Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze della Terra, della Vita e dell'Ambiente

Ciclo XXXIV

Settore Concorsuale: 05/B1 - ZOOLOGIA E ANTROPOLOGIA

Settore Scientifico Disciplinare: BIO/05 - ZOOLOGIA

Mitochondrial Inheritance, Mito-nuclear Coevolution, and Sex-associated Genes in Bivalve Molluscs: Transcriptomics and Molecular Evolution

Presentata da: Ran Xu

Coordinatore Dottorato

Maria Giovanna Belcastro

Supervisore

Fabrizio Ghiselli

Co-supervisore

Liliana Milani

Esame finale anno 2022

Contents

Abstract

Bivalvia represents an ancient taxon including around 25,000 living species that have adapted to a wide range of environmental conditions, and show a great diversity in body size, shell shapes, and anatomic structure. Bivalves are characterized by highly variable genome sizes and extremely high levels of heterozygosity, which obstacle complete and accurate genome assemblies and hinder further genomic studies. Moreover, some bivalve species presented a stable evolutionary exception to the strictly maternal inheritance of mitochondria, namely doubly uniparental inheritance (DUI), making these species a precious model to study mitochondrial biology. During my PhD, I focused on a DUI species, the Manila clam *Ruditapes philippinarum,* and my work was two-folded.

First, taking advantage of a newly assembled draft genome and a large RNA-seq dataset from different tissues of both sexes, I investigated 1) the role of gene expression and alternative splicing in tissue differentiation; 2) the relationship across tissue specificity, regulatory network connectivity, and sequence evolution; 3) sexual contrasting genetic markers potentially associated with sexual differentiation. The detailed information for this part is in Chapter 2. Second, using the same RNA-seq data, I investigated how nuclear oxidative phosphorylation (OXPHOS) genes coordinate with two divergent mitochondrial genomes in DUI species (mito-nuclear coordination and coevolution). To address this question, I compared transcription, polymorphism, and synonymous codon usage in the mitochondrial and nuclear OXPHOS genes of *R. philippinarum* in Chapter 3. To my knowledge, this thesis represents the first study exploring the role of alternative splicing in tissue differentiation, and the first study analyzing both transcriptional regulation and sequence evolution to investigate the coordination of OXPHOS genes in bivalves.

Chapter 1 Introduction

1.1 An usual mitochondrial inheritance system

In most animals, mitochondria follow a strictly maternal inheritance (SMI), from mothers to progeny. Paternal leakage of mitochondrial DNA is usually rare, and even if it happens, it is maintained at low level (Gyllensten et al., 1991; Mastrantonio et al., 2019; Morgan et al., 2013; Nunes et al., 2013). However, in some bivalves, mitochondria can be inherited from both parents, with a mechanism named doubly uniparental inheritance (DUI) (Breton et al., 2007; Passamonti & Ghiselli, 2009; Zouros, 2013). In this system, two highly divergent mitochondrial lineages (F-type and M-type, up to 53% genetic divergence at amino acid level) are present: the F-type is transmitted through eggs while the M-type is transmitted through sperm. Gametes are homoplasmic for the respective sex-linked lineage, while the distribution of F- and M-type mitochondria in adult tissues is variable according to sex, tissue, and species (Ghiselli et al., 2011, 2019). The embryo, on the other hand, is heteroplasmic during the early developmental stages, and two different mitochondrial segregation patterns of sperm mitochondria were observed and were proposed to be linked to the presence of DUI (Cao et al., 2004; Cogswell et al., 2006; Kenchington et al., 2009; Milani et al., 2012; Obata & Komaru, 2005). Sperm mitochondria in male embryos present a non-random segregation and end up into a single blastomere, which is likely the precursor of the male germline, whereas in female embryos, sperm mitochondria are dispersed and appear to be depleted eventually. However, such elimination mechanisms in DUI species may not be that strict, and the situation is probably more complex since M-type mtDNA was detected in female samples and even eggs in *Mytilus galloprovincialis* (Kyriakou et al., 2010; Obata et al., 2006) and *Venustaconcha ellipsiformis* (Breton et al., 2017), and the presence of both F- and M-type proteins in stem cells, germ cells,

oogonia, and spermatogonia, in males and females of *R. philippinarum* (Ghiselli et al., 2019). This means that also females can be heteroplasmic, even if to a lesser extent when compared to males. That said, the large sequence divergence between F and M lineages within the same DUI species can be reached only if the two transmission routes are kept segregated. For this reason, the homoplasmy of gametes for the sex-specific mitochondrial lineage (F-type in eggs, M-type in sperm) must be the most common condition, and any deviation from that can be considered as "leakage" (Ghiselli et al. 2019). A schematic diagram of DUI system is shown in Figure 1.1.

Aside from this particular transmission system, novel sex-specific mitochondrial ORFs (ORFans) have been discovered in DUI species, and strong evidence suggested that these ORFans produce proteins (Breton et al., 2009, 2011; Milani et al., 2013, 2016). Despite that ORFans are among the fastest evolving genes in mitogenomes, ORFans were conserved across DUI species in the same family, and they do not correspond to any known homology or function (Breton et al., 2009, 2011; Milani et al., 2013, 2016). It was hypothesized that these ORFs may arise from endogenization of viral genes, and may link to the establishment of DUI, maintenance of sperm mitochondria, and sexual differentiation (Breton et al., 2009, 2011; Milani et al., 2013, 2016). Others argued that these ORFs may originate from the duplication of the existing gene or from *de novo* origin of the unassigned regions (URs), where ORFs are located (Guerra et al., 2017; Mitchell et al., 2016).

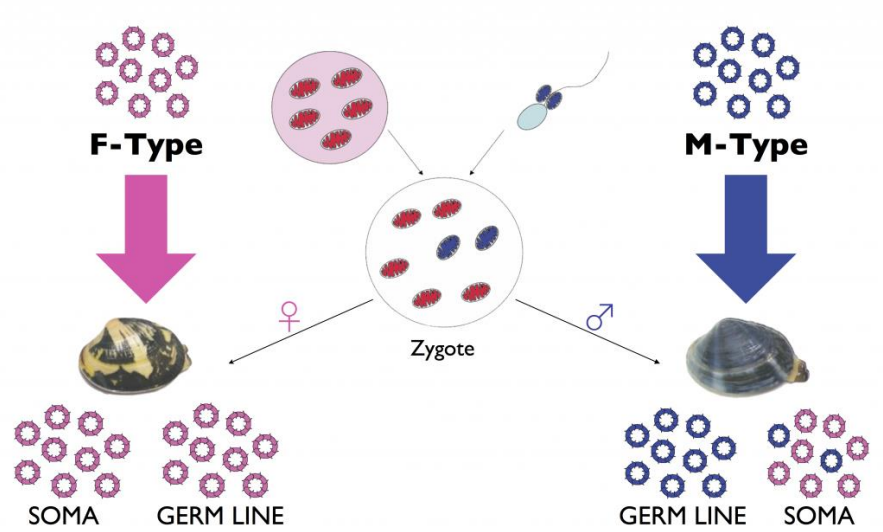


Figure 1.1 A schematic diagram of DUI system in Malina clam. Two divergent mitochondrial genomes were marked pink (F-type) and blue (M-type).

DUI has been reported in ~ 100 bivalves distributed in 13 families: Unionidae, Margaritiferidae, Hyriidae, Mytilidae, Arcticidae, Mactridae, Solenidae, Donacidae, Semelidae, Veneridae, Nuculanidae, Yoldiidae, Tellinidae (Gusman et al., 2016; Pante et al., 2017). Undoubtedly, this number will continue to grow as more bivalve species are studied. To our knowledge, DUI has been only found in bivalves, however, the origin of DUI remains unclear due to incomplete information about the presence/absence of DUI species in bivalves. Some suggested that DUI evolved once in history and was lost in some taxa, while others argued that DUI may have multiple origins (Maeda et al., 2021; Stewart et al., 2020; Zouros, 2013).

Being a stable evolutionary exception to SMI, DUI species have provided valuable insights into mitochondrial biology (Ghiselli et al., 2021; Passamonti & Plazzi 2020). Previous studies suggested a lower number of mitochondria—corresponding to a lower mtDNA copy number—transmitted through sperm with respect to eggs (Cao et al., 2004; Ghiselli et al., 2011; Milani et al., 2012). This provides a precious opportunity to

understand the mitochondrial bottleneck such as in a recent study in the Manila clam (Iannello et al., 2021). Additionally, such a high degree of heteroplasmy also represents an intriguing resource for investigating mito-nuclear coevolution. In DUI species, two divergent mitochondrial genomes need to cooperate with the same nuclear background, which could be challenging for the nuclear compensatory theory (See section 1.2). An unconventional link between DUI and sex have been hypothesized (Breton et al., 2011), although it is not clear if such linkage is causal or coincidental. Considering these unique features, DUI species represent a natural model to understand mitochondrial dynamics, inheritance and evolution.

1.2 Mito-nuclear coordination and coevolution

In most eukaryotes, the primary function of mitochondria is energy production through coupling respiration to ATP generation, a process termed oxidative phosphorylation (OXPHOS). This process involves the cooperation of five complex subunits (Complex I-V), which are encoded by both mitochondrial genes and nuclear genes except complex II (encoded only by nuclear genes). Therefore, to maintain the function of OXPHOS, mitochondrial and nuclear coordination is necessary to ensure the correct and efficient synthesis and assembly of the OXPHOS complexes (Hill, 2020; Isaac et al., 2018; Wolff et al., 2014).

Mito-nuclear co-regulation was found at multiple levels in many lineages, from the coordination of distinct gene expression to the protein transport and assembly (Isaac et al., 2018). In some model organisms, the expression of OXPHOS genes is sequentially co-regulated by numerous programs through mito-nuclear signaling pathways (Isaac et al., 2018; Quirós et al., 2016). At the transcription level, it was observed that nuclear OXPHOS (nuOXPHOS) genes usually cluster together,

indicating a shared transcriptional control mechanism (Barshad et al., 2018; Shyamsundar et al., 2005; van Waveren & Moraes, 2008). Additionally, OXPHOS genes in the same complex tend to show a tighter co-transcription signal than genes from different complexes, indicating a fine-tuning mechanism in each complex beside the shared control mechanism among all the nuOXPHOS genes (van Waveren & Moraes 2008). Recently, a study in human cells showed that the abundance of mitochondrial OXPHOS (mtOXPHOS) and nuOXPHOS RNA corresponds modestly across complexes, while this correspondence was strengthened by translational programs (Soto et al., 2021). In yeast, a remarkable role of translational regulation between mt and nuOXPHOS genes was also observed (Couvillion et al., 2016). By contrast, it was suggested that unassembled oversynthesized OXPHOS proteins can be quickly degraded by protein quality control pathways (Bogenhagen et al., 2020; Song et al., 2021), and that nuOXPHOS subunits in each complex are not in stoichiometric balance (Taggart & Li 2018). This questions the necessity for coordination between cytosolic and mitochondrial protein synthesis. Thus, how mito-nuclear OXPHOS subunit synthesis balance is maintained and fine-tuned is still an open question that requires further studies.

On the other hand, mt and nuOXPHOS subunits from different cellular compartments are experiencing different evolutionary dynamics, while intricate physical interactions between mitochondrial and nuclear subunits provide another perspective of how these subunits coevolve to enable aerobic respiration and energy production, namely mito-nuclear coevolution (Wolff et al., 2014). Mito-nuclear coevolution posits that sequence evolution in one genome causes selection in the other genome for the complementary changes. Three types of mito-nuclear coevolution were proposed: protein-protein interactions within OXPHOS complexes; interactions

between nuclear protein and mitochondrial rRNA/tRNA; interactions between nuclear encoded mtDNA replication/transcription factors and their binding target—mtDNA (Bar-Yaacov et al., 2012). Different mito-nuclear coevolving hypotheses have been proposed to explain the observation that nuclear proteins presented a significantly higher evolutionary rate (estimated by the ratio of non-synonymous to synonymous substitutions rate, dN/dS) compared to the mitochondrial OXPHOS subunits in spite of the high mutation rate in mtDNA (Havird et al., 2015, 2017; Havird & Sloan, 2016; Nabholz et al., 2013; Popadin et al., 2013). Some suggested that highly transcribed mtOXPHOS genes are under strong evolutionary constraints because deleterious effects derived from protein misfolding in highly expressed genes would be more pronounced irrespective of biological functions (Nabholz et al., 2013; Zhang & Yang, 2015). Nevertheless, this hypothesis is challenged by the fact that transcript abundance is not necessarily correlated with the quantity of the protein products. Alternatively, some studies proposed that the nuOXPHOS genes often encoded “peripheral” subunits and might be less functionally important, therefore are subjected to relaxed selective constraints (Popadin et al., 2013; Zhang & Broughton, 2013). However, in some taxa, the elevated sequence evolutionary rate in nuOXPHOS was proposed as a result of positive selection instead of the peripheral role (Havird et al., 2015, 2017; Sloan et al., 2014). Moreover, structure remodeling suggested that the overrepresentation of nuOXPHOS substitutes tend to occur at the positions interacting with mtOXPHOS (Havird et al., 2015). These results are in line with the “nuclear compensation hypothesis”, which states that nuOXPHOS subunits undergo positive selection compensating for the mildly deleterious mutations accumulated in mitochondrial genomes. However, clear nuclear compensation signal is not detected in bivalves, and mtOXPHOS subunits presented similar dN/dS compared to nuOXPHOS subunits

(Piccinini et al., 2021). In particular, this hypothesis could be problematic in DUI species because the same set of nuclear OXPHOS subunits need to interact with two divergent sets of mitochondrial subunits to ensure proper OXPHOS function. If a mutation occurs in one of the two different types of mt genomes, the compensatory mutation in nuOXPHOS gene may disrupt the OXPHOS complex assembly with the other type of mitochondrial subunits. Although several possible solutions for this challenge were proposed (such as in the presence of sex-specific expressed nuOXPHOS paralogs that interact with both type of mtOXPHOS subunits), none of these solutions showed the clues of nuclear compensation (Maeda et al., 2021; Piccinini et al., 2021).

1.3 Regulatory co-expression network

Thanks to the rapid development of new sequencing technologies, it has become possible to investigate large scale transcriptomic data to address specific biological questions. The co-expression network analysis has become a popular tool in recent years due to its capacity to integrate large transcriptional datasets across multiple conditions (Harrison et al., 2021; Rago et al., 2020; Saha et al., 2017; Serin et al., 2016; Shahan et al., 2018). Co-expression network represents an ensemble of genes (nodes) and interactions by links (edges) between pairwise genes. Generally, co-expression network construction would firstly calculate expression similarity or relatedness score (correlation coefficient) between possible pairwise genes, and genes above some threshold would be kept for the network construction. Optionally, modules of highly connected genes can be extracted and associated with phenotypic traits. As a general rule, genes presenting similar functions or involved in the same regulatory pathways tend to form a cluster or module (Shyamsundar et al., 2005; Stuart et al., 2003), which was also evidenced by many cases across a wide range of animals (Harrison et al., 2021;

Rago et al., 2020; Saha et al., 2017; Serin et al., 2016; Shahan et al., 2018). From this perspective, co-expression networks are powerful in finding novel genes that may be involved in the biological processes of interest, and in suggesting the possible biological functions a gene may have. In the network topology, indexes such as connectivity (number of links in the networks), node degree (number of connected nodes), betweenness of the nodes (sum of shortest paths connecting to all the other genes in the network) are commonly used to infer the node ranking (Pavlopoulos et al., 2011). Nodes with higher connectivity, node degree, or betweenness are usually defined as main "hub" genes, which are likely to have essential roles in the network. However, the links (edges) in the correlation-based co-expression network are undirected and no causality can be inferred from the connected genes. Therefore, one commonly used approach to find the regulatory relationships in the network is to focus on the "hub" genes with known gene-gene interactions or binding capacity to the cis-regulatory elements (usually transcription factors) (Serin et al., 2016). While for the general function of genes in the module, enrichment analysis for genes in the same cluster or module is widely used; in particular, it can help to elucidate the potential regulatory genes or biological processes associated with the external phenotypic traits.

Recently, co-expression networks have been widely used to cope with the transcriptional changes across time series or across tissue/cell types, in order to provide insights into biological processes in response to development or environmental simulations (Harrison et al., 2015b; Rago et al., 2020; Saha et al., 2017; Shahan et al., 2018; Song et al., 2021). In bivalves, for example, weighted gene co-expression network (WGCNA) has been used to investigate the cellular process involved in euryhaline adaptation (Zhao et al., 2016), and the role of tissue-specific genes related to inhibitors of apoptosis in innate immune response (Song et al., 2021). In addition to

the comparative sequence information, constructing co-expression network is also a promising approach to understand the relationship between gene regulatory function and evolution (Barua & Mikheyev, 2021; Casasa et al., 2021; Mack et al., 2019; Masalia et al., 2017; Serin et al., 2016; Warner et al., 2019). For example, comparative co-expression analysis in snakes revealed that oral venom systems originated from an ancient and conserved gene regulatory network across amniotes (Barua & Mikheyev 2021). With the advent of the large amount of available sequencing data every year, I anticipate that regulatory co-expression networks would advance our understanding of molecular mechanisms underlying important biological processes, provide insights into the link between genotype and phenotype under gene regulation, and offer a new perspective into evolutionary studies regarding conservation and divergence of regulation genes.

1.4 The role of gene expression and alternative splicing in development and evolution

Regulation of gene activities is essential to the development of multicellular organisms and it makes different cell-types unique in structure and function despite sharing the same genome. Cis- and trans-regulation, as well as alternative splicing are main forms of heritable regulation that can increase proteomic diversity, tissue specificity and potentially phenotypic diversity. Gene expression in multicellular organisms is a tightly regulated process that allows cells to switch on or off the expression of a subset of genes to mediate organ differentiation during development through cell-cell communication. High-throughput sequencing has advanced our understanding between gene expression dynamics and developmental functions (Assou et al., 2011; Peter, 2017; Zeitlinger & Stark, 2010). Alternative splicing can regulate

gene expression through nonsense-mediated decay, an mRNA surveillance pathway to eliminate mRNA transcripts containing premature stop codons. Moreover, it can alter proteome diversity by removing part of coding regions that contain interaction or localization domains (García-Moreno & Romão, 2020; Lee & Rio, 2015; Nilsen & Graveley, 2010). Several types of alternative splicing have been discovered, including exon skipping, alternative 5'/3' sites, intron retention, etc. (Figure 1.2). Now it is well established that alternative splicing contributes to cell differentiation, tissue specificity, organ development, and response to external simulations across a wide range of taxa (Baralle & Giudice, 2017; Bush et al., 2017; Ergun et al., 2013; Li et al., 2016; Rotival et al., 2019; Tian et al., 2020).

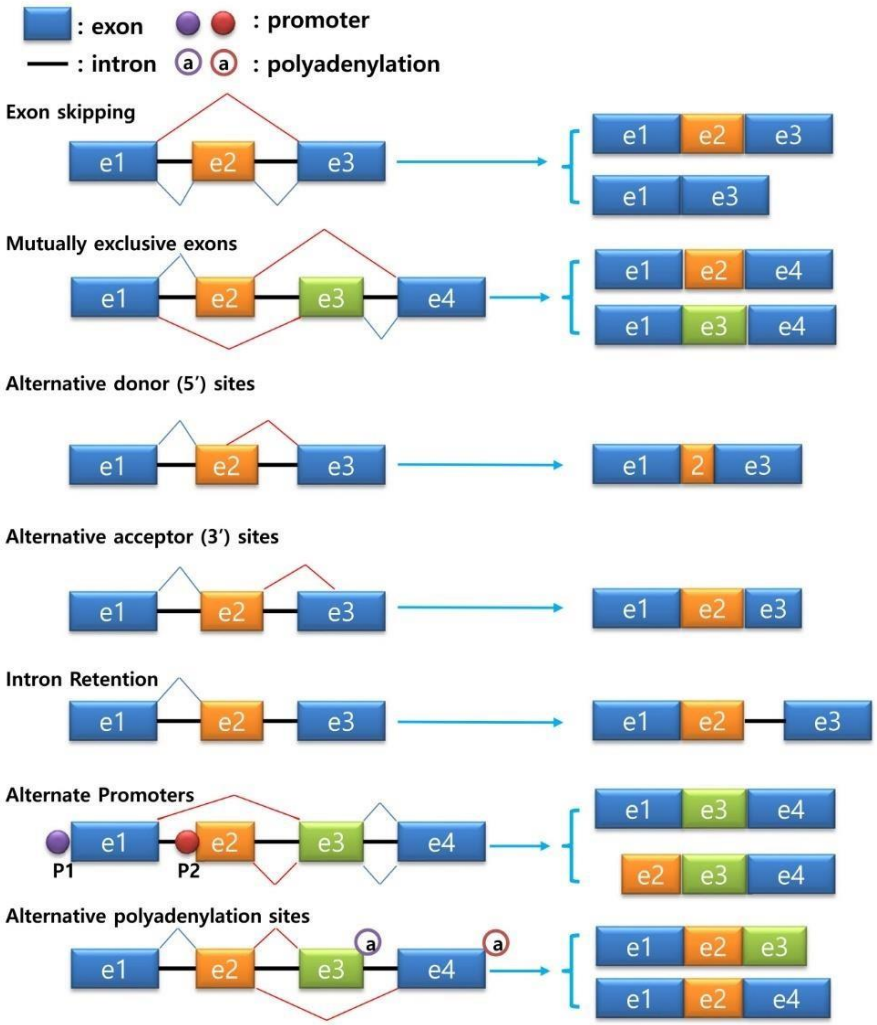


Figure 1.2 The commonly discovered modes of alternative splicing. The structure on the left of the arrows represents the pre-mRNA, and the structure on the right of the arrows represents the mature mRNA after alternative splicing process. Exon skipping: alternative splicing completely skips an exon; Mutually exclusive exons: alternative splicing removes one of two exons, but not both; Alternative donor or acceptor sites: alternative splicing starts from a different 5' and 3' splice junction sites; Intron retention: alternative splicing includes an intron mRNA; Alternate promoters: different promoters are used for different transcripts; Alternative polyadenylation sites: recognition of polyadenylation sites at 3' end caused different spliced transcripts (Kim et al., 2018).

On the other hand, changes in gene expression and splicing patterns contribute to adaptive evolution in many complex traits (Bush et al., 2017; Carroll, 2008; Lee & Rio, 2015; Romero et al., 2012). Comparative genomic studies among multiple organs from different mammals revealed multiple conserved organ-specific co-expression modules across taxa, and also many organ- and lineage-specific gene expression modules, consistent with the role of gene expression in phenotypic differences in mammals (Brawand et al., 2011). A further study suggested that the gene expression patterns across organs from different mammals corresponds to the developmental stages, and that gene expression breadth and purifying selection decrease, while the amount of positive selection and the expression of new genes increase during development (Cardoso-Moreira et al., 2019). In Drosophila, pervasive adaptive evolution of gene expression was observed, particular for genes involved in regulation, sensory perception, sexual behavior and morphology, and sex-specific adaption of gene expression was predominant in males (Nourmohammad et al., 2017). Additionally, extensive sex-biased gene expression due to sexual selection was thought to be the main drive of sexual dimorphism (Harrison et al., 2015a). Marked differences in the occurrence of alternative splicing across lineages have been observed, and the

proportion of alternative splicing genes is consistent with the evolutionary diversification of cell types (Bush et al., 2017; Chen et al., 2014; Lee & Rio, 2015). For example, the current Ensembl annotation (not including RNA-seq information) revealed that the percentage of genes under alternative splicing differs a lot across species: 25% in nematodes, 45% in fruit fly, 63% in mice and 88% in humans (Lee & Rio, 2015). Unlike the evolution of gene expression, alternative splicing patterns are more strongly resemble to the identity of species instead of organ, indicating that alternative splicing diverges faster than gene expression. It was also indicated that alternative splicing events can also occur in highly specific cell types and developmental stages, which could probably facilitate rapid adaptation (Barbosa-

Morais et al., 2012; Reyes et al., 2013). In addition, several studies suggested that gene expression and alternative splicing play complementary roles (acting on different subset of genes) or redundant roles (acting on the same subset of genes) contributing to the local adaptation and rapid phenotypic evolution (Healy & Schulte, 2019; Jacobs & Elmer, 2021; Singh et al., 2017).

References

Assou S, Boumela I, Haouzi D, Anahory T, Dechaud H, De Vos J, et al. 2011. Dynamic changes in gene expression during human early embryo development: from fundamental aspects to clinical applications. *Human Reproduction Update*, **17**: 272–290. DOI: 10.1093/humupd/dmq036

Baralle FE, Giudice J. 2017. Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology*, **18**: 437–451. DOI: 10.1038/nrm.2017.27

Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science*, **338**: 1587–1593. DOI: 10.1126/science.1230612

Barshad G, Marom S, Cohen T, Mishmar D. 2018. Mitochondrial DNA Transcription and Its Regulation: An Evolutionary Perspective. *Trends in Genetics*, **34**: 682–692. DOI: 10.1016/j.tig.2018.05.009

Barua A, Mikheyev AS. 2021. An ancient, conserved gene regulatory network led to the rise of oral venom systems. *Proceedings of the National Academy of Sciences*, **118** DOI: 10.1073/PNAS.2021311118

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature*, **478**: 343–348. DOI: 10.1038/nature10532

Breton S, Beaupré HD, Stewart DT, Hoeh WR, Blier PU. 2007. The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends in Genetics*, **23**: 465–474. DOI: 10.1016/j.tig.2007.05.011

Breton S, Beaupré HD, Stewart DT, Piontkivska H, Karmakar M, Bogan AE, et al. 2009. Comparative Mitochondrial Genomics of Freshwater Mussels (Bivalvia: Unionoida) With Doubly Uniparental Inheritance of mtDNA: Gender-Specific Open Reading Frames and Putative Origins of Replication. *Genetics*, **183**: 1575–1589. DOI: 10.1534/genetics.109.110700

Breton S, Bouvet K, Auclair G, Ghazal S, Sietman BE, Johnson N, et al. 2017. The extremely divergent maternally-and paternally-transmitted mitochondrial genomes are co-expressed in somatic tissues of two freshwater mussel species with

doubly uniparental inheritance of mtDNA. *PLoS ONE*, **12** DOI: 10.1371/journal.pone.0183529

Breton S, Stewart DT, Shepardson S, Trdan RJ, Bogan AE, Chapman EG, et al. 2011. Novel Protein Genes in Animal mtDNA: A New Sex Determination System in Freshwater Mussels (Bivalvia: Unionoida)? *Molecular Biology and Evolution*, **28**: 1645–1659. DOI: 10.1093/molbev/msq345

Bush SJ, Chen L, Tovar-Corona JM, Urrutia AO. 2017. Alternative splicing and the evolution of phenotypic novelty. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**: 20150474. DOI: 10.1098/rstb.2015.0474

Cao L, Kenchington E, Zouros E. 2004. Differential Segregation Patterns of Sperm Mitochondria in Embryos of the Blue Mussel (*Mytilus edulis*). *Genetics*, **166**: 883–894. DOI: 10.1093/genetics/166.2.883

Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, et al. 2019. Gene expression across mammalian organ development. *Nature*, **571**: 505–509. DOI: 10.1038/s41586-019-1338-5

Carroll SB. 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell*, **134**: 25–36. DOI: 10.1016/j.cell.2008.06.030

Casasa S, Biddle JF, Koutsovoulos GD, Ragsdale EJ. 2021. Polyphenism of a Novel Trait Integrated Rapidly Evolving Genes into Ancestrally Plastic Networks. *Molecular Biology and Evolution*, **38**: 331–343. DOI: 10.1093/molbev/msaa235

Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. 2014. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Molecular Biology and Evolution*, **31**: 1402–1413. DOI: 10.1093/molbev/msu083

Cogswell AT, Kenchington EL., Zouros E. 2006. Segregation of sperm mitochondria in two- and four-cell embryos of the blue mussel *Mytilus edulis*: implications for the mechanism of doubly uniparental inheritance of mitochondrial DNA. *Genome*, **49**: 799–807. DOI: 10.1139/g06-036

Couvillion MT, Soto IC, Shipkovenska G, Churchman LS. 2016. Synchronized mitochondrial and cytosolic translation programs. *Nature*, **533**: 499–503. DOI: 10.1038/nature18015

Ergun A, Doran G, Costello JC, Paik HH, Collins JJ, Mathis D, et al. 2013. Differential splicing across immune system lineages. *Proceedings of the National Academy of Sciences*, **110**: 14324–14329. DOI: 10.1073/pnas.1311839110

García-Moreno JF, Romão L. 2020. Perspective in Alternative Splicing Coupled to Nonsense-Mediated mRNA Decay. *International Journal of Molecular Sciences*, **21**: 9424. DOI: 10.3390/ijms21249424

Ghiselli F, Maurizii MG, Reunov A, Ariño-Bassols H, Cifaldi C, Pecci A, et al. 2019. Natural Heteroplasmy and Mitochondrial Inheritance in Bivalve Molluscs. Integrative and comparative biology, 59: 1016-1032. DOI: 10.1093/icb/icz061

Ghiselli F, Milani L, Passamonti M. 2011. Strict sex-specific mtDNA segregation in the germ line of the dui species venerupis philippinarum (Bivalvia: Veneridae). *Molecular Biology and Evolution*, **28**: 949–961. DOI: 10.1093/molbev/msq271

Guerra D, Plazzi F, Stewart DT, Bogan AE, Hoeh WR, Breton S. 2017. Evolution of sex-dependent mtDNA transmission in freshwater mussels (Bivalvia: Unionida). *Scientific Reports*, **7**: 1551. DOI: 10.1038/s41598-017-01708-1

Gusman A, Lecomte S, Stewart DT, Passamonti M, Breton S. 2016. Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtdna. *PeerJ*, **2016** DOI: 10.7717/peerj.2760

Gyllensten U, Wharton D, Josefsson A, Wilson AC. 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature*, **352**: 255–257. DOI: 10.1038/352255a0

Harrison MC, Jaimes LM, No N~, Rodrigues MA, Ryll J, Flatt T, et al. 2021. Gene Co-expression Network Reveals Highly Conserved, Well-Regulated Anti-Ageing Mechanisms in Old Ant Queens. *Genome Biology and Evolution*, **13**: evab093. DOI: 10.1093/gbe/evab093

Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, et al. 2015a. Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **112**: 4393–4398. DOI: 10.1073/pnas.1501339112

Havird JC, Sloan DB. 2016. The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Molecular Biology and Evolution*, **33**: 3042–3053. DOI: 10.1093/molbev/msw185

Havird JC, Trapp P, Miller CM, Bazos I, Sloan DB. 2017. Causes and consequences of rapidly evolving mtDNA in a plant lineage. *Genome Biology and Evolution*, **9**: 323–336. DOI: 10.1093/gbe/evx010

Havird JC, Whitehill NS, Snow CD, Sloan DB. 2015. Conservative and compensatory evolution in oxidative phosphorylation complexes of angiosperms with highly divergent rates of mitochondrial genome evolution. *Evolution*, **69**: 3069–3081. DOI: 10.1111/evo.12808

Healy TM, Schulte PM. 2019. Patterns of alternative splicing in response to cold acclimation in fish. *Journal of Experimental Biology*, **22**2 (5): jeb193516. DOI: 10.1242/jeb.193516

Hill GE. 2020. Mito-nuclear Compensatory Coevolution. *Trends in Genetics*, **36**: 403–414. DOI: 10.1016/j.tig.2020.03.002

Iannello M, Bettinazzi S, Breton S, Ghiselli F, Milani L. 2021. A Naturally Heteroplasmic Clam Provides Clues about the Effects of Genetic Bottleneck on Paternal mtDNA. *Genome Biology and Evolution*, **13**: 1–15. DOI: 10.1093/gbe/evab022

Isaac RS, Mcshane E, Churchman LS. 2018. The Multiple Levels of Mito-nuclear Coregulation. *Annual review of genetics*, **22**: 511–533. DOI: 10.1146/annurev-genet-120417

Jacobs A, Elmer KR. 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Molecular Ecology*, DOI: 10.1111/mec.15817

Kenchington EL, Hamilton L, Cogswell A, Zouros E. 2009. Paternal mtDNA and Maleness Are Co-Inherited but Not Causally Linked in Mytilid Mussels. *PLoS ONE*, **4**: e6976. DOI: 10.1371/journal.pone.0006976

Kim HK, Pham MHC, Ko KS, Rhee BD, Han J. 2018. Alternative splicing isoforms in health and disease. *Pflügers Archiv - European Journal of Physiology*, **470**: 995–1016. DOI: 10.1007/s00424-018-2136-x

Kyriakou E, Zouros E, Rodakis GC. 2010. The atypical presence of the paternal mitochondrial DNA in somatic tissues of male and female individuals of the blue mussel species Mytilus galloprovincialis. *BMC Research Notes*, **3**: 222. DOI: 10.1186/1756-0500-3-222

Lee Y, Rio DC. 2015. Mechanisms and regulation of alternative Pre-mRNA splicing. *Annual Review of Biochemistry*, **84**: 291–323. DOI: 10.1146/annurev-biochem-060614-034316

Li S, Yamada M, Han X, Ohler U, Benfey PN. 2016. High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. *Developmental Cell*, **39**: 508–522. DOI: 10.1016/j.devcel.2016.10.012

Mack KL, Phifer-Rixey M, Harr B, Nachman MW. 2019. Gene expression networks across multiple tissues are associated with rates of molecular evolution in wild house mice. *Genes*, **10** DOI: 10.3390/genes10030225

Maeda GP, Iannello M, McConie HJ, Ghiselli F, Havird JC. 2021. Relaxed selection on male mitochondrial genes in DUI bivalves eases the need for mito-nuclear coevolution. *Journal of Evolutionary Biology*, **34**: 1722–1736. DOI: 10.1111/jeb.13931

Masalia RR, Bewick AJ, Burke JM. 2017. Connectivity in gene co-expression networks negatively correlates with rates of molecular evolution in flowering plants. *PLoS ONE*, **12**: e0182289. DOI: 10.1371/journal.pone.0182289

Mastrantonio V, Urbanelli S, Porretta D. 2019. Ancient hybridization and mtDNA introgression behind current paternal leakage and heteroplasmy in hybrid zones. *Scientific Reports*, **9**: 19177. DOI: 10.1038/s41598-019-55764-w

Milani L, Ghiselli F, Guerra D, Breton S, Passamonti M. 2013. A Comparative Analysis of Mitochondrial ORFans: New Clues on Their Origin and Role in Species with Doubly Uniparental Inheritance of Mitochondria. *Genome Biology and Evolution*, **5**: 1408–1434. DOI: 10.1093/gbe/evt101

Milani L, Ghiselli F, Passamonti M. 2012. Sex-Linked Mitochondrial Behavior During Early Embryo Development in Ruditapes philippinarum (Bivalvia Veneridae) a Species with the Doubly Uniparental Inheritance (DUI) of Mitochondria. *Journal*

of Experimental Zoology Part B: Molecular and Developmental Evolution, **318**: 182–189. DOI: 10.1002/jez.b.22004

Milani L, Ghiselli F, Passamonti M. 2016. Mitochondrial selfish elements and the evolution of biological novelties. *Current Zoology*, **62**: 687–697. DOI: 10.1093/cz/zow044

Mitchell A, Guerra D, Stewart D, Breton S. 2016. In silico analyses of mitochondrial ORFans in freshwater mussels (Bivalvia: Unionoida) provide a framework for future studies of their origin and function. *BMC Genomics*, **17**(1): 1-22. DOI: 10.1186/s12864-016-2986-6

Morgan JAT, Macbeth M, Broderick D, Whatmore P, Street R, Welch DJ, et al. 2013. Hybridisation, paternal leakage and mitochondrial DNA linearization in three anomalous fish (Scombridae). *Mitochondrion*, **13**: 852–861. DOI: 10.1016/j.mito.2013.06.002

Nabholz B, Ellegren H, Wolf JBW. 2013. High levels of gene expression explain the strong evolutionary constraint of mitochondrial protein-coding genes. *Molecular Biology and Evolution*, **30**: 272–284. DOI: 10.1093/molbev/mss238

Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature*, **463**: 457–463. DOI: 10.1038/nature08909

Nourmohammad A, Rambeau J, Held T, Kovacova V, Berg J, Lässig M. 2017. Adaptive Evolution of Gene Expression in Drosophila. *Cell Reports*, **20**: 1385–1395. DOI: 10.1016/j.celrep.2017.07.033

Nunes MDS, Dolezal M, Schlötterer C. 2013. Extensive paternal mtDNA leakage in natural populations of *Drosophila melanogaster*. *Molecular Ecology*, **22**: 2106–2117. DOI: 10.1111/mec.12256

Obata M, Kamiya C, Kawamura K, Komaru A. 2006. Sperm mitochondrial DNA transmission to both male and female offspring in the blue mussel Mytilus galloprovincialis. *Development, Growth and Differentiation*, **48**: 253–261. DOI: 10.1111/j.1440-169X.2006.00863.x

Obata M, Komaru A. 2005. Specific location of sperm mitochondria in mussel Mytilus galloprovincialis zygotes stained by MitoTracker. *Development, Growth and Differentiation*, **47**: 255–263. DOI: 10.1111/j.1440-169X.2005.00801.x

Pante E, Poitrimol C, Saunier A, Becquet V, Garcia P. 2017. Putative sex-linked heteroplasmy in the tellinid bivalve Limecola balthica (Linnaeus, 1758). *Journal of Molluscan Studies*, **83**: 226–228. DOI: 10.1093/mollus/eyw038

Passamonti M, Ghiselli F. 2009. Doubly Uniparental Inheritance: Two Mitochondrial Genomes, One Precious Model for Organelle DNA Inheritance and Evolution. *DNA and Cell Biology*, **28**: 79–89.

Pavlopoulos GA, Secrier M, Moschopoulos CN, Soldatos TG, Kossida S, Aerts J, et al. 2011. Using graph theory to analyze biological networks. *BioData Mining*, **4**: 10. DOI: 10.1186/1756-0381-4-10

Peter IS. 2017. Regulatory states in the developmental control of gene expression. *Briefings in Functional Genomics*, **16**: 281–287. DOI: 10.1093/bfgp/elx009

Piccinini G, Iannello M, Puccio G, Plazzi F, Havird JC, Ghiselli F. 2021. Mito-nuclear Coevolution, but not Nuclear Compensation, Drives Evolution of OXPHOS Complexes in Bivalves. Chang B (ed). *Molecular Biology and Evolution*, DOI: 10.1093/molbev/msab054

Popadin KY, Nikolaev SI, Junier T, Baranova M, Antonarakis SE. 2013. Purifying Selection in Mammalian Mitochondrial Protein-Coding Genes Is Highly Effective and Congruent with Evolution of Nuclear Genes. *Molecular Biology and Evolution*, **30**: 347–355. DOI: 10.1093/molbev/mss219

Quirós PM, Mottis A, Auwerx J. 2016. Mito-nuclear communication in homeostasis and stress. *Nature Reviews Molecular Cell Biology*, **17**: 213–226. DOI: 10.1038/nrm.2016.23

Rago A, Werren JH, Colbourne JK. 2020. Sex biased expression and co-expression networks in development, using the hymenopteran Nasonia vitripennis. *PLoS Genetics*, **16**: e1008518. DOI: 10.1371/journal.pgen.1008518

Reyes A, Anders S, Weatheritt RJ, Gibson TJ, Steinmetz LM, Huber W, et al. 2013. Drift and conservation of differential exon usage across tissues in primate species. *Proceedings of the National Academy of Sciences*, **110**: 15377-15382. DOI: 10.1073/pnas.1307202110

Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, **13**: 505–516. DOI: 10.1038/nrg3229

Rotival M, Quach H, Quintana-Murci L. 2019. Defining the genetic and evolutionary architecture of alternative splicing in response to infection. *Nature Communications*, **10**: 1671. DOI: 10.1038/s41467-019-09689-7

Saha A, Kim Y, Gewirtz ADH, Jo B, Gao C, McDowell IC, et al. 2017. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Research*, **27**: 1843–1858. DOI: 10.1101/gr.216721.116

Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. 2016. Learning from coexpression networks: Possibilities and challenges. *Frontiers in Plant Science*, **7** DOI: 10.3389/fpls.2016.00444

Shahan R, Zawora C, Wight H, Sittmann J, Wang W, Mount SM, et al. 2018. Consensus coexpression network analysis identifies key regulators of flower and fruit development in wild strawberry. *Plant Physiology*, **178**: 202–216. DOI: 10.1104/pp.18.00086

Shyamsundar R, Kim YH, Higgins JP, Montgomery K, Jorden M, Sethuraman A, et al. 2005. Open Access A DNA microarray survey of gene expression in normal human tissues

Singh P, Börger C, More H, Sturmbauer C. 2017. The role of alternative splicing and differential gene expression in cichlid adaptive radiation. *Genome Biology and Evolution*, **9**: 2764–2781. DOI: 10.1093/gbe/evx204

Sloan DB, Triant DA, Wu M, Taylor DR. 2014. Cytonuclear Interactions and Relaxed Selection Accelerate Sequence Evolution in Organelle Ribosomes. *Molecular Biology and Evolution*, **31**: 673–682. DOI: 10.1093/molbev/mst259

Song H, Guo X, Sun L, Wang Q, Han F, Wang H, et al. 2021. The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in Bivalvia. *BMC Biology*, **19**(1): 1-20. DOI: 10.1186/s12915-020-00943-9

Soto I, Couvillion M, Mcshane E, Hansen KG, Moran JC, Barrientos A, et al. 2021. Balanced mitochondrial and cytosolic translatomes underlie the biogenesis of human respiratory complexes. *BioRx*, DOI: 10.1101/2021.05.31.446345

Stewart D, Breton S, Chase E, Robicheau B, Bettinazzi S, Pante E, et al. 2020. An unusual evolutionary strategy: the origins, genetic repertoire, and implications of doubly uniparental inheritance of mitochondrial DNA in bivalves. *Evolutionary Biology-A Transdisciplinary Approach*, 301-323. DOI: 10.1007/978-3-030-57246-4_12ï

Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene co-expression network for global discovery of conserved genetic modules. *Science*, **302**: 249–255. DOI: 10.1126/science.1087447

Tian GG, Li J, Wu J. 2020. Alternative splicing signatures in preimplantation embryo development. *Cell and Bioscience*, **10**: 1-10. DOI: 10.1186/s13578-020-00399-y

Warner MR, Qiu L, Holmes MJ, Mikheyev AS, Linksvayer TA. 2019. Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. *Nature Communications*, **10**: 1-11. DOI: 10.1038/s41467-019-10546-w

van Waveren C, Moraes CT. 2008. Transcriptional co-expression and co-regulation of genes coding for components of the oxidative phosphorylation system. *BMC Genomics*, **9:** 1-15. DOI: 10.1186/1471-2164-9-18

Wolff JN, Ladoukakis ED, Enríquez JA, Dowling DK. 2014. Mito-nuclear interactions: Evolutionary consequences over multiple biological scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1646): 20130443. DOI: 10.1098/rstb.2013.0443

Zeitlinger J, Stark A. 2010. Developmental gene regulation in the era of genomics. *Developmental Biology*, **339**: 230–239. DOI: 10.1016/j.ydbio.2009.12.039

Zhang F, Broughton RE. 2013. Mitochondrial–Nuclear Interactions: Compensatory Evolution or Variable Functional Constraint among Vertebrate Oxidative Phosphorylation Genes? *Genome Biology and Evolution*, **5**: 1781–1791. DOI: 10.1093/gbe/evt129

Zhang J, Yang J-R. 2015. Determinants of the rate of protein sequence evolution. *Nature Reviews Genetics*, **16**: 409–420. DOI: 10.1038/nrg3950

Zhao X, Yu H, Kong L, Li Q. 2016. Gene Co-Expression Network Analysis Reveals the Correlation Patterns Among Genes in Euryhaline Adaptation of Crassostrea gigas. *Marine Biotechnology*, **18**: 535–544. DOI: 10.1007/s10126-016-9715-7

Zouros E. 2013. Biparental Inheritance Through Uniparental Transmission: The Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA. *Evolutionary Biology*, **40**: 1–31. DOI: 10.1007/s11692-012-9195-2

Chapter 2 Long-read-based genome assembly and multi-tissue RNA-Seq analysis reveal complex gene regulation and molecular evolution in the Manila clam

Abstract

Bivalves are characterized by highly variable genome sizes and extremely high levels of heterozygosity, which obstacle the complete and accurate genome assemblies. In this work, we present the first long-read-based *de novo* genome assembly of a wild male Manila clam (*Ruditapes philippinarum*) from North America. Relying on both short and long reads data we found that the estimated heterozygosity (3.7% to 4%) in *R. philippinarum* is, to our knowledge, the highest observed so far in the entire Mollusca phylum. Additionally, taking advantage of large RNA-seq dataset from different tissues of both sexes, we were able to investigate: 1) the role of gene expression and alternative splicing across tissues; 2) the relationship across tissue specificity, regulatory network connectivity, and sequence evolution; 3) sexual contrasting genetic markers potentially associated with sexual differentiation and/or sex-specific functions. We found that gene expression and alternative splicing follow similar patterns across tissues, with alternative splicing playing a central role in gonad distinctiveness. To our knowledge, this is the first study exploring the role of differential splicing in bivalves. We constructed a tissue specific co-expression network and we found that gonads present a higher number of co-expression modules, indicating high complexity in gene regulation in such tissue. Differentially expressed and spliced genes are largely overlapping in pairwise comparisons between tissues and between sexes, and these genes were in the central position of co-expression networks, presenting high tissue specificity and evolutionary rate in sex-related modules, consistent with relaxed functional constraints.

Tissue specificity and the rate of sequence evolution followed similar trends across different modules. Despite that, genes in gonad-associated modules presented relatively high tissue specificity compared to the genes in somatic-associated modules. A male gonad-associated module showed extremely low tissue specificity and sequence evolution, but high intermodular connectivity, indicating the possible role of functional pleiotropy in this subset of genes. Intriguingly, we found that sexual contrasting SNPs are located in the genes overrepresented in mitochondrial related functions, further supporting the potential association between mitochondria and sexual differentiation in bivalves.

Keywords: *Ruditapes philippinarum;* genome assembly; PacBio long reads; co-expression; tissue specificity; gene evolution; alternative splicing.

2.1 Introduction

The Mollusca phylum contains over 93 000 described species, making it the second most species-rich phylum and representing an incredible wealth of diverse life histories, adaptation, and phenotypic plasticity. These attributes have led to numerous species becoming important biological models for monitoring pollution, adaptation to climate change, and the development of biomedical tools (Ahmad et al., 2018; Harris et al., 2020; Krishnakumar et al., 2018). Many molluscan species, namely bivalves, also provide an essential source of protein through aquaculture and fishing, leading to the development of an important aquaculture industry of global economic importance (Figueras et al., 2019; Haszprunar & Wanninger, 2012; Mun et al., 2017). Unfortunately, the genomic resources available for mollusc species (comprising around 7% of animal species) remain very scarce in comparison to other metazoan lineages, counting only 67 partial or complete genome assemblies as of 2021. By contrast, chordates, for

example, comprising just 3.9% of animal species, enriched for more than 50% (1770) available genome assemblies (Hotaling et al., 2021).

The main aspects of bivalve genomes that have hindered the advancement of complete and accurate genome assemblies are their highly variable (and often large) genome sizes and the extremely high levels of heterozygosity observed in these species. This is particularly true in species such as the Manila clam, *R. philippinarum*, whose history of introduction and invasion on a global scale have likely been facilitated by high levels of genetic polymorphism. Two genome assemblies for the Manila clam were recently published from Korean and Chinese laboratories, highlighting genes associated with shell color, the expansion of immune- and stress-related gene families, and providing a basis for future advanced studies in selective breeding and disease resistance applicable to aquaculture (Mun et al., 2017; Yan et al., 2019). While these genome assemblies offer novel resources for a wide range of genetic studies, the Asian populations from which the sequenced specimens were sampled are also the most historically and genetically distant to European and North American populations. Recently, a study using mitochondrial DNA and microsatellite markers assessed genetic differentiation between European, North American, Japanese, and Chinese populations of Manila clam, tracing the genetic heritage of European populations to the 1970s and 80's introductions using North American stock (Cordero et al., 2017).

Another trait of particular interest in this species is the presence of the doubly uniparental inheritance (DUI) of mitochondrial (mt) DNA. DUI has been so far reported exclusively in bivalve molluscs (> 100 species to date), and it is the process through which two distinct lineages of mitochondrial DNA are transferred to offspring: one lineage (F-type) which is transmitted through eggs, and the other (M-type) transmitted through sperm. Numerous authors have sought to elucidate the exact mechanisms of

DUI and speculated on the reasons for maintaining divergent mtDNA lineages in bivalve species, generally converging on the hypothesis that it may be implicated in the evolution of sex, act as a sex-determining factor, and be involved in sexual differentiation (Breton et al., 2018; Capt et al., 2018). Moreover, the Manila clam does not show heteromorphic sex chromosomes, sexual dimorphism, mating behavior, and presents a low number of sex-biased genes in this species (Ghiselli et al., 2018). However, the forces and constraints governing the evolution of sex-biased genes remain largely unknown in bivalves. Unusually rapid sequence evolution of sex-biased genes was observed across the animal world (Dean & Mank, 2016; Ellegren & Parsch, 2007; Grath & Parsch, 2016; Harrison et al., 2015; Lipinska et al., 2015; Mank et al., 2008; Papa et al., 2017; Whittle & Extavour, 2019; Whittle & Johannesson, 2013; Yang et al., 2016, but see Ghiselli et al. 2018). Additionally, several studies revealed that a large proportion of genes were under sex-specific splicing, indicating a possible role of differential splicing in sex-specific development and physiology (Griffin et al., 2013; Rogers et al., 2021; Telonis-Scott et al., 2009). In the species with sexual dimorphism, it was suggested that sexual selection may drive such sex-biased pattern of gene expression and splicing, while gene expression breadth, protein-protein interaction, codon usage and pleiotropy may also contribute to the sex biases (Grath & Parsch, 2016; Harrison et al., 2015; Mank et al., 2007, 2008; Rogers et al., 2021; Whittle & Extavour, 2019; Yang et al., 2016). However, if and how these factors influence the species without sexual dimorphism is largely unexplored.

In this work, we present the first long-read-based *de novo* genome assembly and annotation for a wild, male, North - East American Manila clam sample, and we compared it to the previously published chromosome-level assembly of a Chinese strain sequenced by Yan et al., (2019). Taking advantage of the large RNA-seq dataset in this

study, we analyzed gene expression and alternative splicing patterns across tissues and between sexes. We also built a tissue-associated co-expression network, and we investigated the relationship among tissue specificity, network connectivity, and sequence evolution in the tissue-associated modules. We also performed a comparative SNP analysis using transcriptomic data from male and female clams. The main objectives of this study are 1) to provide a reference genome for North American and Adriatic clam populations, 2) to investigate the role of differential gene expression and differential splicing across tissues, 3) to shed light on the evolution of tissue-specific genes (especially gonad-associated), and 4) to highlight genetic markers potentially associated with sex differentiation.

2.2 Materials & methods

2.2.1 Sample collection and library preparation

2.2.1.1 Genome

Genomic DNA (gDNA) was extracted from a single male individual from the Puget Sound region (Pacific Northwest, USA) using only mantle tissue. The individual was opened, sexed, frozen in liquid nitrogen, and stored at -80ºC. Multiple DNA extractions were performed to obtain the amount of material required for both PacBio and Illumina libraries. The genomic DNA was quantified and its quality was assessed using agarose gel electrophoresis, Nanodrop, and Bioanalyzer; before proceeding, the DNA had to meet the stringent PacBio requirements (i.e.: gDNA size >45 Kb, dsDNA, OD260/OD280 ratio of 1.8 to 2.0, OD260/OD230 ratio of ~2.0, does not contain insoluble material, does not contain RNA contamination, does not contain carryover contamination from the original organism/tissue such as polyphenols and

polysaccharides). The PacBio library was prepared using a SMRTbell template preparation kit, and a 10-50Kb size selection was performed using a BluePippin System. Two types of Illumina libraries were prepared: a “small insert” library (insert size ~500 bp), and a “long insert” library (insert size ~1,500 bp). To minimize batch effects and library preparation biases, we prepared multiple replicates for each library: 9 replicates for the small insert library, and 10 replicates for the large insert library. Replicates were indexed and pooled, and each pool was sequenced in one separated lane.

2.2.1.2 Transcriptome

R. philippinarum specimens used for RNA-Seq were collected from the Northern Adriatic Sea, in the river Po delta region (Sacca di Goro, approximate GPS coordinates: 44º50′06″N, 12º17′55″E) during the spawning season (end of July). The collected individuals were kept in the lab for 48 hours in aerated beakers containing artificial seawater—filtered reverse osmosis water with Red Sea Coral Pro aquariology sea salt (Red Sea Europe, Verneuilsur-Avre, France)—that was changed every 12 hours. Then, clams were opened, sexed by microscope inspection of gonadal tissue, flash-frozen in liquid nitrogen, and stored at −80°C until RNA extraction. Total RNA was extracted with TRIzol, poly-A transcripts were isolated with magnetic beads and used as template for cDNA synthesis following the protocol as in Mortazavi et al. (2008) with modifications as in Ghiselli et al. (2012). The selected insert size was approximately 500 bp. In total, after quality check, 90 samples were obtained from three different tissues (adductor muscle, mantle, and gonad) of 15 males and 15 females.

2.2.2 Genome sequencing and assembly

2.2.2.1 Genome sequencing

The short read libraries were sequenced on an Illumina HiSeq 2500 platform with 2x250 bp reads at the USC Genome Core facility, University of Southern California. The long read libraries were sequenced on a PacBio RSII using a P6-C4 chemistry at the Genomics High-Throughput Facility, University of California, Irvine.

2.2.2.2 Short read trimming and kmer-based genome survey

All Pair-end (PE) libraries were pre-processed with Trimmomatic v0.39 (Bolger et al., 2014) with parameters LEADING:3; TRAILING:3; SLINDINGWINDOW:4:15; MINLEN:36. Quality of filtered reads was checked with FastQC v0.11.8 and summarized with multiqc v1.9.

Genome size, heterozygosity and duplication level were estimated using all filtered short reads and K-Mer Counter v-3.1.1 (KMC; Kokot et al., 2017; parameters: kmer size = 24 - 27 - 30, maximal value of a counter = 5E-9). The resulting kmer histogram was uploaded to the online implementation of Genomescope 2 (Vurture et al., 2017). Additionally, kmercountexact.sh from the BBMap package (Bushnell, 2014) was used as a second tool for kmer-based genome size estimation (with a kmer size of 24, 27, and 30).

2.2.2.3 Assembly pipeline

One of the biggest challenges to obtain high-quality assemblies of bivalve genomes is the high heterozygosity (Sun et al., 2021). The estimated heterozygosity of *R. philippinarum* obtained from this data is, to our knowledge, the highest observed so far in the entire Mollusca phylum (see Results). To overcome this issue and produce an

assembly as haploid as possible, we have based our assembly pipeline on up-to-date software which have been proven to be cost-efficient and to perform well when dealing with high heterozygous and non-model organisms (Guiglielmoni et al., 2020). First, a raw assembly was obtained using all PacBio reads with wtdbg2 (Ruan & Li, 2020) using the following parameters: *-p* 18 *-S* 2 *-g* 1.37 *-t* 0, as suggested by the developers in case of high error rate long sequences. Then, three consecutive rounds of haplotig removal and polishing were performed. In brief, first purge_dups (Guan et al., 2020) was used to identify duplicated contigs mapping back PacBio reads to the assembly and performing a whole genome self-alignment using Minimap2 (Li, 2018). Coverage cutoffs were automatically calculated, while the minimum fraction of haploid/bad/repetitive bases in a sequence (-f parameter) and the minimum alignment score (-a parameter) were decreased from respectively the default values of 0.8 to 0.7 and from 70 to 60 to increase its sensitivity. After that, both long reads and trimmed short reads were re-mapped against the reduced assembly and the resulting bam files used for polishing with Hypo (Kundu et al., 2019; genome size = 1.37; approximate mean short reads coverage = 50). These steps were performed for three consecutive times, using BUSCO v.4 (Seppey et al., 2019) and the Metazoa odb10 core gene set as quality check. Finally, redundans (Pryszcz & Gabaldón, 2016) (default parameters), BUSCO (Seppey et al., 2019), and KAT (Mapleson et al., 2016) were used to assess the quality of the final assembly. This recursive pipeline was chosen because we have noticed that wtdbg2 produced a high contiguous and collapsed raw assembly, even if still larger than expected genome size, but with a high level of errors (See Results). This also influences the ability of purge_dups to identify duplicated contigs. Thus, the polishing runs prior to getting an almost haploid assembly were necessary to correct assembly errors and allow a correct identification and purging of haplotypes. We found

this as being the best solution for our starting raw data and goals, reaching an almost haploid genome assembly with good completeness and contig N50, and a genome size close to the estimated value (see Results).

2.2.2.4 Contaminant Filtering

To assess the presence of contaminants in the final assembly, we used Blobtools v. 2 (Laetsch & Blaxter, 2017). Each contig was blasted against the NCBI nt database with a stringent e-value of 1E-25 and annotated using the bestsum taxrule at the Phylum and species level. Only contigs annotated as Ascomycota, Bacteroidetes, Priapulida, and Zoopagomycota were systematically removed from the genome. Reads coverage and mapping statistics were calculated aligning both filtered short reads and long reads to the final version of the assembly with Minimap2 (Li, 2018). Mosdepth (Pedersen & Quinlan, 2018) was used to calculate per-base and median genome coverage excluding secondary alignments, optical duplicates, and low-quality reads (samtools flag -F 1796).

2.2.3 Whole genome alignment and structural variant detection

Our assembly was aligned to the short-read-only, chromosome-level *R. philippinarum* genome assembly by Yan et al. (2019) using the MUMmer V. 4 package (Marçais et al., 2018). We performed a first whole genome alignment (WGA) using the nucmer function with default parameters and summarizing results with the dnadiff function. A second WGA was performed with the aim to identify structural variations (SVs) between the two assemblies. For this analysis, we have adopted more stringent parameters to increase the specificity of the alignment (-l 100 -c 500) and allow the computation of all maximal matches regardless of their uniqueness (--maxmatch). The resulting delta file was then uploaded to Assemblytics (Nattestad & Schatz, 2016) for

SVs calling with a required unique sequence length of 10,000 bp, a maximum variant size of 100,000 bp and a minimum variant size of 50 bp.

2.2.4 Genome Annotation

For Transposable element (TE) annotation we used a combination of *de novo* and homology-based approaches. In brief, RepeatModeler v2.0.1 (Flynn et al., 2020) with the LTR pipeline extension and MITE Tracker (Crescente et al., 2018) were used with default options for *de novo* mining of repeats. From resulting libraries, non-TE related genes were removed using blastx (E value 1E-10) against the predicted proteomes of *Crassostrea gigas* (GCF_902806645.1), *C. virginica* (GCF_002022765.2), *Lottia gigantea* (GCF_000327385.1), and *Octopus bimaculoides* (GCF_001194135.1) followed by ProtExcluder (Campbell et al., 2014). Tandem repeats were removed with the cleanup_tandem.pl script from the EDTA pipeline (Ou et al., 2019). Then, cleaned consensus libraries were merged with Mollusca repeats present in RepBase and redundancy removed using CD-HIT (Fu et al., 2012) following the 80-80 rule (80% similarity and coverage threshold). As a last step, the repeats library obtained was backblasted against the assembly (Blastn, E value 1E-10; min query coverage 0.7; min identity 70%) and all repeats with less than 5 hits were removed to produce our final, clean set of TE consensus sequences. Genome annotation of repeats was achieved running RepeatMasker v4.1.0 (Tarailo-Graovac & Chen, 2009) in sensitive mode (*-s* parameter).

Gene annotation was carried on using Maker v3.01.03 (Cantarel et al., 2008) following a 3-step procedure: the first round was run by providing the repeat library, three previously assembled transcriptomes of *R. philippinarum*, the Swiss-Prot database, and the same molluscs proteomes used for TE annotation. In the second and third rounds

of annotation, we used gene models produced from each previous Maker run to train gene predictors. Predicted proteins at the end of the third round were annotated via Blastx v2.9.0+ (Altschul et al., 1990) against the full Swiss-Prot database and via InterProScan v 5.38-76.0 (Jones et al., 2014) with default options. Annotation was then included in the genome .gff file by using the "agat_sp_manage_functional_annotation.pl" tool (https://github.com/NBISweden/AGAT). The completeness of the annotation was assessed by running the predicted proteins against the mollusca_db10 from BUSCO v5.1.2 (Simão et al., 2015).

2.2.5 Gene Expression and Co-Expression Analysis

The PE reads were processed with Trimmomatic (Bolger et al., 2014) to remove adaptors and low quality reads with the following setting: LEADING:36 TRAILING:36 SLIDINGWINDOW:10:36 MINLEN:45. Then, clean reads were mapped to the genome assembly using STAR v2.7.7 (Dobin & Gingeras, 2015) in multiple 2-pass modes with the following settings: --outSAMattrIHstart 0 -outSAMstrandField intronMotif. FeatureCounts v2.0.2 (Liao et al., 2014) was used to count the number of reads in the genomic features. Samples with a low number of reads (<50 thousands) were removed and genes with a low expression level were filtered out using NOISeq v2.26.1 (Tarazona et al., 2015) with the following parameters: cpm = 1, cv.cutoff = 200. Differential expression analysis was performed based on the filtered data in DESeq2 v1.22.2 (Love et al., 2014). Genes with adjusted p-values <0.05 and |log2(FoldChange)| > 1 were considered as differentially expressed genes (DEGs). Tissue-specificity for each gene based on Tau method was calculated using tspex v0.6.1 (Camargo et al., 2020). Tissue-specificity was estimated by Tau, which is one of the most widely used methods for determining how specific or broad a gene is expressed

and it ranges from 0 to 1, where 0 indicates broad expression across tissues and 1 indicates tissue-specific expression (Yanai et al., 2005). Therefore, Tau has been used as a proxy of pleiotropy by many studies (Dean & Mank, 2016; Mank et al., 2008; Rogers et al., 2021). The co-expression network was constructed with Weighted Gene Co-expression Network Analysis (WGCNA) v.1.66 (Langfelder & Horvath, 2008). To avoid the noise from lowly expressed genes in the co-expression network, a more stringent filtering was applied with cpm = 5 and cv.cutoff = 200 using NOISeq, and the vst transformed data from DESeq2 were used to build the co-expression network. The network connectivity was retrieved from the co-expression network using the function *intramodularConnectivity* implemented in the WGCNA package. More in detail, for genes in the co-expression network, we measured the connectivity with genes in the same module (intramodular connectivity: kWithin), the connectivity with genes from different modules (intermodular connectivity: kOut) and its global connectivity (kTotal=kWithin+kOut). Therefore, kTotal, kWithin, and kOut in this tissue-specific co-expression network describe different properties: kTotal represents the total network connectivity and is the sum of kWithin and kOut; kWithin represents within module connectivity specific to one or multiple associated tissue types (specific connectivity); kOut represents the connection of one gene to the genes outside the module in the other tissue types (broad connectivity). Moreover, genes ranking in the top 5% of kWithin, representing high connection with the other genes in the module, were defined as the "hub" genes.

2.2.6 Differential Splicing Analysis

To understand the general pattern of splicing across tissues, intron excision ratio was calculated using Leafcutter v0.2.9 (Li et al., 2018), an annotation-free tool for quantification of RNA splicing. A PCA plot based on the intron excision ratio was

produced to visualize the general splicing patterns across tissues. For the pairwise differential splicing analysis between sexes, and between pairwise tissues, we used exon-based limma package v3.42 (Ritchie et al., 2015), which presented general good performances in differential splicing analysis with a large sample size (Mehmood et al., 2020; Merino et al., 2019). Genes with adjusted p-value <0.05 were considered differentially spliced (DS). The bam files generated from STAR were used for genome guided transcriptome assembly in Stringtie v2.2.4 (Pertea et al., 2015), using the parameters: -F 2 -f 0.05. SUPPA v2.3 (Trincado et al., 2018), was used to measure seven alternative splicing events: skipping exon (SE), alternative 5' splicing (A5), alternative 3' splicing (A3), retained intron (RI), alternative first exon (AF) and alternative last exon (AL).

2.2.7 Estimation of the rate of sequence evolution

The protein coding sequences from the closely related species *Cyclina sinensis* (Family Veneridae) were retrieved from Wei et al (2020, https://doi.org/10.5061/dryad.44j0zpcb5). Single-copy orthologs between *C. sinensis* and *R. philippinarum* were identified using OrthoFinder v2.5.1 with the default settings (Emms & Kelly, 2019). The orthologous protein sequences were aligned with Clustal Omega v 1.2.4 (Sievers et al., 2011) and the nucleotide alignments were derived according to the protein alignments using PAL2NAL (Suyama et al., 2006). The protein evolutionary rate was estimated according to the ratio of non-synonymous to synonymous nucleotide changes (Ka/Ks), and it was calculated using KaKs_calculator2 with the GMYN model (Wang et al., 2010a).

2.2.8 Gene Set Enrichment

Gene Ontology (GO) analysis was performed for different sets of genes in this study using topGO v2.44.0 (Alexa 2021). The GO enrichment analysis was performed with Fisher's exact test based on the gene list with a minimum node size of 10 and a pvalue cutoff of 0.01. REVIGO (Supek et al., 2011) was used to reduce redundancy in the enriched GO terms.

2.2.9 SNP Analysis

The quality of the reads from the male/female sequencing runs was assessed using the FastQC quality control tool v0.11.5, before being mapped to the *R. philippinarum* genome assembly using Rsubread v2.4.2 (Liao et al., 2019). Parameters for mapping male/female reads were: maxMismatches = 10, nTrim5 = 5, nTrim3 = 6, unique = FALSE, nBestLocations = 3. The resulting BAM files were used for variant calling with Freebayes v1.2.0 (Garrison & Marth, 2012), a tool for Bayesian haplotype-based genetic polymorphism discovery. Male/female population groups were analyzed with the following parameters: use-best-n-alleles = 4, min-alternate-count = 3, min-alternate-fraction = 0.05, min-mapping-quality = 20. The resulting VCF file was further filtered to retain only biallelic SNPs present in at least 80% of samples using Bcftools v1.11, according to the following criteria: min-alleles = 2, max-alleles = 2, type = snp, min-af = 0.01, exclude-min-quality < 20, exclude-max-missing > 0.2. Next, genotypes (in 0/1 format) were extracted from the two VCF files using the Genome Analysis ToolKit (GATK) v4.1.9.0 (DePristo et al., 2011) and genotype counts by population (n = 2) were used as input for the BayPass package v2.2 (Gautier, 2015), a population genomics software primarily aimed at identifying genetic markers subjected to selection and/or associated to population-specific covariates. SNPs that were identified

by BayPass as significantly contrasted (p-value < 0.001) between the male and female groups were then functionally annotated using Annovar (Wang et al., 2010b). The effect of SNPs was predicted with SnpEff v4.3 (Cingolani et al., 2012) and the PCA plot based on the SNPs across all samples was performed with SNPRelate v1.26 (Zheng et al., 2012).

2.2.10 Statistical analysis

Kruskal-Wallis test followed by Dunn test with FDR correction were used to assess the pairwise difference in kTOtal, kWithin, kOut, Tau, and Ka/Ks. Wilcoxon rank-sum test was used to assess if there was difference for kWithin between DEG and no-DEGs, and between DSGs and no-DSGs. Wilcoxon rank-sum test with HolmBonferroni correction was used to compare module-specific kTotal, kWithin, kOut, Tau, and Ka/Ks to the overall values across all the modules. The correlation between pairwise two indexes was performed with Spearman's rank-sum test. All the tests and data visualization described above were performed in the R v4.1.4 (https://www.Rproject.org/).

2.3 Results

2.3.1 Genome sequencing and assembly

PacBio sequencing consisted of 54 SMRT cells that yielded ~4M reads (36.5 Gb) of raw sequences with a median length of ~45Kb. The Illumina sequencing resulted in ~145 M reads (~75 Gb) for the short insert library, and ~48 M reads (~25 Gb) for the long insert library. After trimming both short reads libraries a total of ~ 180M PE reads were kept.

Table 2.1 Kmer based genome size estimations results.

Software	Kmer size	Genome size (bp)
Kmc + Genomescope	24	1,334,086,371
kmercountexact	24	1,377,873,056
Kmc + Genomescope	27	1,337,859,973
kmercountexact	27	1,395,519,877
Kmc + Genomescope	30	1,339,004,055
kmercountexact	30	1,377,970,397

Table 2.2 Genomescope heterozygosity and repetitive content estimation.

Kmer size	Heterozygosity (%)	Repetitive content (%)
24	3.97174	61
27	3.82226	49.4
30	3.69349	48.2

2.3.1.1 Genome survey, assembly, and contamination detection

We estimated a genome size ranging from 1.34 to 1.40 Gb depending on the kmer size and on the utilized tool (**Table 2.1**). This estimation is much lower than the 1.92 Gb estimation from flow cytometry, but coherent with previous kmer-based results of 1.32 (Yan et al., 2019) and 1.37 (Mun et al., 2017). Genomescope highlights a highly heterozygous and repetitive genome, with a heterozygosity ranging from 3.7% to 4% (**Table 2.2**). The repetitive content was estimated between 61.2% and 48%, depending on the kmer size (**Table 2.2**). Our preliminary version of the assembly showed a greater than expected genome size of 1.61 Gb, a good contig N50 of 144 Kb, but a low level of BUSCO completeness (64.2%; **Table 2.3**). These results are in line with recent

published assembler benchmarks which show that wtdbg2, compared to other assemblers, has a good ability to collapse haplotigs but lower accuracy (Guiglielmoni et al., 2020). Our purging-polishing steps successfully increased all genome evaluation statistics with a very little decrease in BUSCO completeness (**Table 2.3**). After three rounds, the final version of the assembly consisted in 15,908 contigs with a N50 of 183Kb, a total genome size of 1.41 Gb and a mean GC content of 0.32. We identified 884 out of 954 Metazoa BUSCO orthologs (92.7%), of which 802 were present as single copy (84.1%), and 82 as duplicates (8.6%). Missing genes represent 4.7% of the core gene set, while only 2.6% were identified as fragmented (**Table 2.4**). KAT analyses show a kmer completeness of 52,48% (**Table 2.4**), and 95% and 98% of the short and long reads were successfully remapped on the assembly respectively, with a median coverage depth of 53.42 and 22.69 (**Table 2.4)**. Blobtools identified 20 contigs as possible bacterial contaminations. These show a GC content and short read coverage that clearly deviates from the rest of the genome. Other six contigs were annotated as belonging to Priapulida, while only one to Zoopagomycota. In total, 937,293 bp (0.0007%) were removed from the assembly for downstream analyses. The 174 Chordata contigs were kept since they show the same GC content and coverage of Mollusca contigs. Moreover, at the species level, eight of them were annotated as belonging to *R. philippinarum*, and 104 (60%) as *Pseudochaenichthys georgianus*, a fish with a distribution (Antarctic Peninsula and Scotia Sea) that does not overlap with our sampling area. In summary, even if we cannot exclude the absolute absence of contamination in the final assembly, if present it should be in negligible amounts.

Table 2.3 Summary statistics of the different versions of the assembly

Assembly	No. contigs	N50(kb)	Busco results	Genome size (Gb)
Raw	28,484	144	C:64.2%[S:61.4%,D:2.8%],F:10.7%,M:25.1%,n:954	1.68
Purge1	21,831	159	C:64%[S:61.3%,D:2.7%],F:10.7%,M:25.3%,n:954	1.58
purge1_Hypo1	21,831	159	C:89.8%[S:77.9%, D:11.9%],F:3.7%,M:6.5%,n:954	1.58
Purge2_Hypo1	17,686	176	C:88.5%[S:81.4%,D:7.1%],F:3.7%,M:7.8%,n:954	1.46
purge2_Hypo2	17,686	176	C:92.6%[S:83.2%,D:9.4%],F:2.7%,M:4.7%,n:954	1.46
purge3_Hypo2	15,984	182	C:91.6%[S:82.9%,D:8.6%],F:3.1%,M:5.3%,n:954	1.41
purge3_Hypo3	15,984	182	C:92.7%[S:84.0%,D:8.7%],F:2.7%,M:4.6%,n:954	1.41
Redundans	15,908	183	C:92.7%[S:84.1%,D:8.6%],F:2.6%,M:4.7%,n:954	1.41

Table 2.4 Statistics of the final version of the assembly.

Assembly genome size	1,409,123,410 bp
No. contigs	15,908
Average contig length	88,579.55 bp
Largest contig	1,574,940 bp
N50	182,737 bp
N90	37,082 bp
Busco	C:92.7% [S:84.1%, D:8.6%], F:2.6%, M:4.7%, n:954
Mapped short reads	343,975,629 (95%)
Mapped long reads	12,691,865 (98%)
Median short reads depth	53.42
Median long reads depth	22.69
Kmer completness	52,48%
GC content	0.32

2.3.1.2 Genome comparison

We also compared our assembly to the recently published *R. philippinarum* chromosome-level assembly by Yan et al. (2019), hereinafter referred to as "reference genome". Out of the 15,908 contigs that composed our assembly, 15,781 (99.2%) had at least one alignment block to the reference genome. In total, all alignment blocks represent 80% of the assembly (1,129,816,977 bp) (**Table 2.5**). Of these, ~80% were aligned to assembled chromosomes, while only the remaining 20% to unplaced or unlocalized scaffolds. Interestingly, out of the 26,963 scaffolds—of which 19 represent the assembled chromosomes—that compose the reference genome, the 10.4% (2,810) do not show any alignment block with our assembly (**Table 2.5**), with all of them being classified as unlocalized scaffolds (*i.e* scaffolds that could not be placed in any chromosome). Totally, the 77.4% (868,865,079 bp) of the reference genome got aligned, while one-to-one alignments cover 750,261,296 bp (**Table 2.5**). Overall one-to-one aligned blocks had a mean nucleotide identity of 93% (**Table 2.5**).

Table 2.5 Results of WGA alignment

	Reference	Query
Total number of sequences	26,963	15,908
Aligned sequences	24,153 (89.6%)	15,781 (99.2%)
Unaligned sequences	2810 (10.4%)	127(0.8%)
Total number of bases	1,123,164,463	1,409,123,410
Aligned bases	868,865,079 (77.4%)	1,129,816,977 (80%)
Unaligned bases	254,299,384 (22.6%)	279,306,433 (19.8%)
1-to-1 aligned bases	750,261,296	750,261,296
Average identity	93%	

Considering our genome, tandem expansion (*i.e.* SVs that occur between overlapping alignments in the reference genome), repeat expansion, and repeat contraction (i.e. SVs that occur within unmappable gaps between alignments) comprise most of the total bases affected by SVs (93%; **Table 2.6**). Particularly, 923 tandem expansions affected 12,232,280 bp, with the 6% (63) of these having a size ranging from 50,000 to 100,000 bp and covering the 40% (4,404,773 bp) of the total affected bases. On the contrary, only 7 tandem contractions affecting 18,610 bp were identified. With regards to repeat expansions and contractions, Assemblytics called respectively 14,790 and 9,983 variants, covering 45,596,721 bp and 29,682,647 bp.

2.3.2 Genome Annotation

Using *de novo* approaches, we build up a starting consensus library composed of 5,600 sequences (3,197 and 2,403 by RepeatModeler and MITE_Tracker respectively). We added another 1,031 TEs already characterized in molluscs and retrieved from RepBase. After removal of genes/genes fragments, tandem, and low copy number repeats (< 5 good hits on the genome) we used a total of 2,332 nonredundant consensus sequences to annotate *R. philippinarum* repeatome. The 39.7% of the genome was masked by repetitive elements with a prevalence of cut and paste (DNA + MITEs) and Rolling Circle TEs (14.7% Unknown elements; 9.23% MITEs; 6.1% Rolling circle; 3.5% DNA; 2.95% LINE; 1.84% LTR; 1.25% SINE) (**Table 2.7)**.

The annotation pipeline generated 34,505 genes (36,076 mRNAs), with an average length of 8,053 bp. The completeness of the annotation was tested with BUSCO on the predicted proteins and showed that among the 954 genes included in the metazoa_db10 set, 83.4% were complete (74.5% as single copy) while 8.8% were fragmented. Of the

34,505 genes, 22,103 (64%) had a positive match by blastx against the swiss-prot database (**Table 2.7**).

Table 2.6 Assemblytics results

SV type	Size range (bp)	Count	bp	Total Count	Total bp
Insertion	50-500	2,291	431,323	4,449	4,936,132
Insertion	500-1000	2,114	3,803,513		
Insertion	10,000-50,000	43	649,859		
Insertion	50,000-100,000	1	51,437		
Deletion	50-500	2,104	343,503	2,706	1,589,257
Deletion	500-1000	585	913,812		
Deletion	10,000-50,000	17	331,942		
Deletion	50,000-100,000	0	0		
Tandem Expansion	50-500	153	44,415	923	12,232,280
Tandem Expansion	500-1000	451	1,530,103		
Tandem Expansion	10,000-50,000	256	6,252,989		
Tandem Expansion	50,000-100,000	63	4,404,773		
Tandem Contraction	50-500	3	393	7	18,610
Tandem Contraction	500-1000	4	18,217		
Tandem Contraction	10,000-50,000	0	0		
Tandem Contraction	50,000-100,000	0	0		
Repeat Expansion	50-500	3,919	894,227	14,790	45,596,721
Repeat Expansion	500-1000	9,984	28,469,497		
Repeat Expansion	10,000-50,000	865	14,883,699		
Repeat Expansion	50,000-100,000	22	1,349,298		
Repeat Contraction	50-500	3,470	785,870	9,983	29,682,647
Repeat Contraction	500-1000	5,870	14,539,245		
Repeat Contraction	10,000-50,000	600	11,545,708		
Repeat Contraction	50,000-100,000	43	2,811,824		

Table 2.7 Annotation statistics of the Manila clam genome

Repeats genome coverage	39.70%
LINE	1.84%
DNA	2. 95%
SINE	1.25%
LTR	1.84%
MITE	6.10%
Unknown	14.70%
No. of predicted genes	34,505
Mean gene length	8,053 bp
Mean exons per gene	6.4
Mean exon length	212 bp
Mean introns per gene	5.2
Mean intron length	185,869 bp
No. of genes with a significant hit against Swiss - Prot	22,103

2.3.3 Differential Expression and Co-Expression Network

To investigate the global expression patterns in all tissues of both sexes, a PCA analysis was performed in DESeq2. As shown in **Figure 2.1a**, different tissues presented distinct expression profiles, and expression patterns between female and male somatic tissues were quite similar, while large differences were found in gonads. Consistently, the number of DEGs between females and males in adductors and mantles were low (578 and 22, respectively), while the number of DEGs between gonads were 6,167, including 3,024 female-biased DEGs and 3,143 male-biased DEGs (**Figure 2.2a**). The comparisons of DEGs between pairwise tissues were performed for males and females separately. Generally, the number of DEGs between somatic tissues (adductor muscle vs. mantle) was less than the number of DEGs between somatic tissue and gonad (e.g. gonad vs mantle) (**Table 2.8**). A large proportion of DEGs in females

in pairwise tissue comparisons overlapped with the corresponding DEGs in males. In all pairwise tissue comparisons, 1,787 and 2,277 genes were differentially expressed across all three tissues in females and males, respectively, and 1,009 of these DEGs were shared between females and males.

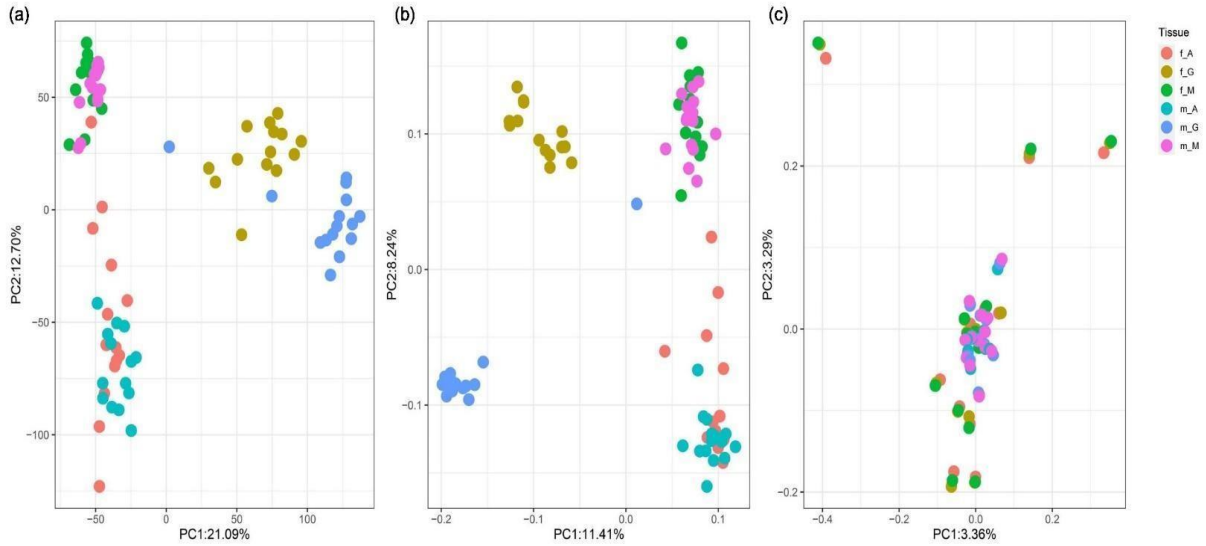


Figure 2.1 PCA plot for gene expression (a), alternative splicing (b) and genotype (c). Each dot represents a sample and each color represents a tissue type; f_A: female adductor; f_G: female gonad; f_M; female mantle; m_A: male adductor; m_G: male gonad; m_M: male mantle.

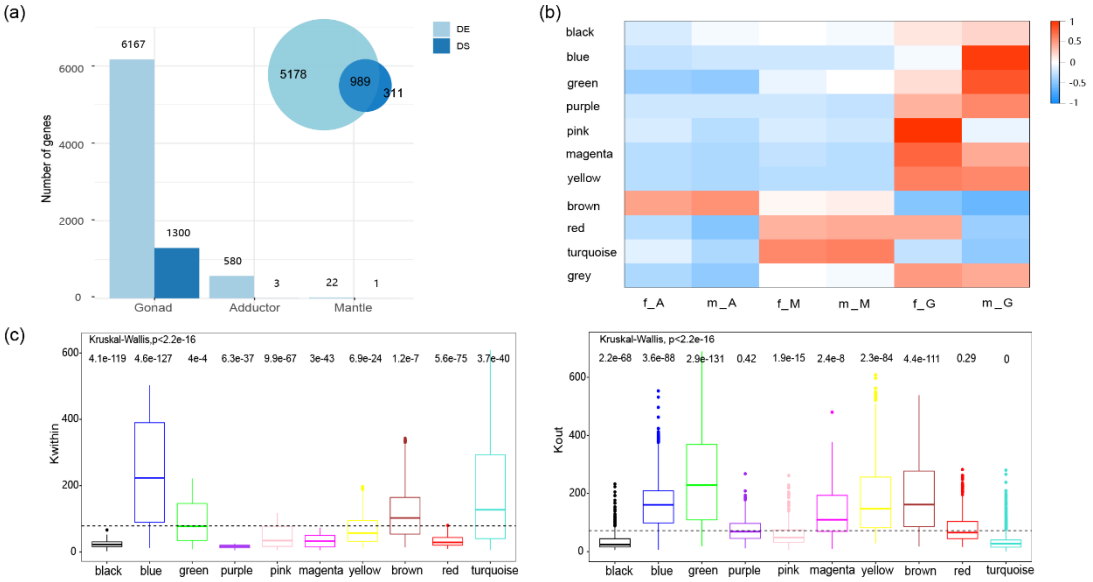


Figure 2.2 Differentially expressed, spliced and co-expressed genes across tissue types. (a) The number of differentially expressed (DE) and spliced (DS) genes between females and males in each tissue. The Venn plot on the top-left represents the overlap between DE and DS genes in the gonad. (b) Module-tissue association based on the gene expression. Each row corresponds to a co-expression module, and each column represents a tissue-type. Each co-expressed module was named with a specific color. The correlations and p values between module and tissue are shown in each cell. (c) The distribution of within (kWithin) module connectivity and outside (kOut) module connectivity for genes in the co-expression modules. Wilcoxon rank-sum test with FDR corrections was used to compare the distribution of kWithin and kOut in each module to the overall distribution and the pvalues were shown on the top of the boxplot. The dash line indicates the median of the overall distribution.

Table 2.8 The number of differentially expressed and spliced genes in pairwise tissue comparisons

Genes	f_A vs f_G	f_A vs f_M	f_G vs f_M	m_A vs m_G	m_A vs m_M	m_G vs m_M
DEGs	10094	4915	11062	9855	6612	11986
DSGs	1981	954	1439	1901	1320	2181
DEGs&DSGs	1614(81.5%)	672(70.4%)	1262(87.7%)	1693(89.1%)	1051(79.6%)	1981(90.8%)

A tissue-specific gene co-expression network was constructed to investigate the gene regulatory relationships in tissue-associated modules. A total number of 8,640 genes were assigned to 10 modules (**Figure 2.2b**). The blue module (1,334 genes) and the green module (790 genes) showed high association with male gonads, while the pink module (417 genes) was positively associated with female gonads. Moreover, yellow (977 genes), magenta (232 genes), and purple (80 genes) modules were associated with both female and male gonads, and turquoise (2,718 genes) and brown (1,749 genes) were associated with somatic tissues (Supplementary Figure 2.1). The distribution of intramodular connectivity (kWithin) and intermodular connectivity (kOut) for genes in each module is shown in **Figure 2.2c.** Generally, the gonad-associated blue module and mantleassociated turquoise module presented significantly higher kWithin compared with the overall distribution, while another gonad-associated green module presented significantly higher kOut (Figure 2.2c). A predominant number of 1,253 (93.9%), 579 (73.3%), and 397 (95.2%) genes in the blue, green, and pink modules respectively, were also DEGs between female and male gonads. Besides, the kWithin for DEGs in gonadassociated blue, green and pink modules were significantly higher than non-DEGs between female and male gonads (**Figure 2.3a**). Moreover, "hub" genes in the female/male gonad-associated modules were retrieved and those with annotation are listed in **Supplementary Table 2.1**. These genes included male-gonad specific SRYbox transcription factor 30 (sox30), and female-gonad specific mating-type-like protein ALPHA2 (mtlalpha2).

GO enrichment analysis was applied to explore the predicted functions of different subsets of genes, and the results are shown in **Supplementary Table 2.2**. Considering the low number of DEGs between female and male mantles, we did not perform the enrichment analysis on this subset of genes. The significantly enriched GO terms in

adductor muscles between males and females were related to microtubule-based process, motor activity (**Supplementary Table 2.2**). Reproduction and cell cycle related processes were significantly enriched between females and males in gonads (**Supplementary Table 2.2**). Reproductive related processes were also enriched in the male-gonad associated blue module (**Supplementary Table 2.2**). The genes co-expressed in the other male-gonad associated (green) module appeared to overrepresent some general functions, with processes like “organelle assembly”, “cell project”, and “catalytic activity” being enriched. In the female-gonad associated pink module, processes related to transferase and protein metabolic activities were significantly enriched. Different functional processes were enriched in the three gonad-associated modules (magenta, purple and yellow), such as “cell adhesion” (purple and magenta modules), homeostatic related processes (purple module), and process related to tissue development (magenta module) (**Supplementary Table 2.2**). Intriguingly, for genes in the yellow module, processes related to DNA repair, DNA replication, and gene expression were significantly enriched. In mantle-associated turquoise modules, genes were overrepresented in the immune-related process and metal ion binding.

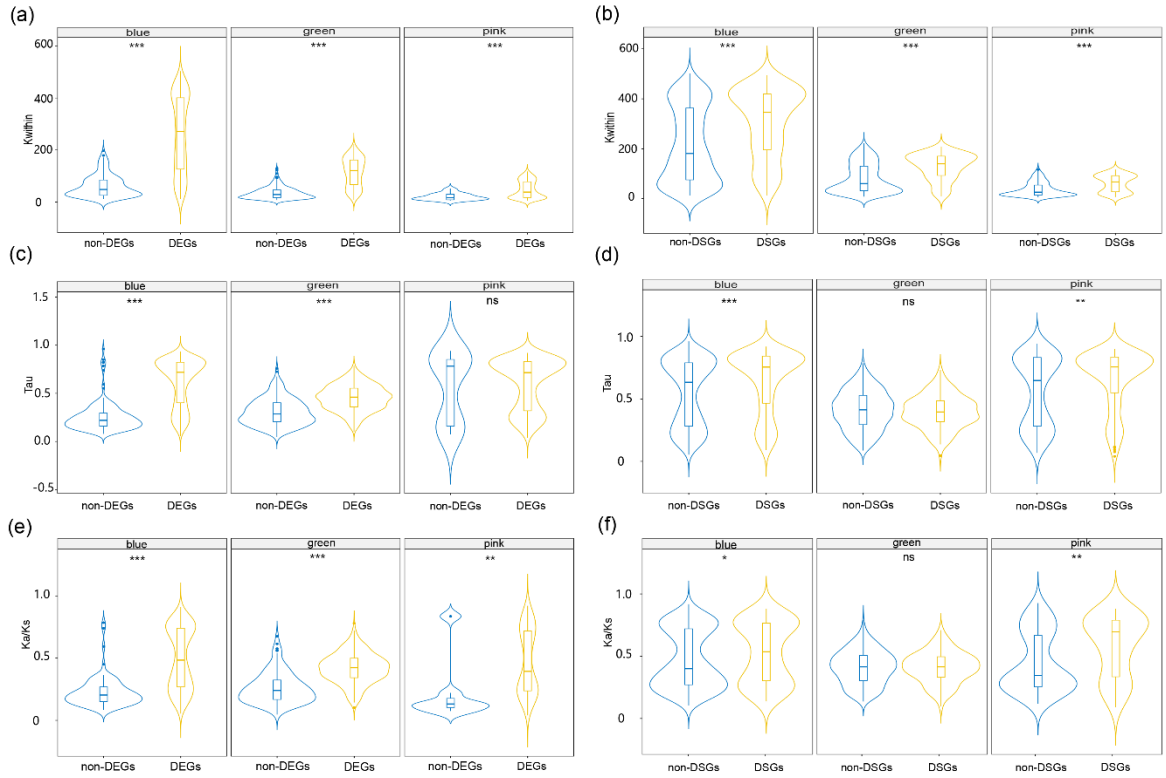


Figure 2.3 Comparisons between differentially expressed and non-differentially expressed genes, and between differentially spliced genes and non-differentially spliced genes in female and male gonad-associated modules. (a), (c), and (e) represent comparisons of the connectivity, tissue specificity and sequence evolutionary rate between differentially expressed genes and non-differentially expressed genes. (b), (d), (f) represent the comparisons of the connectivity, tissue specificity and sequence evolutionary rate between differentially spliced genes and non-differentially spliced genes.

2.3.4 Differential Splicing Analysis

Consistently with the expression profiles, splicing patterns also differed across tissues, and differences between females and males were observed in gonads but not in somatic tissues (**Figure 2.1b**). Global alternative splicing events for each tissue is shown in **Figure 2.4**. Generally, SE, A5, and AF accounted for a large proportion in all tissues, while RI and MXE were the least represented events in all tissues. Moreover, alternative splicing in gonads and mantles seems to be more frequent than in adductor muscles. Despite the pervasiveness of alternative splicing in all tissues, the number of

genes showing differential splicing between females and males in each tissue, and between pairwise tissues were far less compared to DEGs. The number of DSGs between females and males in adductor muscles, mantles, and gonads were 3, 1, and 1,300, respectively (**Figure 2.2a**). Notably, among all the 1,300 DSGs between female and male gonads, 989 (76%) of them were also differentially expressed between female and male gonads. We also retrieved these DSGs in three sex associated co-expression modules (blue, green, and pink) and we found that the DSGs in these modules showed significantly higher kWithin than non-DSGs (**Figure 2.3b**). The number of DSGs between gonads and somatic tissues was higher than that between two somatic tissues (Table 2.8). Moreover, in all these comparisons between different tissues, DEGs and DSGs were largely overlapped for both females and males, and around 80-90% DSGs between gonads and somatic tissues were also DEGs (**Table 2.8)**. Some of these DSGs overlapped with DEGs or sexassociated modules (listed in **Supplementary Table 2.1)**, and the large amount of overlapping genes between DSGs and DEGs in gonads also resulted to have many

However, “chromatin remodeling” was enriched in DSGs but not between female and male gonads. Additionally, we investigated the function of the DSGs that did not overlap with DEGs, and such genes were involved in chromatin remodeling and mRNA catabolic processes that were not overrepresented in DEGs.

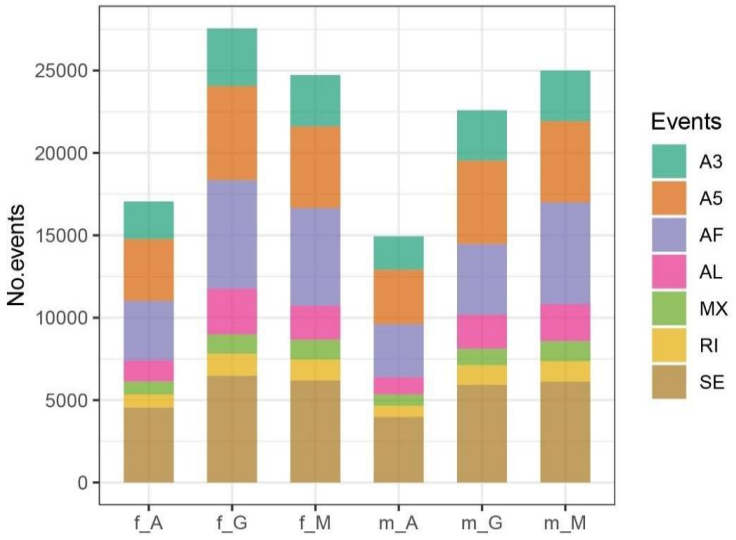


Figure 2.4 The distribution of different splicing events in each tissue. SE: skipping exon, A5: alternative 5' splicing, A3: alternative 3' splicing, RI: retained intron, AF: alternative first exon, AL: alternative last exon.

2.3.5 Tissue specificity in the co-expression Network

In this study, the tissue specificity index Tau ranged from 0.2 to 0.8 for most genes, while only a small proportion of genes showed extremely tissue-specific (> 0.8) or broad (< 0.2) expression. Tau values in different co-expression modules varied markedly (Kruskal-Wallis test: $p < 0.001$). Generally, in somatic associated red, brown modules, genes showed relatively low Tau values, indicating low tissue-specificity (**Figure 2.5**). By contrast, we found relatively high and variable Tau values in most gonad-associated modules, except for the male-gonad-associated green module and gonad-associated yellow module, which had relatively low Tau values, with median values at around 0.4 and 0.3, respectively (**Figure 2.5**). Interestingly, we found that the yellow and green modules also showed relatively high intermodular connectivities, indicating that genes in these two modules showed also high connections with other tissue types (**Figure 2.2c**).

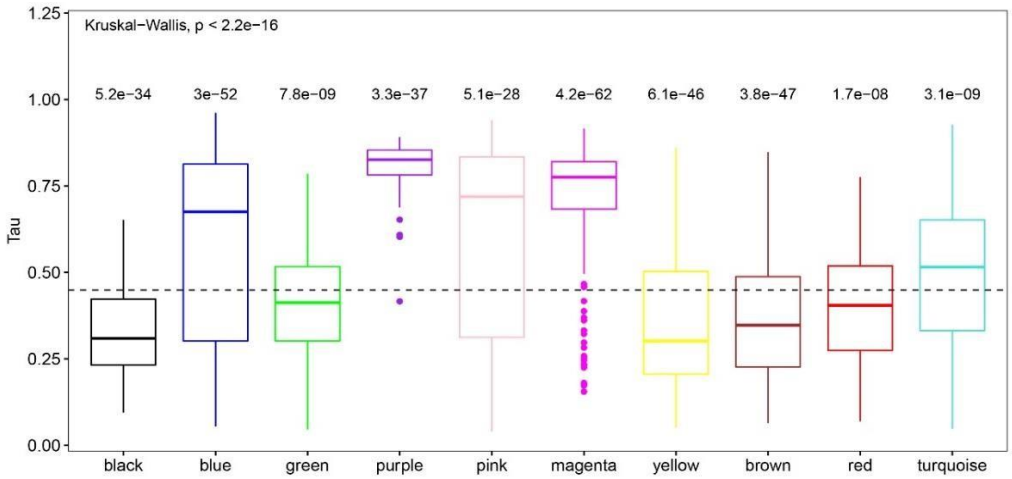


Figure 2.5 The tissue specificity (Tau) for each module. Different color indicates the corresponding modules in the co-expression network. Wilcoxon rank-sum test with FDR corrections was used to compare the distribution of Tau in each module to the overall distribution and the p-values were shown on the top of the boxplot. The dash line indicates the median of the overall distribution.

We further investigated the Tau distribution across the co-expression network and we found significantly positive correlation between whole network connectivity (kTotal) and tissue specificity (Tau) (Spearman's **R = 0.24, p < 2.2E-16**) and between intramodular connectivity (kWithin) and tissue specificity (Spearman's **R = 0.34, p < 2.2E-16**), but a weak correlation between intermodular connectivity (kOut) and tissue specificity (Spearman's **R = -0.07, p = 6.433E-11**). Moreover, we found significant positive correlation between tissue specificity Tau and kWithin, kTotal in most tissue associated modules such as blue, pink, and turquoise modules, indicating that genes with high tissue specificity also presented high connection in the specific tissue type (**Figure 2.6a and Supplementary Figure 2.2a**). Additionally, the negative correlation between kOut and Tau was also observed in most modules except for blue, green, and yellow modules, where a positive correlation was observed (**Supplementary Figure 2.2b**).

We also investigated the tissue-specificity for DEGs and DSGs in the co-expression network, mainly focusing on gonad-associated modules because of the low number of DEGs and DSGs in somatic tissues. In the male-gonad-associated blue module, DSGs and DEGs presented significantly higher Tau values than non-DSGs and non-DEGs (**Figure 2.3b**). In the green module, DEGs also presented significantly higher Tau values than non-DEGs, while Tau values between DSGs and non-DSGs were not significantly different from each other (**Figure 2.3b**). However, in the pink module, Tau showed no significant difference between DEGs and non-DEGs, but DSGs presented slightly higher Tau values than non-DSGs (**Figure 2.3b**).

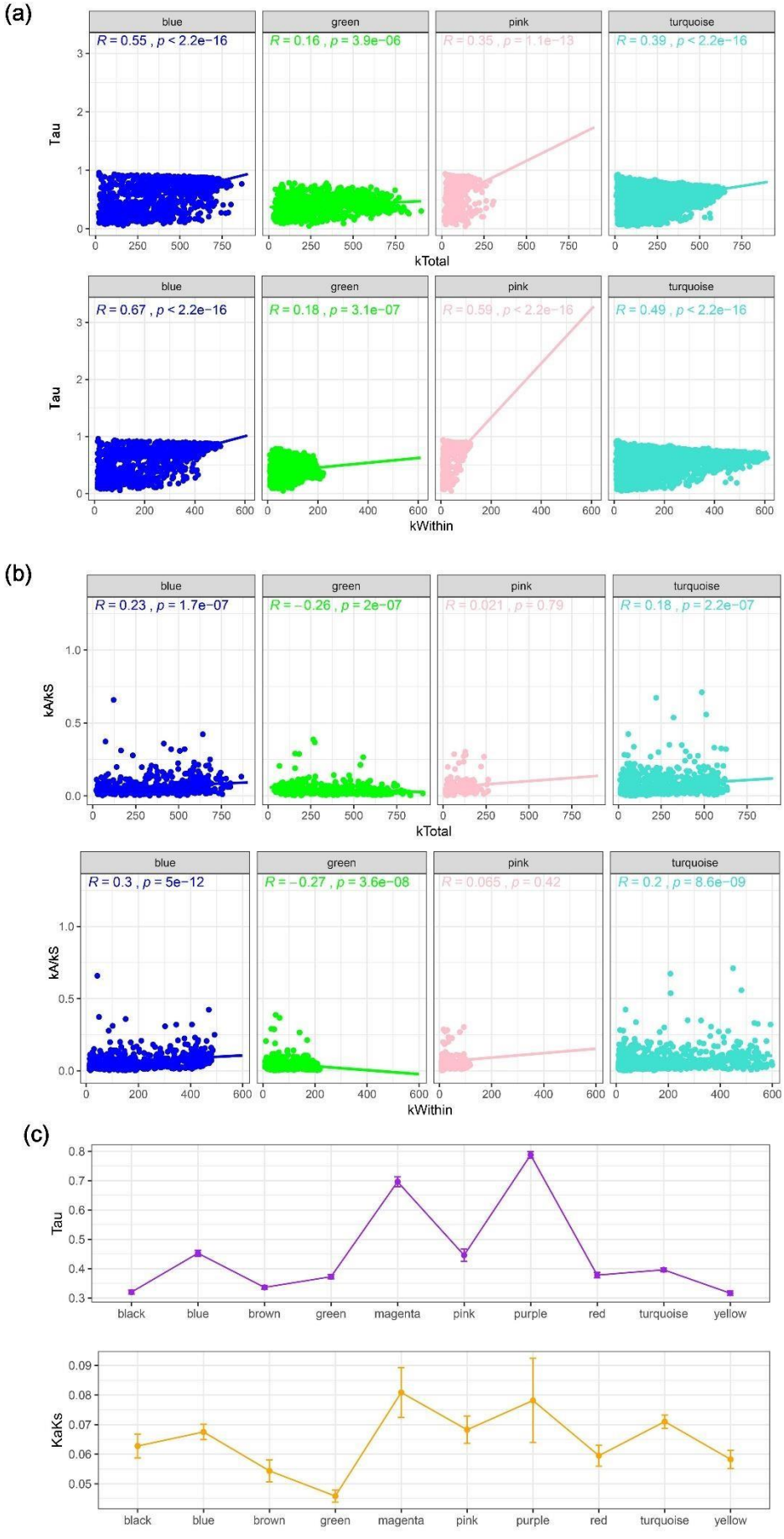

(a)
blue
green
pink
turquoise
R = 0.55 , p < 2.2e−16
R = 0.16 , p = 3.9e−06
R = 0.35 , p = 1.1e−13
R = 0.39 , p < 2.2e−16
Tau
kTotal
R = 0.67 , p < 2.2e−16
R = 0.18 , p = 3.1e−07
R = 0.59 , p < 2.2e−16
R = 0.49 , p < 2.2e−16
kWithin
(b)
R = 0.23 , p = 1.7e−07
R = −0.26 , p = 2e−07
R = 0.021 , p = 0.79
R = 0.18 , p = 2.2e−07
kA/kS
R = 0.3 , p = 5e−12
R = −0.27 , p = 3.6e−08
R = 0.065 , p = 0.42
R = 0.2 , p = 8.6e−09
(c)
black
blue
brown
green
magenta
pink
purple
red
turquoise
yellow
KaKs

Figure 2.6 The relationship between network connectivity and tissue specificity, evolution rate. (a) The correlation between tissue specificity index (Tau) and total connectivity (kTotal), within module connectivity (kWithin) in four tissue-associated modules. (b) The correlation between evolutionary rate and total network connectivity (kTotal), within module connectivity (kWithin) for four tissue-associated modules. The Spearman's correlation (R) and p values were shown on the top. (c) The trends of tissue specificity index (Tau) and evolutionary rate (Ka/Ks) in the co-expression modules. Average value (each dot) and standard error (error bar) was used for each module.

2.3.6 Variations in the rate of sequence evolution across co-expression network

Ka/Ks distribution also varied in different co-expression modules, with the male gonad-associated blue module and mantle-associated turquoise module presenting a significantly higher than overall values, and the green module presenting a significantly lower than overall values (**Figure 2.7**). We found no significant correlation between general network connectivity (kTotal) and evolutionary rate Ka/Ks (kTotal: Spearman's $R = -0.0044$, $p = 0.78$). However, when we investigated this relationship in each module, we found that genes in male gonad-associated blue module and mantleassociated turquoise module showed significantly positive correlation between network connectivities (both kTotal and kWithin) and evolutionary rates, while genes in the other male gonad associated green module showed significantly negative correlation between connectivities and evolutionary rates (**Figure 2.6b and Supplementary Figure 2.2a**). Most modules presented no significant correlation between intermodular connectivity and evolutionary rate (**Supplementary Figure 2.2b**).

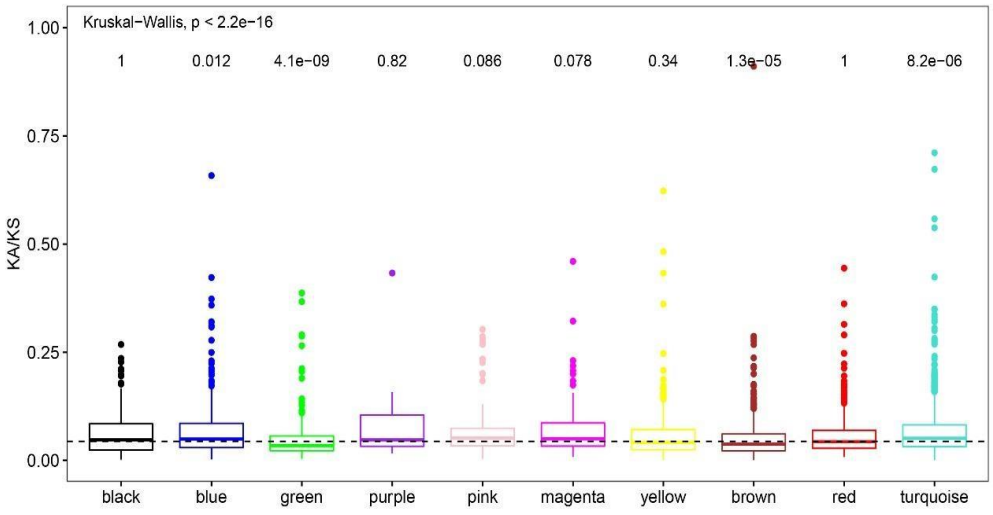


Figure 2.7 The distribution of Ka/Ks in each module. Different color indicates the corresponding modules in the co-expression network. Wilcoxon rank-sum test with FDR corrections was used to compare Ka/Ks values in each module to the overall distribution, and the p-values were shown on the top of the boxplot. The dash line indicates the median of the overall distribution.

Tau was positively correlated with Ka/Ks (Spearman's R = 0.17, p < 2.2E-16), nevertheless such correlation was not detected in all the tissue-associated modules. Similar to the correlation between connectivity and Ka/Ks, significant positive correlation between Tau and Ka/Ks was also detected in blue and turquoise modules (**Supplementary Figure 2.3**). In spite of the lack of correlations in most modules, the Tau and Ka/Ks values showed similar trends across different modules (**Figure 2.6c**). Combined with the tissue specificity analysis above, It appears that genes in the male gonad-associated blue module and mantle-associated turquoise module with high intramodular connectivity and tissue-specificities also presented high evolutionary rates, while genes in the green module with high connections to outside the modules had a lower evolutionary rate.

We also observed the differences in evolutionary rate between DEGs and non-DEGs, and between DSGs and non-DSGs in the female and male gonad-associated

modules (**Figure 2.3c**). In all the three gonad-associated modules, DEGs presented significantly higher Ka/Ks than non-DEGs. Likewise, we found that DSGs in blue and pink modules also showed significantly higher Ka/Ks than non-DSGs, but such result was not detected in the male gonad-associated green module.

2.3.7 Contrasting SNPs

We first retrieved SNPs for each sample separately and we found that the polymorphism in different tissues of the same individual was extremely low (**Figure 2.1c**). Thus, to retrieve sex-specific SNPs, we divided all the samples into female and male groups merging the three tissues of the same individual together. We detected 750,790 total variants between male and female groups, of which 676,009 were SNPs. Of these, 252,858 SNPs were present in at least 80% of individual samples with a minimum quality score of 20. Filtered SNPs from male and female groups were analyzed using Baypass for contrast based on genotype counts, yielding 614 SNPs significantly contrasting between groups ($p < 0.001$), respectively. Annovar merged the selected SNPs with the genome assembly annotation to identify the locations of each marker, specifying that of the 614 significantly contrasting SNPs. Eighty-nine contrasting SNPs located in the intergenic regions, while the other SNPs located in 448 genes. We assigned SNPs located in the genes to the co-expression network, and we found that the proportion of genes containing SNPs in each module were significantly differed from expected proportion of genes for each module in the co-expression network (chi-square test: $p < 1E\text{-}5$). Instead, some modules such as green and yellow modules enriched more genes containing SNPs than expected, while blue and turquoise module enriched less genes (**Figure 2.8**).

Finally, exonic SNPs from male and female groups were searched against a set of SNPs from a DNA pooled sequencing experiment of Mediterranean and Atlantic *R. philippinarum* populations, revealing that the two datasets contained 260 exonic SNPs in common. Genes containing contrasting SNPs are listed in **Supplementary Table 2.1**, and some of them were also identified in DEGs, DSGs or tissue-associated modules such as ankyrin repeat domain-containing protein, double-strand-break repair protein rad21-like protein, folliculin, transcriptional regulator ATRX. Functional enrichment indicated that genes containing contrasting SNPs involved in the process such as "mitochondrial transmembrane transport", "protein localization to organelle" and "chromatin remodeling" (**Supplementary Table 2.2**).

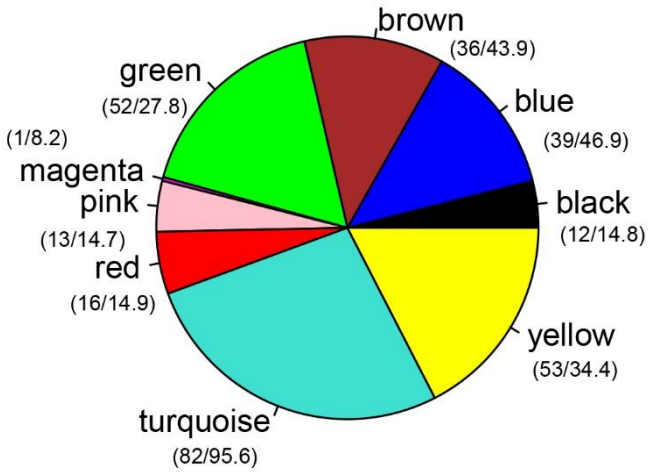


Figure 2.8 The number of genes containing SNPs in the co-expression modules. The numbers in the bracket suggest the number of genes containing SNPs in each module (in front of slash) and expected number of genes containing SNPs in each module (behind slash). The expected values were calculated by the percentage of genes in each module multiply the number of SNP-containing genes assigned in the co-expression network.

2.4 Discussion

2.4.1 Genome characterization and comparison

In the present work we sequenced and assembled a new long-read-based draft genome *R. philippinarum*. Notably, this represents the first effort to sequence and assemble a wild (i.e. not inbred) specimen relying both on short and long read data, and the first long read genome assembly for this species. In line with previous *R. philippinarum* sequencing projects (Mun et al., 2017; Yan et al., 2019), we estimated the genome size to range between 1.34 and 1.40 Gb with an extremely high level of heterozygosity (3.7-4%) and, to a lesser extent, repetitive content (48-61.2%). Even if such features are common in molluscan genomes (Sun et al., 2021), to our knowledge our sample represents the most heterozygous mollusc genome sequenced so far. Such a high heterozygosity clearly represents one of the biggest obstacles in obtaining high quality assemblies (Sun et al., 2021). Despite these challenges, our pipeline produced an assembly with a size close to the expected kmer based estimations (1.41 Gb), a BUSCO completeness in line with previous published bivalve genomes (95.3 % [Complete + Fragmented]) such as *Pinctada fucata* (Du et al., 2017) and *M. galloprovincialis* (Gerdol et al., 2020), and a good contig N50 of 144 Kb.

Compared to the recently published *R. philippinarum* genome by Yan et al. (2019), our assembly resulted in a significantly higher assembly size (1.41. Gb vs 1.12 Gb in Yan et al., 2019) and contig N50 (144 kb vs 28.1 kb in Yan et al., 2019), in similar BUSCO completeness levels (92.7% in our assembly, 93.8% in Yan et al., 2019), but higher levels of duplications (8.6% in our assembly, only 1.9% in Yan et al., 2019) (**Table 2.4**). In our opinion, the higher level of duplication in our assembly depends on a combination of three not mutually exclusive factors: (a) the use of a wild sample

compared to a highly inbred one; (b) the use of a more stringent removal of suspicious duplicated contigs in combination with only short reads data in the reference genome assembly; (c) a high genetic and genomic diversity between Chinese and North - American populations. Therefore, even if some duplications in our assembly could be due to the persistence of haplotigs due to its high heterozygosity, it is also possible that, at least partially, they represent true duplications which have been collapsed in the reference, or specific to our sample. The presence of extensive intraspecific genome duplications would not be a novelty inside Bivalvia, with *Crassostrea virginica* showing the 16.5% of its genome duplicated, with a reduced percentage in inbred samples and a great prevalence of tandem duplications (Modak et al., 2021). Moreover, intraspecific gene presence/absence variation (Gerdol et al., 2020) and chromosomal rearrangements (Insua et al., 1998; Thiriot-Quiévreux & Insua, 1992), as well as hemizygosity (Calcino et al., 2021), seem to be widespread across different bivalve species and we can speculate that similar features are shared also by *R. philippinarum*. Coherently with these observations, repeat and tandem expansion were found to be the most impacting SVs (*i.e* affecting the greatest number of base pairs) between the two assemblies, but while the former were partially counterbalanced by numerous repeat contraction with a ratio of 0.65 between affected bps, the latter were not (ratio of 0.0015). Moreover, tandem expansion represents the class with the highest number of SVs falling in the highest size bin (50,000 - 100,000 bp; 63 SVs) (**Table 2.5**). Finally, even if most of both genomes successfully aligned, it is worth noting the low nucleotide identity (93%) between one-to-one alignment blocks and the relatively low number of bases involved, highlighting a possible high genetic and genomic diversity between clam populations from North America and China. Unfortunately, we could not perform a deep gene-based comparison due to the absence of publicly available resources from

Yan et al. (2019), however, we predicted a similar number of genes and repetitive content (respectively 22103 and 39.7% in our genome; respectively 27,652 and 38.30% in Yan et al., 2019) (**Table 2.7**).

Overall, these data indicates that our assembly is fairly complete and collapsed, and that it can be a source of new and complementary genomic information for the Manila clam *R. philippinarum.*

2.4.2 Tissue-associated co-expression network and tissue specificity

We built a tissue specific co-expression network and we investigated the relationships among network connectivity, tissue-specificity, and evolutionary rates. In a co-expression network, the intramodular connectivity can be used to identify "hub" genes, and usually a higher intramodular connectivity represents a tighter association between gene and trait especially in the modules with high module-trait correlations. On the other hand, intermodular connectivity indicates the degree of connection of a module with other tissue-associated modules (Langfelder & Horvath, 2008; Mauro & Ghalambor, 2020). Consequently, genes with high intramodular connectivity in one module might be more peripheral in the others, while genes with high connection to the other modules may be less tolerant to the sequence evolution because of high pleiotropy (Mack et al., 2019). Likewise, tissue specificity was used as a proxy for pleiotropy: the lower the tissue specificity, the higher the pleiotropy (Rogers et al., 2021; Yanai et al., 2005). Therefore, in the tissue-specific co-expression network, the moduletissue correlation represents the degree to which these genes are specific to the relevant tissues, which can explain the significant positive correlation between tissue specificity and intramodular connectivity in gonad-associated modules with relatively high module-trait correlations (**Figure 2.2 and Supplementary Figure 2.2**). Additionally, somatic

tissues usually are associated with one or two co-expression modules, while gonads are associated with multiple modules with different co-expression patterns (**Figure 2.2**), revealing the complexity of transcriptional regulation in gonads. Intermodular connectivity indicates a role of module coordinator associating different traits, therefore it is no wonder to see the negative correlation between tissue specificity and intermodular connectivities in some modules (**Supplementary Figure 2.2b**).

Intriguingly, blue, green, and yellow modules presented significantly positive correlation between both inter- and intramodular connectivity and tissue specificity (**Supplementary Figure 2.2**), indicating that genes with high tissue specificity also presented high overall network connections. Therefore, these three modules may include some genes that play the "coordinator" role of the tissue specific co-expression network. Consistently, the wide variation of intermodular and intramodular connectivities in different modules also revealed the different position of these modules in the co-expression network (**Figure 2.2c**). The male gonad-associated green module was characterized by the highest intermodular connectivity and may represent the central coordination position in the co-expression network, while female-gonad associated pink module was characterized by relatively low intermodular and intramodular connectivities, indicating probably a more particular and independent role.

2.4.3 Large overlap between DEGs and DSGs in gonad

Different expression patterns between females and males have been investigated by several studies in gonochoric and sequential hermaphroditic bivalves, but mainly focused on the reproductive tissue (gonad), or across development stages (Broquard et al., 2021; Capt et al., 2019; Ghiselli et al., 2018; Yue et al., 2018). A comprehensive information for the general expression across tissues is lacking. In this study, we found

distinct transcription patterns across tissues, with a huge difference being observed between female and male gonads, but not between female and male somatic tissues (**Figure 2.1a**). Intriguingly, we found that the alternative splicing showed a similar pattern as the differential expression, with even more difference between female and male in gonads (**Figure 2.1b**), indicating that alternative splicing also plays a central role in tissue distinctiveness, particularly in gonads. Additionally, DEGs and DSGs between females and males in gonads (also between pairwise tissues) are largely overlapping (**Figure 2.2a**, **Table 2.8 and Table 2.9**). Previous studies revealed that transcriptional regulation and alternative splicing can be two parallel processes playing contrasting roles by involving different genes in response to the changing environment (Grantham & Brisson, 2018; Jacobs & Elmer, 2021; Wang et al., 2019). On the other hand, large overlaps between DEGs and DSGs have also been reported and DSGs were thought to play "redundant" roles as DEGs (Healy & Schulte, 2019; Zhang et al., 2019). However, in this study, we found that both DEGs and DSGs showed generally higher tissue-specificities and higher connectivities in the sexual related co-expression modules (**Figure 2.2 and 2.3**), indicating that these genes play central roles in the corresponding co-expression modules and have strong functional impacts on gonad differentiation. Therefore, instead of having a "redundant role", we propose that the two processes are somewhat "complementary", as alternative splicing is able to extend the protein diversity favoring tissue differentiation without changing gene expression. In this study, both DEGs and DSGs were involved in microtubule-related dynamics, which was also observed in other bivalves and is likely related to spermatogenesis (Boulais et al., 2019; Capt et al., 2019; Dunleavy et al., 2019). Enriched GO terms related to reproduction for DEGs and chromatin remodeling for DSGs could be a sign of an "independent" role for the two processes.

2.4.5 Evolution of differentially expressed genes, differentially spliced genes and tissue-associated genes

It was suggested that genes with broad expression patterns are subject to greater functional constraints and are less likely to be differentially expressed (Mank et al., 2008; Meisel 2011). Consistently, we also observed that DEGs in three gonadassociated modules presented significantly higher tissue specificity than non-DEGs. These DEGs also showed elevated evolutionary rates compared to non-DEGs (Figure 2.3), which was also observed in other organisms and is supposed to be the consequence of relaxed constraints or genetic drift (Gershoni & Pietrokovski, 2014; Harrison et al., 2015; Rogers et al., 2021). In contrast, splicing was supposed to be less constrained than expression variation, and it was observed to be common in genes with high pleiotropy (Jacobs & Elmer, 2021; McGirr & Martin, 2018; Papakostas et al., 2014; Rogers et al., 2021). However, in this study we found that DSGs also presented higher tissue specificity and higher evolutionary rate than non-DSGs in male-gonad associated blue and female-gonad associated pink modules (**Figure 2.3**). Interestingly, the other male-gonad associated green module with relatively low tissue specificity and high intermodular connectivity presented no difference in tissue specificity and evolutionary rates between DSGs and non-DSGs, which is in line with splicing being common in highly pleiotropic genes (Jacobs et al., 2021; Roger et al., 2021). Combined with the large overlap between differential expression and splicing, our results indicated that DEGs and DSGs may be both under strong constraints in gonads depending on the gene sets.

Many studies have suggested a negative correlation between protein evolution and gene expression, and it was proposed that tissue-specific proteins have lower constraints (Meisel et al., 2011; Moyle et al., 2021). In line with this hypothesis, our

results indicated that tissue specificity is significantly correlated to the rate of sequence evolution (Spearman's R = 0.17, p < 2.2E-16). Additionally, we found that genes in gonad-associated modules were characterized by relatively high sequence evolution and tissue specificity, while somatic associated modules had relatively low tissuespecificity and sequence evolution even if we did not find significant differences in mantle-associated modules (see discussion below). Gonad specific genes are thought to be important in evolution as they contribute to reproductive success, fitness, and speciation (Whittle & Extavour, 2017). Several studies reported that genes expressed in reproductive tissues are evolving faster than genes expressed in somatic tissues, as a result of possible conflicting sex-specific selection pressures (sexually antagonistic genes), mate-choice, and intrasexual competition (Ellegren & Parsch, 2007; Harrison et al., 2015; Moyle et al., 2021; Rago et al., 2020; Turner & Hoekstra, 2008; Whittle & Extavour, 2019). Moreover, higher evolutionary rates in male-biased genes were observed in a wide range of animals (Grath & Parsch, 2012; Harrison et al., 2015; Parsch & Ellegren, 2013). However, we found that genes in the female gonadassociated co-expression network presented slightly higher average Ka/Ks compared to genes in the male gonad and somatic associated networks. Higher evolutionary rate in females have also been detected in other species and it was hypothesized that higher evolutionary rates in both female and male biased genes are due to external fertilization causing strong gamete competition (Dean & Mank, 2016; Lipinska et al., 2015; Papa et al., 2017; Whittle & Johannesson, 2013; Yang et al., 2016). By contrast, two distinct patterns were observed for genes in the male gonad-associated networks: genes in the blue module were characterized by relatively high tissue specificity, high within-module connectivity, and high protein evolutionary rate, while genes in the green module were characterized by relatively low tissue specificity, low intramodular

connectivity, low evolutionary rate, and high intermodular connectivity. Moreover, we found different trends between protein evolution and intramodular connectivity, with significant positive correlation in the blue module and negative correlation in the green module. Blue-module genes are overrepresented in processes related to reproduction, chromosome segregation, and therefore higher evolutionary rates in this module are consistent with faster evolution of male-biased genes involved in fertilization in a wide range of species (Ellegren & Parsch, 2007; Grath & Parsch, 2012; Harrison et al., 2015; Parsch & Ellegren, 2013). While the low evolutionary rate in the green module is consistent with the higher pleiotropic constraints in this module, which was also observed in some species (Dean & Mank, 2016; Mank et al., 2008; Meisel, 2011; Moyle et al., 2021; Whittle & Extavour, 2017; Zhang et al., 2007).

Notably, we found that the mantle-associated turquoise module presented slightly higher tissue specificity and connectivity compared to the adductor-associated modules. Moreover, tissue specificity was positively correlated to the intramodular connectivity but negatively correlated to the intermodular connectivity, indicating that these genes are involved in mantle specific functions. Interestingly, mantle is involved in shell biogenesis and immune responses in bivalves (Addadi et al., 2006; Allam & Raftos, 2015; Bettencourt et al., 2017; Song et al., 2019). Indeed, the processes related to metal cluster binding and innate immune response were significantly enriched for genes in the turquoise module. The sequence evolution in this module is also higher than the other somatic related modules, which is consistent with the fast evolving of immune related genes in other organisms (Lazzaro & Clark, 2012; Sackton et al., 2017; Viljakainen et al., 2009).

2.4.6 Contrasting SNPs and genes potentially involved in sex determination

Heteromorphic sex chromosomes are absent in bivalves, and sex determination is thought to be polygenic with the additional influence of environmental factors that can also trigger sex changes (Breton et al., 2018). Recognizing sex of bivalves can be problematic in the early development stages without sex-linked markers. Identification of sex-specific SNPs is crucial for accurate sex diagnosis, breeding, and understanding sex-determination mechanisms (Han et al., 2021; Shi et al., 2018). In this study, we revealed 614 high-confidence contrasting SNPs between males and females, which provide potential genetic markers for sex identification in bivalves. We investigated the functions of genes containing these contrasting SNPs. Interestingly, we found that such genes were overrepresented in the processes like protein targeting, and protein localization to the mitochondrion. These genes included coiled-coil-helix-coiled-coil-helix domain containing 2 (cdchd2), mitochondrial carrier protein Rim2, mitochondrial import inner membrane translocase subunit Tim16, mitochondrial import inner membrane translocase subunit TOM22 (**Supplementary Table 2.1**), and the latter three genes being involved in translocation of nuclear-encoded proteins into mitochondria (Herrmann & Neupert, 2013). While cdchd2 was found to be involved in diverse functions in model animals, including mediating oxidative phosphorylation, responding to hypoxic stress, regulating cell migration, and mitochondrial apoptosis (Kee et al., 2021). Additionally, the gonad-associated green and yellow modules presented higher number of genes containing contrasting SNPs than expected. It has been proposed that DUI bivalves might have an unconventional sex determination/differentiation system that involves mitochondrial genomes, and such process in these organisms may require an appropriate recognition/distinction between mitochondrial and nuclear factors

(Breton et al., 2018; Ghiselli et al., 2013; Milani et al., 2013; Zouros, 2020). Our results are consistent with this hypothesis, and further, more specific, experiments will be necessary to provide direct evidence for this unconventional sex determination.

We also identified some candidate genes that has been associated with the sex determination/differentiation mechanism in other animals (Capt et al., 2019; Evensen et al., 2022; Ghiselli et al., 2012; Svingen & Koopman, 2007; Zhang et al., 2014; Zhou et al., 2019), such as "hub" genes SRY-box transcription factor 30 (sox30), and PBX homeobox 4 (pbx4). Sox30—a putative homolog to mammal sexderteming gene Sry—has been found to be differentially expressed between females and males in many bivalve species (Capt et al., 2019; Ghiselli et al., 2012; Zhang et al., 2014; Zhou et al., 2019), and we found that this gene showed a male-gonad specific transcription in this study. Pbx4 was a "hub" gene in the female gonad-associated module and this gene also presented differential splicing between female and males (**Supplementary Table 2.1**). Pbx4 in mammals have been found to be associated with gametogenesis (Kawai et al., 2018; Svingen & Koopman, 2007; Wagner et al., 2001), however, in this study we found that this gene was specifically transcribed in female gonads. Pbx genes are characterized as hox gene co-activator, and have been found to be associated with oogenesis, embryonic development, germ cell maturation (Svingen & Koopman, 2007).

2.5 Conclusions

In this study, we provided a long-read-based *de novo* genome assembly of a Manila clam from the North American Pacific Coast, providing the insights into the role of differential expression and splicing in bivalve tissues. Although differential splicing was largely overlapping with differential gene expression, it was preferentially involved

in gonad functions, and also presented some “independent” roles. Taking advantage of tissue-specific co-expression network, we were able to detect different features of sex-associated genes. Generally, our results indicated that both differentially expressed and spliced genes were under less constraints, whereas differentially splicing was also detected for genes in the highly pleiotropic module, supporting the hypothesis that splicing is widespread in highly pleiotropic genes. We observed similar trends between tissue-specificity and sequence evolution across tissue-associated modules, and we found that highly tissue-specific genes were characterized by higher sequence evolution. Generally, our results were consistent with a faster sequence evolution in gonad-specific genes, however, we found one of the male-gonad associated modules to be highly connected with other tissues and showing an extremely low sequence evolution, indicating a high pleiotropy effect in this module. Contrasting SNPs between males and females were overrepresented in mitochondrial related functions, further supporting the connection between sex and mitochondrial function in bivalves. Together, this study advances our understanding of the role of differential expression, differential splicing, and sequence evolution of sex-specific genes. We also provided resourceful genomic data for further studies regarding sex diagnosis and breeding.

References

Addadi L, Joester D, Nudelman F, Weiner S. 2006. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. *Chemistry–A European Journal*, **12**: 980–987.

Ahmad TB, Liu L, Kotiw M, Benkendorff K. 2018. Review of anti-inflammatory, immune-modulatory and wound healing properties of molluscs. *Journal of Ethnopharmacology*, **210**: 156–178. DOI: 10.1016/j.jep.2017.08.008

Alexa, A., Rahnenfuhrer, J., 2021. topGO: Enrichment Analysis for Gene Ontology. R package version 2.44.0

Allam B, Raftos D. 2015. Immune responses to infectious diseases in bivalves. *Journal of Invertebrate Pathology*, **131**: 121–136. DOI: 10.1016/j.jip.2015.05.005

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, **215**: 403–410. DOI: 10.1016/S0022-

2836(05)80360-2

Bettencourt R, Barros I, Martins E, Martins I, Cerqueira T, Colaço A, et al. 2017. An Insightful Model to Study Innate Immunity and Stress Response in Deep-Sea Vent Animals: Profiling the Mussel Bathymodiolus azoricus. In *Organismal and Molecular Malacology*. InTech; DOI: 10.5772/68034

Bolger AM, Lohse M, Usadel B. 2014. Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data. **30**: 2114–2120. DOI: 10.1093/bioinformatics/btu170

Boulais M, Demoy-Schneider M, Alavi SMH, Cosson J. 2019. Spermatozoa motility in bivalves: Signaling, flagellar beating behavior, and energetics. *Theriogenology*, **136**: 15–27.

Breton S, Capt C, Guerra D, Stewart D. 2018. Sex-Determining Mechanisms in Bivalves. In *Transitions Between Sexual Systems*. Springer International Publishing:

Cham; 165–192. DOI: 10.1007/978-3-319-94139-4_6

Broquard C, Saowaros S, Lepoittevin M, Degremont L, Lamy J-B, Morga B, et al. 2021. Gonadal transcriptomes associated with sex phenotypes provide potential male

and female candidate genes of sex determination or early differentiation in Crassostrea gigas, a sequential hermaphrodite mollusc. *BMC Genomics*, **22**: 609. DOI: 10.1186/s12864-021-07838-1

Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Calcino AD, Kenny NJ, Gerdol M. 2021. Single individual structural variant detection uncovers widespread hemizygosity in molluscs. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **376**: 20200153. DOI: 10.1098/rstb.2020.0153

Camargo AP, Vasconcelos AA, Fiamenghi MB, Pereira GAG, Carazzolle MF. 2020. tspex: a tissue-specificity calculator for gene expression data. *Preprint available at Research Square*

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, et al. 2014. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiology*, **164**: 513–524. DOI: 10.1104/pp.113.230144

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, et al. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, **18**: 188–196. DOI: 10.1101/gr.6743907

Capt C, Renaut S, Ghiselli F, Milani L, Johnson NA, Sietman BE, et al. 2018. Deciphering the Link between Doubly Uniparental Inheritance of mtDNA and Sex Determination in Bivalves: Clues from Comparative Transcriptomics. *Genome Biology and Evolution*, **10**: 577–590. DOI: 10.1093/gbe/evy019

Capt C, Renaut S, Stewart DT, Johnson NA, Breton S. 2019. Putative Mitochondrial

Sex Determination in the Bivalvia: Insights From a Hybrid Transcriptome Assembly in Freshwater Mussels. *Frontiers in Genetics*, **10.** DOI: 10.3389/fgene.2019.00840

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3 DOI: 10.4161/fly.19695

Cordero D, Delgado M, Liu B, Ruesink J, Saavedra C. 2017. Population genetics of the Manila clam (Ruditapes philippinarum) introduced in North America and Europe. *Scientific Reports*, **7**: 1–13.

Crescente JM, Zavallo D, Helguera M, Vanzetti LS. 2018. MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC Bioinformatics*, **19**: 348. DOI: 10.1186/s12859-018-2376-y

Dean R, Mank JE. 2016. Tissue specificity and sex-specific regulatory variation permit the evolution of sex-biased gene expression. *American Naturalist*, **88**: E74–E84. DOI: 10.1086/687526

DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**: 491–498. DOI: 10.1038/ng.806

Dobin A, Gingeras TR. 2015. Mapping RNA-seq Reads with STAR. *Current Protocols in Bioinformatics*, **51**(1): 11-14. DOI: 10.1002/0471250953.bi1114s51

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

Dunleavy JEM, O'Bryan MK, Stanton PG, O'Donnell L. 2019. The cytoskeleton in spermatogenesis. *Reproduction*, **157**: R53–R72. DOI: 10.1530/REP-18-0457

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, **8**: 689–698. DOI: 10.1038/nrg2167

Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20:**1-14. DOI: 10.1186/s13059-019-1832-y

Evensen KG, Robinson WE, Krick K, Murray HM, Poynton HC. 2022. Comparative phylotranscriptomics reveals putative sex differentiating genes across eight diverse bivalve species. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, **41**: 100952. DOI: 10.1016/j.cbd.2021.100952

Figueras A, Moreira R, Sendra M, Novoa B. 2019. Genomics and immunity of the Mediterranean mussel *Mytilus galloprovincialis* in a changing environment. *Fish & Shellfish Immunology*, **90**: 440–445. DOI: 10.1016/j.fsi.2019.04.064

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, **117**: 9451–9457. DOI: 10.1073/pnas.1921046117

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. Sequence analysis CD-HIT: accelerated for clustering the next-generation sequencing data. **28**: 3150–3152. DOI: 10.1093/bioinformatics/bts565

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907. DOI: 10.48550/arXiv.1207.3907

Gautier M. 2015. Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. Genetics, 201: 1555–1579. DOI: 10.1534/genetics.115.181453

Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. 2020. Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. *Genome Biology*, **21**: 275. DOI: 10.1186/s13059-020-021803

Gershoni M, Pietrokovski S. 2014. Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. *Nature Communications*, **5**: 4438. DOI: 10.1038/ncomms5438

Ghiselli F, Iannello M, Puccio G, Chang PL, Plazzi F, Nuzhdin S V, et al. 2018. Comparative Transcriptomics in Two Bivalve Species Offers Different Perspectives on the Evolution of Sex-Biased Genes. *Genome Biology and Evolution*, **10**: 1389–1402. DOI: 10.6084/m9.figshare.5398618.v1

Ghiselli F, Milani L, Chang PL, Hedgecock D, Davis JP, Nuzhdin S V., et al. 2012. De novo assembly of the Manila clam Ruditapes philippinarum transcriptome provides new insights into expression bias, mitochondrial doubly uniparental inheritance and sex determination. *Molecular Biology and Evolution*, **29**: 771–786. DOI: 10.1093/molbev/msr248

Ghiselli F, Milani L, Guerra D, Chang PL, Breton S, Nuzhdin S V., et al. 2013. Structure, transcription, and variability of metazoan mitochondrial genome:

Perspectives from an unusual mitochondrial inheritance system. *Genome Biology and Evolution*, **5**: 1535–1554. DOI: 10.1093/gbe/evt112

Grantham ME, Brisson JA. 2018. Extensive Differential Splicing Underlies Phenotypically Plastic Aphid Morphs. *Molecular Biology and Eolution*, **35**: 1934–1946. DOI: 10.1093/molbev/msy095

Grath S, Parsch J. 2012. Rate of amino acid substitution is influenced by the degree and conservation of male-biased transcription over 50 myr of Drosophila evolution. *Genome Biology and Evolution*, **4**: 346–359.

Grath S, Parsch J. 2016. Sex-Biased Gene Expression. *Annual Review of Genetics*, **50**: 29–44. DOI: 10.1146/annurev-genet-120215-035429

Griffin RM, Dean R, Grace JL, Rydén P, Friberg U. 2013. The Shared Genome Is a Pervasive Constraint on the Evolution of Sex-Biased Gene Expression. *Molecular Biology and Evolution*, **30**: 2168–2176. DOI: 10.1093/molbev/mst121

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, **36**: 2896–2898. DOI: 10.1093/bioinformatics/btaa025

Guiglielmoni N, Derzelle A, van Doninck K, Flot JF. 2020. Overcoming uncollapsed haplotypes in long-read assemblies of non-model organisms. *bioRxiv*, DOI: 10.1101/2020.03.16.993428

Han Y-L, Sun Z-H, Chang S, Wen B, Song J, Zuo R-T, et al. 2021. Application of SNP in Genetic Sex Identification and Effect of Estradiol on Gene Expression of SexRelated Genes in Strongylocentrotus intermedius. *Frontiers in Endocrinology*, **12** DOI: 10.3389/fendo.2021.756530

Harris RMB, Loeffler F, Rumm A, Fischer C, Horchler P, Scholz M, et al. 2020. Biological responses to extreme weather events are detectable but difficult to formally attribute to anthropogenic climate change. *Scientific Reports*, **10**: 14067. DOI: 10.1038/s41598-020-70901-6

Harrison PW, Wright AE, Zimmer F, Dean R, Montgomery SH, Pointer MA, et al. 2015. Sexual selection drives evolution and rapid turnover of male gene expression.

Proceedings of the National Academy of Sciences of the United States of America, **112**: 4393–4398. DOI: 10.1073/pnas.1501339112

Haszprunar G, Wanninger A. 2012. Molluscs. *Current Biology*, **22**: R510–R514. DOI: 10.1016/j.cub.2012.05.039

Healy TM, Schulte PM. 2019. Patterns of alternative splicing in response to cold acclimation in fish. *Journal of Experimental Biology*, **222** DOI: 10.1242/jeb.193516

Herrmann JM, Neupert W. 2013. Protein Import into Mitochondria. In *Encyclopedia of Biological Chemistry*. Elsevier; 632–636. DOI: 10.1016/B978-0-12-378630-2.00203-6

Hotaling S, Kelley JL, Frandsen PB. 2021. Toward a genome sequence for every animal: Where are we now? Proceedings of the National Academy of Sciences, 118(52). DOI: 10.1073/pnas.2109019118

Insua A, López-Piñón MJ, Méndez J. 1998. Characterization of Aequipecten opercularis (Bivalvia: Pectinidae) chromosomes by different staining techniques and fluorescent in situ hybridization. *Genes & Genetic Systems*, **73**: 193–200.

Jacobs A, Elmer KR. 2021. Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Molecular Ecology*, DOI: 10.1111/mec.15817

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**: 1236–1240. DOI: 10.1093/bioinformatics/btu031

Kawai Y, Oda A, Kanai Y, Goitsuka R. 2018. Germ cell-intrinsic requirement for the homeodomain transcription factor PKnox1/Prep1 in adult spermatogenesis. *PLOS ONE*, **13**: e0190702. DOI: 10.1371/journal.pone.0190702

Kee TR, Espinoza Gonzalez P, Wehinger JL, Bukhari MZ, Ermekbaeva A, Sista A, et al. 2021. Mitochondrial CHCHD2: Disease-Associated Mutations, Physiological Functions, and Current Animal Models. *Frontiers in Aging Neuroscience*, **13** DOI: 10.3389/fnagi.2021.660843

Kokot M, Długosz M, Deorowicz S. 2017. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics*, **33**: 2759–2761. DOI: 10.1093/bioinformatics/btx304 Krishnakumar PK, Qurban MA, Sasikumar G. 2018. Biomonitoring of Trace Metals in the Coastal Waters Using Bivalve Molluscs. In *Trace Elements - Human Health and Environment*. InTech; DOI: 10.5772/intechopen.76938

Kundu R, Joshua C, Sung W-K. 2019. HyPo: Super Fast & Accurate Polisher for Long Read Genome Assemblies. *Biorxiv*. DOI: 10.1101/2019.12.19.882506

Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research*, **6**: 1287. DOI: 10.12688/f1000research.12232.1

Langfelder P, Horvath S. 2008. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**:1-13**.** DOI: 10.1186/1471-2105-9-559

Lazzaro BP, Clark AG. 2012. Rapid evolution of innate immune response genes. In *Rapidly Evolving Genes and Genetic Systems*. Oxford University Press; 203–210. DOI: 10.1093/acprof:oso/9780199642274.003.0020

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**: 3094–3100. DOI: 10.1093/bioinformatics/bty191

Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, et al. 2018. Annotation-free quantification of RNA splicing using LeafCutter. *Nature Genetics*, **50**: 151–158. DOI: 10.1038/s41588-017-0004-9

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**: 923–930. DOI: 10.1093/bioinformatics/btt656

Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, **47**: e47–e47. DOI: 10.1093/nar/gkz114

Lipinska A, Cormier A, Luthringer R, Peters AF, Corre E, Gachon CMM, et al. 2015. Sexual Dimorphism and the Evolution of Sex-Biased Gene Expression in the Brown Alga Ectocarpus. *Molecular Biology and Evolution*, **32**: 1581–1597. DOI: 10.1093/molbev/msv049

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**: 550. DOI: 10.1186/s13059-014-0550-8

Mack KL, Phifer-Rixey M, Harr B, Nachman MW. 2019. Gene expression networks across multiple tissues are associated with rates of molecular evolution in wild house mice. *Genes*, **10**: 225. DOI: 10.3390/genes10030225

Mank JE, Hultin-Rosenberg L, Axelsson E, Ellegren H. 2007. Rapid Evolution of Female-Biased, but Not Male-Biased, Genes Expressed in the Avian Brain. *Molecular Biology and Evolution*, **24**: 2698–2706. DOI: 10.1093/molbev/msm208

Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. 2008. Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. *American Naturalist*, **171**: 35–43. DOI: 10.1086/523954

Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2016. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, 33: 574-576. DOI: 10.1093/bioinformatics/btw663

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology*, **14**: e1005944. DOI: 10.1371/journal.pcbi.1005944

Mauro AA, Ghalambor CK. 2020. Trade-offs, Pleiotropy, and Shared Molecular Pathways: A Unified View of Constraints on Adaptation. *Integrative and Comparative Biology*, **60**: 332–347. DOI: 10.1093/icb/icaa056

McGirr JA, Martin CH. 2018. Parallel evolution of gene expression between trophic specialists despite divergent genotypes and morphologies. *Evolution Letters*, **2**: 62–75. DOI: 10.1002/evl3.41

Mehmood A, Laiho A, Venäläinen MS, Mcglinchey AJ, Wang N, Elo LL. 2020. Systematic evaluation of differential splicing tools for RNA-seq studies. *Briefings in Bioinformatics*, **21**: 2052–2065. DOI: 10.1093/bib/bbz126

Meisel RP. 2011. Towards a More Nuanced Understanding of the Relationship between Sex-Biased Gene Expression and Rates of Protein-Coding Sequence Evolution. *Molecular Biology and Evolution*, **28**: 1893–1900. DOI: 10.1093/molbev/msr010

Merino GA, Conesa A, Ferná Ndez EA. 2019. A benchmarking of workflows for detecting differential splicing and differential expression at isoform level in human

RNA-seq studies. *Briefings in Bioinformatics*, **20**: 471–481. DOI: 10.1093/bib/bbx122

Milani L, Ghiselli F, Nuzhdin S V., Passamonti M. 2013. Nuclear genes with sex bias in Ruditapes philippinarum (Bivalvia, veneridae): Mitochondrial inheritance and sex determination in DUI species. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **320**: 442–454. DOI: 10.1002/jez.b.22520

Modak TH, Literman R, Puritz JB, Johnson KM, Roberts EM, Proestou D, et al. 2021. Extensive genome-wide duplications in the eastern oyster (*Crassostrea virginica*). *Philosophical Transactions of the Royal Society B: Biological Sciences*, **376**: 20200164. DOI: 10.1098/rstb.2020.0164

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, **5**: 621–628. DOI: 10.1038/nmeth.1226

Moyle LC, Wu M, Gibson MJS. 2021. Reproductive Proteins Evolve Faster Than Nonreproductive Proteins Among Solanum Species. *Frontiers in Plant Science*, **12** DOI: 10.3389/fpls.2021.635990

Mun S, Kim Y-J, Markkandan K, Shin W, Oh S, Woo J, et al. 2017. The WholeGenome and Transcriptome of the Manila Clam (Ruditapes philippinarum). *Genome Biology and Evolution*, **9**: 1487–1498. DOI: 10.1093/gbe/evx096

Nattestad M, Schatz MC. 2016. Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics*, **32**: 3021–3023. DOI: 10.1093/bioinformatics/btw369

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, **20**: 275. DOI: 10.1186/s13059-019-1905-y

Papa F, Windbichler N, Waterhouse RM, Cagnetti A, D'Amato R, Persampieri T, et al. 2017. Rapid evolution of female-biased genes among four species of Anopheles

malaria mosquitoes. *Genome Research*, **27**: 1536–1548. DOI: 10.1101/gr.217216.116

Papakostas S, Vøllestad LA, Bruneaux M, Aykanat T, Vanoverbeke J, Ning M, et al. 2014. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. *Nature Communications*, **5**:1-9. DOI: 10.1038/NCOMMS5071

Parsch J, Ellegren H. 2013. The evolutionary causes and consequences of sex-biased gene expression. *Nature Reviews Genetics*, **14**: 83–87. DOI: 10.1038/nrg3376

Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*, **34**: 867–868. DOI: 10.1093/bioinformatics/btx699

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**: 290–295. DOI: 10.1038/nbt.3122

Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, **44**: e113–e113. DOI: 10.1093/nar/gkw294

Rago A, Werren JH, Colbourne JK. 2020. Sex biased expression and co-expression networks in development, using the hymenopteran Nasonia vitripennis. *PLoS Genetics*, **16**: e1008518. DOI: 10.1371/journal.pgen.1008518

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**: e47–e47. DOI: 10.1093/nar/gkv007

Rogers TF, Palmer DH, Wright AE. 2021. Sex-Specific Selection Drives the Evolution of Alternative Splicing in Birds. *Molecular biology and evolution*, **38**: 519–530. DOI: 10.1093/molbev/msaa242

Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. *Nature Methods*,

17: 155–158. DOI: 10.1038/s41592-019-0669-3

Sackton TB, Lazzaro BP, Clark AG. 2017. Rapid expansion of immune-related gene families in the house fly, *Musca domestica*. *Molecular Biology and Evolution*, 34: 857-872. DOI: 10.1093/molbev/msw285

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. In gene prediction. Humana, New York, NY. p227–245. DOI: 10.1007/978-1-4939-9173-0_14

Shi X, Waiho K, Li X, Ikhwanuddin M, Miao G, Lin F, et al. 2018. Female-specific SNP markers provide insights into a WZ/ZZ sex determination system for mud crabs *Scylla paramamosain, S. tranquebarica* and *S. serrata* with a rapid method for genetic sex identification. *BMC Genomics*, **19**: 981. DOI: 10.1186/s12864-018-5380-8

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**: 539. DOI: 10.1038/msb.2011.75

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**: 3210–3212. DOI: 10.1093/bioinformatics/btv351

Song X, Liu Z, Wang L, Song L. 2019. Recent advances of shell matrix proteins and cellular orchestration in marine molluscan shell biomineralization. *Frontiers in Marine Science*, **6**: 41. DOI: 10.3389/fmars.2019.00041

Sun J, Li R, Chen C, Sigwart JD, Kocot KM. 2021. Benchmarking Oxford Nanopore read assemblers for high-quality molluscan genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **376**: 20200160. DOI: 10.1098/rstb.2020.0160

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE*, **6**: e21800. DOI: 10.1371/journal.pone.0021800

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**: W609–W612. DOI: 10.1093/nar/gkl315

Svingen T, Koopman P. 2007. Involvement of Homeobox Genes in Mammalian Sexual Development. *Sexual Development*, **1**: 12–23. DOI: 10.1159/000096235

Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current Protocols in Bioinformatics*, **25** DOI: 10.1002/0471250953.bi0410s25

Tarazona S, FuríO-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, et al. 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Research*, **43**: 140. DOI: 10.1093/nar/gkv711

Telonis-Scott M, Kopp A, Wayne ML, Nuzhdin S V, McIntyre LM. 2009. Sex-Specific Splicing in Drosophila: Widespread Occurrence, Tissue Specificity and Evolutionary Conservation. *Genetics*, **181**: 421–434. DOI: 10.1534/genetics.108.096743

Thiriot-Quiévreux C, Insua A. 1992. Nucleolar organiser region variation in the chromosomes of three oyster species. *Journal of Experimental Marine Biology and Ecology*, **157**: 33–40. DOI: 10.1016/0022-0981(92)90072-I

Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, et al. 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biology*, **19**: 40. DOI: 10.1186/s13059-018-1417-1

Turner LM, Hoekstra HE. 2008. Causes and consequences of the evolution of reproductive proteins. *The International Journal of Developmental Biology*, **52**: 769–780. DOI: 10.1387/ijdb.082577lt

Viljakainen L, Evans JD, Hasselmann M, Rueppell O, Tingek S, Pamilo P. 2009. Rapid Evolution of Immune Proteins in Social Insects. *Molecular Biology and Evolution*, **26**: 1791–1801. DOI: 10.1093/molbev/msp086

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, **33**: 2202–2204. DOI: 10.1093/bioinformatics/btx153

Wagner K, Mincheva A, Korn B, Lichter P, Pöpperl H. 2001. Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. *Mechanisms of Development*, **103**: 127–131. DOI: 10.1016/S0925-4773(01)00349-5

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010a. KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics & Bioinformatics*, **8**: 77–80. DOI: 10.1016/S1672-0229(10)60008-3

Wang K, Li M, Hakonarson H. 2010b. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, **38**: e164–e164. DOI: 10.1093/nar/gkq603

Wang X, Chen S, Shi X, Liu D, Zhao P, Lu Y, et al. 2019. Hybrid sequencing reveals insight into heat sensing and signaling of bread wheat. *Plant Journal*, **98**: 1015–1032. DOI: 10.1111/tpj.14299

Wei M, Ge H, Shao C, Yan X, Nie H, Duan H, et al. 2020. Chromosome-Level Clam Genome Helps Elucidate the Molecular Basis of Adaptation to a Buried Lifestyle. *iScience*, 23(6): 101148. DOI: 10.1016/j.isci.2020.101148

Whittle CA, Extavour CG. 2017. Rapid evolution of ovarian-biased genes in the yellow fever mosquito (Aedes Aegypti). *Genetics*, **206**: 2119–2137. DOI: 10.1534/genetics.117.201343

Whittle CA, Extavour CG. 2019. Selection shapes turnover and magnitude of sex-biased expression in Drosophila gonads. *BMC Evolutionary Biology*, **19**: 1–20. DOI: 10.1186/s12862-019-1377-4

Whittle CA, Johannesson H. 2013. Evolutionary Dynamics of Sex-Biased Genes in a Hermaphrodite Fungus. *Molecular Biology and Evolution*, **30**: 2435–2446. DOI: 10.1093/molbev/mst143

Yan X, Nie H, Huo Z, Ding J, Li Z, Yan L, et al. 2019. Clam Genome Sequence Clarifies the Molecular Basis of Its Benthic Adaptation and Extraordinary Shell Color Diversity. *iScience*, **19**: 1225–1237. DOI: 10.1016/j.isci.2019.08.049

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*, **21**: 650–659. DOI: 10.1093/bioinformatics/bti042

Yang L, Zhang Z, He S. 2016. Both Male-Biased and Female-Biased Genes Evolve Faster in Fish Genomes. *Genome Biology and Evolution*, **8**: 3433–3445. DOI: 10.1093/gbe/evw239

Yue C, Li Q, Yu H. 2018. Gonad Transcriptome Analysis of the Pacific Oyster Crassostrea gigas Identifies Potential Genes Regulating the Sex Determination and Differentiation Process. *Marine Biotechnology*, **20**: 206–219. DOI: 10.1007/s10126-018-9798-4

Zhang N, Xu F, Guo X. 2014. Genomic Analysis of the Pacific Oyster (*Crassostrea gigas*) Reveals Possible Conservation of Vertebrate Sex Determination in a Mollusc. *G3 Genes|Genomes|Genetics*, **4**: 2207–2217. DOI: 10.1534/g3.114.013904

Zhang W-J, Zhou J, Li Z-F, Wang L, Gu X, Zhong Y. 2007. Comparative Analysis of Codon Usage Patterns Among Mitochondrion, Chloroplast and Nuclear Genes in Triticum aestivum L. *Journal of Integrative Plant Biology*, **49**: 246–254. DOI: 10.1111/j.1672-9072.2007.00404.x

Zhang Y, Nyong'A TM, Shi T, Yang P. 2019. The complexity of alternative splicing and landscape of tissue-specific expression in lotus (Nelumbo nucifera) unveiled by Illumina- and single-molecule real-time-based RNA-sequencing. *DNA research: an international journal for rapid publication of reports on genes and genomes*, **26**: 301– 311. DOI: 10.1093/dnares/dsz010

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, **28**: 3326–3328. DOI: 10.1093/bioinformatics/bts606

Zhou Q, Wang T, Leng L, Zheng W, Huang J, Fang F, et al. 2019. Single-cell RNAseq reveals distinct dynamic behavior of sex chromosomes during early human embryogenesis. *Molecular Reproduction and Development*, **86**: 871–882. DOI: 10.1002/mrd.23162

Zouros E. 2020. Doubly uniparental inheritance of mitochondrial DNA: Might it be simpler than we thought? DOI: 10.1111/jzs.12364

Supplementary Files

Supplementary Table 2.1 Genes belonging to different gene sets.

Gene ID	Annotation	DEG	DSG	module	contrasting SNP
RPHIG00000016546	regulator of G protein signaling like 1	+	-	blue(hub)	-
RPHIG00000025695	BTB domain containing 9	+	-	blue(hub)	-
RPHIG00000009761	membrane associated ring-CH-type finger 2	+	+	blue(hub)	-
RPHIG00000011963	membrane associated ring-CH-type finger 8	+	-	blue(hub)	-
RPHIG00000007409	BTB domain containing 2	+	+	blue(hub)	-
RPHIG00000030063	a disintegrin and metalloproteinase with thrombospondin motif 18	+	-	blue(hub)	-
RPHIG00000025334	ST14 transmembrane serine protease matriptase	+	-	blue(hub)	-
RPHIG00000031121	solute carrier family 44, member 1	+	-	blue(hub)	-
RPHIG00000008958	zinc finger protein 83-like	+	-	blue(hub)	-
RPHIG00000014300	cell division cycle 25 A	+	-	blue(hub)	-
RPHIG00000019692	zinc finger DHHC-type palmitoyltransferase 14	+	-	blue(hub)	-
RPHIG00000003222	PTI1-like tyrosine-protein kinase 3	+	-	blue(hub)	-
RPHIG00000010879	Von Willebrand Factor C Domain Containing 2 Like	+	-	blue(hub)	-
RPHIG00000020064	kelch like family member 10	+	-	blue(hub)	-
RPHIG00000030775	testis specific serine kinase 4	+	-	blue(hub)	-
RPHIG00000017413	SRY-box transcription factor 30	+	-	blue(hub)	-
RPHIG00000025696	BTB domain containing 1	+	-	blue(hub)	-
RPHIG00000003978	MOS proto-oncogene, serine/threonine kinase	+	-	pink(hub)	-
RPHIG00000017162	mating-type-like protein ALPHA2 homolog	+	+	pink(hub)	-
RPHIG00000007500	PBX homeobox 4	+	+	pink(hub)	-
RPHIG00000021761	mannose receptor C-type 1	+	+	pink(hub)	-
RPHIG00000011124	hes related family bHLH transcription factor with YRPW motif	+	-	pink(hub)	-
RPHIG00000011387	aminomethyltransferase 1	+	+	pink(hub)	-
RPHIG00000031224	cell division cycle 20	+	+	pink(hub)	-
RPHIG00000016153	cyclin B2	+	-	pink(hub)	-
RPHIG00000013311	RIB43A domain with coiled-coils 2	+	-	green(hub)	-
RPHIG00000007499	radial spoke head component 4A	+	-	green(hub)	-
RPHIG00000000500	stathmin 1	+	-	green(hub)	-
RPHIG00000010829	EF-hand domain containing 2	+	-	green(hub)	-
RPHIG00000022450	coiled-coil domain-containing protein 87	+	-	green(hub)	-
RPHIG00000008004	cilia- and flagella-associated protein 53	+	-	green(hub)	-
RPHIG00000002522	tektin-1	+	-	green(hub)	-
RPHIG00000012886	dynein heavy chain 7, axonemal-like	+	-	green(hub)	-
RPHIG00000000097	adenylate kinase 9	+	+	green(hub)	+
RPHIG00000004633	maats1	+	-	green(hub)	-

RPHIG00000010180	radial spoke head 3	+	-	green(hub)	-
RPHIG00000004987	growth arrest specific 8	+	-	green(hub)	-
RPHIG00000005899	dynein regulatory complex subunit 7	+	+	green(hub)	-
RPHIG00000020354	dynein regulatory complex subunit 1	+	-	green(hub)	-
RPHIG00000017669	IQ and ubiquitin-like domain-containing	+	-	green(hub)	-
RPHIG00000010862	spermatogenesis associated 17	+	-	green(hub)	-
RPHIG00000022292	tektin2	+	-	green(hub)	-
RPHIG00000012449	EF-hand calcium binding domain 5	+	+	green(hub)	+
RPHIG00000023582	enkurin, TRPC channel interacting protein	+	-	green(hub)	+
RPHIG00000010250	radial spoke head component 1	+	-	green(hub)	-
RPHIG00000030762	ankyrin repeat domain-containing protein 17	+	+	pink	+
RPHIG00000003170	double-strand-break repair protein rad21-like protein	+	+	green	+
RPHIG00000004696	folliculin	-	+	green	+
RPHIG00000011840	protein polybromo-1	-	+	green	+
RPHIG00000019988	transcriptional regulator ATRX	-	-	pink	+
RPHIG00000002994	tetratricopeptide repeat protein 30A	-	+	yellow	+
RPHIG00000027276	minichromosome maintenance complex component 3 associated protein	-	-	-	+
RPHIG00000004284	mitochondrial import inner membrane translocase subunit Tim16, mitochondrial	+	-	-	+
RPHIG00000023473	regulation of nuclear pre-mRNA domain-containing protein 2-like	-	-	-	+
RPHIG00000003170	double-strand-break repair protein rad21-like protein	+	+	green	+
RPHIG00000001490	chromodomain-helicase-DNA-binding protein 9	+	-	pink	-
RPHIG00000012597	mitochondrial carrier protein Rim2, mitochondrial	-	-	-	+
RPHIG00000003491	general transcription factor IIA subunit 1	-	-	-	+
RPHIG00000006937	far upstream element-binding protein 1	-	-	-	+
RPHIG00000015953	coiled-coil-helix-coiled-coil-helix domain containing 2, mitochondrial	-	-	-	+
RPHIG00000004222	mitochondrial import inner membrane translocase subunit TOM22	-	-	-	+

Supplementary Table 2.2 GO enrichment for different subset of genes.

Gene sets	Type	GO_id	GO Terms	P value
DE: f_G vs. m_G	BP	GO:0007017	microtubule-based process	< 1e-30
		GO:0006928	movement of cell or subcellular componen...	2.8E-30
		GO:0007018	microtubule-based movement	1.6E-27
		GO:0030030	cell projection organization	1.4E-09
		GO:0120036	plasma membrane bounded cell projection ...	1.4E-09
		GO:0070925	organelle assembly	2.9E-07
		GO:0007049	cell cycle	5E-07
		GO:0040011	locomotion	2.2E-06
		GO:0051674	localization of cell	7.4E-06
		GO:0022402	cell cycle process	0.000012
		GO:0007010	cytoskeleton organization	0.000017
		GO:0032886	regulation of microtubule-based process	0.00016
		GO:0006091	generation of precursor metabolites and ...	0.00019
		GO:0009987	cellular process	0.00042
		GO:0006090	pyruvate metabolic process	0.00046
		GO:0000003	reproduction	0.00055
		GO:0022414	reproductive process	0.00055
		GO:0046390	ribose phosphate biosynthetic process	0.00056
		GO:0017144	drug metabolic process	0.00077
		GO:0051726	regulation of cell cycle	0.00077
		GO:0051656	establishment of organelle localization	0.0015
		GO:0009201	Ribonucleoside triphosphate biosynthetic...	0.00194
		GO:0072521	purine-containing compound metabolic pro...	0.00204
		GO:0016052	carbohydrate catabolic process	0.00294
		GO:0030705	cytoskeleton-dependent intracellular tra...	0.00294
		GO:0019693	ribose phosphate metabolic process	0.00303
		GO:0032787	monocarboxylic acid metabolic process	0.00329
		GO:0006261	DNA-dependent DNA replication	0.00348
		GO:0051298	centrosome duplication	0.00348
		GO:0046034	ATP metabolic process	0.00352
		GO:0006836	neurotransmitter transport	0.00446
		GO:0051493	regulation of cytoskeleton organization	0.00449
		GO:0022008	neurogenesis	0.00504
		GO:0051128	regulation of cellular component organiz...	0.00621
		GO:0009123	nucleoside monophosphate metabolic proce...	0.00716
		GO:0009141	nucleoside triphosphate metabolic proces...	0.0076
		GO:0006996	organelle organization	0.00813
		GO:0032270	positive regulation of cellular protein ...	0.00817
	MF	GO:0003774	motor activity	1.5E-23
		GO:0003777	microtubule motor activity	1.5E-23

		GO:0017111	nucleoside-triphosphatase activity	0.000011
		GO:0016817	hydrolase activity, acting on acid anhyd...	0.000013
		GO:0008092	cytoskeletal protein binding	0.000037
		GO:0019205	nucleobase-containing compound kinase ac...	0.00124
		GO:0051015	actin filament binding	0.00134
		GO:0031267	small GTPase binding	0.00326
		GO:0016301	kinase activity	0.00525
		GO:0044877	protein-containing complex binding	0.00547
		GO:0016491	oxidoreductase activity	0.0058
		GO:0009055	electron transfer activity	0.0078
	CC	GO:0005856	cytoskeleton	9.6E-11
		GO:0015630	microtubule cytoskeleton	4E-08
		GO:0099080	supramolecular complex	1.3E-07
		GO:0099081	supramolecular polymer	1.3E-07
		GO:0042995	cell projection	0.00014
		GO:0120025	plasma membrane bounded cell projection	0.00014
		GO:0043228	non-membrane-bounded organelle	0.00484
		GO:0043232	intracellular non-membrane-bounded organ...	0.00484
		GO:0030286	dynein complex	0.00613
DS female vs male gonads	BP	GO:0007018	microtubule-based movement	9E-07
		GO:0006928	movement of cell or subcellular componen...	5.7E-06
		GO:0007017	microtubule-based process	0.000037
		GO:0030030	cell projection organization	0.00142
		GO:0120036	plasma membrane bounded cell projection ...	0.00142
		GO:0070925	organelle assembly	0.00306
		GO:0006338	chromatin remodeling	0.00559
		GO:0051347	positive regulation of transferase activ...	0.00918
	MF	GO:0044877	protein-containing complex binding	0.000041
		GO:0003774	motor activity	0.000075
		GO:0003777	microtubule motor activity	0.000075
		GO:0008092	cytoskeletal protein binding	0.00053
		GO:0051015	actin filament binding	0.00083
		GO:0003690	double-stranded DNA binding	0.00162
	CC	GO:0042995	cell projection	0.0036
		GO:0120025	plasma membrane bounded cell projection	0.0036
Contrasting SNPs	BP	GO:0033365	protein localization to organelle	0.00031
		GO:0017038	protein import	0.00089
		GO:1990542	mitochondrial transmembrane transport	0.00161
		GO:0044743	protein transmembrane import into intrac...	0.00211
		GO:0070585	protein localization to mitochondrion	0.0027
		GO:0072655	establishment of protein localization to...	0.0027
		GO:0045184	establishment of protein localization	0.00291
		GO:0034613	cellular protein localization	0.00363

		GO:0070727	cellular macromolecule localization	0.00363
		GO:0006605	protein targeting	0.00375
		GO:0006839	mitochondrial transport	0.00416
		GO:0071806	protein transmembrane transport	0.00416
		GO:0006886	intracellular protein transport	0.00647
	CC	GO:0044441	ciliary part	0.0079
		GO:0044463	cell projection part	0.0088
		GO:0120038	plasma membrane bounded cell projection ...	0.0088
Module_Pink	BP	GO:0051347	positive regulation of transferase activ...	0.000078
		GO:0031401	positive regulation of protein modificat...	0.0001
		GO:0051338	regulation of transferase activity	0.00026
		GO:0051247	positive regulation of protein metabolic...	0.00032
		GO:0032268	regulation of cellular protein metabolic...	0.00043
		GO:0031399	regulation of protein modification proce...	0.00046
		GO:0051246	regulation of protein metabolic process	0.00048
		GO:0006259	DNA metabolic process	0.00108
		GO:0043085	positive regulation of catalytic activit...	0.00168
		GO:0016051	carbohydrate biosynthetic process	0.00189
		GO:0044093	positive regulation of molecular functio...	0.00262
		GO:0051173	positive regulation of nitrogen compound...	0.0029
		GO:0051171	regulation of nitrogen compound metaboli...	0.00379
		GO:0060255	regulation of macromolecule metabolic pr...	0.00398
		GO:0080090	regulation of primary metabolic process	0.00402
		GO:0006261	DNA-dependent DNA replication	0.00519
		GO:0031323	regulation of cellular metabolic process	0.00589
		GO:0019222	regulation of metabolic process	0.00597
		GO:0009893	positive regulation of metabolic process	0.00803
	MF	GO:0044877	protein-containing complex binding	0.0021
Module_Magenta	BP	GO:0007155	cell adhesion	0.000094
		GO:0022610	biological adhesion	0.000094
		GO:0031099	regeneration	0.0016
		GO:0042246	tissue regeneration	0.0016
		GO:0048589	developmental growth	0.0016
		GO:0040007	growth	0.0019
		GO:0006364	rRNA processing	0.0052
		GO:0009888	tissue development	0.0052
		GO:0016072	rRNA metabolic process	0.0064
		GO:0034470	ncRNA processing	0.0089
Module_Purple	BP	GO:0007160	cell-matrix adhesion	8.9E-06
		GO:0031589	cell-substrate adhesion	9.9E-06
		GO:0006826	iron ion transport	0.00013
		GO:0006879	cellular iron ion homeostasis	0.00013
		GO:0055072	iron ion homeostasis	0.00013

		GO:0007155	cell adhesion	0.00015
		GO:0022610	biological adhesion	0.00015
		GO:0055076	transition metal ion homeostasis	0.00018
		GO:0000041	transition metal ion transport	0.00023
		GO:0050801	ion homeostasis	0.0003
		GO:0019725	cellular homeostasis	0.00045
		GO:0042592	homeostatic process	0.00115
		GO:0065008	regulation of biological quality	0.00894
	MF	GO:0046872	metal ion binding	0.000029
		GO:0043169	cation binding	0.00003
		GO:0043167	ion binding	0.00078
		GO:0005509	calcium ion binding	0.00083
	CC	GO:0005576	extracellular region	0.000004
Module_Red	BP	GO:0044877	protein-containing complex binding	0.00103
		GO:0099003	vesicle-mediated transport in synapse	0.00145
		GO:0099504	synaptic vesicle cycle	0.00145
		GO:0001505	regulation of neurotransmitter levels	0.00244
		GO:0007269	neurotransmitter secretion	0.00244
		GO:0099643	signal release from synapse	0.00244
		GO:0006836	neurotransmitter transport	0.00305
		GO:0007010	cytoskeleton organization	0.00447
		GO:0065008	regulation of biological quality	0.00535
		GO:0019722	calcium-mediated signaling	0.00591
		GO:0019932	second-messenger-mediated signaling	0.00591
		GO:0003008	system process	0.00643
		GO:0032787	monocarboxylic acid metabolic process	0.00751
		GO:0007268	chemical synaptic transmission	0.0087
		GO:0099536	synaptic signaling	0.0087
		GO:0006952	defense response	0.00999
	MF	GO:0051015	actin filament binding	3.8E-06
		GO:0003779	actin binding	0.000032
		GO:0008092	cytoskeletal protein binding	0.00012
		GO:0005509	calcium ion binding	0.00199
		GO:0043169	cation binding	0.00726
	CC	GO:0016021	integral component of membrane	0.0073
		GO:0031224	intrinsic component of membrane	0.0081
Module_Turquoise	BP	GO:0098660	inorganic ion transmembrane transport	0.00016
		GO:0015698	inorganic anion transport	0.00041
		GO:0031347	regulation of defense response	0.0006
		GO:0007154	cell communication	0.0006
		GO:0098609	cell-cell adhesion	0.00061
		GO:0023052	signaling	0.00071
		GO:0032101	regulation of response to external stimu...	0.00088

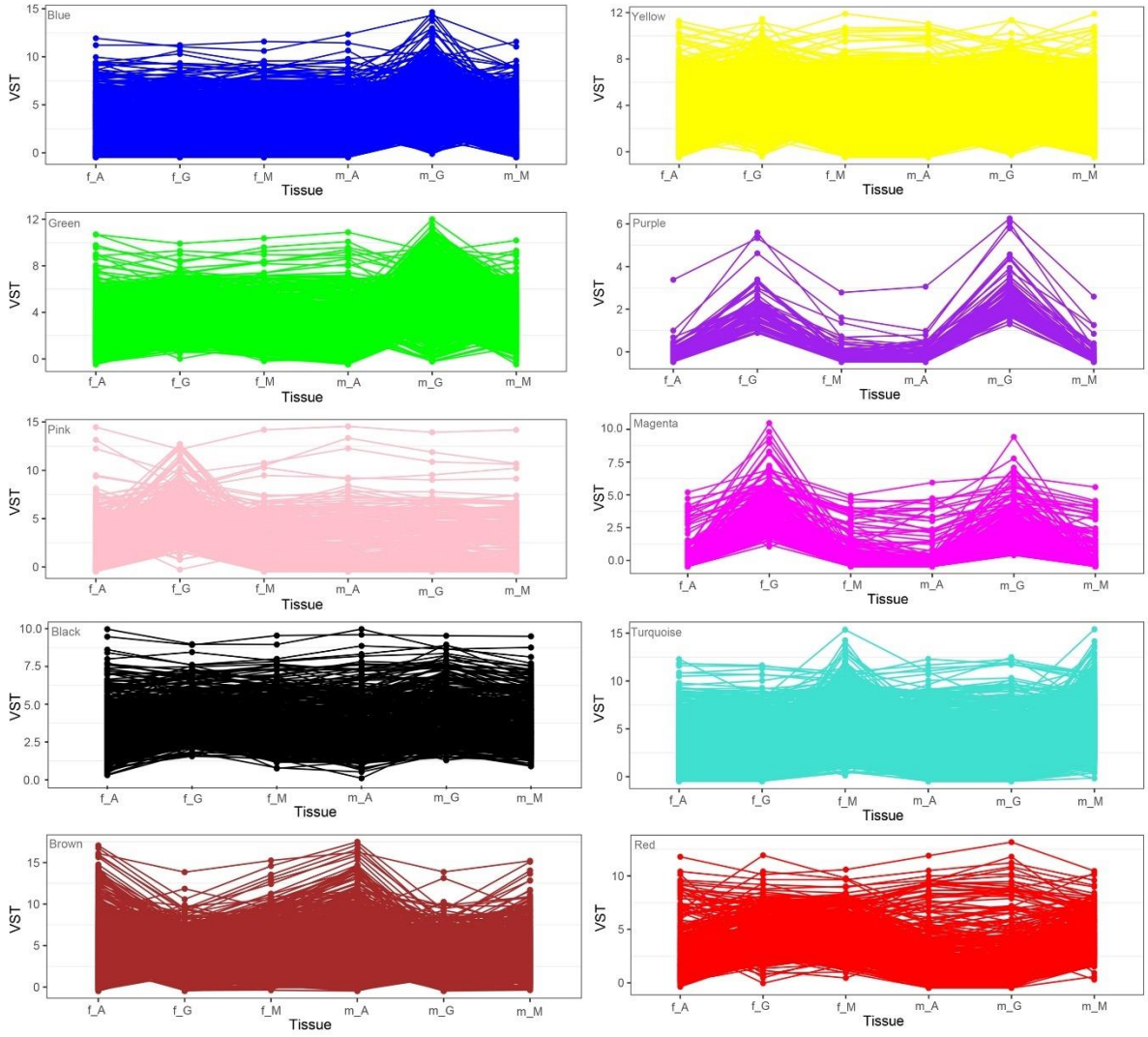
		GO:0006508	proteolysis	0.00136
		GO:0007165	signal transduction	0.00143
		GO:0009308	amine metabolic process	0.00198
		GO:0001816	cytokine production	0.00228
		GO:0001817	regulation of cytokine production	0.00228
		GO:0002831	regulation of response to biotic stimulu...	0.00228
		GO:0045087	innate immune response	0.00228
		GO:0002682	regulation of immune system process	0.003
		GO:0006952	defense response	0.003
		GO:0006811	ion transport	0.00327
		GO:0051239	regulation of multicellular organismal p...	0.00388
		GO:0055085	transmembrane transport	0.00392
		GO:0000041	transition metal ion transport	0.00441
		GO:0009605	response to external stimulus	0.00494
		GO:0009141	nucleoside triphosphate metabolic proces...	0.00602
		GO:0048583	regulation of response to stimulus	0.00697
		GO:0009607	response to biotic stimulus	0.00981
		GO:0044419	biological process involved in interspec...	0.00981
	MF	GO:0030414	peptidase inhibitor activity	2.9E-08
		GO:0016491	oxidoreductase activity	4.6E-06
		GO:0005507	copper ion binding	0.00003
		GO:0015103	inorganic anion transmembrane transporte...	0.000035
		GO:0005044	scavenger receptor activity	0.000078
		GO:0051536	iron-sulfur cluster binding	0.00029
		GO:0051540	metal cluster binding	0.00029
		GO:0048038	quinone binding	0.0004
		GO:0016641	oxidoreductase activity, acting on the C...	0.00041
		GO:0016638	oxidoreductase activity, acting on the C...	0.00058
		GO:0046914	transition metal ion binding	0.00119
		GO:0043169	cation binding	0.00121
		GO:0051213	dioxygenase activity	0.00146
		GO:0046872	metal ion binding	0.00187
		GO:0016702	oxidoreductase activity, acting on singl...	0.00191
		GO:0015291	secondary active transmembrane transport...	0.00208
		GO:0016701	oxidoreductase activity, acting on singl...	0.00298
		GO:0004089	carbonate dehydratase activity	0.00502
		GO:0022804	active transmembrane transporter activit...	0.00708
		GO:0008233	peptidase activity	0.0076
	CC	GO:0016020	membrane	0.0023
		GO:0005576	extracellular region	0.0002
Module_Yellow	BP	GO:0051276	chromosome organization	2.2E-14
		GO:0090304	nucleic acid metabolic process	3.5E-13
		GO:0022402	cell cycle process	6.8E-13

		GO:0006139	nucleobase-containing compound metabolic...	3.7E-12
		GO:0006996	organelle organization	6.4E-12
		GO:0046483	heterocycle metabolic process	7.2E-12
		GO:0006725	cellular aromatic compound metabolic pro...	1.9E-11
		GO:1901360	organic cyclic compound metabolic proces...	3.5E-11
		GO:0006259	DNA metabolic process	4.7E-11
		GO:0007049	cell cycle	9.7E-11
		GO:0007059	chromosome segregation	2.5E-10
		GO:0071840	cellular component organization or bioge...	2.4E-09
		GO:0006260	DNA replication	5.8E-09
		GO:0048285	organelle fission	8.5E-09
		GO:0006281	DNA repair	9.4E-09
		GO:0034641	cellular nitrogen compound metabolic pro...	1.5E-08
		GO:0009987	cellular process	1.4E-06
		GO:0006950	response to stress	1.7E-06
		GO:0016070	RNA metabolic process	7.9E-06
		GO:0030261	chromosome condensation	0.000012
		GO:0044260	cellular macromolecule metabolic process	0.000018
		GO:0043170	macromolecule metabolic process	0.000021
		GO:0071824	protein-DNA complex subunit organization	0.000023
		GO:0032774	RNA biosynthetic process	0.000081
		GO:0006323	DNA packaging	0.00011
		GO:0006261	DNA-dependent DNA replication	0.00016
		GO:0007017	microtubule-based process	0.00018
		GO:0034654	nucleobase-containing compound biosynthe...	0.00025
		GO:0006338	chromatin remodeling	0.00043
		GO:0018130	heterocycle biosynthetic process	0.00062
		GO:0010564	regulation of cell cycle process	0.00067
		GO:0006351	transcription, DNA-templated	0.00074
		GO:0019438	aromatic compound biosynthetic process	0.00081
		GO:0070925	organelle assembly	0.00093
		GO:1901362	organic cyclic compound biosynthetic pro...	0.00099
		GO:0065004	protein-DNA complex assembly	0.00114
		GO:0000226	microtubule cytoskeleton organization	0.00116
		GO:0000724	double-strand break repair via homologou...	0.00258
		GO:0000725	recombinational repair	0.00258
		GO:0044770	cell cycle phase transition	0.00258
		GO:0044237	cellular metabolic process	0.00353
		GO:0010467	gene expression	0.00397
		GO:0016071	mRNA metabolic process	0.00404
		GO:0006397	mRNA processing	0.00417
		GO:0034645	cellular macromolecule biosynthetic proc...	0.00594
		GO:0009059	macromolecule biosynthetic process	0.00614

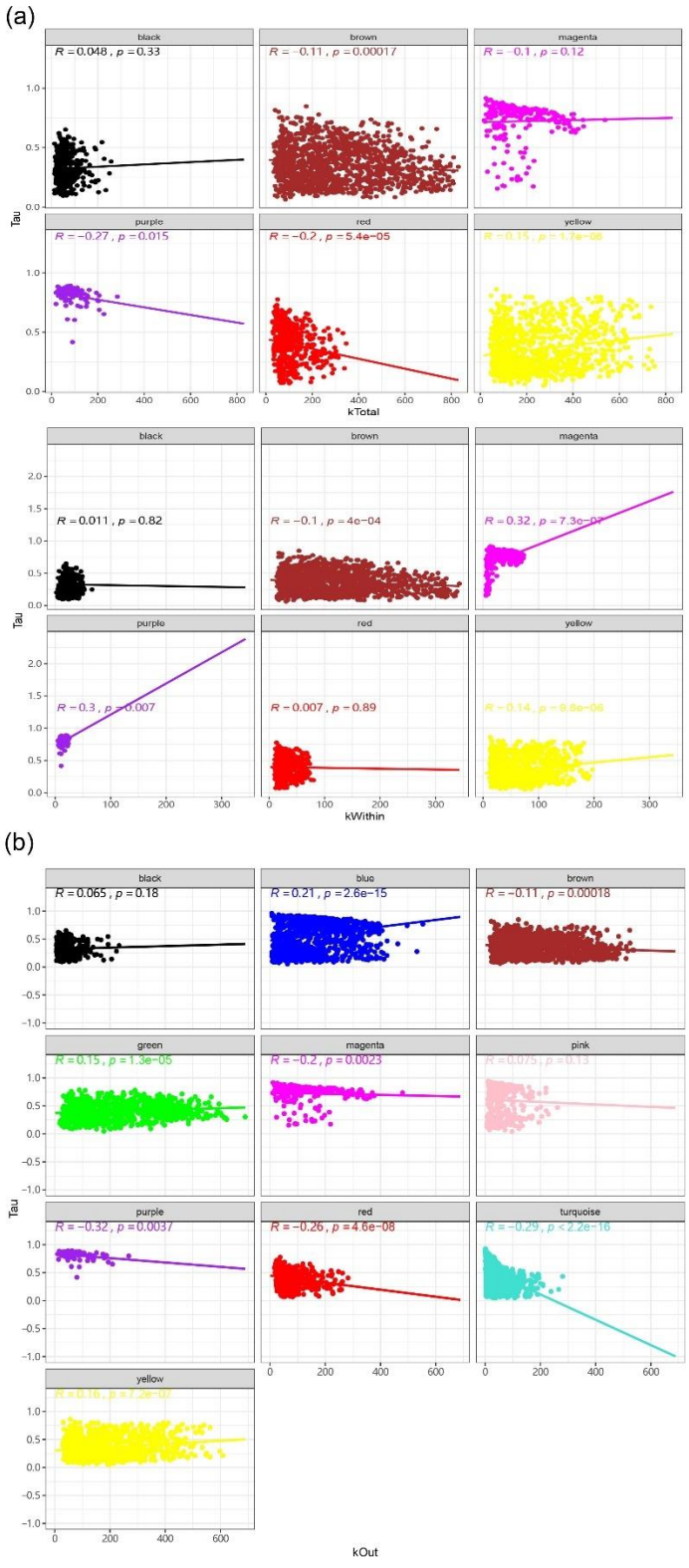
		GO:0006928	movement of cell or subcellular componen...	0.00746
		GO:0051726	regulation of cell cycle	0.00896
	MF	GO:1901363	heterocyclic compound binding	1E-12
		GO:0097159	organic cyclic compound binding	1.5E-12
		GO:0003676	nucleic acid binding	6.8E-09
		GO:0003899	DNA-directed 5'-3' RNA polymerase activi...	7.6E-08
		GO:0016779	nucleotidyltransferase activity	2.2E-07
		GO:0005488	binding	2.5E-06
		GO:0003677	DNA binding	8.2E-06
		GO:0097367	carbohydrate derivative binding	0.000017
		GO:0036094	small molecule binding	0.000049
		GO:0032559	adenyl ribonucleotide binding	0.00007
		GO:0003723	RNA binding	0.0001
		GO:0003678	DNA helicase activity	0.00013
		GO:0016772	transferase activity, transferring phosp...	0.00014
		GO:0140097	catalytic activity, acting on DNA	0.00016
		GO:0004386	helicase activity	0.00041
		GO:0003774	motor activity	0.00052
		GO:0003777	microtubule motor activity	0.00052
		GO:0017076	purine nucleotide binding	0.00091
		GO:1901265	nucleoside phosphate binding	0.00165
		GO:0016817	hydrolase activity, acting on acid anhyd...	0.00214
		GO:0008135	translation factor activity, RNA binding	0.00243
		GO:0017111	nucleoside-triphosphatase activity	0.00263
		GO:0035639	purine ribonucleoside triphosphate bindi...	0.00283
		GO:0043168	anion binding	0.00438
		GO:0003697	single-stranded DNA binding	0.00751
		GO:0017056	structural constituent of nuclear pore	0.01
	CC	GO:0005694	chromosome	3E-18
		GO:0005634	nucleus	1.6E-14
		GO:0005622	intracellular anatomical structure	8.3E-13
		GO:0043229	intracellular organelle	1.8E-12
		GO:0043228	non-membrane-bounded organelle	1.3E-11
		GO:0043232	intracellular non-membrane-bounded organ...	1.3E-11
		GO:0032991	protein-containing complex	1.8E-11
		GO:0043226	organelle	1.9E-11
		GO:0140513	nuclear protein-containing complex	6.6E-10
		GO:0000228	nuclear chromosome	8.1E-10
		GO:0043231	intracellular membrane-bounded organelle	4.6E-08
		GO:0043227	membrane-bounded organelle	3.2E-07
		GO:0031974	membrane-enclosed lumen	1.5E-06
		GO:0000775	chromosome, centromeric region	5.9E-06
		GO:1902494	catalytic complex	0.000059

		GO:1904949	ATPase complex	0.00015
		GO:0070603	SWI/SNF superfamily-type complex	0.00027
		GO:1902562	H4 histone acetyltransferase complex	0.00044
		GO:0005667	transcription regulator complex	0.00129
		GO:0099080	supramolecular complex	0.00193
		GO:0033202	DNA helicase complex	0.00322
		GO:0090575	RNA polymerase II transcription regulato...	0.00416
Module_Blue	BP	GO:0000003	reproduction	3.8E-09
		GO:0022414	reproductive process	3.8E-09
		GO:0051704	multi-organism process	0.000054
		GO:0007049	cell cycle	0.000056
		GO:0000280	nuclear division	0.000091
		GO:0098813	nuclear chromosome segregation	0.000091
		GO:0007017	microtubule-based process	0.00012
		GO:0048285	organelle fission	0.00017
		GO:0007059	chromosome segregation	0.00029
		GO:0010970	transport along microtubule	0.00043
		GO:0030705	cytoskeleton-dependent intracellular tra...	0.00043
		GO:0051656	establishment of organelle localization	0.00078
		GO:0030031	cell projection assembly	0.00095
		GO:0051640	organelle localization	0.00151
		GO:0006928	movement of cell or subcellular componen...	0.00319
		GO:0030030	cell projection organization	0.00552
		GO:0070925	organelle assembly	0.00741
	MF	GO:0008092	cytoskeletal protein binding	0.0013
		GO:0005488	binding	0.0079
	CC	GO:0005856	cytoskeleton	0.000026
		GO:0043228	non-membrane-bounded organelle	0.00028
		GO:0043232	intracellular non-membrane-bounded organ...	0.00028
		GO:0099080	supramolecular complex	0.00036
		GO:0015630	microtubule cytoskeleton	0.00036
		GO:0099081	supramolecular polymer	0.00042
		GO:0000502	proteasome complex	0.00047
		GO:0000793	condensed chromosome	0.0013
Module_Green	BP	GO:0007018	microtubule-based movement	< 1e-30
		GO:0006928	movement of cell or subcellular componen...	< 1e-30
		GO:0007017	microtubule-based process	< 1e-30
		GO:0070925	organelle assembly	3.7E-08
		GO:0030030	cell projection organization	5.6E-08
		GO:0120036	plasma membrane bounded cell projection ...	5.6E-08
		GO:0022402	cell cycle process	0.00001
		GO:0007049	cell cycle	0.00013
		GO:0007010	cytoskeleton organization	0.00034

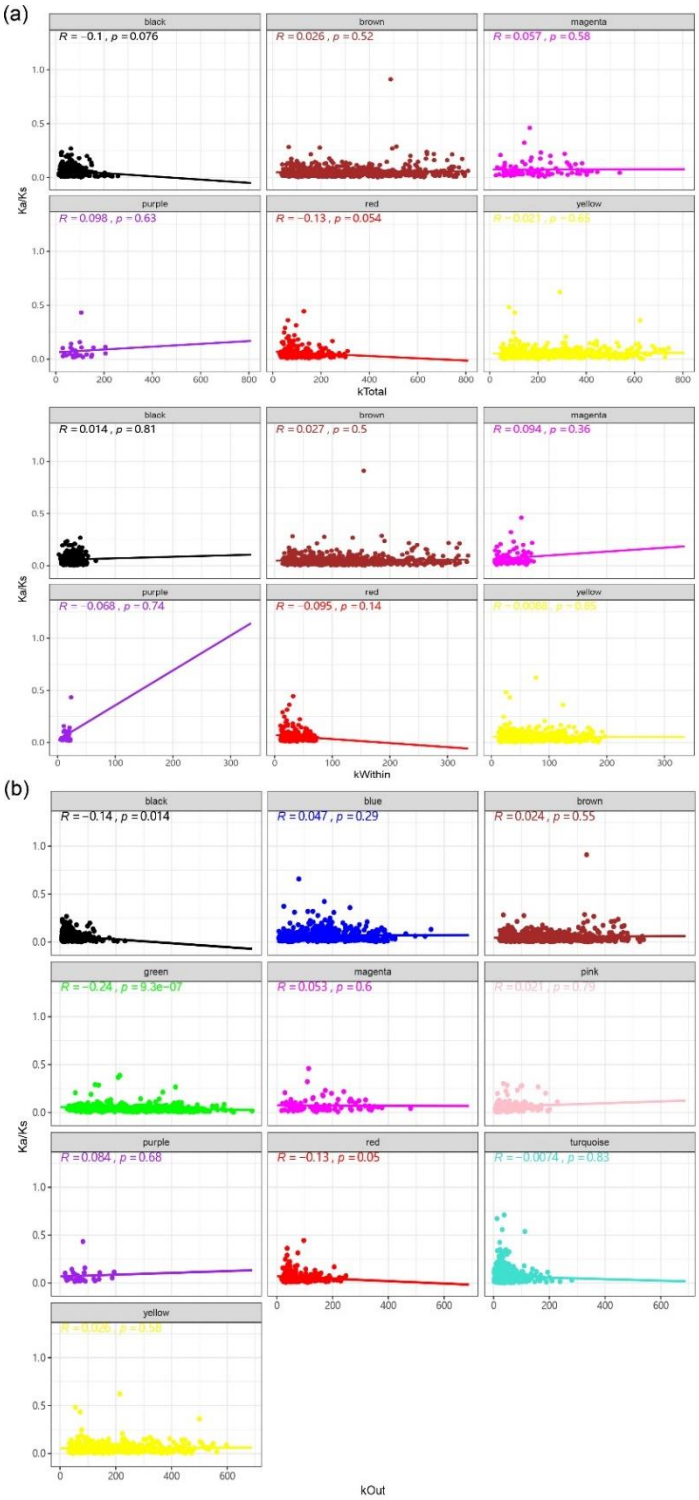
		GO:0051674	localization of cell	0.00064
		GO:0040011	locomotion	0.00084
		GO:0006996	organelle organization	0.00104
		GO:0010564	regulation of cell cycle process	0.00108
		GO:1905515	non-motile cilium assembly	0.0068
	MF	GO:0017111	nucleoside-triphosphatase activity	2.2E-21
		GO:0016817	hydrolase activity, acting on acid anhyd...	1.5E-20
		GO:0016787	hydrolase activity	6.4E-10
		GO:0019205	nucleobase-containing compound kinase ac...	0.000016
		GO:0003824	catalytic activity	0.000058
		GO:0008017	microtubule binding	0.00477
		GO:0003774	motor activity	< 1e-30
		GO:0003777	microtubule motor activity	< 1e-30
	CC	GO:0042995	cell projection	2.5E-11
		GO:0120025	plasma membrane bounded cell projection	2.5E-11
		GO:0005815	microtubule organizing center	4.7E-09
		GO:0005856	cytoskeleton	9.6E-08
		GO:0043228	non-membrane-bounded organelle	6.3E-06
		GO:0043232	intracellular non-membrane-bounded organ...	6.3E-06
		GO:0043226	organelle	0.000042
		GO:0043229	intracellular organelle	0.00088
		GO:0005622	intracellular anatomical structure	0.00585



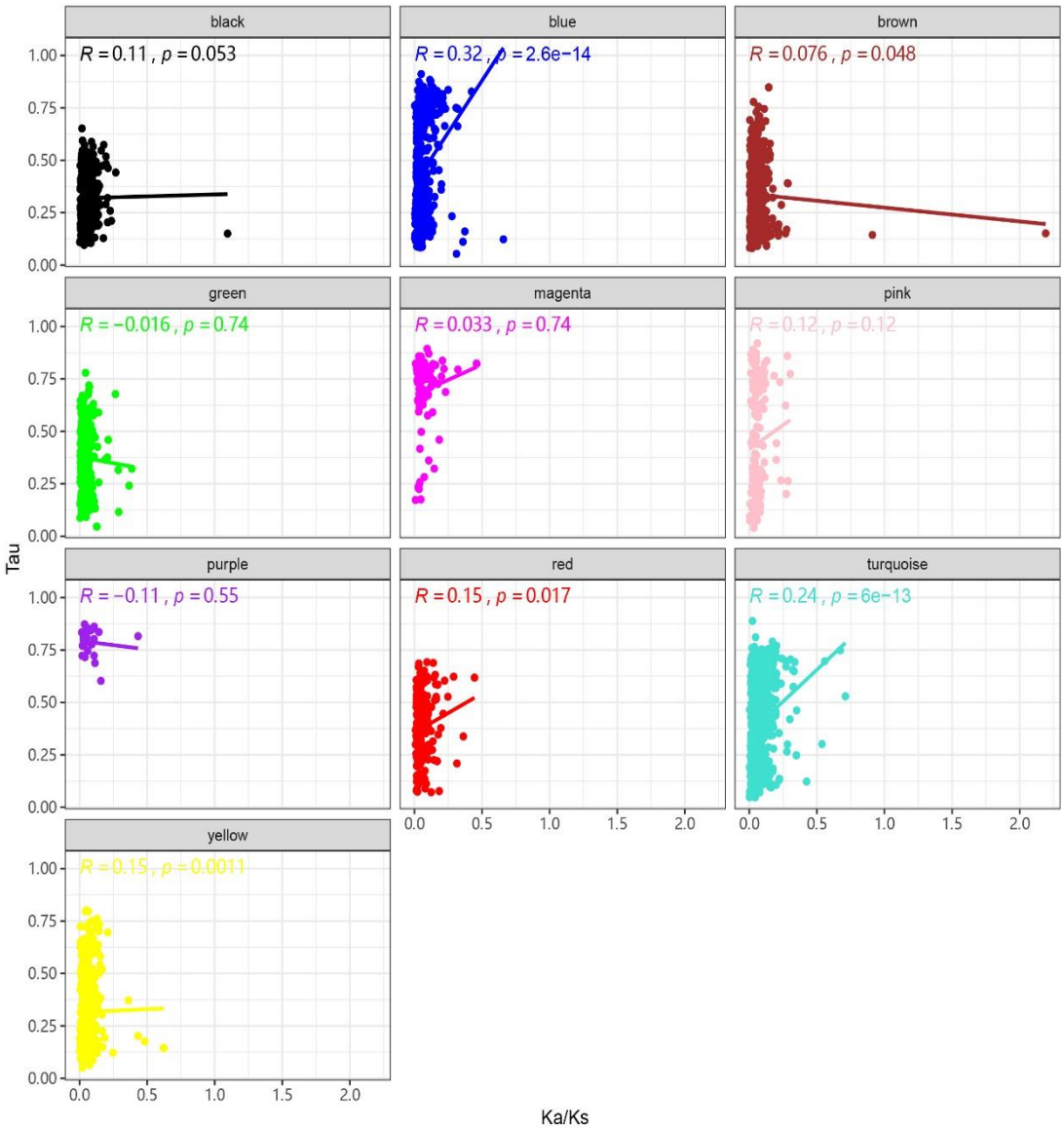
Supplementary Figure 2.1 Gene expression profiles for the co-expressed modules. The expression level was indicated by variance-stabilizing-transformation (VST) normalized reads.



Supplementary Figure 2.2 The correlation between Tau and connectivity. (a) The correlation between Tau and total network connectivity (kTotal), and between Tau and intramodular connectivity (kWithin) in non-gonad associated modules. (b) The correlation between Tau and intermodular (kOut) connectivity in all the modules. The Spearman's correlation (R) and p values were shown on the top.



Supplementary Figure 2.3 The correlation between Ka/Ks and network for each module. (a) The correlation between Ka/Ks and total network connectivity (kTotal), and between Ka/Ks and intramodular connectivity (kWithin) in non-gonad associated modules. (b) The correlation between Ka/Ks and intermodular connectivity (kOut) in all the modules. The Spearman's correlation (R) and p values were shown on the top.



Supplementary Figure 2.4 The correlation between Tau and Ka/Ks in each module. The Spearman's correlation (R) and p values were shown on the top.

Chapter 3 Lack of transcriptional coordination between mitochondrial and nuclear oxidative phosphorylation genes in the presence of two divergent mitochondrial genomes

Abstract

In most eukaryotes, oxidative phosphorylation (OXPHOS) is the main energy production process and it involves both mitochondrial and nuclear genomes. The close interaction between the two genomes is critical for the coordinated function of the OXPHOS process. Some bivalves show doubly uniparental inheritance (DUI) of mitochondria, where two highly divergent mitochondrial genomes, one inherited through eggs (F-type) and the other through sperm (M-type), coexist in the same individual. However, it remains a puzzle how nuclear OXPHOS genes coordinate with two divergent mitochondrial genomes in DUI species. In this study, we compared transcription, polymorphism, and synonymous codon usage in the mitochondrial and nuclear OXPHOS genes of the DUI species *Ruditapes philippinarum* using sex- and tissue-specific transcriptomes. Mitochondrial and nuclear OXPHOS genes showed different transcription profiles. Strong co-transcription signal was observed within mitochondrial (separate for F- and M-type) and within nuclear OXPHOS genes but the signal was weak or absent between mitochondrial and nuclear OXPHOS genes, suggesting that the coordination between mitochondrial and nuclear OXPHOS subunits is not achieved transcriptionally. McDonald-Kreitman and frequency-spectrum based tests indicated that M-type OXPHOS genes deviated significantly from neutrality, and that F-type and M-type OXPHOS genes undergo different selection patterns. Codon usage analysis revealed that mutation bias and translational selection were the major

factors affecting the codon usage bias in different OXPHOS genes, nevertheless, translational selection in mitochondrial OXPHOS genes appears to be less efficient than nuclear OXPHOS genes. Therefore, we speculate that the coordination between OXPHOS genes may involve post-transcriptional/translational regulation.

Keywords: Oxidative phosphorylation; Doubly uniparental inheritance; Co-transcription; Polymorphism; Codon usage bias; Translational selection

3.1 Introduction

In most animals, mitochondria are maternally inherited and one consequence of this kind of inheritance is a limitation of the genetic variance in the mitochondrial population within an individual (Lane, 2012). Until recently, it was assumed that mitochondrial DNA (mtDNA) is present in a state of homoplasmy, namely that identical copies of the mtDNA are found within an individual. The presence of different genetic variants, termed heteroplasmy, was thought to be mainly associated with unfavorable conditions such as ageing and disease (e.g., James White et al., 2008; Lane, 2012; Stewart & Chinnery, 2015). More recently, high-throughput sequencing revealed that mtDNA heteroplasmy (at least at low levels) is much more common than previously thought (Barrett et al., 2019; Dowling, 2014; Zhang et al., 2018). Heteroplasmy is a central issue in mitochondrial biology because genetic variation can lead to within-individual selection which can negatively affect coordination with nuclear-encoded mitochondrial components (Lane, 2011).

In Metazoa, the primary function of mitochondria is energy production through the process named oxidative phosphorylation (OXPHOS), which involves the tight interaction between mitochondrial (mt) and nuclear (nu) encoded subunits. Therefore,

to ensure the correct and efficient synthesis and assembly of the OXPHOS system, proper coordination between mt and nuOXPHOS genes is required. In mice and humans, van Waveren & Moraes (2008) reported a shared transcriptional control mechanism of nuOXPHOS genes and strongly correlated transcriptional signal within the same complex of nuOXPHOS genes. Barshad et al. (2018) investigated the OXPHOS transcription regulatory landscape across multiple tissues in humans and found a strong co-regulation signal between mt and nuOXPHOS genes across tissues. These findings make intuitive sense as gene products that must cofunction may be transcriptionally coregulated. However, Couvillion et al. (2016) showed that transcription levels in mt and nuOXPHOS genes were not concordant during mitochondrial biogenesis in yeast, while translational responses in both mt and nuOXPHOS genes were rapid and synchronously regulated, indicating the coordination between mt and nuOXPHOS genes is at the translation, not transcriptional level. A recent study in human cells showed that the average synthesis of mt and nuOXPHOS subunits for each complex was also highly correlated, although coordinated cytosolic and mitochondrial translation may require a nu-encoded mt protein—leucine rich pentatricopeptide repeat containing protein (LRPPRC)—to maintain cellular proteostasis (Soto et al., 2021).

Transcriptional coordination between mt and nuOXPHOS genes could be particularly complex in some bivalve species. Some bivalve species exhibit an evolutionarily stable exception to the strictly maternal mtDNA inheritance (SMI), a condition referred to as doubly uniparental inheritance (DUI; see Zouros, 2013 for a review). In DUI species, two mitochondrial lineages (F-type and M-type) are present: the F-type is transmitted through eggs while the M-type is transmitted through sperm. Because of the strict segregation between F and M lineages, F- and M-type mtDNA can accumulate an impressive genetic divergence (10%–43% at the nucleotide level, up to

53% at the amino acid level). Gametes are homoplasmic for the respective sex-linked lineage, while the distribution of F- and M-type in adult tissues is variable according to sex, tissue, and species. Generally, heteroplasmic females are less common (and if heteroplasmy is present it is usually at low levels), whereas males are always heteroplasmic. According to previous studies, the M-type mtDNA in the DUI species is predominant in male gonads and present in variable number (and can also be absent) in male somatic tissues, being generally absent (or rare) in female samples. F-type mtDNA is present in all the somatic tissues of both females and males (Ghiselli et al., 2019, 2021a). Moreover, the transcription of M-type mtDNA was also detected in male somatic tissues, but the frequency and the percentage of its presence seem to vary across different species (e.g., Breton et al., 2017; Dalziel & Stewart, 2002; Ghiselli et al., 2011; Milani et al., 2014; Mioduchowska et al., 2016; Obata et al., 2011). For example, the M-type mtRNA was detected in 60% and 89.5% of male somatic samples in *Utterbackia peninsularis* and *Venustaconcha ellipsiformis*, respectively (Breton et al., 2017). However, sex- and tissue-specific transcriptomic resources are lacking for DUI species, making many inferences on the abundance and expression of F-type, M-type, and related nuclear genes difficult.

While mito-nuclear coregulation is likely important, it is therefore unclear when and where selection has acted on coordination across different eukaryotes. Sequence coevolution between mt and nuOXPHOS genes provides another way to coordinate OXPHOS across the genomes. Mito-nuclear coevolution implies that sequence evolution within one genome could exert selection on the other genome for complementary changes (Hill, 2020; Hill et al., 2019; Rand et al., 2004), and mito-nuclear coevolution has been observed across a wide range of eukaryotic lineages (Barreto et al., 2018; Barreto & Burton, 2013; Havird et al., 2017; Havird & Sloan,

2016; Yan et al., 2019), including bivalves (Piccinini et al., 2021, which included 4 DUI species). In DUI species, two highly divergent mt genomes have to cofunction with the same nuclear background, which may be challenging for mito-nuclear coevolution. Previously studies indicated that F- and M-type genomes might have evolved separately multiple times (Gusman et al., 2016; Zouros, 2013), and two types of mt genomes might be under different selective pressure. It has been reported that Mtype mt genomes show higher nonsynonymous to synonymous substitution rates (dN/dS) than F-type, and several studies proposed that the M-type mt genome might be under relaxed selection (Hoeh et al., 1996, 1997, 2002; Liu et al., 1996; Ort & Pogson, 2007; Śmietanka et al., 2009, 2013; Soroka & Burzyński, 2010; Stewart et al., 1995, 1996; Zbawicka et al., 2010; Zouros, 2013). However, other studies have hypothesized that M-type may have undergone adaptive evolution optimizing sperm/male gonad functions (Bettinazzi et al., 2019, 2020; Burt & Trivers, 2006; Ghiselli et al., 2013, 2021a, 2021b; Iannello et al., 2019; Skibinski et al., 1994, 2017). Therefore, studying the selections on different OXPHOS genes would be critical to understanding how two divergent mt genomes evolve with the same nuOXPHOS genes (Iannello et al., 2019; Maeda et al., 2021; Piccinini et al., 2021).

An assumption of dN/dS metrics is that variations at synonymous sites are neutral. However, recent studies have shown selection on synonymous sites in the form of preferential codon usage, without changing the protein sequence (reviewed in Gingold & Pilpel, 2011; Plotkin & Kudla, 2011). Natural selection acting on synonymous codons to increase protein synthesis speed and accuracy is known as translational selection. Translational selection combined with mutational bias can create synonymous codon usage bias (CUB), in which codons are used in different frequencies in the coding regions across the genome (Gingold & Pilpel, 2011; Hershberg & Petrov,

2009; Plotkin & Kudla, 2011). CUB can influence various cellular processes, including gene expression (Jeacock et al., 2018), protein folding and function (Yu et al., 2015; Zhou et al., 2009), and exon splicing (Parmley & Hurst, 2007), therefore it can play an important role in genome evolution. In addition, it has been shown that translational selection is pervasive and detectable in a wide range of vertebrates (de Oliveira et al., 2021; Doherty & McInerney, 2013; Machado et al., 2020). Translational selection has also been detected in mitochondria in a wide range of species (Jia & Higgs, 2008; Sun et al., 2008; Wang et al., 2011; Wei et al., 2014). Furthermore, several studies reported that the rate of mRNA translation into protein (translational efficiency) in mt genes is lower than nuclear genes (Adrion et al., 2016; Havird & Sloan, 2016; Pett & Lavrov, 2015; Sloan et al., 2013), leading to the hypothesis that differences in translational selection for efficiency between mt and nu genes might be associated with the different evolution rates in mt and nuOXPHOS genes.

DUI species, with the stable and natural occurrence of two very divergent mitochondrial genomes in the same individual, represent an interesting evolutionary puzzle, and provide a unique model to study heteroplasmy and mito-nuclear interactions. Taking advantage of RNA-Seq data on three different tissues of 15 females and 15 males, we compared transcription, polymorphism, divergence, and codon usage in mt and nuOXPHOS genes in the DUI species *Ruditapes philippinarum* (the Manila clam). We observed lack of co-transcriptional coordination among F-type, M-type and nuOXPHOS genes. Furthermore, three genomes were constrained by different selection patterns, occurring at both synonymous and nonsynonymous substitutions. In particular, transcriptional selection shapes codon bias differently in mt and nuOXPHOS genes. Considering our results, we predict that mito-nuclear coordination does not occur at transcriptional level, but it is achieved by post-trascriptional/translational mechanisms

in DUI species. To our knowledge, this is the first study analyzing both transcriptional regulation and sequence evolution to investigate the coordination of OXPHOS genes in mollusks.

3.2 Materials & Methods

3.2.1 Dataset and reference transcriptome

Raw reads of *Ruditapes philippinarum* were downloaded from NCBI (BioProject PRJNA672267). All the clams were collected during the spawning season (end of July) from the same population in the Northern Adriatic Sea (Italy), in the river Po delta region (Sacca di Goro, approximate GPS coordinates: 44°50′06″N, 12°17′55″E). By visual inspection at optical microscope, gonads contained either eggs or sperm in late developmental stages, and clam were sexed concordantly. Adductor muscle, mantle, and gonad from 15 males and 15 females (with the exception of a missing female mantle; 89 samples in total) were sequenced using Illumina HiSeq 2500 with insert size of 500 bp to generate 150 bp paired-end reads. Detailed information about RNA extraction, library preparation, sequencing and *de novo* transcriptome assembly can be found in Maeda et al. (2021). The *de novo* reference transcriptome assembly for *R. philippinarum* is available on the Transcriptome Shotgun Assembly Sequence Database (TSA) of NCBI (accession No. GIVW00000000).

3.2.2 Transcriptome analysis

We used CD-HIT-EST (Fu et al., 2012) to reduce transcriptome redundancy (the presence of multiple transcripts belonging to the same gene), with a similarity threshold of 0.9. To retrieve F- and M-type mt genes from the transcriptome, we downloaded the complete F- and M-type mt genomes of *R. philippinarum* from NCBI (Accession Nos.: AB065374, AB065375). Considering some inaccuracies in the NCBI original annotations, we reannotated the mt genomes, by using BLAST (Camacho et al., 2009) against the non-redundant protein database (nr) to confirm protein coding regions, and

by manually curating start and stop codons. Then all reads were mapped to the transcriptome (without mt transcripts) and reference mt genes. The filtered reads were mapped to the transcriptome using bowtie2 (Langmead & Salzberg, 2012) with the default settings and only reads with mapping quality >10 were included in the following analyses. SAMtools (Li et al., 2009) was used to retrieve reads that were properly paired and uniquely mapped. Samples having <1 thousand reads mapped to mt protein coding genes and <1 million total mapped reads, were excluded from the analysis.

3.2.3 Transcriptome annotation

Nuclear OXPHOS transcripts were retrieved from Maeda et al. (2021) (BioProject PRJNA672267). To get coding sequences from nuOXPHOS transcripts, we ran TransDecoder (https://github.com/TransDecoder/TransDecoder/wiki) using homology searches against nr and Pfam databases with the minimum length of the open reading frame of 150 bp and only the longest ORF for each OXPHOS gene was kept.

We additionally performed the annotation of the whole R. philippinarum transcriptome as follows. First, contaminations from non-metazoans were filtered out by using a BLASTX (Altschul et al., 1997) search (with default parameters and adding information about taxon id) against the nr database of NCBI. We therefore extracted the full taxonomic lineage for each BLAST hit and we kept only transcripts having a best BLAST hit against Metazoa. To predict open reading frames (ORF) in the transcriptome, we used findorf (Krasileva et al., 2013); the prediction was performed using both a BLASTX search against the nr database and an HMMER (Mistry et al., 2013) search against the Pfam database (Finn et al., 2016). Annotation of predicted proteins was performed by using both a BLASTP search against the Swiss-Prot database and an

HMMER search against the Pfam database. We used Argot2 (Falda et al., 2012) to obtain Gene Ontology (GO) terms from BLASTP and HMMER outputs.

3.2.4 Transcription and co-transcriptional analysis

We used the gene length-corrected TMM (GeTmm) normalization method to allow both intra- and inter-sample comparisons (Smid et al., 2018). Before normalization, transcripts with low number of counts were filtered out using the NOIseq R package (Tarazona et al., 2015), with the following parameters: CPM=1, cv.cutoff=300. Genes that failed to pass this threshold were defined as lowly transcribed genes. It is worth mentioning that F-type and M-type genomes both contain a lineagespecific ORF (FORF and MORF), which might play functional roles in DUI species (Breton et al., 2011; Milani et al., 2013a; Minoiu et al., 2016). Therefore, although we still do not know if they belong to any complex subunits, we included them among the mtOXPHOS genes if not specified in the context. The transcription for mt and nuOXPHOS genes were plotted for each tissue, and Wilcoxon rank-sum test was used to compare pairwise transcriptional differences between mt and nuOXPHOS genes in each tissue. Kruskal-Wallis rank-sum test followed by a Dunn's test with Bonferroni correction was used to compare transcriptional differences across tissues.

To retrieve the general correlation trend of transcription across tissues, we calculated Spearman's rank-sum correlation with FDR correction across all samples using psych R package (http://CRAN.R-project.org/package=psych). The same process was also performed separately for each tissue to obtain the tissue-specific correlation trend. The correlation was considered significant with adjusted $P<0.05$. To test if the correlation strength between mt and nuOXPHOS genes was higher than genes involved in the different biochemical activities, we compared the differences in Spearman's

correlation coefficients (rho) between OXPHOS genes and a set of randomly selected same number of nuclear genes (56 genes) using Wilcoxon rank-sum test with Bonferroni correction. To test the hypothesis that genes within the same complex (intracomplex) presented a stronger correlation than the genes between different complexes (intercomplex), we compared the intracomplex correlation to intercomplex correlation using Wilcoxon rank-sum test with Bonferroni correction.

To uncover genes co-transcribed with OXPHOS genes, we retrieved the nuclear genes that were significant highly (rho>0.6) correlated with mtOXPHOS genes, and with nuOXPHOS genes. Functional enrichment for the nuclear genes that were co-transcribed with the OXPHOS genes was performed in topGO v2.34.0 (Alexa & Rahnenführer, 2018), using the classic Fisher's test, with a nodeSize of 5 and a *P*-value cutoff of 0.01.

3.2.5 SNP calling and McDonald-Kreitman test

The F- and M-type mt genomes were used as references for SNP calling on mtOXPHOS genes. For the nuOXPHOS genes, SNPs were called based on the *de novo* assembled transcriptome. We used Freebayes v1.2.0 (Garrison & Marth, 2012) to call the SNPs from all the samples simultaneously. VCFtools (Danecek et al., 2011) was run to calculate the rate of missing SNPs and to filter the SNPs for each sample using the following parameters: --minGQ 20 --minQ 30 --minDP 30. Finally, the number of SNPs in each gene was normalized by the gene length and the total number of SNPs in the sample to enable comparison across different genes. Considering the uneven coverage and the different rates of missing SNPs across sexes and tissues, allele frequency and PCA classification were performed with a genotype likelihood approach implemented in ANGSD (Korneliussen et al., 2014), which is particularly suited for low and medium

depth data. The Kolmogorov-Smirnov test with Bonferroni correction was used to assess if there is difference in the distribution of allele frequencies between different set of genes. SnpEFF (Cingolani et al., 2012) was used to predict the effects of SNPs on mt and nuOXPHOS genes. Samples with the rate of missing data > 40% were filtered out for the SNP effect prediction. Statistical differences between the proportion of synonymous and nonsynonymous SNPs in each component of OXPHOS (F-type, M-type and nuOXPHOS) genes were tested by Wilcoxon rank-sum test with Bonferroni correction.

We performed McDonald-Kreitman (MK) tests (McDonald & Kreitman, 1991) and frequency-based tests—Tajima's *D* (Tajima, 1989), Fu & Li's *D*, and Fu & Li's *F* (Fu & Li, 1993)—on mt and nuOXPHOS genes and randomly selected nuclear proteincoding genes (30 genes), using DnaSP v6.12 (Rozas et al., 2017). For mtOXPHOS genes, the MK test was performed between F-type and M-type OXPHOS, and also separately for F-type and M-type using the closely related species *Ruditapes decussatus* (SMI) as an outgroup. For nuOXPHOS and randomly selected genes, the MK test was performed with *R. decussatus* as an outgroup. The OXPHOS orthologues in *R. decussatus* were extracted from Iannello et al. (2019) and Piccinini et al. (2021), while the random orthologues were retrieved with OrthoFinder2 (Emms & Kelly, 2019). Among the nuOXPHOS genes identified above, 32 nuOXPHOS orthologues were retrieved for the MK analysis. VCFtools (Danecek et al., 2011) and SeqKit (Shen et al., 2016) were used to retrieve consensus nucleotide sequences and amino acid sequences, respectively. Clustal Omega (Sievers & Higgins, 2018) was used for multiple sequence alignment and PAL2NAL (Suyama et al., 2006) was used to retrieve the homologous nucleotide region.

3.2.6 Codon usage analysis

The codon frequencies of OXPHOS genes were calculated using the EMBOSS cusp tool (http://emboss.bioinformatics.nl/cgi-bin/emboss/cusp). The genetic codes 1 and 5 (https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi) were used for the nuclear genes and mitochondrial genes, respectively. The GC composition (at first and second codons position: GC12, at the third codon position: GC3), relative synonymous codon usage (RSCU), the effective number of codons (ENC), and codon adaptation index (CAI) were calculated using CAIcal (Puigbò et al., 2008). RSCU is defined as the ratio of the observed frequency of codons to the expected frequency given that all the synonymous codons for the same amino acid are used equally (Sharp & Li, 1987). The ENC quantifies the extent to which the codon usage in a gene or genome departs from the equal usage and it ranges from 20 (if only one codon is used for each amino acid) to 61 (if all codons are used equally) (Wright, 1990). CAI is another commonly used statistic which requires a set of highly expressed genes as reference and it presumes translational selection in highly expressed genes, therefore it can assess the extent to which selection has driven the pattern of codon usage (Sharp & Li, 1987). For the mitochondrial genes, the reference database for CAI estimation is available. Considering M-type OXPHOS genes were present mainly and predominantly in male gonads, we therefore only used the average transcription levels in male gonads for Mtype OXPHOS genes and in female gonads for F-type OXPHOS genes to calculate the correlation between CAI and the transcription levels. For nuclear genes, no reference database was available, so 30 highly transcribed (average transcription in gonads) nu-encoded genes: top 15 transcribed nuOXPHOS genes and top 15 transcribed nu-encoded genes (ribosomal-related genes were excluded to avoid bias) were selected to build the reference database. We also calculated ENC and CAI for the whole

transcriptome, and Kruskal-Wallis rank-sum test followed by a Dunn's test with Bonferroni correction was performed to assess if ENC and CAI presented differences between OXPHOS genes and nuclear genes. The ENC-GC3 relationship (Nc-plot, Wright, 1990) and neutrality test between GC12 and GC3 (Sueoka, 1999) were performed to assess the factors influencing the CUB in OXPHOS genes. Spearman's rank-sum test was used to assess the relationship between ENC and CAI. The comparison between ENC and CAI was also used to demonstrate the relationship between mutation and natural selection on codon usage bias (Behura & Severson, 2012). The correspondence analysis (COA) based on both codon counts and RSCU was performed in R FactoMineR package to detect the factors affecting CUB. We also performed Chi-square tests for context-dependent mutations (the rate of mutation from any one base to any other is influenced by the neighboring bases) in each set of OXPHOS genes according to the procedures described in Jia and Higgs (2008). The mutation equilibrium was calculated according to Lynch (2007). Briefly, we inferred the minor allele according to the allele frequency and treated the minor alleles as the new mutations (Hildebrand et al., 2010). Sites with more than two alleles or with two alleles at an equal frequency were discarded. Then the expected mutation GC equilibrium was calculated as the following formula: $GCeq=1/(1+m)$ (1) where $m=v/u$, u is the mutation rate of A/T to G/C and v is the mutation rate of G/C to A/T (Johri et al., 2019; Lynch, 2007). Consequently, m close to 1 indicates little mutation bias, while GCeq close to the percentage of GC3 means that codon usage bias is majorly determined by mutation bias (Johri et al., 2019; Lynch, 2007).

3.3 Results

3.3.1 Different transcription patterns of mitochondrial and nuclear OXPHOS genes in different tissues

Five samples were filtered out due to a low number of reads or contamination (f_67_G, f_67_M, m_70_A, m_70_G, m_70_M). A total number of 84 samples was used for the following analysis. The percentage of reads mapped to the F- and M-type mtOXPHOS genes were reported in Figure 3.1: F-type was predominant (~100% in females; >95% in male somatic tissues) in all the tissues except male gonads, while Mtype accounted for more than 90% of reads in male gonads. Small traces of M-type reads were also detected in male adductor muscles (average: 0.92%) and male mantles (average: 0.58%) with five samples presenting more than 1% of M-type reads in male somatic tissues (Figure 3.1). No read was retrieved for the *ATP8* gene, possibly due to its short length (120 bp in F-type and 84 bp in M-type).

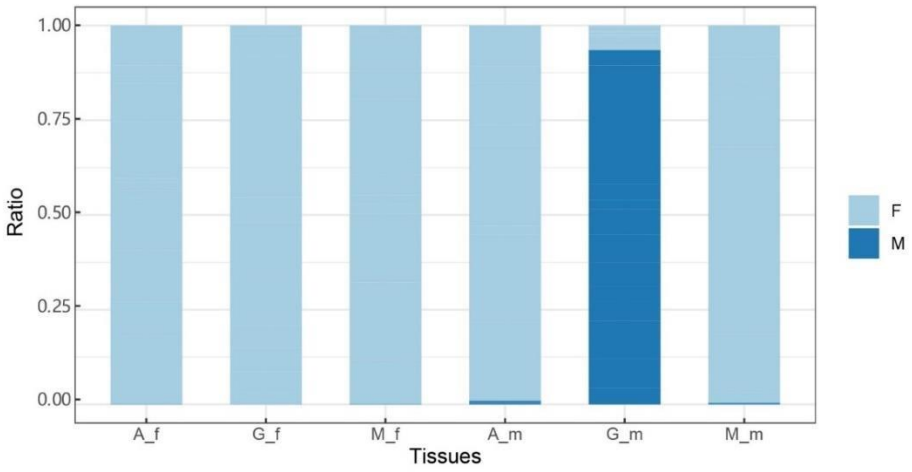


Figure 3.1 The ratio of reads mapped to the F- and M-type genomes in each tissue. A_f: female adductor; G_f: female gonad; M_f: female mantle; A_m: male adductor; G_m: male gonad; M_m: male mantle.

After filtering lowly transcribed genes, 27 mitochondrial protein-coding genes (14 from F-type and 13 from M-type), 56 (out of 67) nuOXPHOS genes, and 20,214 nuclear encoded genes were kept. To assess transcriptional differences of OXPHOS genes across different components (F-type, M-type and nuOXPHOS), a PCA analysis was performed based on the transcription level of all OXPHOS genes (Figure 3.2 A–C), only nuOXPHOS genes (Figure 3.2 D) and only mtOXPHOS genes (Figure 3.2 D-F). We found that F-type, M-type, and nuOXPHOS formed three distinct clusters taking account of transcription of all the OXPHOS genes across tissues (Figure 3.2 A).

Notably, M-type OXPHOS genes were clustered remarkably apart from the F-type and nuOXPHOS genes, in line with our expectations because the transcription of M-type in somatic tissues is rare or absent. However, transcriptional profiles of M-type in male gonads can also contribute to this departure. To test this, we conducted the PCA analysis for all the OXPHOS genes only in male gonads. M-type presented an extremely wide distribution despite the overlap between different OXPHOS genes in male gonads (Figure 3.2 B), indicating that transcriptional profiles of M-type in male gonads also contributes partially to the departure of M-type in Figure 3.2 A. Moreover, the OXPHOS genes in different tissues also showed different patterns, with male gonads departing from the other tissues and adductors showing relatively wider variation (Figure 3.2 C). To further investigate transcriptional differences across tissues (Figure 3.2 C), we focused on the nu and mtOXPHOS genes separately. Interestingly, we found that the distribution of nuOXPHOS genes showed an overlap in different tissues despite the wider variation in adductor muscles (Figure 3.2 D). By contrast, if we consider all the mtOXPHOS genes together, we found that male gonads were clearly apart from the other tissues (Figure 3.2 E). However, if we only look at F-type OXPHOS genes, the male gonads did not deviate from the other tissues (Figure 3.2 F). Thus, it seems that

the deviation of male gonads in Figure 3.2 C and Figure 3.2 D was due to the presence of M-type OXPHOS genes.

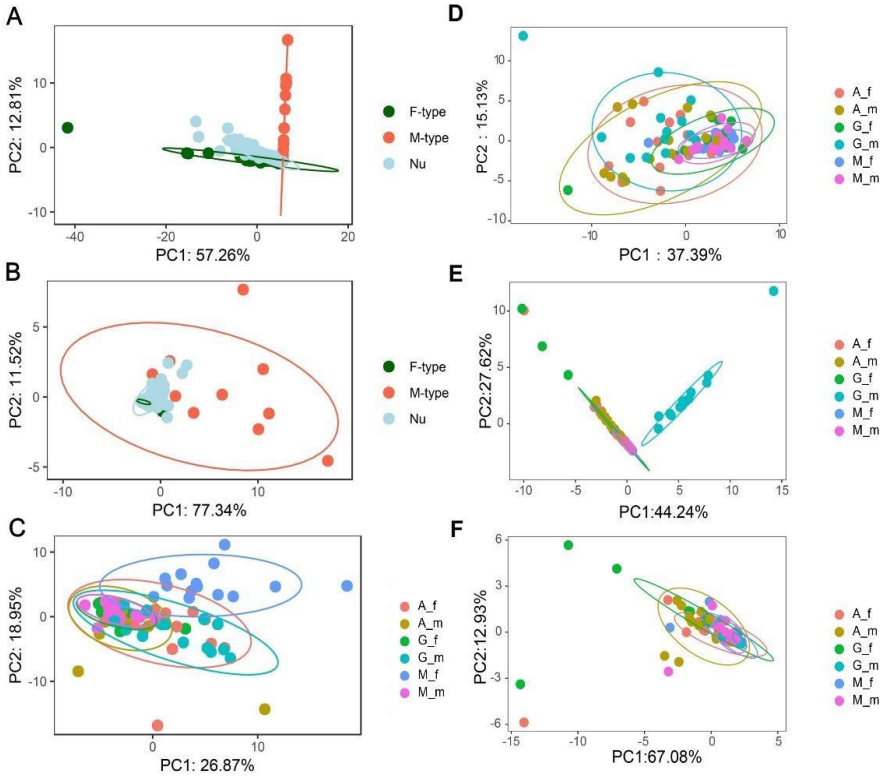


Figure 3.2 The PCA plot based on the transcription of OXPHOS genes. A: PCA plot for all OXPHOS genes in all samples, each point is a gene. B: PCA plot for all OXPHOS genes in only male gonads, each point is a gene. C: PCA plot for all the OXPHOS genes in each tissue, each point is a sample. D: PCA plot based on nuclear OXPHOS genes; E: PCA plot based on all the mitochondrial genes; F: PCA plot based on Ftype mitochondrial genes. The circle indicates the 95% confidence interval.

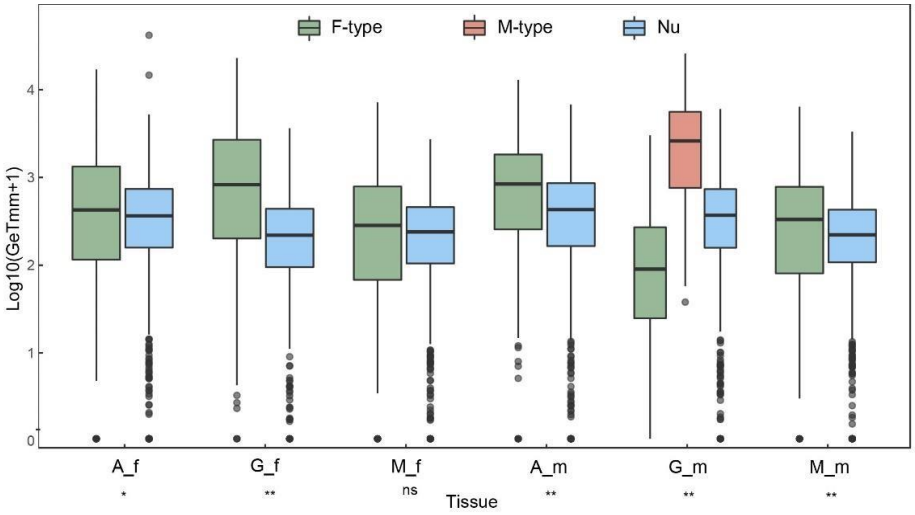


Figure 3.3 The transcription level of mitochondrial and nuclear OXPHOS genes in each tissue. The significance of Wilcoxon rank-sum test between F-type OXPHOS and nuOXPHOS was reported below the x axis. X axis represents six tissue types; Y axis represents transcription level. The transcription of M-type OXPHOS was significantly different from the F-type and nuOXPHOS in male gonads. *: –P<0.05; **: –P<0.001; ns: not significant.

Figure 3.3 compares the transcription level of mt and nuOXPHOS genes in different tissues. Because we do not know whether the two lineage-specific ORFs (*FORF* and *MORF;* see Ghiselli et al., 2013; Milani et al., 2013a) and F-type cytochrome c oxidase subunit 2 duplication (*COX2B*) have any OXPHOS function, they were not included here. The transcription of both mt and nuOXPHOS genes showed significant differences across tissues (in both cases, Kruskal-Wallis test: P<0.001), with nuOXPHOS genes presenting higher transcription in most pairwise comparisons between adductor muscles and other tissues and mtOXPHOS genes presenting higher transcription in most pairwise comparisons between gonads and somatic tissues (Significance for pairwise comparisons in Supplementary Table 3.1). Moreover, mtOXPHOS genes showed an overall significantly higher transcription than nuOXPHOS in all the tissues except female mantles (Figure 3.3). M-type OXPHOS

genes presented remarkably higher transcription in male gonads than F-type and nuOXPHOS genes, which is consistent with the deviation of M-type OXPHOS genes in Figure 3.2. The transcription level for each OXPHOS gene is shown in Supplementary Figure 3.1 (A and B, respectively). Intriguingly, the nuOXPHOS succinate dehydrogenase cytochrome b560 (*SDHC*) had two divergent sequences (*SDHC-1* and *SDHC-2*) and one of them (*SDHC*-2) presented a gonad-specific transcription (in both males and females).

3.3.2 Strong co-transcription signal within mitochondrial and within nuclear OXPHOS genes across tissues, but weak or absent across genomes

To test the hypothesis that genes involved in the same biochemical activity tend to be co-transcribed in *R. philippinarum* (Shyamsundar et al., 2005; Stuart et al., 2003), we calculated Spearman's rho between pairwise OXPHOS genes. According to the transcriptional correlations across tissues, all OXPHOS genes were clustered into four major distinct groups: F-type, M-type, nuOXPHOS1 and nuOXPHOS2. Genes from two nuOXPHOS subclusters showed a less pronounced, but still positive correlation between each other (Figure 3.4 A). To further investigate the correlation strength within and between different gene components, we plotted the correlation coefficients (rho) of OXPHOS genes separately for each component, and we randomly selected a subset of nuclear genes as a control to evaluate our observation. We found that the correlation within mtOXPHOS genes (F-type or M-type) is higher than the correlation within the nuOXPHOS, which in turn is higher than the correlation within randomly selected nuclear genes (Figure 3.4 B). The correlation between F-type and nuOXPHOS genes is slightly higher than the correlation between F-type and random nuclear genes,

indicating a weak co-transcription signal between F-type and nuOXPHOS genes, while the correlation between M-type and nuOXPHOS genes is remarkably higher than the correlation between M-type and random nuclear genes (Figure 3.4 C). Considering that the M-type transcription was primarily detected in male gonads, it is unclear whether the high co-transcription signal between M-type and nuOXPHOS results from a malegonad specific transcription or reflects actual co-transcription. Therefore, we investigated the correlation between mt and nuOXPHOS genes separately for each tissue. We found that the correlation between M-type and nuOXPHOS genes in male gonads was not significantly different from the correlation between M-type and randomly selected genes (Wilcoxon rank-sum test with Bonferroni correction, $P>0.05$), indicating that the significant co-transcription signal across tissues between M-type and nuOXPHOS genes in Figure 3.4 C was due to the gonad-specific transcription of Mtype in male gonads (Figure 3.5). Moreover, the co-transcription signal between F-type and nuOXPHOS genes was not consistent in different tissues (Figure 3.5). Weak co-transcription was detected in the female mantle between F-type and nuOXPHOS genes, but the signal disappeared in other tissues (Figure 3.5). Taken together, our results indicated a weak or absent co-transcription signal between mt and nuOXPHOS genes, but the strong co-transcription signal within F-type, within M-type, and within nuOXPHOS genes.

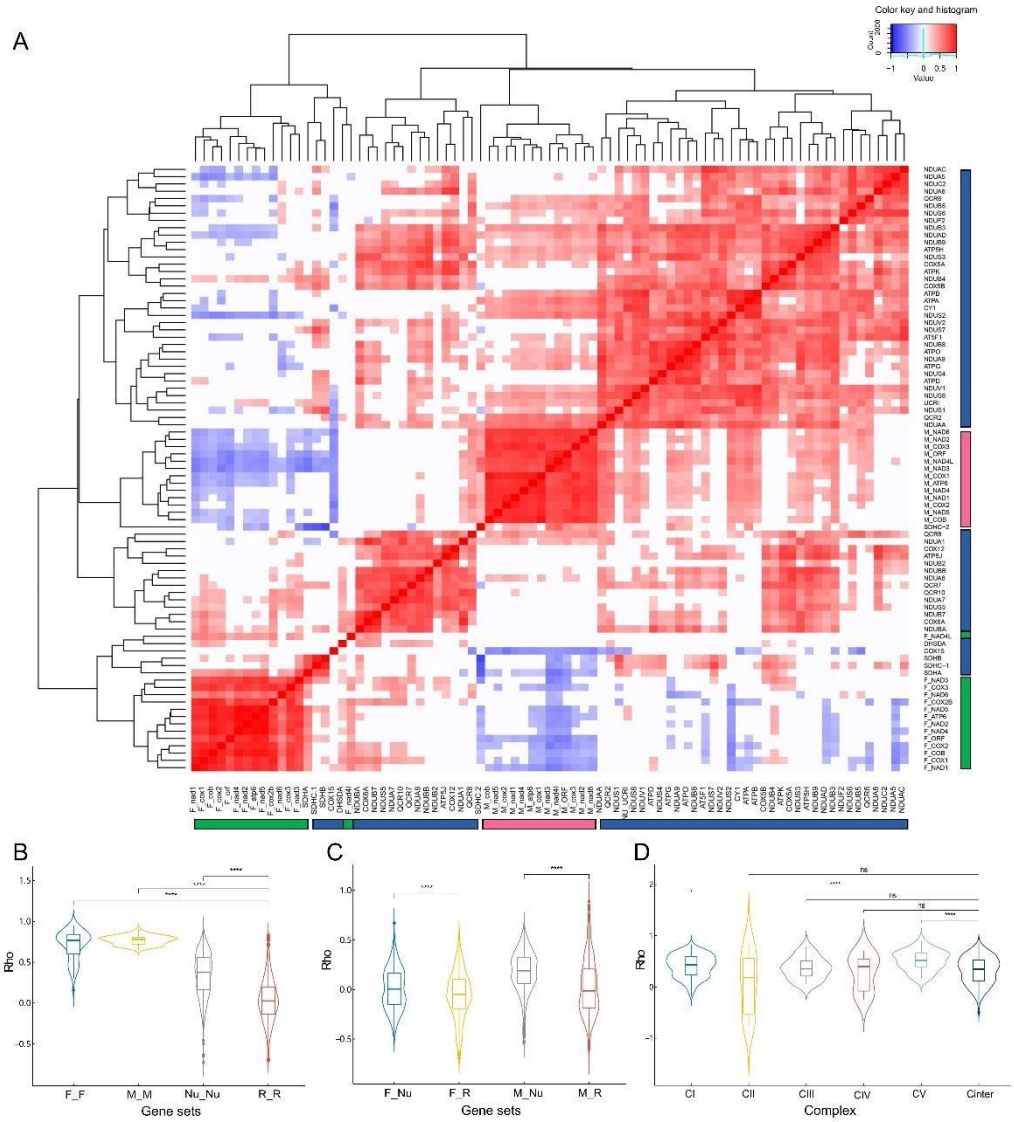


Figure 3.4 The co-transcription between mitochondrial and nuclear OXPHOS genes. A: The overall co-transcription across all the OXPHOS genes; blue, red and green color represent nuclear OXPHOS genes, M-type and F-type OXPHOS genes, respectively. B: The correlation coefficient (rho) distributions within F-type (F-F), M-type (M-M), nuOXPHOS genes (Nu-Nu), and random nuclear genes (R-R). C: The correlation coefficients distribution between mt and nuOXPHOS genes, and between mtOXPHOS and random nuclear genes. D: The correlation coefficient distribution of nuOXHPOS genes within each complex (intracomplex: CI-CV) and between different complexes (intercomplex: Cinter). Statistical significance in 3B-3D was performed using Wilcoxon rank-sum test with Bonferroni correction. *: – P<0.05; **: –P<0.001; ****: – P<1e-5; ns: not significant.

To test the hypothesis that genes within the same complex (intracomplex) presented a stronger correlation than the genes between different complexes (intercomplex) (Garbian et al., 2010; van Waveren & Moraes, 2008), we plotted the correlation coefficient of nuOXPHOS genes separately for each complex and compared them with intercomplex correlation coefficients ("Cinter" in Figure 3.4 D). Such analysis was not performed for mtOXPHOS because of a few genes in each complex. The nuOXPHOS genes belonging to the same complex were not clustered together (Figure 3.4 A) and the correlation coefficients within the same complex were not significantly different from intercomplex correlation coefficients except for complex I and V (Wilcoxon rank-sum test with Bonferroni correction, *P*<1-e5), in which intracomplex coefficients were slightly higher than the intercomplex coefficients (Figure 3.4 D).

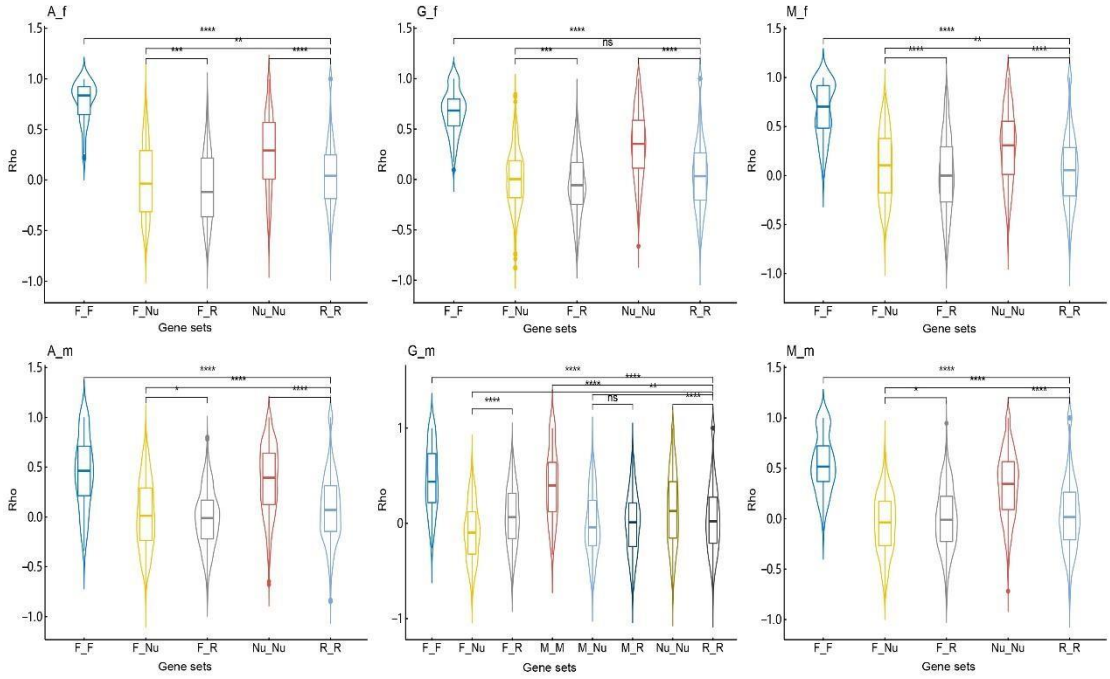


Figure 3.5 The co-transcription between and within OXPHOS genes in each tissue. Statistical significance was performed using Wilcoxon rank-sum test with Bonferroni correction. *: –*P*<0.05, **: –*P*<0.001, ****: –*P*<1-e-5, ns–not significant.

3.3.3 Nuclear genes co-transcribed with OXPHOS genes are enriched for mitochondrial processes

The exceptional heteroplasmic condition in DUI bivalves raises some questions: how is the compatibility between nuOXPHOS and two highly divergent mtDNA populations maintained? And which genes or pathways could be possibly involved in coordinating the OXPHOS process? To address these questions, we retrieved the nuclear genes that were co-transcribed with OXPHOS genes. We plotted the distribution of *P*-values from Spearman's correlation and the corresponding rho for all the co-transcribed nuclear genes, and a strict cutoff (rho>0.6) was used to ensure reliability. In this way, a total number of 136, 1 077 and 3 468 nuclear genes showed a significantly positive correlation with the F-type, M-type, and nuclear OXPHOS genes, respectively. Many nuclear genes co-transcribed with mtOXPHOS genes involved in the assembly of OXPHOS complexes, mitochondrial stability, and quality control. Also, a large number of nuOXPHOS was co-transcribed with genes encoding ribosomal proteins and genes involved in the TCA cycle. To further investigate the function of these co-transcribed nuclear genes, we performed a GO enrichment analysis and the results are shown in Supplementary Table 3.2. The overrepresented nuclear genes co-transcribed with F-type OXPHOS genes were associated with homeostatic process, mitochondrial respiratory chain complex assembly, and regulation of cellular pH (Supplementary Table 3.2). On the other hand, the nuclear genes co-transcribed with M-type OXPHOS genes presented a different situation, with reproductive process, nucleotide phosphorylation, and cell cycle being enriched (Supplementary Table 3.2). The nuclear genes correlated with nuOXPHOS involved in the biosynthetic process, protein metabolic process, mitochondrion organization, gene expression, and translational initiation (Supplementary Table 3.2). To identify candidates possibly

involved in the transcriptional regulation of mt and nuOXPHOS, we focused on the co-transcribed nuclear genes annotated as transcription factors, or that contain DNA or RNA binding sites. The candidate genes are listed in Supplementary Table 3.3.

3.3.4 Polymorphism and divergence in OXPHOS genes

The average number of SNPs across all samples identified in F-type, M-type, and nuOXPHOS were 50, 118, and 201, respectively. Figure 3.6 A shows the percentage of synonymous and nonsynonymous SNPs in the different gene components, with F-type presenting a significantly higher percentage of nonsynonymous SNPs, M-type and nuOXPHOS genes presenting a higher percentage of synonymous SNPs (Figure 3.6 A). However, a high percentage of nonsynonymous SNPs were found in one *COX2* copy, named *COX2B,* in the F-type (Supplementary Figure 3.2 A). If we exclude *COX2B*, the percentage of synonymous and nonsynonymous SNPs in F-type OXPHOS genes were not significantly different from the respective categories in the M-type (Figure 3.6 A). Ghiselli et al. (2013) observed a markedly different transcription level between the two *COX2* copies, with *COX2B* showing a lower transcription. They hypothesized that *COX2B* might be undergoing a pseudogenization process. The high number of nonsynonymous variants in *COX2B* resulting from this work is consistent with such a hypothesis. The relative ratio of synonymous and nonsynonymous SNPs in mt and nuOXPHOS genes are shown in Supplementary Figure 3.2 A, B.

To assess patterns of selection in OXPHOS genes, we applied two approaches: frequency spectrum-based test, and McDonald-Kreitman (MK) test. Allele frequency was calculated for three different components of OXPHOS genes and a set of randomly selected genes. Four distinct distributions were observed: one for the F-type in gonads (note that the distributions for the F-type in female and male gonads were not

significantly different from each other), one for the M-type, one for random genes in gonads, and one for nuOXPHOS genes in gonads (Kolmogorov-Smirnov test with Bonferroni correction: $P<0.001$, Figure 3.6 B). M-type OXPHOS genes presented a remarkably high intermediate allele frequency. Tajima's *D*, Fu & Li's *D* and Fu & Li's *F* showed negative values for most F-type, nuOXPHOS genes, and randomly selected nuclear genes (data not shown), but positive values for most M-type OXPHOS genes (Table 3.1).

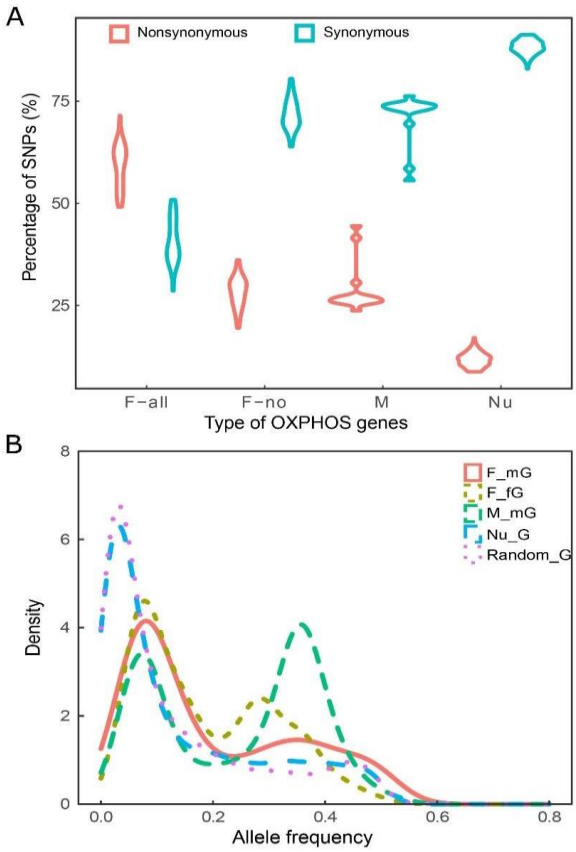


Figure 3.6 The SNP effects and allele frequency in OXPHOS genes. A: The percentage of synonymous and nonsynonymous SNPs in F-type, M-type and nuOXPHOS. The Ftype without *COX2B* was shown as F_no. B: Allele frequency for F-type OXPHOS genes in female gonads (F_fG), male gonads (F_mG), Mtype OXPHOS genes in male gonads (M_mG), nuOXPHOS genes in gonads (Nu_G) and randomly selected genes in gonads (Random_G). Allele frequency in this study is defined as the percentage of individuals with SNPs in the population.

For mtOXPHOS genes, we firstly compared the polymorphism (Pn and Ps) and divergence (Dn and Ds) between F-type and M-type OXPHOS genes. As shown in Table 3.1, except for the *ATP8*, *COX3, NAD3 and NAD4L,* all the rest of mtOXPHOS genes showed significant neutrality index (NI). NI is derived from the MK test and it quantifies the direction and degree of departure from the neutrality: NI=1 indicates the neutrality; NI>1 indicates negative selection; NI<1 indicates positive selection (Rand & Kann, 1996). Therefore, most mtOXPHOS genes showed a signal of positive selection between F-type and M-type. We also used the direction of selection (DOS) to evaluate the data in the MK test. The positive value of DOS could be consistent with positive selection, whereas the negative value indicates the presence of slightly deleterious mutations segregating in the population (James et al., 2016; Stoletzki & Eyre-Walker, 2011). Similarly, the DOS test also indicated possible positive selection in most mtOXPHOS genes. To test if the positive selection signal is present in the F-type or M-type OXPHOS genes or both, we also performed the MK test using *R. decussatus* (SMI species) as an outgroup. Interestingly, most F-type OXPHOS genes showed extremely low polymorphic differences which yield the excess of non-significant NI, while most M-type OXPHOS genes presented relatively high polymorphic differences and significant NI<1, which could be consistent with positive selection acting on M-type OXPHOS genes (Figure 3.7 A; Table 3.1). The MK test was also performed on nuOXPHOS and randomly selected nuclear genes. A considerable number of nuOXPHOS and randomly selected nuclear genes presented non-significant NI and a nearly equal ratio of polymorphic and divergent differences, consistent with neutrality (Figure 3.7 B). However, a large proportion of nuOXPHOS genes and some randomly selected genes also presented relatively high divergence, indicating the signature of positive selection.

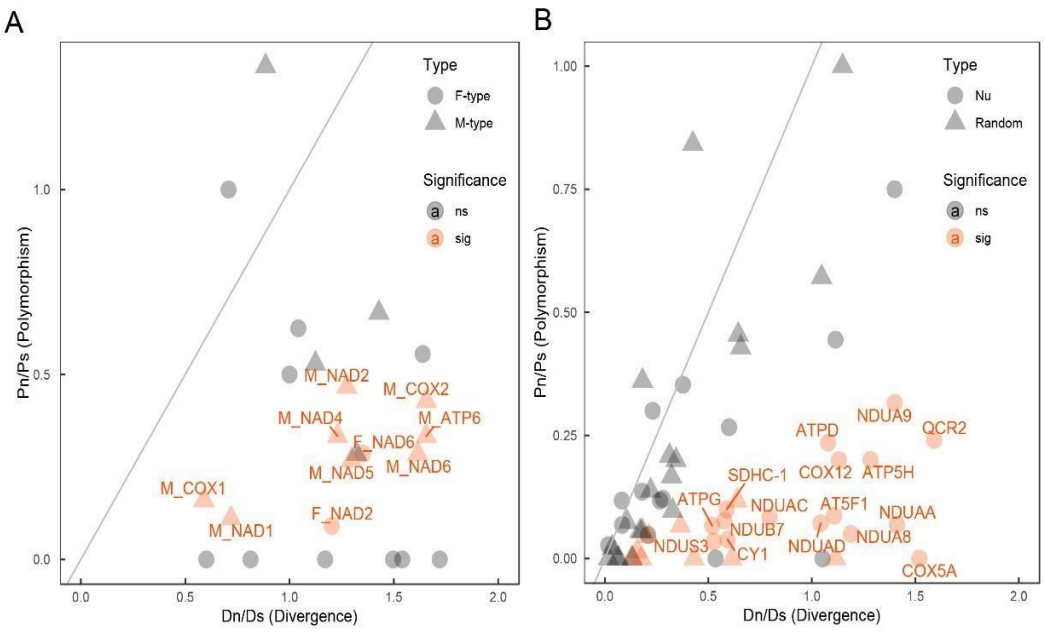


Figure 3.7 The results for McDonald-Kreitman test. A: McDonald-Kreitman test in mitochondrial OXPHOS genes. B: McDonald-Kreitman test in nuclear OXPHOS genes and random nuclear genes. Genes marked orange indicate significant (sig) departure from the neutrality, while genes marked grey indicate non-significant (ns) results. Gene names were added only for genes with significant results. The grey line on the figure indicates the equal ratio between polymorphism and divergence (under neutrality).

Table 3.1 The McDonald-Kreitman test and neutrality test in OXPHOS genes.

Comparison	Gene	Ds	Ps	Dn	Pn	NI	DOS	Tajima'D		Fu & Li's D		Fu & Li's F	
								F	M	F	M	F	M
F-type vs. M-type (*Ruditapes philippinarum*)	ATP6	107	16	104	4	0.26*	0.29	0.83	0.38	1.07	-0.06	1.15	0.06
	ATP8	6	1	11	3	1.64	-0.10	0.84	-0.14	0.72	0.02	0.85	-0.03
	COB	189	28	143	6	0.27**	0.25	-0.54	0.15	-0.88	-0.45	-0.90	-0.33
	COX1	241	43	65	0	0**	0.20	-0.69	0.87	-0.57	0.30	-0.69	0.52
	COX2	212	33	299	16	0.34**	0.26	-0.89	0.79	-0.86	0.27	-1.00	0.48
	COX3	121	16	104	6	0.44	0.18	-0.32	0.41	-0.35	-0.27	-0.39	-0.10
	NAD1	133	24	76	3	0.22*	0.25	-0.56	1.33	-0.92	0.64	-0.94	0.94
	NAD2	162	29	190	5	0.15**	0.39	-0.39	1.06	-0.34	0.31	-0.40	0.59
	NAD3	60	4	51	4	1.18	-0.54	-1.16	0.25	-1.40	-0.35	-0.15	-0.22
	NAD4	202	32	194	9	0.29**	0.27	-0.89	0.80	-0.87	0.42	-1.00	0.60
	NAD4L	48	8	55	2	0.22	0.28	-0.16	0.22	-1.40	0.02	-1.51	0.15
	NAD5	246	54	223	10	0.20**	0.32	-0.62	0.14	-1.06	-0.41	-1.08	-0.30
	NAD6	75	15	108	4	0.19**	0.38	0.20	0.45	0.67	0.36	0.62	0.44
	All	1796	303	1623	72	0.26**	0.28	-0.59	0.63	-0.65	0.05	-0.73	0.24
F-type vs. *Ruditapes decussatus* mt genes	ATP6	124	3	191	0	0.00	0.61	0.83		1.07		1.15	
	ATP8	12	1	28	0	0.00	0.70	0.84		0.72		0.85	
	COB	205	10	240	0	0.00	0.54	-0.54		-0.88		-0.90	
	COX1	256	15	154	0	0.00	0.38	-0.69		-0.57		-0.69	
	COX2	235	9	385	5	0.34	0.26	-0.89		-0.86		-1.00	
	COX3	141	8	211	0	0.00	0.60	-0.32		-0.35		-0.39	
	NAD1	171	3	121	3	1.41	-0.09	-0.56		-0.92		-0.94	
	NAD2	183	11	220	1	0.08**	0.46	-0.39		-0.34		-0.40	
	NAD3	75	1	61	0	0.00	0.45	-1.16		-1.40		-0.15	
	NAD4	237	8	247	5	0.60	0.13	-0.89		-0.87		-1.00	
	NAD4L	50	1	86	0	0.00	0.63	-0.16		-1.40		-1.51	
	NAD5	280	8	280	4	0.50	0.17	-0.62		-1.06		-1.08	
	NAD6	88	7	119	2	0.21*	0.35	0.20		0.67		0.62	
M-type vs. *Ruditapes decussatus* mt genes	ATP6	116	12	192	4	0.20*	0.37	0.38		-0.06		0.06	
	ATP8	9	0	16	3	ns	ns	-0.14		0.02		-0.03	
	COB	233	17	262	9	0.47	0.18	0.15		-0.45		-0.33	
	COX1	249	25	147	4	0.27*	0.16	0.87		0.30		0.52	
	COX2	218	14	361	6	0.26*	0.32	0.79		0.27		0.48	
	COX3	154	9	220	6	0.47	0.19	0.41		-0.27		-0.10	
	NAD1	157	18	113	2	0.15*	0.32	1.33		0.64		0.94	
	NAD2	193	15	246	7	0.37*	0.24	1.06		0.31		0.59	
	NAD3	70	3	62	4	1.51	-0.10	0.25		-0.35		-0.22	
	NAD4	234	21	288	7	0.27*	0.30	0.80		0.42		0.60	
	NAD4L	56	7	74	2	0.2	0.35	0.22		0.02		0.15	
	NAD5	256	41	332	11	0.21**	0.35	0.14		-0.41		-0.30	
	NAD6	91	7	147	2	0.18*	0.40	0.45		0.36		0.44	
	ATPB	61	37	1	1	1.65	-0.01	-0.96		-1.02		-1.18	

Ruditapes philippinarum nuOXPHOS genes vs. *Ruditapes decussatus* nuOXPHOS genes	ATPA	129	42	27	2	0.23	0.13	-1.45	-2.48	-2.53
	NDUBB	35	9	39	4	0.40	0.22	-0.45	0.20	0.00
	NDUA5	43	1	23	0	0.00	0.35	-1.15	-1.66	-1.75
	DHSDA	43	0	12	0	ns	ns	ns	ns	ns
	NDUS6	48	0	24	0	ns	ns	ns	ns	ns
	ATP5H	60	25	77	5	0.16**	0.40	-0.64	-0.65	-0.76
	AT5F1	111	23	123	2	0.08**	0.45	0.89	0.67	0.87
	NDUS3	76	29	40	1	0.07**	0.31	0.09	0.21	0.20
	QCR2	166	54	264	13	0.15**	0.42	-0.78	-0.84	-0.97
	NDUV2	68	33	19	4	0.43	0.11	-0.74	-0.56	-0.73
	NDUAA	158	29	223	2	0.05**	0.52	-0.36	-0.36	-0.43
	NDUAC	44	12	35	1	0.11*	0.37	-1.14	-0.66	-0.95
	SDHB	73	34	6	4	1.43	-0.03	-0.56	-0.67	-0.75
	NDUB7	45	13	26	1	0.13*	0.29	-0.47	-0.52	-0.59
	COX12	46	5	52	1	0.18*	0.36	-0.31	-1.07	-0.98
	COX15	91	10	21	3	1.30	-0.04	-1.69	-2.39	-2.54
	COX5A	52	16	79	0	0.00**	0.60	0.40	0.46	0.52
	ATPG	81	30	42	2	0.13*	0.28	-1.18	-1.92	-1.98
	NDUV1	156	44	13	3	0.82	0.01	-0.44	0.22	0.01
	ATP5J	25	8	35	6	0.54	0.15	-0.85	-0.11	-0.40
	NDUS1	161	81	29	11	0.75	0.03	-0.25	-0.19	-0.25
	NDUBA	58	7	61	0	0.00	0.51	-0.65	-1.46	-1.42
	ATPD	51	17	55	4	0.22*	0.33	-1.53	-1.29	-1.60
	NDUA8	53	20	63	1	0.04**	0.50	-0.95	-0.70	-0.92
	NDUAD	47	14	49	1	0.07*	0.44	-1.40	-1.54	-1.75
	ATPO	60	15	36	4	0.44	0.16	-1.05	-0.30	-0.63
	NDUS2	109	34	29	4	0.44	0.10	-0.48	0.02	-0.16
	NDUS7	53	17	20	6	0.94	0.01	-1.21	-1.59	-1.72
	SDHC-1	56	20	33	2	0.17*	0.28	1.15	0.80	1.07
	NDUA9	65	19	91	6	0.23*	0.34	-0.12	-1.09	-0.92
	CY1	86	26	51	1	0.07**	0.34	-0.24	0.05	-0.05

Note: Ds: the number of fixed synonymous substitutions; Dn: the number of fixed nonsynonymous substitutions; Ps: the number of synonymous polymorphism; Pn: the number of nonsynonymous polymorphism; NI: Neutrality index (*, $p<0.05$; **, $p<0.001$); DOS: direction of selection (DOS>0, positive selection; DOS<0, the presences of slightly deleterious). The nuOXPHOS genes that interacted with the mtOXPHOS were marked yellow.

To test whether nuOXPHOS subunits that are predicted to be in contact with mtOXPHOS subunits presented a different percentage of synonymous and nonsynonymous SNPs, we divided the nuOXPHOS genes into two groups (see Piccinini et al., 2021 for details): the "contact" group which was supposed to physically contact mtOXPHOS subunits; and the "non-contact" group which was predicted to have no direct interaction with mtOXPHOS subunits. Although the contact group presented a higher percentage of both synonymous and nonsynonymous SNPs than the non-contact group (Supplementary Figure 3.2 C), the MK test indicated that NI index in contact and non-contact groups were not significantly different from each other (Wilcoxon rank-sum test: $P>0.05$; Table 3.1).

3.3.5 Codon usage bias in OXPHOS genes

Mt and nuOXPHOS genes showed remarkably different GC compositions (Table 3.2), with extremely high AT skew in mtOXPHOS, as also found in other eukaryotes. Interestingly, we also found significant differences in GC composition between F- and M-type OXPHOS genes. F-type and M-type OXPHOS genes showed a similar percentage of GC12 (F-type: 34.42%; M-type: 34.83%), while a significantly higher percentage of GC3 was found in M-type (F-type: 21.69%; M-type: 26.91%) (Table 3.2; Figure 3.8). Moreover, M-type OXPHOS genes presented generally slightly higher ENC and lower CAI values compared to F-type (ENC: $P>0.05$; CAI: $P<0.05$; Table 3.2; Figure 3.8), indicating alleviated CUB in M-type OXPHOS genes. NuOXPHOS genes displayed relatively high ENC and CAI values that are comparable to those of the other nuclear genes in the transcriptome (Wilcoxon rank-sum test with Bonferroni correction, $P>0.05$). Although under different CUB, the heatmap based on RSCU values revealed that both mt and nuOXPHOS presented a shared usage bias towards A/U-ending codons (Supplementary Figure 3.3).

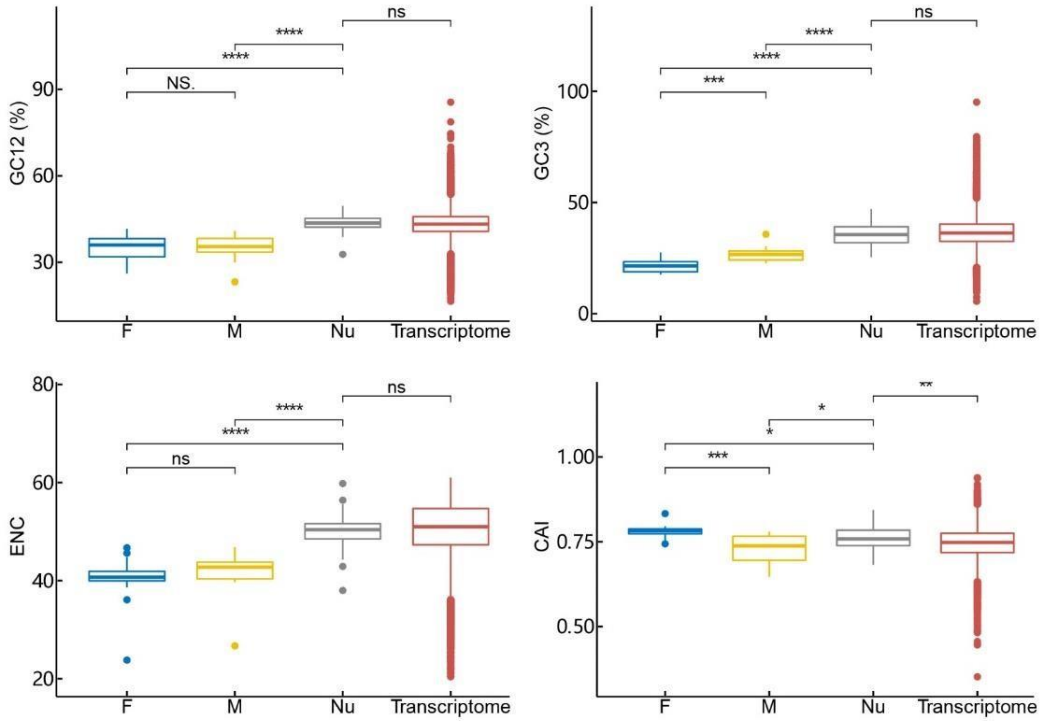


Figure 3.8 The comparison of codon usage parameters in different set of genes. GC12: GC composition at first and second codons position, GC3: at the third codon position, ENC: the effective number of codons, CAI: codon adaptation index. F, M and Nu indicate F-type, M-type and nuclear OXPHOS genes, respectively. Transcriptome represents all the other genes in the transcriptome. Statistical significance was performed using Wilcoxon rank-sum test with Bonferroni correction. *: –P<0.05, **: –P<0.001, ****: –P<1-e-5, ns–not significant.

Table 3.2 Statistics for codon usage in different OXPHOS gene components.

Analysis	Index and correlation	F-type OXPHOS	M-type OXPHOS	Nu OXPHOS	Transcriptome
Codon usage index	CAI	0.78	0.73	0.76	0.75
	ENC	40.22	41.71	50.08	50.67
	GC eq	25.71%	32.19%	31.36%	-
	GC3	21.69%	26.91%	35.59%	36.58%
	GC12	34.42%	34.83%	43.60%	43.20%
Correlation	CAI ~ ENC	0.05	**-0.60***	-0.2	**-0.43****
	ENC ~ GC3	0.18	**0.70***	0.11	**0.39****
	CAI ~ GC3	**-0.59***	**-0.83****	0.02	**0.34****
	GC3 ~ GC12	-0.02	-0.04	-0.21	0.01
	CAI ~ Transcription	-0.26	0.14	**0.51****	**0.07****
Correspondence based on codon usage	Axis1 ~ CAI	0.51	**0.85****	**0.62****	**0.35****
	Axis1 ~ ENC	0.006	-0.45	-0.08	**-0.22****
	Axis1 ~ GC3	-0.1	**-0.70***	**0.49****	**-0.31****
	Axis1 ~ Transcription	**-0.88****	0.2	**0.37***	**0.03***
	Axis2 ~ Transcription	-0.4	**-0.86****	**0.34***	0.005
Correspondence based on RSCU	Axis1 ~ CAI	-0.14	**0.71****	**-0.35***	**0.62****
	Axis1 ~ ENC	0.16	-0.12	0.09	**-0.50****
	Axis1 ~ GC3	**0.64***	**-0.59***	-0.16	**-0.67****
	Axis1 ~ Transcription	-0.06	0.22	-0.23	**-0.03****
	Axis2 ~ Transcription	0.31	0.01	-0.14	**-0.11****

ENC and GC3 relation plot (Nc-plot) compares the actual distribution of genes to an expected distribution which assumes no selection, therefore the departure from the expected curve indicates that these genes are under selective pressure (Wright, 1990). Likewise, the neutrality test plots the GC12 against GC3 to reflect the equilibrium between mutation pressure and natural selection (Sueoka, 1999). In this study, the Nc-plot revealed that only a small number of OXPHOS genes laid on the expected Nc curve, with most genes departing from the corresponding predictions (Figure 3.9). The neutrality test showed no correlation between GC12 and GC3 in both mt and nuOXPHOS (Table 3.2), indicating that mutation bias was not the only factor affecting CUB. Negative correlation between CAI and ENC was observed in M-type OXPHOS

genes. CAI indicates the selection for the preferred codon, while ENC is a non-directional parameter for either selection or mutation bias, therefore the correlation between CAI and ENC could indicate the role of selection, but mutation would lessen this correlation (Behura & Severson, 2012). Significant negative correlations between GC3 and CAI were observed in both F-type and M-type OXPHOS genes (Table 3.2), indicating that GC3 may be associated with CUB in mtOXPHOS genes. Moreover, GCeq in both F-type and M-type OXPHOS was slightly higher than GC3, while GCeq in nuOXPHOS is lower than the GC3 (Table 3.2). All these tests indicated that mutation alone cannot explain the CUB in OXPHOS genes. In nuOXPHOS genes, a significant correlation between CAI and transcription level was observed, indicating the possible role of translational selection. Similarly, for the protein-coding genes in the transcriptome, significant correlations were also detected between GC3 and CUB, and between CAI and transcription.

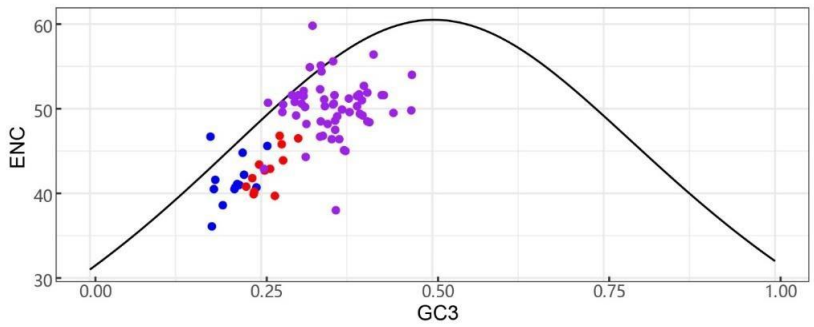


Figure 3.9 ENC and GC3 relation plot (Nc-plot) for mitochondrial genes and nuclear OXPHOS. The blue, red, purple points represented F-type, M-type and nuOXPHOS genes, respectively. The black curve represented the expected relation between GC3 and ENC.

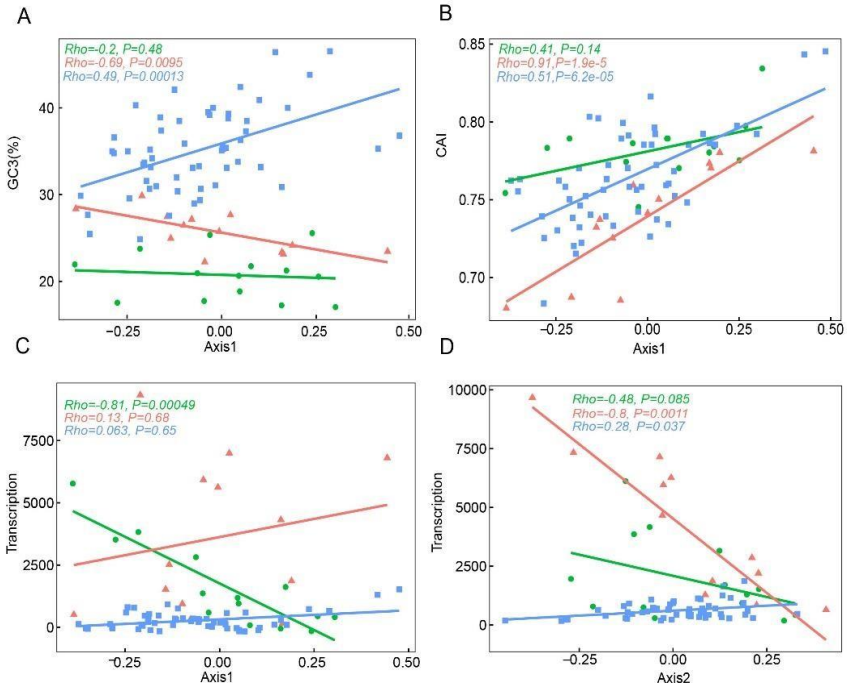


Figure 3.10 The relationship between codon usage and GC3 (A), CAI (B) and transcription. Only the first two highly representative axes (Axis 1 and 2) were considered here. Each point represents a gene. Green dots, red triangles, and blue squares represent F-type, M-type and nuOXPHOS genes, respectively. Spearman's correlation (rho) and corresponding *P*-values were on the top left.

The COA analysis based on the codon counts indicated that Axis 1 was significantly correlated with the CAI and GC3 in both M-type and nuOXPHOS genes, while no such correlation was observed in F-type OXPHOS genes (Figure 3.10; Table 3.2). Besides, the correlation between transcription and Axis 1, and between transcription and Axis 2 were observed in three different OXPHOS components. The COA analysis based on RSCU also showed similar results (Table 3.2). Consistently, these results indicated the GC3 composition and transcription levels could be responsible for the CUB for both OXPHOS genes and proteincoding genes in the transcriptome. Moreover, the chi-square test for the context-dependent mutations in F-type, M-type, and nuOXPHOs genes all indicated that the third position bases were not

independent of the second position bases (F: X-square=143.25, *P*<0.001; M: Xsquare=111.2, *P*<0.001; NuOXPHOS: X-square=328.08, *P*<0.001).

3.4 Discussion

3.4.1 Distinct transcriptional dynamics and regulatory mechanism in OXPHOS genes

Although the quantification of mtDNA and mtRNA in DUI species has been investigated before (Breton et al., 2017; Dalziel & Stewart, 2002; Ghiselli et al., 2011; Milani et al., 2014; Obata et al., 2011), the transcription patterns in DUI species across tissues in both sexes are still largely unknown. In the present study, we assessed the transcription of F-type and M-type genes in adductors, mantles, and gonads of a DUI species *R. philippinarum*. We found that the transcription levels of F-type OXPHOS genes were significantly different across tissues, while M-type OXPHOS genes were highly transcribed only in male gonads (Figure 3.3). Traces of M-type mtRNA (>1% reads mapped to M-type in the sample) were also detected in 5 (out of 28) male somatic samples analyzed (Figure 3.1). The presence of M-type mtRNA in somatic tissues seems to vary across DUI species, for example 60% male somatic samples in *U. peninsularis* and 89.5% male somatic samples in *V. ellipsiformis* presented somatic transcription of M-type mt genes (Breton et al., 2017). In *R. philippinarum*, previous work reported variable number of M-type DNA and RNA in somatic tissues, depending on the individual (Ghiselli et al., 2011; Iannello et al., 2021; Milani et al., 2014). In this work, we found that M-type is barely transcribed in the somatic tissues of most males. Such a pattern could be due to a low number of M-type mtDNA in these samples or tissue-specific transcription of M-type in males, which leads the M-type to be highly transcribed in male gonads, but poorly transcribed in male somatic tissues. Ghiselli et al. (2011) detected M-type mtDNA in most of (~87%) male somatic tissues, and Milani

et al. (2014) reported no correlation between the transcription level of cytochrome *b* (*cyt b* and its DNA copy number in males of *R. philippinarum*. These evidences were in line with a tissue-specific regulation of transcription of M-type, which is independent of the M-type DNA copy number. On the other hand, the extremely low transcription levels observed in somatic tissues of many males could reflect an absence of M-type mtDNA in such samples. Future investigations including sequencing of both mtDNA and mtRNA will help to clarify this point.

Transcription levels clearly distinguished F-type, M-type, and nuOXPHOS genes into three groups (Figure 3.2), indicating distinct transcription dynamics. Tissue-specific transcription was observed in F-type and nuOXPHOS genes, with nuOXPHOS genes presenting higher transcription in adductors and mtOXPHOS genes presenting higher transcription in gonads (Figure 3.3). Similar tissue-specific transcription of OXPHOS was widely reported and quantified in humans and mice, and this transcription pattern was proposed to be associated with differences in metabolic profiles and variable energetic demands in different tissues (Barshad et al., 2018; van Waveren & Moraes, 2008). However, different from the results in van Waveren & Moraes (2008), which showed stronger correlations within than among complexes—in this study we observed slightly higher intracomplexes correlation only in complex I and V (Figure 3.4 D). This discrepancy may be explained by several reasons: (1) In DUI species, nuclear OXPHOS genes must cooperate with two mitochondrial genomes, which might loosen the correlations between different complexes; (2) *R. philippinarum* is a sedentary living bivalve that has lower energy needs, therefore it may be subject to weaker selection in maintaining OXPHOS processes compared to the other taxa (Piccinini et al., 2021; Iannello et al., 2019); (3) Normalization methods may also influence co-transcription results, and recent study on large human dataset indicated

that normalization techniques based on total read count (such as TPM or FPKM) may lead to artefactual positive correlations (Perez & Sarkies, 2021).

Significant positive co-transcription was observed separately in F-type, in M-type, and in nuOXPHOS genes, but co-transcription signal between mtOXPHOS and nuOXPHOS wasweak/absent and not consistent across tissues (Figure 3.4 and 3.5). Considering the distinct transcriptional trends and the separate co-transcription patterns, our data indicate that the transcriptional difference of nuOXPHOS genes was independent of the M- or F-type genome and that the transcription of F-type, M-type, and nuOXPHOS genes might be under different regulatory mechanisms, including co-translational regulation (Couvillion et al., 2016).

Nuclear OXPHOS genes were subdivided into two positively correlated clusters (Figure 3.3). While the reason behind this split is unknown, one possibility could be the presence of supercomplexes within nuOXPHOS genes, with a tighter co-regulation inside each. Several different types of supercomplexes (such as CI+CIII+CIV, CIII+CV, CI+CIII) have been established in model organisms and several genes (*COX7*, *COX6A*, *NDUFB4, NDUFB9, UQCRC1, UQCRQ*) are involved in supercomplex formation (reviewed in Chaban et al., 2014; Milenkovic et al., 2017). However, the mechanisms and functional roles of supercomplexes are still largely unknown, especially in non-model organisms.

3.4.2 Candidate pathways and genes associated with OXPHOS co-transcription

In the present study, the mt and nuOXPHOS genes were co-transcribed with nu-encoded genes involved in many biological processes, such as mitochondrial respiratory chain complex assembly, cellular homeostasis, and translation (Supplementary Table S3.2). Mitochondrial respiratory chain complex assembly is an

intricate process that requires tightly orchestrated coregulation from both mitochondrial and nuclear genomes (Tang et al., 2020). Different nu-encoded assembly factors were observed to be co-transcribed with mt and nuOXPHOS genes (Supplementary Table S3.2) and these factors have been previously shown to be essential for the proper assembly and function of the OXPHOS system (van Waveren & Moraes, 2008). Genes involved in cellular homeostasis were overrepresented for genes co-transcribed with the F-type OXPHOS genes, along with many genes involved in mitophagy and ubiquitination processes (Supplementary Tables S3.3). This is in line with our expectations, as mitochondria are also essential for cellular homeostasis, calcium signaling, and metabolite synthesis. The proper function of mitochondria also requires balanced coordination between mitochondrial biogenesis and mitophagy through complex signaling pathways (Willems et al., 2015).

Notably, the translation process was overrepresented for genes co-transcribed with nuOXPHOS genes, indicating that nuOXPHOS genes might also be under translational coregulation (Supplementary Table S3.2). By contrast, nu-encoded genes co-transcribed with Mtype OXPHOS genes were overrepresented in the reproductive process and spermatogenesis (Supplementary Table S3.2). One reason could be that the M-type OXPHOS genes were found almost only in male gonads among our samples and thus the nuclear genes co-transcribed with M-type OXPHOS genes might also co-express with the genes responsible for the development of male gonads. Alternatively, M-type OXPHOS genes might be directly associated with reproduction in DUI species. According to previous studies, M-type mitochondria might be involved in some aspects of sex differentiation in DUI species as suggested by several authors (Ghiselli et al., 2012; Milani et al., 2013b; Zouros, 2013). Since M-type is limited to male gonads in these samples, it is not surprising to see the co-transcription between gonad-specific

genes and M-type OXPHOS genes. Therefore, co-transcription could reflect either functional interaction between M-type and gonad-specific nuclear genes, or similar independent transcription profiles.

It is worth mentioning that Maeda et al. (2021) found two divergent *SDHC* sequences in their study and one of them (*SDHC*-2) showed a gonad-specific transcription (Supplementary Figure S3.1). Around half of the nuclear genes co-transcribed with nuOXPHOS were correlated only with the *SDHC*-2, which might explain the presence of enriched GO terms involved in the cell cycle (Supplementary Table S3.2). Indeed, *SDHC* is important for both OXPHOS and Krebs cycle, and studies indicated that deficiency in this gene would increase ROS production and induce metabolic stress, genomic instability, and hypoxia (Slane et al., 2006; Tretter et al., 2016). However, the reason behind this remarkable tissue-specific expression is not clear. A recent study in humans indicated that tissue-dependent splice variants and OXPHOS subunit paralogs may also be involved in retaining OXPHOS activities (Barshad et al., 2018).

Considering the strong co-transcription within F-type, within M-type and within nuOXPHOS genes, it is reasonable to speculate that nuclear regulators may be responsible for transcribing each group. Here we identified a group of candidate regulator genes that are either transcription factors or had a DNA/RNA binding domain (Supplementary Table S3.3). These genes included zinc finger protein 341, which can activate transcription factor STAT1, a gene previously shown to regulate the OXPHOS process (Pitroda et al., 2009). Pentatricopeptide repeat domain-containing protein 3 is a mitochondrial RNA-binding protein and proteins containing PPR domains are known to play a role in transcription, RNA processing, splicing, stability, editing, and translation (Miglani et al., 2021; Schmitz-Linneweber & Small, 2008). Particularly,

PPR genes in plant are predominantly linked to the cytoplasmic male sterility (CMS, see discussion below) (Kim & Zhang 2018). Moreover, the roles of PPR proteins in mitochondrial gene expression and OXPHOS process has also been reported (Lightowlers & Chrzanowska-Lightowlers, 2008; Soto et al., 2021). Recently, a ribosome profiling study in human cells revealed that balanced mito-nuclear OXPHOS synthesis requires a nuclear-encoded mt protein LRPPRC (Soto et al., 2021).

3.4.3 Selection acts on both nonsynonymous and synonymous sites of OXPHOS genes

In DUI species, interspecific comparisons have found that M-type genes accumulate more mutations and have a higher evolutionary rate than F-type (Hoeh et al., 1996, 1997, 2002; Liu et al., 1996; Ort & Pogson, 2007; Śmietanka et al., 2009; Soroka & Burzyński, 2010; Stewart et al., 1995, 1996; Zbawicka et al., 2010), which led to the hypothesis that F- and M-type genomes experienced different selective pressures (reviewed in Zouros, 2013). Here, we found that MK tests between F-type and M-type in *R. philippinarum*, and between mt genes in *R. decussatus* and M -type in *R. philippinarum* genes consistently indicated positive selection on most M-type OXPHOS genes (Figure 3.7). M-type *COX3* and *NAD3* did not show departure from the neutral expectations, indicating that selection on M-type might be variable in different genes. On the other hand, intermediate allele frequency and positive Tajima's *D* in M-type OXPHOS genes suggested the possibility of population bottleneck or balancing selection (Figure 3.6). Native to the Pacific coast of east Asia, *R. philippinarum* was first transported to

America in the 1930s, and then was transferred to Europe to cope with the production decline of local clam species during 1970s–1980s (Chiesa et al., 2017; Cordero et al., 2017). Several studies revealed that reduced genetic diversity and genetic

differentiation compared to the American population were consistent with a strong founder effect in the European population (Chiesa et al., 2017; Cordero et al., 2017). Thus, in line with these studies, the founder effect can also explain why F-type and nuclear genes presented generally negative Tajima's *D* and excess of rare alleles. However, the founder effect was not sufficient to explain why M-type showed such different patterns both in the Italian population and also in the American population (Ghiselli et al., 2013). One possibility could be the narrower germline bottleneck of M-type mtDNA. Past studies indicated that the F-type mtDNA copy number in eggs is on average 10 times higher than the copy number of M-type mtDNA in sperm (Ghiselli et al., 2011), and that the narrower genetic bottleneck in M-type mtDNA could lead to the segregation of mtDNA variants in different tissues, causing remarkable within-individual variation and therefore also higher variability between samples (Iannello et al., 2021). Alternatively, the presence of high intermediate frequency alleles and positive Tajima's *D* could be a signal of balancing selection. Balancing selection on mtDNA has been found in gynodioecious plants showing cytoplasmic male sterility (CMS), in which a population consisting of both females and hermaphrodites and the sex is determined by the interaction between mitochondrial male-sterility genes and nuclear restorer-of-fertility genes (reviewed in Chase, 2007; Delph & Kelly, 2014). Under balancing selection, restorer genes are not fixed in the population because of the "cost of restoration" and CMS genes are under negative frequency-dependent selection to maintain the long-term balanced sex ratio in the population (Delph & Kelly 2014). DUI and CMS show some common features (Breton et al., 2010, 2011; Ghiselli et al., 2013; Milani et al., 2016; Mitchell et al., 2016): (1) the presence of novel lineage-specific mt-ORFs (Ghiselli et al., 2013; Milani et al., 2013a; Mitchell et al., 2016), which allows potential interaction between mitochondria and nuclear genes; (2) an

excess of mid-frequency polymorphism in M-type mtDNA (Ort & Pogson, 2007; Quesada et al., 1998), which might lead to match/mismatch between mitotype and nuclear genes; (3) the hypothesized association between mtDNA and sex determination in DUI species (reviewed in Breton et al., 2017; Zouros, 2013); (4) the presence of biased sex ratios (the proportion of males in populations range from 8% to 83%) (Ghiselli et al., 2012; Yusa et al., 2013); (5) recombination of M-type mtDNA in male gonads(Burzyński et al., 2003; Ladoukakis & Zouros, 2001), which allows for the emergence of divergent mitotypes. Under balancing selection, M-type polymorphisms in the population will not be fixed because two major mitotypes might have different fitness to the environment and mitotypes might interact with the nuclear genes to determine/differentiate the sex. Certainly, this is just speculation and more studies are needed to shed light on such aspects.

It is worth mentioning that positive selection on M-type has also been reported in other DUI species (Ort & Pogson, 2007; Śmietanka et al., 2009), and that selection on M-type inferred by population genetic tests (e.g., McDonald-Kreitman test) and phylogeny-based method (e.g., dN/dS) have been inconsistent in many cases (Ort & Pogson, 2007; Śmietanka et al., 2009; Zbawicka et al., 2010). In *R. philippinarum*, phylogenetic analysis indicated the relaxed selection on most M-type OXPHOS genes (Maeda et al., 2021), whereas population tests in this study showed a signal of positive selection or balancing selection. Although demographic events or bottleneck differences may influence population-based methods, several genes such as COX3 and NAD3 showed the same signal between the two methods. Moreover, these two genes showed different signals compared to the other genes in population-based tests (see discussion above), which cannot be explained by demographic events because we

should see the same signal across all genes if demographic changes were the major forces.

Selection on synonymous codon usage in different components of OXPHOS genes was also detected. Although mt genomes usually encode only one tRNA for each codon family, more and more evidences from numerous organisms indicated that pools of tRNA in mitochondria include both locally encoded and imported tRNAs (see Rubio & Hopper, 2011; Salinas-Giegé et al., 2015 for a review), which enables selection for both efficiency and accuracy on mt genes. The deviation from the expected ENC values in the Nc plot (Figure 3.9) and lack of correlation between GC12 and GC3 (Table 3.2) indicated a role of natural selection in CUB. Moreover, the significant and high correlation between transcription level and CUB may reflect the selective pressure to optimize the codon usage in highly transcribed mRNA to avoid sequestration of ribosomes and slow down the elongation rate (Gingold & Pilpel, 2011; Plotkin & Kudla, 2011). Similarly, translational selection was also detected in nuclear OXPHOS genes (Table 3.2; Figure 3.10). However, translational selection in mt and nuOXPHOS seems to favor codons with different endings. In both mt and nuOXPHOS genes, GC3 showed a significant impact on codon usage. Whereas GC3 in mtOXPHOS genes was lower than GCeq, GC3 in nuOXPHOS genes was higher than Gceq, suggesting that translational selection in AT-rich mt genomes drives the codons into an A/U ending, while selection in nuOXPHOS genes drives the codons into G/C ending. Despite the presence of translational selection in OXPHOS genes, we argue that the mutational bias is still a major force in OXPHOS genes. Our results revealed that CUB in OXPHOS genes is shaped by the balance between selection favoring preferred codons and mutation bias coupled with random drift.

3.4.4 The coordination of OXPHOS genes may involve translational regulation

Although selection on silent sites does not result in changes to the protein sequence, it can still drive protein evolution in terms of expression regulation. With the advent of highthroughput sequencing, increasing evidence shows that translational selection on CUB facilitates the regulation of gene expression and the generation of differential protein abundance (Camiolo et al., 2012; Horn, 2008; Jeacock et al., 2018; Najafabadi et al., 2009). It was hypothesized that codon usage may be selected during evolution to synchronize the efficiency of translation with functional requirements for the expression of specific proteins at certain times, in a specific tissue (Camiolo et al., 2012; Najafabadi et al., 2009).

The tight interaction between mt and nuOXPHOS requires the coordinated regulation of gene expression to ensure the demands for cellular energy are met. In the present study, we found separate co-transcription within F-type, M-type, and nuOXPHOS genes but weak or absent co-transcription between OXPHOS genes across genomes (Figure 3.4, 3.5), suggesting that co-regulation of OXPHOS genes is not at transcription stage. Instead, translational selection was detected for both mt and nuOXHPOS genes (Table 3.2; Figure 3.10), suggesting the possibility of co-regulation at the translation level. Although translational selection was detected in all three different components of OXPHOS genes, the selective strength seems to be different. A significant correlation was detected between CAI and transcription in nuOXPHOS genes and also weakly in nuclear protein-coding genes, but not in mtOXPHOS genes, indicating translational selection in nuOXPHOS genes may be stronger than mtOXPHOS genes. Selection on F- and M-type mt genes may also be different. Translational selection might be stronger in M-type, indicated by the significant

positive correlation between CAI and axis 1 in correspondence analysis and significant negative correlation between CAI and ENC. Therefore, our results are consistent with previous hypotheses that translation in mt genes is less efficient than nuclear genes (Adrion et al., 2016; Havird & Sloan, 2016; Pett & Lavrov, 2015; Sloan et al., 2013; Woodson & Chory, 2008). Combined with previous studies in yeast and humans that indicated translational regulation during OXPHOS complex synthesis (Couvillion et al., 2016; Soto et al., 2021), we speculate that the different strengths of translational selection on OXPHOS genes may be responsible for regulating protein abundance of OXPHOS genes and that the coordination of expression of OXPHOS genes may involve translational regulation in DUI species.

3.5 Conclusions

In addition to the common knowledge of co-transcriptional coordination between mt and nuOXPHOS genes in mammals, our study revealed that coordination in other species, particularly in DUI species, could be different and might involve post-transcriptional/translational regulation. We found a clear co-transcription signal within F-type, within M-type and within nuOXPHOS genes, but the signal is weak or absent between mt and nuOXPHOS genes, suggesting that coordination between mt and nuOXPHOS genes may not occur at the transcription level in DUI species. It will be interesting to assess if such situation is due to a peculiarity of the DUI system, or if it is more widespread across bivalves and/or other invertebrates. Translational selection on synonymous codon usage of both mt and nuOXPHOS genes further indicated the possible role of translational regulation in coordinating the OXPHOS genes. Together, these results advance our understanding of the coordination between mt and nuOXPHOS gene, and provide a new perspective of diverse and complex coordination mechanisms of OXPHOS genes in the animal world.

References

Adrion JR, White PS, Montooth KL. 2016. The roles of compensatory evolution and constraint in aminoacyl tRNA synthetase evolution. *Molecular Biology and Evolution*, **33**(1): 152–161.

Alexa A, Rahnenführer J. 2018. topGO: enrichment analysis for gene ontology. R package version 2.34.0.

Altschul SF, Madden TL, Schäffer AA, Zhang JH, Zhang Z, Miller W, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**(17): 3389–3402.

Barreto FS, Burton RS. 2013. Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Molecular Biology and Evolution*, **30**(2): 310–314.

Barreto FS, Watson ET, Lima TG, Willett CS, Edmands S, Li WZ, et al. 2018. Genomic signatures of mito-nuclear coevolution across populations of *Tigriopus californicus*. *Nature Ecology & Evolution*, **2**(8): 1250–1257.

Barrett A, Arbeithuber B, Zaidi A, Wilton P, Paul IM, Nielsen R, et al. 2019. Pronounced somatic bottleneck in mitochondrial DNA of human hair. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **375**(1790): 20190175.

Barshad G, Blumberg A, Cohen T, Mishmar D. 2018. Human primitive brain displays negative mitochondrial-nuclear expression correlation of respiratory genes. *Genome Research*, **28**(7): 952–967.

Behura SK, Severson DW. 2012. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One*, **7**(8): e43111.

Bettinazzi S, Nadarajah S, Dalpé A, Milani L, Blier PU, Breton S. 2020. Linking paternally inherited mtDNA variants and sperm performance. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **375**(1790): 20190177.

Bettinazzi S, Rodríguez E, Milani L, Blier PU, Breton S. 2019. Metabolic remodelling associated with mtDNA: insights into the adaptive value of doubly uniparental

inheritance of mitochondria. *Proceedings of the Royal Society B: Biological Sciences*, **286**(1896): 20182708.

Breton S, Bouvet K, Auclair G, Ghazal S, Sietman BE, Johnson N, et al. 2017. The extremely divergent maternally-and paternally-transmitted mitochondrial genomes are co-expressed in somatic tissues of two freshwater mussel species with doubly uniparental inheritance of mtDNA. *PLoS One*, **12**(8): e0183529.

Breton S, Ghiselli F, Passamonti M, Milani L, Stewart DT, Hoeh WR. 2011. Evidence for a Fourteenth mtDNA-encoded protein in the female-transmitted mtDNA of marine mussels (Bivalvia: Mytilidae). *PLoS One*, **6**(4): e19365.

Breton S, Stewart DT, Hoeh WR. 2010. Characterization of a mitochondrial ORF from the gender-associated mtDNAs of *Mytilus* spp. (Bivalvia: Mytilidae): identification of the "missing" ATPase 8 gene. *Marine Genomics*, **3**(1): 11–18.

Burt A, Trivers R. 2006. Genes in Conflict: The Biology of Selfish Genetic Elements. Cambridge, MA: Belknap Press of Harvard University Press.

Burzyński A, Zbawicka M, Skibinski DOF, Wenne R. 2003. Evidence for recombination of mtDNA in the marine mussel *Mytilus trossulus* from the Baltic. *Molecular Biology and Evolution*, **20**(3): 388–392.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1): 421.

Camiolo S, Farina L, Porceddu A. 2012. The relation of codon bias to tissue-specific gene expression in *Arabidopsis thaliana*. *Genetics*, **192**(2): 641–649.

Chaban Y, Boekema EJ, Dudkina NV. 2014. Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation. *Biochimica et Biophysica Acta (BBA)- Bioenergetics*, **1837**(4): 418–426.

Chase CD. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. *Trends in Genetics*, **23**(2): 81–90.

Chiesa S, Lucentini L, Freitas R, Nonnis Marzano F, Breda S, Figueira E, et al. 2017. A history of invasion: *COI* phylogeny of Manila clam *Ruditapes philippinarum* in Europe. *Fisheries Research*, **186**: 25–35.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118} ; *iso*-2; *iso*-3. Fly, **6**(2): 80–92.

Cordero D, Delgado M, Liu BZ, Ruesink J, Saavedra C. 2017. Population genetics of the Manila clam (*Ruditapes philippinarum*) introduced in North America and Europe. *Scientific Reports*, **7**(1): 39745.

Couvillion MT, Soto IC, Shipkovenska G, Churchman LS. 2016. Synchronized mitochondrial and cytosolic translation programs. *Nature*, **533**(7604): 499–503.

Dalziel AC, Stewart DT. 2002. Tissue-specific expression of male-transmitted mitochondrial DNA and its implications for rates of molecular evolution in *Mytilus* mussels (Bivalvia: Mytilidae). *Genome*, **45**(2): 348–355.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, et al. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**(15): 2156–2158.

de Oliveira JL, Morales AC, Hurst LD, Urrutia AO, Thompson CRL, Wolf JB. 2021. Inferring adaptive codon preference to understand sources of selection shaping codon usage bias. *Molecular Biology and Evolution*, **38**(8): 3247–3266.

Delph LF, Kelly JK. 2014. On the importance of balancing selection in plants. *New Phytologist*, **201**(1): 45–56.

Doherty A, McInerney JO. 2013. Translational selection frequently overcomes genetic drift in shaping synonymous codon usage patterns in vertebrates. *Molecular Biology and Evolution*, **30**(10): 2263–2267.

Dowling DK. 2014. Evolutionary perspectives on the links between mitochondrial genotype and disease phenotype. *Biochimica et Biophysica Acta (BBA)- General Subjects*, **1840**(4): 1393–1403.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**(1): 238.

Falda M, Toppo S, Pescarolo A, Lavezzo E, Di Camillo B, Facchinetti A, et al. 2012. Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics*, **13**(4): S14.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, **44**(D1): D279–D285.

Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**(23): 3150–3152.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics*, **133**(3): 693–709.

Garbian Y, Ovadia O, Dadon S, Mishmar D. 2010. Gene expression patterns of oxidative phosphorylation complex I subunits are organized in clusters. *PLoS One*, **5**(4): e9985.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907.

Ghiselli F, Gomes-dos-Santos A, Adema CM, Lopes-Lima M, Sharbrough J, Boore JL. 2021a. Molluscan mitochondrial genomes break the rules. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **376**(1825): 20200159.

Ghiselli F, Iannello M, Piccinini G, Milani L. 2021b. Bivalve molluscs as model systems for studying mitochondrial biology. *Integrative and Comparative Biology*, doi: 10.1093/icb/icab057.

Ghiselli F, Maurizii MG, Reunov A, Ariño-Bassols H, Cifaldi C, Pecci A, et al. 2019. Natural heteroplasmy and mitochondrial inheritance in bivalve molluscs. *Integrative and Comparative Biology*, **59**(4): 1016–1032.

Ghiselli F, Milani L, Chang PL, Hedgecock D, Davis JP, Nuzhdin SV, et al. 2012. De novo assembly of the Manila clam *Ruditapes philippinarum* transcriptome provides new insights into expression bias, mitochondrial doubly uniparental inheritance and sex determination. *Molecular Biology and Evolution*, **29**(2): 771–786.

Ghiselli F, Milani L, Guerra D, Chang PL, Breton S, Nuzhdin SV, et al. 2013. Structure, transcription, and variability of metazoan mitochondrial genome: perspectives from an unusual mitochondrial inheritance system. *Genome Biology and Evolution*, **5**(8): 1535–1554.

Ghiselli F, Milani L, Passamonti M. 2011. Strict sex-specific mtDNA segregation in the germ line of the dui species *Venerupis philippinarum* (Bivalvia: Veneridae). *Molecular Biology and Evolution*, **28**(2): 949–961.

Gingold H, Pilpel Y. 2011. Determinants of translation efficiency and accuracy. *Molecular Systems Biology*, **7**(1): 481.

Gusman A, Lecomte S, Stewart DT, Passamonti M, Breton S. 2016. Pursuing the quest for better understanding the taxonomic distribution of the system of doubly uniparental inheritance of mtdna. *PeerJ*, **4**: e2760.

Havird JC, Sloan DB. 2016. The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial versus nuclear genomes. *Molecular Biology and Evolution*, **33**(12): 3042–3053.

Havird JC, Trapp P, Miller CM, Bazos I, Sloan DB. 2017. Causes and consequences of rapidly evolving mtDNA in a plant lineage. *Genome Biology and Evolution*, **9**(2): 323–336.

Hershberg R, Petrov DA. 2009. General rules for optimal codon choice. *PLoS Genetics*, **5**(7): e1000556.

Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content in bacteria. *PLoS Genetics*, **6**(9): e1001107.

Hill GE. 2020. Mito-nuclear compensatory coevolution. *Trends in Genetics*, **36**(6): 403–414.

Hill GE, Havird JC, Sloan DB, Burton RS, Greening C, Dowling DK. 2019. Assessing the fitness consequences of mito-nuclear interactions in natural populations. *Biological Reviews*, **94**(3): 1089–1104.

Hoeh WR, Stewart DT, Guttman SI. 2002. High fidelity of mitochondrial genome transmission under the doubly uniparental mode of inheritance in freshwater mussels (Bivalvia: Unionoidea). *Evolution*, **56**(11): 2252–2261.

Hoeh WR, Stewart DT, Saavedra C, Sutherland BW, Zouros E. 1997. Phylogenetic evidence for role-reversals of gender-associated mitochondrial DNA in mytilus (Bivalvia: Mytilidae). *Molecular Biology and Evolution*, **14**(9): 959–967.

Hoeh WR, Stewart DT, Sutherland BW, Zouros E. 1996. Cytochrome c oxidase sequence comparisons suggest an unusually high rate of mitochondrial DNA evolution in mytilus (Mollusca: Bivalvia). *Molecular Biology and Evolution*, **13**(2): 418–421.

Horn D. 2008. Codon usage suggests that translational selection has a major impact on protein expression in trypanosomatids. *BMC Genomics*, **9**(1): 2.

Iannello M, Bettinazzi S, Breton S, Ghiselli F, Milani L. 2021. A naturally heteroplasmic clam provides clues about the effects of genetic bottleneck on paternal mtDNA. *Genome Biology and Evolution*, **13**(3): evab022.

Iannello M, Puccio G, Piccinini G, Passamonti M, Ghiselli F. 2019. The dynamics of mito-nuclear coevolution: a perspective from bivalve species with two different mechanisms of mitochondrial inheritance. *Journal of Zoological Systematics and Evolutionary Research*, **57**(3): 534–547.

James JE, Piganeau G, Eyre-Walker A. 2016. The rate of adaptive evolution in animal mitochondria. *Molecular Ecology*, **25**(1): 67–78.

James White D, Nikolai Wolff J, Pierson M, John Gemmell N. 2008. Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, **17**(23): 4925–4942.

Jeacock L, Faria J, Horn D. 2018. Codon usage bias controls mRNA and protein abundance in trypanosomatids. *elife*, **7**: e32496.

Jia WL, Higgs PG. 2008. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. *Molecular Biology and Evolution*, **25**(2): 339–351.

Johri P, Marinov GK, Doak TG, Lynch M. 2019. Population genetics of *Paramecium* mitochondrial genomes: recombination, mutation spectrum, and efficacy of selection. *Genome Biology and Evolution*, **11**(5): 1398–1416.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, **15**(1): 356.

Krasileva KV, Buffalo V, Bailey P, Pearce S, Ayling S, Tabbita F, et al. 2013. Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biology*, **14**(6): R66.

Ladoukakis ED, Zouros E. 2001. Recombination in animal mitochondrial DNA: evidence from published sequences. *Molecular Biology and Evolution*, **18**(11): 2127–2131.

Lane N. 2011. Mito-nuclear match: optimizing fitness and fertility over generations drives ageing within generations. *BioEssays*, **33**(11): 860–869.

Lane N. 2012. The problem with mixing mitochondria. *Cell*, **151**(2): 246–248.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357–359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078–2079.

Lightowlers RN, Chrzanowska-Lightowlers ZMA. 2008. PPR (pentatricopeptide repeat) proteins in mammals: Important aids to mitochondrial gene expression. *Biochemical Journal*, **416**(1): e5–e6.

Liu HP, Mitton JB, Wu SK. 1996. Paternal mitochondrial DNA differentiation far exceeds maternal mitochondrial DNA and allozyme differentiation in the freshwater mussel, *Anodonta grandis grandis*. *Evolution*, **50**(2): 952–957.

Lynch M. 2007. The Origins of Genome Architecture. Sunderland, MA: Sinauer Associates.

Machado HE, Lawrie DS, Petrov DA. 2020. Pervasive strong selection at the level of codon usage bias in *Drosophila melanogaster*. *Genetics*, **214**(2): 511–528.

Maeda GP, Iannello M, McConie HJ, Ghiselli F, Havird JC. 2021. Relaxed selection on male mitochondrial genes in DUI bivalves eases the need for mito-nuclear coevolution. *Journal of Evolutionary Biology*, **34**(11): 1722–1736.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, **351**(6328): 652–654.

Miglani GS, Kaur M, Manchanda P. 2021. Pentatricopeptide repeat proteins: I. Involvement in RNA editing in plants. *Journal of Crop Improvement*, **35**(5): 605–653.

Milani L, Ghiselli F, Guerra D, Breton S, Passamonti M. 2013a. A comparative analysis of mitochondrial ORFans: new clues on their origin and role in species with doubly

uniparental inheritance of mitochondria. *Genome Biology and Evolution*, **5**(7): 1408–1434.

Milani L, Ghiselli F, Iannello M, Passamonti M. 2014. Evidence for somatic transcription of male-transmitted mitochondrial genome in the DUI species *Ruditapes philippinarum* (Bivalvia: Veneridae). *Current Genetics*, **60**(3): 163–173.

Milani L, Ghiselli F, Nuzhdin SV, Passamonti M. 2013b. Nuclear genes with sex bias in *Ruditapes philippinarum* (Bivalvia, veneridae): mitochondrial inheritance and sex determination in DUI species. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, **320**(7): 442–454.

Milani L, Ghiselli F, Passamonti M. 2016. Mitochondrial selfish elements and the evolution of biological novelties. *Current Zoology*, **62**(6): 687–697.

Milenkovic D, Blaza JN, Larsson NG, Hirst J. 2017. The enigma of the respiratory chain supercomplex. *Cell Metabolism*, **25**(4): 765–776.

Minoiu I, Burzyński A, Breton S. 2016. Analysis of the coding potential of the ORF in the control region of the female-transmitted *Mytilus* mtDNA. *Gene*, **576**(1): 586–588.

Mioduchowska M, Kaczmarczyk A, Zając K, Zając T, Sell J. 2016. Gender-associated mitochondrial DNA heteroplasmy in somatic tissues of the endangered freshwater mussel *Unio crassus* (Bivalvia: Unionidae): implications for sex identification and phylogeographical studies. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, **325**(9): 610–625.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*, **41**(12): e121.

Mitchell A, Guerra D, Stewart D, Breton S. 2016. *In silico* analyses of mitochondrial ORFans in freshwater mussels (Bivalvia: Unionoida) provide a framework for future studies of their origin and function. *BMC Genomics*, **17**(1): 597.

Najafabadi HS, Goodarzi H, Salavati R. 2009. Universal function-specificity of codon usage. *Nucleic Acids Research*, **37**(21): 7014–7023.

Obata M, Sano N, Komaru A. 2011. Different transcriptional ratios of male and female transmitted mitochondrial DNA and tissue-specific expression patterns in the blue

mussel, *Mytilus galloprovincialis*. *Development*, *Growth & Differentiation*, **53**(7): 878–886.

Ort BS, Pogson GH. 2007. Molecular population genetics of the male and female mitochondrial DNA molecules of the California Sea Mussel, *Mytilus californianus*. *Genetics*, **177**(2): 1087–1099.

Parmley JL, Hurst LD. 2007. Exonic splicing regulatory elements skew synonymous codon usage near intron-exon boundaries in mammals. *Molecular Biology and Evolution*, **24**(8): 1600–1603.

Perez MF, Sarkies P. 2021. Malignancy and NF-kB signalling strengthen coordination between the expression of mitochondrial and nuclear-encoded oxidative phosphorylation genes. *BioRxiv*, doi: 10.1101/2021.06.30.450588.

Pett W, Lavrov DV. 2015. Cytonuclear interactions in the evolution of animal mitochondrial tRNA metabolism. *Genome Biology and Evolution*, **7**(8): 2089–2101.

Piccinini G, Iannello M, Puccio G, Plazzi F, Havird JC, Ghiselli F. 2021. Mito-nuclear coevolution, but not nuclear compensation, drives evolution of OXPHOS complexes in Bivalves. *Molecular Biology and Evolution*, **38**(6): 2597–2614.

Pitroda SP, Wakim BT, Sood RF, Beveridge MG, Beckett MA, Macdermed DM, et al. 2009. STAT1-dependent expression of energy metabolic pathways links tumour growth and radioresistance to the Warburg effect. *BMC Medicine*, **7**(1): 68.

Plotkin JB, Kudla G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, **12**(1): 32–42.

Puigbò P, Bravo IG, Garcia-Vallve S. 2008. CAIcal: a combined set of tools to assess codon usage adaptation. *Biology Direct*, **3**(1): 38.

Quesada H, Gallagher C, Skibinski DAG, Skibinski DOF. 1998. Patterns of polymorphism and gene flow of gender-associated mitochondrial DNA lineages in European mussel populations. *Molecular Ecology*, **7**(8): 1041–1051.

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends in Ecology & Evolution*, **19**(12): 645–653.

Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. *Molecular Biology and Evolution*, **13**(6): 735–748.

Rozas J, Ferrer-Mata A, Carlos Sánchez-DelBarrio JS, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, **34**(12): 3299–3302.

Rubio MAT, Hopper AK. 2011. Transfer RNA travels from the cytoplasm to organelles. *WIREs RNA*, **2**(6): 802–817.

Salinas-Giegé T, Giegé R, Giegé P. 2015. tRNA biology in mitochondria. *International Journal of Molecular Sciences*, **16**(3): 4518–4559.

Schmitz-Linneweber C, Small I. 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends in Plant Science*, **13**(12): 663–670.

Sharp PM, Li WH. 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, **15**(3): 1281–1295.

Shen W, Le S, Li Y, Hu FQ. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*, **11**(10): e0163962.

Shyamsundar R, Kim YH, Higgins JP, Montgomery K, Jorden M, Sethuraman A, et al. 2005. A DNA microarray survey of gene expression in normal human tissues. *Genome Biology*, **6**(3): R22.

Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, **27**(1): 135–145.

Skibinski DOF, Gallagher C, Beynon CM. 1994. Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. *Genetics*, **138**(3): 801–809.

Skibinski DOF, Ghiselli F, Diz AP, Milani L, Mullins JGL. 2017. Structure-related differences between cytochrome oxidase I proteins in a stable heteroplasmic mitochondrial system. *Genome Biology and Evolution*, **9**(12): 3265–3281.

Slane BG, Aykin-Burns N, Smith BJ, Kalen AL, Goswami PC, Domann FE, et al. 2006. Mutation of succinate dehydrogenase subunit C results in increased $O_2^{\cdot -}$, oxidative stress, and genomic instability. *Cancer Research*, **66**(15): 7615–7620.

Sloan DB, Triant DA, Wu M, Taylor DR. 2013. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Molecular Biology and Evolution*, **31**(3): 673–682.

Smid M, Coebergh van den Braak RRJ, van de Werken HJG, van Riet J, van Galen A, de Weerd V, et al. 2018. Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics*, **19**(1): 236.

Soroka M, Burzyński A. 2010. Complete sequences of maternally inherited mitochondrial genomes in mussels *Unio pictorum* (Bivalvia, Unionidae). *Journal of Applied Genetics*, **51**(4): 469–476.

Soto I, Couvillion M, Mcshane E, Hansen KG, Moran JC, Barrientos A, et al. 2021. Balanced mitochondrial and cytosolic translatomes underlie the biogenesis of human respiratory complexes. *BioRxiv*, doi: 10.1101/2021.05.31.446345.

Stewart DT, Kenchington ER, Singh RK, Zouros E. 1996. Degree of selective constraint as an explanation of the different rates of evolution of gender-specific mitochondrial DNA Lineages in the Mussel Mytilus. *Genetics*, **143**(3): 1349–1357.

Stewart DT, Saavedra C, Stanwood RR, Ball AO, Zouros E. 1995. Male and female mitochondrial DNA lineages in the blue mussel (*Mytilus edulis*) Species Group. *Molecular Biology and Evolution*, **12**(5): 735–747.

Stewart JB, Chinnery PF. 2015. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature Reviews Genetics*, **16**(9): 530–542.

Stoletzki N, Eyre-Walker A. 2011. Estimation of the neutrality index. *Molecular Biology and Evolution*, **28**(1): 63–70.

Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene co-expression network for global discovery of conserved genetic modules. *Science*, **302**(5643): 249–255.

Sueoka N. 1999. Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of *A=T* and *G=C*. *Journal of Molecular Evolution*, **49**(1): 49–62.

Sun Z, Ma L, Murphy RW, Zhang XS, Huang DW. 2008. Factors affecting mito-nuclear codon usage interactions in the OXPHOS system of *Drosophila melanogaster*. *Journal of Genetics and Genomics*, **35**(12): 729–735.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**(S2): W609–W612.

Śmietanka B, Burzyński A, Wenne R. 2009. Molecular population genetics of male and female mitochondrial genomes in European mussels *Mytilus*. *Marine Biology*, **156**(5): 913–925.

Śmietanka B, Zbawicka M, Sańko T, Wenne R, Burzyński A. 2013. Molecular population genetics of male and female mitochondrial genomes in subarctic *Mytilus trossulus*. *Marine Biology*, **160(7)**: 1709–1721.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**(3): 585–595.

Tang JX, Thompson K, Taylor RW, Oláhová M. 2020. Mitochondrial OXPHOS biogenesis: Co-regulation of protein synthesis, import, and assembly pathways. *International Journal of Molecular Sciences*, **21**(11): 3820.

Tarazona S, Furió-Tarí P, Turrà D, Di Pietro A, Nueda MJ, Ferrer A, et al. 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Research*, **43**(21): e140.

Tretter L, Patocs A, Chinopoulos C. 2016. Succinate, an intermediate in metabolism, signal transduction, ROS, hypoxia, and tumorigenesis. *Biochimica et Biophysica Acta (BBA)- Bioenergetics*, **1857**(8): 1086–1101.

van Waveren C, Moraes CT. 2008. Transcriptional co-expression and co-regulation of genes coding for components of the oxidative phosphorylation system. *BMC Genomics*, **9**(1): 18.

Wang B, Yuan J, Liu J, Jin L, Chen JQ. 2011. Codon usage bias and determining forces in green plant mitochondrial genomes. *Journal of Integrative Plant Biology*, **53**(4): 324–334.

Wei L, He J, Jia X, Qi Q, Liang ZS, Zheng H, et al. 2014. Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evolutionary Biology*, **14**(1): 14.

Willems PHGM, Rossignol R, Dieteren CEJ, Murphy MP, Koopman WJH. 2015. Redox homeostasis and mitochondrial dynamics. *Cell Metabolism*, **22**(2): 207–218.

Woodson JD, Chory J. 2008. Coordination of gene expression between organellar and nuclear genomes. *Nature Reviews Genetics*, **9**(5): 383–395.

Wright F. 1990. The 'effective number of codons' used in a gene. *Gene*, **87**(1): 23–29.

Yan ZC, Ye GY, Werren JH. 2019. Evolutionary rate correlation between mitochondrial-encoded and mitochondria-associated nuclear-encoded proteins in insects. *Molecular Biology and Evolution*, **36**(5): 1022–1036.

Yu CH, Dang YK, Zhou ZP, Wu C, Zhao FZ, Sachs MS, et al. 2015. Codon usage influences the local rate of translation elongation to regulate Co-translational protein folding. *Molecular Cell*, **59**(5): 744–754.

Yusa Y, Breton S, Hoeh WR. 2013. Population genetics of sex determination in *Mytilus* mussels: reanalyses and a model. *Journal of Heredity*, **104**(3): 380–385.

Zbawicka M, Burzyński A, Skibinski D, Wenne R. 2010. Scottish *Mytilus trossulus* mussels retain ancestral mitochondrial DNA: complete sequences of male and female mtDNA genomes. *Gene*, **456**(1–2): 45–53.

Zhang HX, Burr SP, Chinnery PF. 2018. The mitochondrial DNA genetic bottleneck: inheritance and beyond. *Essays in Biochemistry*, **62**(3): 225–234.

Zhou T, Weems M, Wilke CO. 2009. Translationally optimal codons associate with structurally sensitive sites in proteins. *Molecular Biology and Evolution*, **26**(7): 1571–1580.

Zouros E. 2013. Biparental inheritance through uniparental transmission: the Doubly Uniparental Inheritance (DUI) of Mitochondrial DNA. *Evolutionary Biology*, **40**(1): 1–31.

Supplementary Files

Supplementary Table 3.1 Statistical test for the transcriptional differences across tissues.

		f_A	f_G	f_M	m_A	m_G
Acorss tissue comparison for nuOXPHOS (Kruskal-Wallis p<0.001)	f_G	**				
	f_M	**	ns			
	m_A	ns	**	**		
	m_G	ns	**	**	ns	
	m_M	**	ns	ns	**	**
		f_A	f_G	f_M	m_A	m_G
Across tissue comparison for mtOXPHOS (Kruskal-Wallis p<0.001)	f_G	*				
	f_M	ns	**			
	m_A	ns	ns	**		
	m_G	**	**	**	**	
	m_M	ns	**	ns	**	**

Note: Kruskal-Wallis tests followed by Dunn test with Bonferroni corrections was performed to test transcriptional differences for both mt and nuOXPHOS across tissues. * - p<0.05, **p<0.001, ns -non significant.

Supplementary Table 3.2 The enriched GO terms for the genes co-transcribed with the F-type, M-type and nuOXPHOS genes.

	GO.ID	Term	Annotated	Significant	Expected	Fisher
F_BP	GO:0019725	cellular homeostasis	15	3	0.14	0.00029
	GO:0042592	homeostatic process	22	3	0.2	0.00093
	GO:0045454	cell redox homeostasis	6	2	0.06	0.00119
	GO:0033108	mitochondrial respiratory chain complex assembly	9	2	0.08	0.0028
	GO:0065008	regulation of biological quality	43	3	0.4	0.0066
M_BP	GO:0022414	reproductive process	27	13	1.76	2.20E-09
	GO:0000003	reproduction	31	13	2.02	1.80E-08
	GO:0051704	multi-organism process	43	13	2.8	1.60E-06
	GO:0007276	gamete generation	16	7	1.04	3.10E-05
	GO:0007283	spermatogenesis	16	7	1.04	3.10E-05
	GO:0019953	sexual reproduction	16	7	1.04	3.10E-05
	GO:0044703	multi-organism reproductive process	16	7	1.04	3.10E-05
	GO:0048232	male gamete generation	16	7	1.04	3.10E-05
	GO:0048609	multicellular organismal reproductive process	16	7	1.04	3.10E-05
	GO:0051321	meiotic cell cycle	9	5	0.59	0.00011
	GO:0007017	microtubule-based process	98	17	6.39	0.00013
	GO:0032504	multicellular organism reproduction	20	7	1.3	0.00017
	GO:0016310	phosphorylation	110	18	7.17	0.00018
	GO:0044419	biological process involved in interspecies interaction between organisms	17	6	1.11	0.00047
	GO:0007127	meiosis I	7	4	0.46	0.00052
	GO:0061982	meiosis I cell cycle process	7	4	0.46	0.00052
	GO:0140013	meiotic nuclear division	7	4	0.46	0.00052
	GO:0007018	microtubule-based movement	31	8	2.02	0.00059
	GO:0006928	movement of cell or subcellular component	48	10	3.13	0.00079
	GO:0046939	nucleotide phosphorylation	19	6	1.24	0.00093
	GO:1903046	meiotic cell cycle process	8	4	0.52	0.00099
	GO:0016032	viral process	9	4	0.59	0.00169
	GO:0006793	phosphorus metabolic process	213	25	13.88	0.00212
	GO:0006796	phosphate-containing compound metabolic process	213	25	13.88	0.00212
	GO:0000280	nuclear division	22	6	1.43	0.00217
	GO:0044403	symbiont process	10	4	0.65	0.00268
	GO:0051716	cellular response to stimulus	229	26	14.92	0.00278
	GO:0006310	DNA recombination	23	6	1.5	0.00278
	GO:0048285	organelle fission	23	6	1.5	0.00278
	GO:0072527	pyrimidine-containing compound metabolic process	13	4	0.85	0.00781
	GO:0003006	developmental process involved in reproduction	7	3	0.46	0.00783
	GO:0022402	cell cycle process	55	9	3.58	0.00802
	GO:0007049	cell cycle	75	11	4.89	0.00815

	GO:0006281	DNA repair	66	10	4.3	0.00917
	GO:0034404	nucleobase-containing small molecule biosynthetic process	21	5	1.37	0.00957
	GO:0050896	response to stimulus	302	30	19.68	0.00987
Nu_BP	GO:0006412	translation	196	90	36.36	3.00E-20
	GO:0043043	peptide biosynthetic process	198	90	36.73	7.20E-20
	GO:0043604	amide biosynthetic process	200	90	37.11	1.70E-19
	GO:0006518	peptide metabolic process	202	90	37.48	3.90E-19
	GO:0043603	cellular amide metabolic process	213	92	39.52	1.90E-18
	GO:1901566	organonitrogen compound biosynthetic process	309	111	57.33	5.30E-15
	GO:0044267	cellular protein metabolic process	414	130	76.81	2.30E-12
	GO:0019538	protein metabolic process	510	143	94.62	2.00E-09
	GO:0044271	cellular nitrogen compound biosynthetic process	508	132	94.25	1.90E-06
	GO:1901564	organonitrogen compound metabolic process	690	169	128.01	2.50E-06
	GO:0005977	glycogen metabolic process	7	7	1.3	7.30E-06
	GO:0006112	energy reserve metabolic process	7	7	1.3	7.30E-06
	GO:0007005	mitochondrion organization	51	23	9.46	1.00E-05
	GO:0034645	cellular macromolecule biosynthetic process	485	124	89.98	1.10E-05
	GO:0009059	macromolecule biosynthetic process	488	124	90.54	1.60E-05
	GO:1901576	organic substance biosynthetic process	613	149	113.73	2.30E-05
	GO:0009058	biosynthetic process	629	152	116.7	2.60E-05
	GO:0044249	cellular biosynthetic process	609	147	112.99	4.20E-05
	GO:0044262	cellular carbohydrate metabolic process	11	8	2.04	0.00013
	GO:0044260	cellular macromolecule metabolic process	776	177	143.97	0.00018
	GO:0006073	cellular glucan metabolic process	9	7	1.67	0.00019
	GO:0044042	glucan metabolic process	9	7	1.67	0.00019
	GO:0044264	cellular polysaccharide metabolic process	9	7	1.67	0.00019
	GO:0071806	protein transmembrane transport	11	7	2.04	0.0012
	GO:0031032	actomyosin structure organization	9	6	1.67	0.002
	GO:0005976	polysaccharide metabolic process	12	7	2.23	0.00242
	GO:0010467	gene expression	598	135	110.94	0.00256
F_CC	GO:0005737	cytoplasm	1171	18	9.02	0.0004
	GO:0044444	cytoplasmic part	670	11	5.16	0.0075
M_CC	GO:0044430	cytoskeletal part	214	36	13.85	4.20E-08
	GO:0015630	microtubule cytoskeleton	167	29	10.81	5.10E-07
	GO:0005856	cytoskeleton	262	38	16.96	9.60E-07
	GO:0042995	cell projection	74	16	4.79	1.30E-05
	GO:0031514	motile cilium	32	10	2.07	1.90E-05
	GO:0005929	cilium	62	14	4.01	2.70E-05
	GO:0019012	virion	8	5	0.52	5.20E-05
	GO:0120025	plasma membrane bounded cell projection	68	14	4.4	8.10E-05
	GO:0005815	microtubule organizing center	56	12	3.62	0.00018
	GO:0018995	host	3	3	0.19	0.00027
	GO:0033643	host cell part	3	3	0.19	0.00027

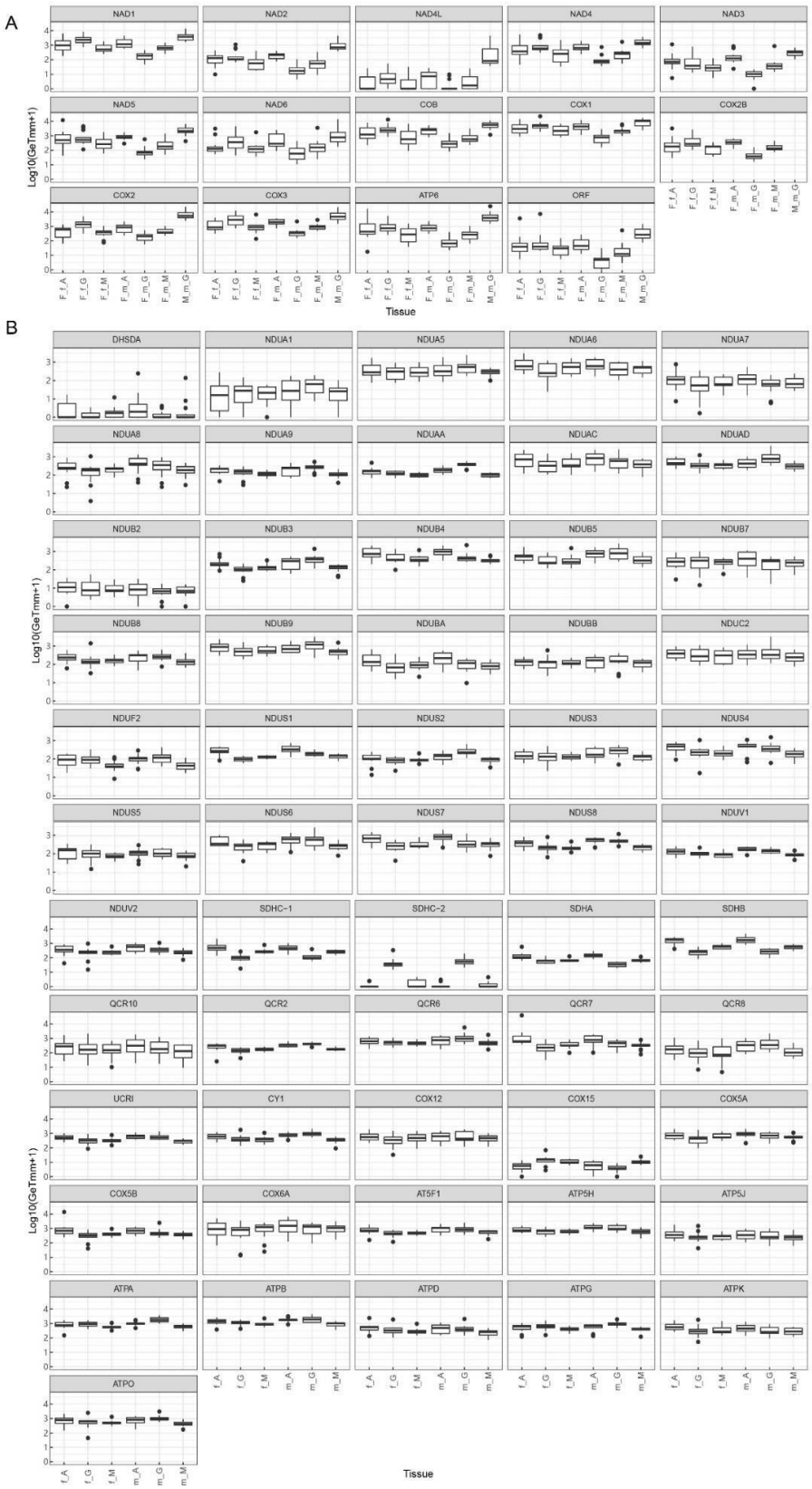
	GO:0033646	host intracellular part	3	3	0.19	0.00027
	GO:0033647	host intracellular organelle	3	3	0.19	0.00027
	GO:0033648	host intracellular membrane-bounded organelle	3	3	0.19	0.00027
	GO:0042025	host cell nucleus	3	3	0.19	0.00027
	GO:0043656	intracellular region of host	3	3	0.19	0.00027
	GO:0043657	host cell	3	3	0.19	0.00027
	GO:0044215	other organism	3	3	0.19	0.00027
	GO:0044216	other organism cell	3	3	0.19	0.00027
	GO:0044217	other organism part	3	3	0.19	0.00027
	GO:0000794	condensed nuclear chromosome	7	4	0.45	0.00051
	GO:0044423	virion part	7	4	0.45	0.00051
	GO:0005930	axoneme	9	4	0.58	0.00166
	GO:0097014	ciliary plasm	9	4	0.58	0.00166
	GO:0032838	plasma membrane bounded cell projection cytoplasm	10	4	0.65	0.00263
	GO:1990716	axonemal central apparatus	2	2	0.13	0.00417
	GO:0019028	viral capsid	6	3	0.39	0.00462
	GO:0005875	microtubule associated complex	45	8	2.91	0.00727
Nu_CC	GO:0005840	ribosome	133	78	24.2	2.00E-26
	GO:0005737	cytoplasm	1171	296	213.03	2.90E-14
	GO:1990904	ribonucleoprotein complex	242	90	44.03	2.90E-13
	GO:0044444	cytoplasmic part	670	188	121.89	1.20E-12
	GO:0044391	ribosomal subunit	38	22	6.91	4.50E-08
	GO:0005739	mitochondrion	134	50	24.38	7.50E-08
	GO:0005759	mitochondrial matrix	20	13	3.64	4.70E-06
	GO:0032991	protein-containing complex	871	202	158.45	9.50E-06
	GO:0098798	mitochondrial protein complex	39	19	7.09	1.20E-05
	GO:0015934	large ribosomal subunit	22	13	4	2.10E-05
	GO:0044429	mitochondrial part	85	31	15.46	4.10E-05
	GO:0000313	organellar ribosome	14	9	2.55	0.00017
	GO:0005761	mitochondrial ribosome	14	9	2.55	0.00017
	GO:0005838	proteasome regulatory particle	10	7	1.82	0.00046
	GO:0000502	proteasome complex	38	16	6.91	0.0005
	GO:1905369	endopeptidase complex	38	16	6.91	0.0005
	GO:0043232	intracellular non-membrane-bounded organelle	603	139	109.7	0.00054
F_MF	GO:0016782	transferase activity, transferring sulfur-containing groups	17	2	0.14	0.0081
M_MF	GO:0016301	kinase activity	250	40	15.12	5.10E-09
	GO:0016773	phosphotransferase activity, alcohol group as acceptor	208	33	12.58	1.60E-07
	GO:0016772	transferase activity, transferring phosphorus-containing groups	304	40	18.38	1.30E-06
	GO:0016849	phosphorus-oxygen lyase activity	6	5	0.36	4.40E-06
	GO:0004672	protein kinase activity	186	28	11.25	4.50E-06
	GO:0017076	purine nucleotide binding	661	61	39.97	0.00026
	GO:0032555	purine ribonucleotide binding	661	61	39.97	0.00026
	GO:0032553	ribonucleotide binding	666	61	40.27	0.00032
	GO:0030554	adenyl nucleotide binding	529	51	31.99	0.00033

	GO:0032559	adenyl ribonucleotide binding	529	51	31.99	0.00033
	GO:0042578	phosphoric ester hydrolase activity	83	14	5.02	0.00037
	GO:0004112	cyclic-nucleotide phosphodiesterase activity	7	4	0.42	0.0004
	GO:0004114	3',5'-cyclic-nucleotide phosphodiesterase activity	7	4	0.42	0.0004
	GO:0097367	carbohydrate derivative binding	704	63	42.57	0.00048
	GO:0035639	purine ribonucleoside triphosphate binding	652	59	39.43	0.00057
	GO:0043168	anion binding	800	69	48.37	0.0007
	GO:0005524	ATP binding	520	49	31.44	0.00075
	GO:0008144	drug binding	576	53	34.83	0.00078
	GO:0000166	nucleotide binding	819	70	49.52	0.00082
	GO:1901265	nucleoside phosphate binding	819	70	49.52	0.00082
	GO:0036094	small molecule binding	882	74	53.33	0.00096
	GO:0000287	magnesium ion binding	29	7	1.75	0.00134
	GO:0016740	transferase activity	628	55	37.97	0.00201
	GO:0140096	catalytic activity, acting on a protein	577	50	34.89	0.00416
	GO:0008081	phosphoric diester hydrolase activity	19	5	1.15	0.00448
	GO:0003774	motor activity	76	11	4.6	0.00544
	GO:0043167	ion binding	1683	121	101.77	0.0069
	GO:0097159	organic cyclic compound binding	1564	113	94.57	0.00857
	GO:1901363	heterocyclic compound binding	1564	113	94.57	0.00857
u_MF	GO:0003735	structural constituent of ribosome	121	72	21.26	9.00E-26
	GO:0005198	structural molecule activity	236	94	41.46	6.70E-17
	GO:0019843	rRNA binding	63	34	11.07	4.60E-11
	GO:0005267	potassium channel activity	21	15	3.69	7.80E-08
	GO:0015079	potassium ion transmembrane transporter activity	25	15	4.39	2.30E-06
	GO:0005261	cation channel activity	42	20	7.38	6.50E-06
	GO:0022838	substrate-specific channel activity	77	29	13.53	1.90E-05
	GO:0005216	ion channel activity	74	28	13	2.40E-05
	GO:0015267	channel activity	80	29	14.05	4.40E-05
	GO:0022803	passive transmembrane transporter activity	80	29	14.05	4.40E-05
	GO:0003723	RNA binding	304	79	53.4	8.60E-05
	GO:0005516	calmodulin binding	22	12	3.86	9.10E-05
	GO:0008135	translation factor activity, RNA binding	80	28	14.05	0.00012
	GO:0003779	actin binding	123	38	21.61	0.00018
	GO:0022839	ion gated channel activity	59	22	10.36	0.00023
	GO:0022836	gated channel activity	60	22	10.54	0.0003
	GO:0046873	metal ion transmembrane transporter activity	79	26	13.88	0.00063
	GO:0005244	voltage-gated ion channel activity	26	12	4.57	0.00069
	GO:0022832	voltage-gated channel activity	26	12	4.57	0.00069
	GO:0004339	glucan 1,4-alpha-glucosidase activity	4	4	0.7	0.00095
	GO:0022843	voltage-gated cation channel activity	14	8	2.46	0.00097
	GO:0003743	translation initiation factor activity	57	20	10.01	0.00108
	GO:0005249	voltage-gated potassium channel activity	12	7	2.11	0.00175
	GO:0015926	glucosidase activity	7	5	1.23	0.00254

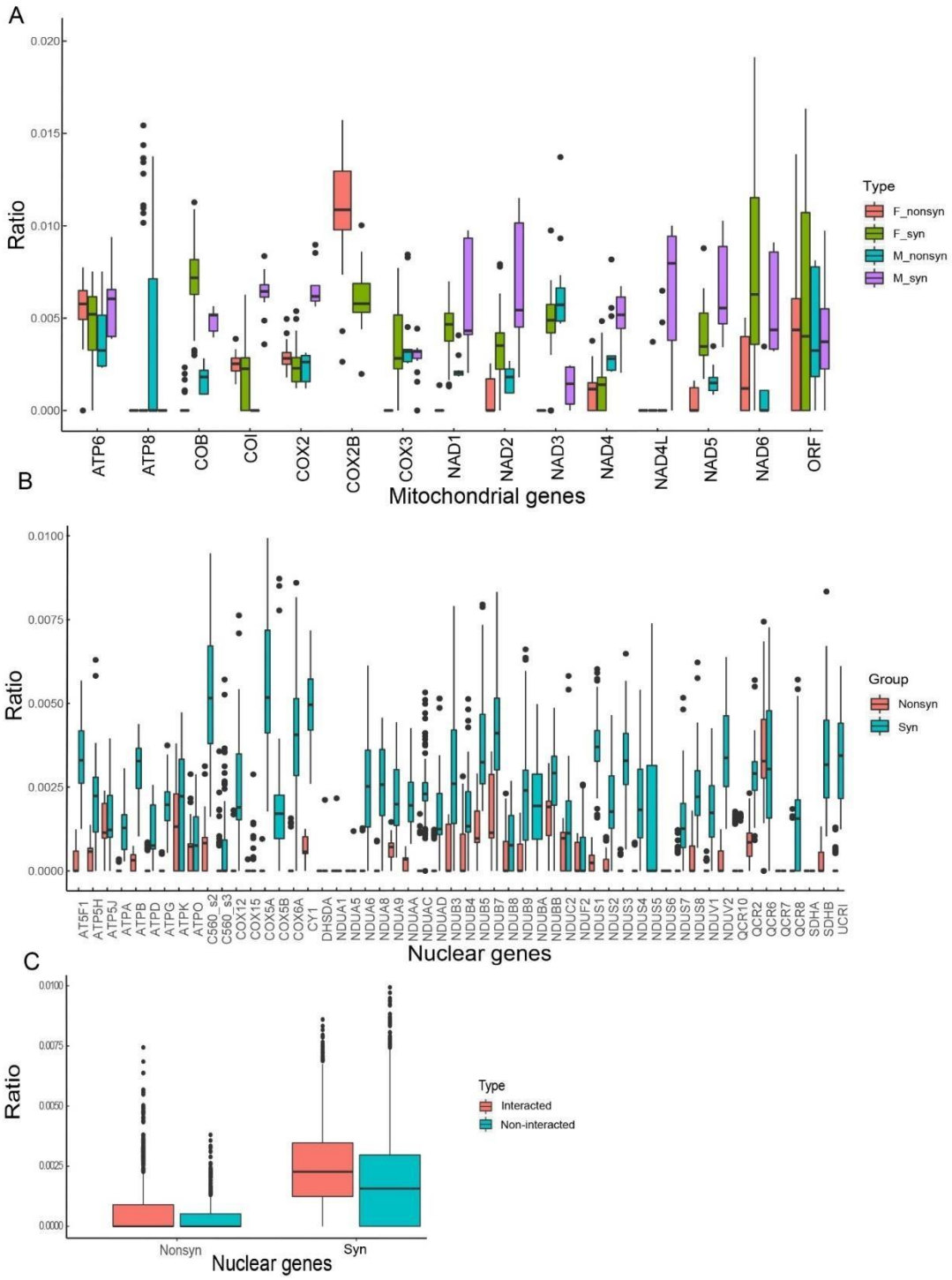
GO:0019200	carbohydrate kinase activity	7	5	1.23	0.00254
GO:0015077	inorganic cation transmembrane transporter activity	74	23	13	0.003
GO:0004396	hexokinase activity	3	3	0.53	0.0054
GO:0005231	excitatory extracellular ligand-gated ion channel activity	3	3	0.53	0.0054
GO:0005234	extracellularly glutamate-gated ion channel activity	3	3	0.53	0.0054
GO:0005536	glucose binding	3	3	0.53	0.0054
GO:0030943	mitochondrion targeting sequence binding	3	3	0.53	0.0054
GO:0071933	Arp2/3 complex binding	3	3	0.53	0.0054
GO:0015450	P-P-bond-hydrolysis-driven protein transmembrane transporter activity	8	5	1.41	0.00579
GO:0043022	ribosome binding	8	5	1.41	0.00579
GO:0044877	protein-containing complex binding	70	21	12.3	0.00705
GO:0022890	inorganic cation transmembrane transporter activity	115	31	20.2	0.00722
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	42	14	7.38	0.00978

Supplementary Table S3.3 The candidate genes that might be involved in the co-transcriptional regulation.

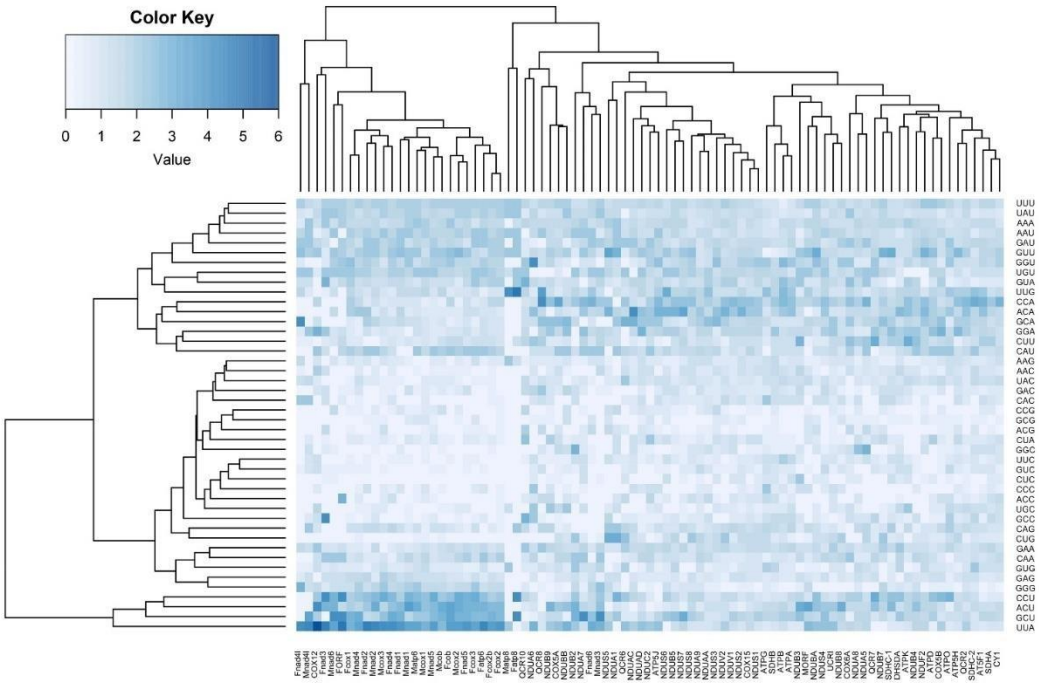
OXPHOS	Locus	No. of gene	Annotation in NR db
F-type	Locus1895123	1	Zinc finger protein 558
	Locus226176	3	DNA-directed RNA polymerase II subunit RPB9
M-type	Locus8350352	2	Zinc finger protein 341
	Locus7670362	1	Dynein light chain 1
	Locus572990	4	APOBEC1 complementation factor
	Locus382505	3	Eukaryotic translation initiation factor 3 subunit G
M-type & Nu	Locus1618449	1	Transcriptional activator protein Pur-beta
Nu	Locus4607497	4	Pentatricopeptide repeat domain-containing protein 3
	Locus2398964	3	DNA-directed RNA polymerase III subunit RPC3
	Locus1549169	3	Muscle blind-like protein 3
	Locus1065041	4	BRCA1-A complex subunit BRE
	Locus4177360	4	Transcription factor RFX3
	Locus7072300	3	Transcriptional coactivator YAP1
	Locus7055868	3	Y-box factor homolog
	Locus1872614	2	Cyclic AMP-dependent transcription factor ATF-4



Supplementary Figure S3.1 The transcriptional levels of mitochondrial (A) and nuclear OXPHOS genes in each tissue (B).



Supplementary Figure S3.2 The distribution of synonymous and nonsynonymous SNPs in OXPHOS genes. A: Mitochondrial genes; B: Nuclear OXPHOS genes; C: Interacted and non-interacted nuclear OXPHOS genes.



Supplementary Figure S3.3 The heatmap based on the RSCU values for four-fold degenerate families. Considering the Met, Trp, Ser, Arg and Ile were different between mt and nuOXPHOS, they were not included here.

Acknowledgements

Light travels like an arrow, and time like a shuttle. Three years passed so fast and I can still remember the first day I came to Italy, for me just like it was yesterday! Despite the hard time due to the COVID-19 pandemic, the past three years are the most precious time in the first half of my school life. I have learned the skills I wanted; I experienced interesting differences between cultures; And I met a group of great guys who influenced me a lot by their actions, their perception and inclusiveness during daily life, and their enthusiasm towards scientific work.

Sincerely thanks to my supervisor and co-supervisor, Fabrizio Ghiselli and Liliana Milani, who always give me their support and encouragement. They guided me to the fascinating DUI world and taught me how to be serious and rigorous towards scientific works. They helped me polish the manuscript once and once again with great patience. They made all these three years possible and their effort made me what i am today.

I also want to express my gratitude to Mariangella Iannello, who taught the first bash and R command. She gave me countless help whenever I encountered problems. She also influences me a lot by the way she thinks. Thanks also to Giovanni Piccinini, the one who brought me to lunch on the first school day in Italy and always showed his kindness and gentleness to me; to Alex Cussigh, who is always patient to explain me everything; to Jacopo Martelossi, who is sharing with me office, jokes, doubts and thinkings; to Giobbe Forni and Filippo Castellucci, who influenced me a lot by his enthusiasm towards academic career; and to the Professors and Researchers in Mozoo lab, who are always kind to me.

Without the support from my family, I would never be able to pursue the PhD degree. And great thanks also to my friends in China, and in America, who share my joy and sorrow without time differences. Finally, I want to acknowledge the China Scholarship Council for their financial support and for giving me the opportunity to see the world.