the abalone (*Haliotis discus hannai*) have been published. These studies briefly report statistics of the assembly and predicted gene models [14, 16, 19, 21, 22]. Rapidly accumulating whole genome data will contribute further understanding of the molecular biology of molluscs.

## Genome-wide studies for molluscan aquaculture

Beside providing fundamental insights into biological features of molluscs, whole genome data are essential for the aquaculture industry to develop genetic markers for economically valuable traits. High-throughput sequencing technology is effective not only for whole genome shotgun sequencing, but also for genome-wide genetic marker discovery. Massive parallel, short read sequencing combined with a reduced representation library is an optimal strategy for this purpose. Various genotyping methods for reduced representation sequencing have been developed, such as restriction-site-associated DNA sequencing (RAD-seq) [66], genotyping by sequencing (GBS) [67], 2b-restriction site-associated DNA (2b-RAD) sequencing [68], and specific-locus amplified fragment sequencing (SLAF-seq) [69]. These techniques have become common for genotyping commercially valuable molluscs. In principle, all of these methods use one or more restriction enzymes to prepare DNA libraries for sequencing. Genomic DNA is fragmented with restriction enzymes and adapters containing sequencing-initiation sites are ligated at the cohesive ends. As a result, genomic regions close to the restriction enzyme recognition sites are selectively sequenced so that high sequence coverage sufficient for genotyping can be obtained. Furthermore, by adding sample-specific index sequences (barcodes) to the adapters, multiple individuals can be sequenced in a sequencing run. The reduced representation sequencing method discovers thousands of single-nucleotide

polymorphisms (SNP) within populations. In order to establish high-density linkage maps, 1,000-10,000 SNP markers are identified from hundreds of individuals. Table 1 lists the high-density linkage map studies of commercial mollusc species [70-76]. For instance, 96 full-sib progeny were sequenced and 3,806 markers were identified from the Chinese scallop, *Chlamys farreri*, using the 2b-RAD method [70]. Once a sufficient number of SNP markers have been established, an SNP array is an alternative method of genome-wide genotyping. Medium- to high-density SNP arrays for *Crassostrea gigas*, a *Crassostrea gigas* x *Ostrea edulis*, cross, and *Pinctada maxima* have been tested for genotyping [77-79].

Linkage maps are used to identify quantitative trait loci (QTLs). Genotypes in QTLs are correlated with particular phenotypes; therefore, they are used as markers for selection. Growth-related traits, such as shell size and body weight, are of major research interest for mollusc aquaculture [70-76]. The triangle sail mussel, *Hyriopsis cumingii*, which is cultured for fresh water pearl production, was analyzed for QTLs associated with nacre color [73, 80]. QTLs for shell color and resistance to disease in *C. gigas* have also been investigated [81, 82]. In cases where genome assemblies are available, QTL regions in the physical map or associated genes can be identified. In the *C. farreri* genome, the transcription factor gene, PROP1, that regulates animal growth, is associated with a growth QTL [70]. A shell matrix protein gene, N16, is also reported to be linked to a growth-related QTL in the *Pinctada fucata* genome [72]. These results of QTL analyses will provide genetic markers correlated with economically valuable traits. Then individuals can be efficiently selected for breeding programs using marker-assisted selection (MAS). MAS is efficient if the desired trait or phenotype is controlled by a small number of genes or QTLs. Genome-wide association studies (GWAS) may contribute significantly to mollusc aquaculture because GWA does not require family information. Therefore, individuals captured in the wild can be analyzed as potential genetic resources. Genomic selection (GS) based on GWA is a more powerful genetic tool when the trait of interest is weakly associated with a large number of QTLs. Although GWA combined with GS is more costly than QTL analysis, because in general, tens of thousands of SNPs and a ≥1,000 individuals must be analyzed, this technology will become standard as sequencing costs continue to drop.

Linkage maps are also used for anchoring genome scaffolds to linkage groups. Theoretically, if at least one genetic marker is mapped on each scaffold, the genome scaffolds can be clustered

into linkage groups or chromosomes. Using this method, genome assemblies of *Patinopecten yessoensis*, *C. gigas*, and *P. fucata* were enhanced to chromosome level assembly [18, 20]. Furthermore, linkage maps are available to assess assembly errors in genome scaffolds. Based on linkage maps generated from high-density genetic markers, about 40% of *C. gigas* genome scaffolds with more than one marker were mapped to different linkage groups, indicating that the scaffolds were misassembled [83]. Linkage analysis can correct and improve continuity of genome assemblies.

## Conclusion

By virtue of fast-growing sequencing technology, molluscan genome sequencing projects are proceeding at an astonishing rate. Challenging issues still remain for decoding molluscan genomes, such as their huge genome sizes and the high heterozygosity of bivalve genomes. Therefore, sequencing strategies and assembly methods should be carefully considered. Constructing linkage maps is an efficient way to evaluate assembly errors and to construct chromosomal-level assemblies.

Molluscan genomes provide fundamental molecular information to address their unique biological features. Lineage-specific, expanded gene families related to shell formation, immunity, the nervous system, etc. are evident. Functional analyses such as gene knockdown, RNA-Seq, and proteomics enrich our understanding of their significance. Genome sequence data can be used to develop genetic tools for aquaculture. Conventionally, in order to select individuals with valuable traits such as high growth rate and resistance to disease, costly long-term rearing or infectivity assays are necessary. Using DNA markers, characteristics of each individual can be estimated efficiently. Whole-genome assembly and gene annotation information can identify genes located near genetic markers correlated with specific traits. If gene functions are already known, biological evidence corroborates the selection program. Alternatively, function of unknown genes can be inferred based on the presence of markers associated with the research interest. Biological knowledge from molluscan genomics facilitates data-driven breeding programs, and accumulation of genotypic and phenotypic information can assist functional genomic studies. Modern genomic studies facilitate molluscan biology, genetics, and aquaculture.

## Acknowledgements

I am grateful to all members of Marine Genomics Unit at OIST for their support. I also thank Dr. Steven D. Aird for editing the manuscript.

## Compliance with Ethical Standards

## Conflict of Interest

This research was supported by grants from the Project to Advance Institutional Bio-oriented Technology Research, NARO (special project on advanced research and development for next-generation technology), and by internal funds from the Okinawa Institute of Science and Technology (OIST).

## Human and Animal Rights and Informed Consent

This article does not contain any studies with human or animal subjects performed by the author.

## References

1. Rosenberg G. A new critical estimate of named species-level diversity of the recent Mollusca. American Malacological Bulletin. 2014;32(2):308-22. doi:10.4003/006.032.0204.

2. FAO. The State of World Fisheries and Aquaculture 2016: Contributing to food security and nutrition for all. Rome: Food and Agriculture Organization; 2016.

3. FAO. The State of World Fisheries and Aquaculture 2006. Rome: Food Agriculture Organization; 2007.

4. FAO. FAO Yearbook. Fishery and Aquaculture Statistics 2011. Rome: Food and Agriculture Organization; 2013.

5. Saavedra C, Bachère E. Bivalve genomics. Aquaculture. 2006;256(1–4):1-14. doi:10.1016/j.aquaculture.2006.02.023.

6. Meuwissen T, Hayes B, Goddard M. Genomic selection: A paradigm shift in animal breeding. Animal Frontiers. 2016;6(1):6-14. doi:10.2527/af.2016-0002.

7. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proceedings of the National Academy of Sciences. 2016;113(28):E3995-E4004. doi:10.1073/pnas.1519061113.

8. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-51. doi:10.1038/nrg.2016.49.

** 9. Takeuchi T, Kawashima T, Koyanagi R, Gyoja F, Tanaka M, Ikuta T et al. Draft genome of the pearl oyster *Pinctada fucata*: a platform for understanding bivalve biology. DNA Res. 2012;19(2):117-30. doi:10.1093/dnares/dss005.

This is the first published molluscan genome article. They showed high heterozygosity in the bivalve genome.

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

In this study of the Pacific oyster genome, expansion of specific gene family such as HSP70 and IAPs, reinforces their adaptation ability to harsh environmental stresses in the intertidal zone.

11. Simakov O, Marletaz F, Cho S-J, Edsinger-Gonzales E, Havlak P, Hellsten U et al. Insights into bilaterian evolution from three spiralian genomes. Nature. 2013;493(7433):526-31. doi:10.1038/nature11696.

12. Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E et al. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature. 2015;524(7564):220-4. doi:10.1038/nature14668.

13. Takeuchi T, Koyanagi R, Gyoja F, Kanda M, Hisata K, Fujie M et al. Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. Zoological Letters. 2016;2(1):3. doi:10.1186/s40851-016-0039-2.

14. Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C. A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. PLoS One. 2016;11(3):e0151561. doi:10.1371/journal.pone.0151561.

15. Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. Nature Communications. 2017;8:15451. doi:10.1038/ncomms15451.

16. Schell T, Feldmeyer B, Schmidt H, Greshake B, Tills O, Truebano M et al. An annotated draft genome for *Radix auricularia* (Gastropoda, Mollusca). Genome Biol Evol.

2017;9(3):585-92. doi:10.1093/gbe/evx032.

17. Sun J, Zhang Y, Xu T, Zhang Y, Mu H, Zhang Y et al. Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. Nature Ecology & Evolution. 2017;1:0121. doi:10.1038/s41559-017-0121.

** 18. Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y et al. Scallop genome provides insights into evolution of bilaterian karyotype and development. Nature Ecology & Evolution. 2017;1:0120. doi:10.1038/s41559-017-0120.

In this study, three bivalve genome assemblies including the scallop, the Pacific oyster, and the pearl oyster were reconstructed to chromosomal level by using genetic maps. They showed the scallop genome retains bilaterian ancestral state.

19. Du X, Song K, Wang J, Cong R, Li L, Zhang G. Draft genome and SNPs associated with carotenoid accumulation in adductor muscles of bay scallop (*Argopecten irradians*). Journal of Genomics. 2017;5:83-90. doi:10.7150/jgen.19146.

20. Du X, Fan G, Jiao Y, Zhang H, Guo X, Huang R et al. The pearl oyster *Pinctada fucata martensii* genome and multi-omic analyses provide insights into biomineralization. GigaScience. 2017;6(8):1-12. doi:10.1093/gigascience/gix059.

21. Mun S, Kim Y-J, Markkandan K, Shin W, Oh S, Woo J et al. The whole-genome and transcriptome of the Manila clam (*Ruditapes philippinarum*). Genome Biol Evol. 2017;9(6):1487-98. doi:10.1093/gbe/evx096.

22. Nam B-H, Kwak W, Kim Y-O, Kim D-G, Kong HJ, Kim W-J et al. Genome sequence of pacific abalone (*Haliotis discus hannai*): the first draft genome in family Haliotidae. GigaScience. 2017;6(5):1-8. doi:10.1093/gigascience/gix014.

23. Consortium TGP. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73. doi:10.1038/nature09534.

24. Guo X, He Y, Zhang L, Lelong C, Jouaux A. Immune and stress responses in oysters with insights on adaptation. Fish Shellfish Immunol. 2015;46(1):107-19. doi:10.1016/j.fsi.2015.05.018.

25. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature. 2014;515(7526):261-3. doi:10.1038/nature13685.

26. Curole JP, Hedgecock D. Bivalve genomics: complications, challenges, and future perspectives. Aquaculture genome technologies. Oxford, UK: Blackwell Publishing Ltd; 2007.

27. Compeau PEC, Pevzner PA, Tesler G. How to apply de Bruijn graphs to genome assembly.

Nat Biotech. 2011;29(11):987-91. doi:10.1038/nbt.2023.

28. Henson J, Tischler G, Ning Z. Next-generation sequencing and large genome assemblies. Pharmacogenomics. 2012;13(8):901-15. doi:10.2217/pgs.12.72.

29. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience. 2013;2(1):10. doi:10.1186/2047-217X-2-10.

30. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 2014;24(8):1384-95. doi:10.1101/gr.170720.113.

31. Kelley DR, Salzberg SL. Detection and correction of false segmental duplications caused by genome mis-assembly. Genome Biol. 2010;11(3):R28. doi:10.1186/gb-2010-11-3-r28.

32. Zheng H, Li L, Zhang G. Inbreeding depression for fitness-related traits and purging the genetic load in the hermaphroditic bay scallop *Argopecten irradians irradians* (Mollusca: Bivalvia). Aquaculture. 2012;366:27-33. doi:10.1016/j.aquaculture.2012.08.029.

33. Kocot KM, Jeffery NW, Mulligan K, Halanych KM, Gregory TR. Genome size estimates for Aplacophora, Polyplacophora and Scaphopoda: small solenogasters and sizeable scaphopods. Journal of Molluscan Studies. 2016;82(1):216-9. doi:10.1093/mollus/eyv054.

34. Ieyama H, Ogaito H. Chromosomes and nuclear DNA contents of two subspecies in the Diplommatinidae. Venus: the Japanese journal of malacology. 1998;57(2):133-6.

35. Bonnaud L, Ozouf-Costaz C, Boucher-Rodoni R. A molecular and karyological approach to the taxonomy of *Nautilus*. Comptes Rendus Biologies. 2004;327(2):133-8. doi:10.1016/j.crvi.2003.12.004.

36. Hallinan NM, Lindberg DR. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. Genome Biol Evol. 2011;3:1150-63. doi:10.1093/gbe/evr087.

37. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA. 2015;6(1):11. doi:10.1186/s13100-015-0041-9.

38. Nagai K. A history of the cultured pearl industry. Zoolog Sci. 2013;30(10):783-93. doi:10.2108/zsj.30.783.

39. Gyoja F, Satoh N. Evolutionary aspects of variability in bHLH orthologous families: Insights from the pearl oyster, *Pinctada fucata*. Zoolog Sci. 2013;30(10):868-76. doi:10.2108/zsj.30.868.

40. Koga H, Hashimoto N, Suzuki DG, Ono H, Yoshimura M, Suguro T et al. A genome-wide

survey of genes encoding transcription factors in Japanese pearl oyster *Pinctada fucata*: II. Tbx, Fox, Ets, HMG, NFκB, bZIP, and C2H2 zinc fingers. Zoolog Sci. 2013;30(10):858-67. doi:10.2108/zsj.30.858.

41. Morino Y, Okada K, Niikura M, Honda M, Satoh N, Wada H. A genome-wide survey of genes encoding transcription factors in the Japanese pearl oyster, *Pinctada fucata*: I. homeobox genes. Zoolog Sci. 2013;30(10):851-7. doi:10.2108/zsj.30.851.

42. Setiamarga DHE, Shimizu K, Kuroda J, Inamura K, Sato K, Isowa Y et al. An in-silico genomic survey to annotate genes coding for early development-relevant signaling molecules in the pearl oyster, *Pinctada fucata*. Zoolog Sci. 2013;30(10):877-88. doi:10.2108/zsj.30.877.

43. Funabara D, Watanabe D, Satoh N, Kanoh S. Genome-wide survey of genes encoding muscle proteins in the pearl oyster, *Pinctada fucata*. Zoolog Sci. 2013;30(10):817-25. doi:10.2108/zsj.30.817.

44. Matsumoto T, Masaoka T, Fujiwara A, Nakamura Y, Satoh N, Awaji M. Reproduction-related genes in the pearl oyster genome. Zoolog Sci. 2013;30(10):826-50. doi:10.2108/zsj.30.826.

45. Miyamoto H, Endo H, Hashimoto N, limura K, Isowa Y, Kinoshita S et al. The diversity of shell matrix proteins: genome-wide investigation of the pearl oyster, *Pinctada fucata*. Zoolog Sci. 2013;30(10):801-16. doi:10.2108/zsj.30.801.

46. Kinoshita S, Ning W, Inoue H, Maeyama K, Okamoto K, Nagai K et al. Deep sequencing of ESTs from nacreous and prismatic layer producing tissues and a screen for novel shell formation-related genes in the pearl oyster. PLoS One. 2011;6(6):e21238. doi:10.1371/journal.pone.0021238.

47. Marie B, Joubert C, Tayalé A, Zanella-Cléon I, Belliard C, Piquemal D et al. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proceedings of the National Academy of Sciences. 2012;109(51):20986-91. doi:10.1073/pnas.1210552109.

48. Liu C, Li S, Kong J, Liu Y, Wang T, Xie L et al. In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. Sci Rep. 2015;5:17269. doi:10.1038/srep17269.

49. Suzuki M, Saruwatari K, Kogure T, Yamamoto Y, Nishimura T, Kato T et al. An acidic matrix protein, Pif, is a key macromolecule for nacre formation. Science. 2009;325(5946):1388-90.

50. Funabara D, Ohmori F, Kinoshita S, Koyama H, Mizutani S, Ota A et al. Novel genes participating in the formation of prismatic and nacreous layers in the pearl oyster as revealed by

their tissue distribution and RNA interference knockdown. PLoS One. 2014;9(1):e84706. doi:10.1371/journal.pone.0084706.

51. Marie B, Jackson DJ, Ramos-Silva P, Zanella-Cléon I, Guichard N, Marin F. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. FEBS Journal. 2013;280(1):214-32. doi:10.1111/febs.12062.

52. Feng D, Li Q, Yu H, Kong L, Du S. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. Sci Rep. 2017;7:45754. doi:10.1038/srep45754.

53. Mann K, Edsinger-Gonzales E, Mann M. In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. Proteome Sci. 2012;10(1):28. doi:10.1186/1477-5956-10-28.

54. Mann K, Edsinger E. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. Proteome Sci. 2014;12(1):28. doi:10.1186/1477-5956-12-28.

55. McDougall C, Aguilera F, Degnan BM. Rapid evolution of pearl oyster shell matrix proteins with repetitive, low-complexity domains. Journal of The Royal Society Interface. 2013;10(82):20130041. doi:10.1098/rsif.2013.0041.

56. Rohfritsch A, Bierne N, Boudry P, Heurtebise S, Cornette F, Lapègue S. Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, *Crassostrea gigas*. Evolutionary Applications. 2013;6(7):1064-78. doi:10.1111/eva.12086.

57. Ruesink JL, Lenihan HS, Trimble AC, Heiman KW, Micheli F, Byers JE et al. Introduction of non-native oysters: ecosystem effects and restoration implications. Annu Rev Ecol Evol Syst. 2005;36:643-89.

58. Orensanz JM, Schwindt E, Pastorino G, Bortolus A, Casas G, Darrigran G et al. No longer the pristine confines of the world ocean: A survey of exotic marine species in the Southwestern Atlantic. Biol Invasions. 2002;4(1):115-43. doi:10.1023/A:1020596916153.

59. Guo X. Use and exchange of genetic resources in molluscan aquaculture. Reviews in Aquaculture. 2009;1(3-4):251-9. doi:10.1111/j.1753-5131.2009.01014.x.

60. Galil BS. A sea under siege – alien species in the Mediterranean. Biol Invasions. 2000;2(2):177-86. doi:10.1023/A:1010057010476.

61. FAO. FAO Yearbook. Fishery and Aquaculture Statistics 2014. Rome: Food and Agriculture Organization; 2016.

62. Gerdol M, Venier P, Pallavicini A. The genome of the Pacific oyster *Crassostrea gigas* brings new insights on the massive expansion of the C1q gene family in Bivalvia. Dev Comp Immunol. 2015;49(1):59-71. doi:10.1016/j.dci.2014.11.007.

63. Zhang L, Li L, Guo X, Litman GW, Dishaw LJ, Zhang G. Massive expansion and functional divergence of innate immune genes in a protostome. Sci Rep. 2015;5:8693. doi:10.1038/srep08693.

64. He Y, Jouaux A, Ford SE, Lelong C, Sourdaine P, Mathieu M et al. Transcriptome analysis reveals strong and complex antiviral response in a mollusc. Fish Shellfish Immunol. 2015;46(1):131-44. doi:10.1016/j.fsi.2015.05.023.

65. Rosenthal Joshua JC, Seeburg Peter H. A-to-I RNA editing: effects on proteins key to neural excitability. Neuron. 2012;74(3):432-9. doi:10.1016/j.neuron.2012.04.010.

** 66. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 2008;3(10):e3376. doi:10.1371/journal.pone.0003376.

They developed reduced representation sequencing method with high-throuput sequencing technology. The method called restriction-site-associated DNA sequencing or RAD-seq is modified and actively utilized for SNP discovery in non-model aquacultuer animals.

67. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One. 2011;6(5):e19379. doi:10.1371/journal.pone.0019379.

68. Wang S, Meyer E, McKay JK, Matz MV. 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat Meth. 2012;9(8):808-10. doi:10.1038/nmeth.2023.

69. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W et al. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PLoS One. 2013;8(3):e58700. doi:10.1371/journal.pone.0058700.

70. Jiao W, Fu X, Dou J, Li H, Su H, Mao J et al. High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: Building up an integrative genomic framework for a bivalve mollusc. DNA Res. 2014;21(1):85-101. doi:10.1093/dnares/dst043.

71. Li Y, He M. Genetic mapping and QTL analysis of growth-related traits in *Pinctada fucata* using restriction-site associated DNA sequencing. PLoS One. 2014;9(11):e111707. doi:10.1371/journal.pone.0111707.

72. Shi Y, Wang S, Gu Z, Lv J, Zhan X, Yu C et al. High-density single nucleotide polymorphisms linkage and quantitative trait locus mapping of the pearl oyster, *Pinctada fucata*

*martensii* Dunker. Aquaculture. 2014;434:376-84. doi:10.1016/j.aquaculture.2014.08.044.

73. Bai Z-Y, Han X-K, Liu X-J, Li Q-Q, Li J-L. Construction of a high-density genetic map and QTL mapping for pearl quality-related traits in *Hyriopsis cumingii*. Sci Rep. 2016;6:32608. doi:10.1038/srep32608.

74. Ren P, Peng W, You W, Huang Z, Guo Q, Chen N et al. Genetic mapping and quantitative trait loci analysis of growth-related traits in the small abalone *Haliotis diversicolor* using restriction-site-associated DNA sequencing. Aquaculture. 2016;454:163-70. doi:10.1016/j.aquaculture.2015.12.026.

75. Nie H, Yan X, Huo Z, Jiang L, Chen P, Liu H et al. Construction of a high-density genetic map and quantitative trait locus mapping in the Manila clam *Ruditapes philippinarum*. Sci Rep. 2017;7:229. doi:10.1038/s41598-017-00246-0.

76. Wang J, Li L, Zhang G. A high-density SNP genetic linkage map and QTL analysis of growth-related traits in a hybrid family of oysters *(Crassostrea gigas* × *Crassostrea angulata*) using genotyping-by-sequencing. G3: Genes|Genomes|Genetics. 2016;6(5):1417.

77. Gutierrez AP, Turner F, Gharbi K, Talbot R, Lowe NR, Peñaloza C et al. Development of a Medium Density Combined-Species SNP Array for Pacific and European Oysters (*Crassostrea gigas* and & *Ostrea edulis*). G3: Genes|Genomes|Genetics. 2017;7(7):2209.

78. Qi H, Song K, Li C, Wang W, Li B, Li L et al. Construction and evaluation of a high-density SNP array for the Pacific oyster (*Crassostrea gigas*). PLoS One. 2017;12(3):e0174007. doi:10.1371/journal.pone.0174007.

79. Jones DB, Jerry DR, Forêt S, Konovalov DA, Zenger KR. Genome-wide SNP validation and mantle tissue transcriptome analysis in the silver-lipped pearl oyster, *Pinctada maxima*. Mar Biotechnol. 2013;15(6):647-58. doi:10.1007/s10126-013-9514-3.

80. Bai Z, Han X, Luo M, Lin J, Wang G, Li J. Constructing a microsatellite-based linkage map and identifying QTL for pearl quality traits in triangle pearl mussel (*Hyriopsis cumingii*). Aquaculture. 2015;437:102-10. doi:10.1016/j.aquaculture.2014.11.008.

81. Zhong X, Li Q, Guo X, Yu H, Kong L. QTL mapping for glycogen content and shell pigmentation in the Pacific oyster *Crassostrea gigas* using microsatellites and SNPs. Aquac Int. 2014;22(6):1877-89. doi:10.1007/s10499-014-9789-z.

82. Sauvage C, Boudry P, De Koning DJ, Haley CS, Heurtebise S, Lapègue S. QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). Anim Genet. 2010;41(4):390-9. doi:10.1111/j.1365-2052.2009.02018.x.

83. Hedgecock D, Shin G, Gracey AY, Den Berg DV, Samanta MP. Second-generation linkage

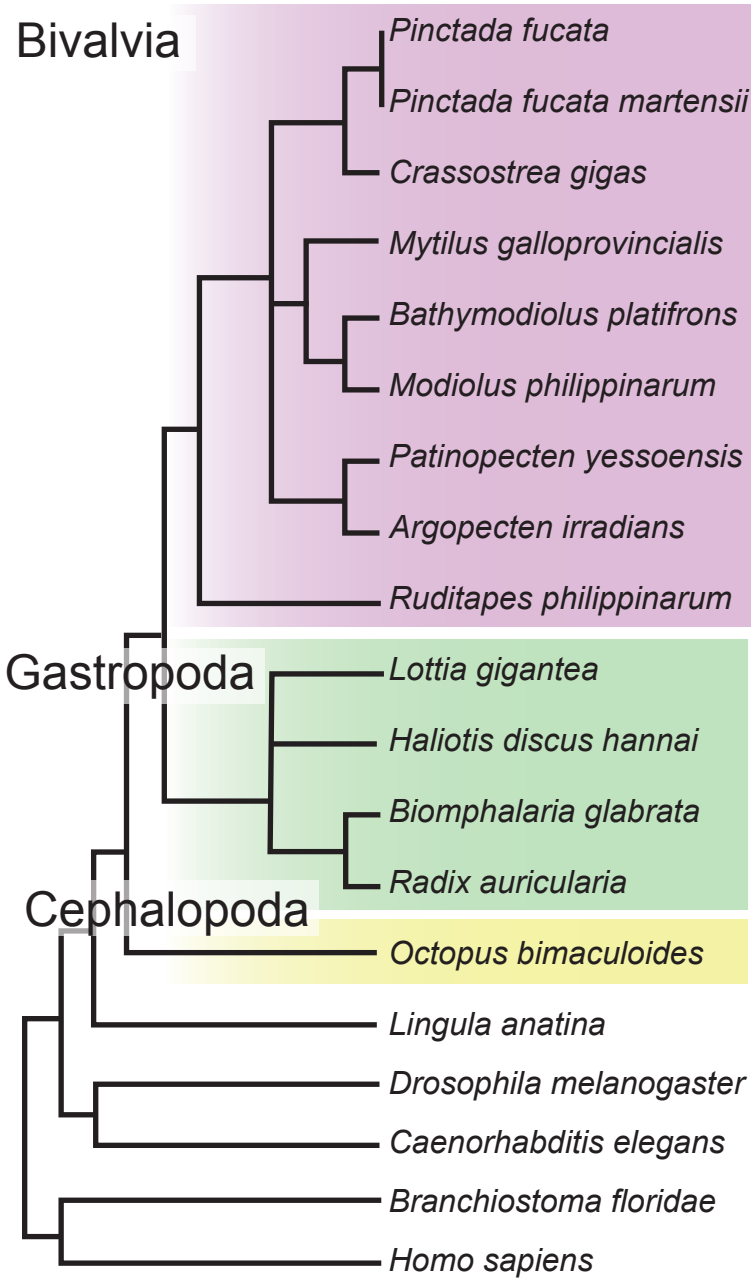
maps for the Pacific oyster *Crassostrea gigas* reveal errors in assembly of genome scaffolds. G3: Genes|Genomes|Genetics. 2015;5(10):2007.

Figure legend

**Figure 1. Published molluscan genome assemblies and their statistics.** Evolutionary relationship of the molluscs is show at the left.

Table1. Linkage map with high density SNPs

| Class | Species | Genotyping method | Number of linkage groups | Number of markers | Total size (cM) | Average distance (cM) | Reference |
|---|---|---|---|---|---|---|---|
| Bivalvia | *Chlamys farreri* | 2b-RAD | 19 | 3,806 | 1543.36 | 0.41 | Jiao *et al.* (2014) [70] |
| | *Pinctada fucata martensii* | 2b-RAD | 14 | 3,117 | 990.74 | 0.39 | Shi *et al.* (2014) [72] |
| | *Pinctada fucata* | RAD-seq | 14 | 1373 | 1091.81 | 1.41 | Li and He (2014) [71] |
| | *Pinctada fucata martensii* | RAD-seq | 14 | 4,463 | 4287.61 | 0.96 | Du *et al.* (2017) [20] |
| | *Crassostrea gigas* x *C. angulata* | GBS | 10 | 1,695 | 1084.3 | 0.80 | Wang *et al.* (2016) [76] |
| | *Hyriopsis cumingii* | SLAF-seq | 19 | 4,920 | 2713.17 | 1.81 | Bai *et al.* (2016) [80] |
| | *Ruditapes philippinarum* | GBS | 18 | 9,658 | 1926.98 | 0.42 | Nie *et al.* (2017) [75] |
| | *Patinopecten yessoensis* | 2b-RAD | 19 | 7,489 | 1918.65 | 0.26 | Wang *et al.* (2017) [18] |
| Gastropoda | *Haliotis diversicolor* | RAD-seq | 16 | 3,717 | 2190.1 | 0.59 | Ren *et al.* (2016) [74] |




| Common name | Genome size | Total scaffold length | Number of scaffolds | Scaffold N50 | Reference |
|---|---|---|---|---|---|
| pearl oyster | 1.15 Gb [a)] | 815.3 Mb | 29,306 | 167.0 kb | Takeuchi *et al.* (2012) [9]<br>Takeuchi *et al.* (2016) [13] |
| pearl oyster | – | 990.6 Mb | 8,621 | 324.3 kb | Du *et al.* (2017) [20] |
| Pacific oyster | 637 Mb [a)] | 558.6 Mb | 11,969 | 401.3 kb | Zhang *et al.* (2012) [10] |
| Mediterranean mussel | 1.6 Gb [b)] | 1,599 Mb | 1,746,447 | 2.6 kb | Murgarella *et al.* (2016) [14] |
| deep sea mussle | 1.63 Gb [b)] | 1,660 Mb | 65,664 | 343.4 kb | Sun *et al.* (2017) [17] |
| mussel | 2.21 Gb [b)] | 2,630 Mb | 74,575 | 100.2 kb | Sun *et al.* (2017) [17] |
| scallop | 1.44 Gb [a)] | 987.6 Mb | 82,731 | 803.6 kb | Wang *et al.* (2017) [18] |
| bay scallop | 990 Mb [b)] | 700.3 Mb | 217,310 | 6.8 kb | Du *et al.* (2017) [19] |
| Manila clam | 1.37 Gb [b)] | 2,561 Mb | 223,851 | 48.4 kb | Mun *et al.* (2017) [21] |
| owl limpet | 420 Mb [c)] | 359.5 Mb | 4,475 | 1,870 kb | Simakov *et al.* (2013) [11] |
| abalone | 1.8 Gb [a)] | 1,860 Mb | 35,450 | 211.3 kb | Nam *et al.* (2017) [22] |
| fresh water snail | 916 Mb [d)] | 916.3 Mb | 331,400 | 48 kb | Adema *et al.* (2017) [15] |
| fresh water snail | 1.58 Gb [a)] | 910 Mb | 4,823 | 578.7 kb | Schell *et al.* (2017) [16] |
| octopus | 2.68 Gb [a)] | 2,371 Mb | 379,696 | 1,369 kb | Albertin *et al.* (2015) [12] |

a) Genome size measured by flow cytometry
b) Genome size estimated by k-mer frequency analysis
c) Genome size measured by fluorometric assay
d) Genome size measured by Feulgen image analysis densitometry