HOKKAIDO UNIVERSITY

Title	Elucidation of mitogenomic adaptation and structural characteristics of big defensin in molluscs using bioinformatics and computational biology
Author(s)	DHAR, Dipanjana
Degree Grantor	北海道大学
Degree Name	博士(理学)
Dissertation Number	甲第14790号
Issue Date	2022-03-24
DOI	https://doi.org/10.14943/doctoral.k14790
Doc URL	https://hdl.handle.net/2115/88179
Type	doctoral thesis
File Information	Dipanjana_Dhar.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Doctoral Dissertation

Elucidation of mitogenomic adaptation and structural characteristics of big defensin in molluscs using bioinformatics and computational biology

(バイオインフォマティクスと計算生物学を用いた軟体動物におけるミトゲノムの適応とビッグディフェンシンの構造特性の解明)

Dipanjana Dhar

Graduate School of Science, Hokkaido University

Department of Natural History Sciences

2022 March

INDEX

Acknowledgements 4

List of Tables 5

List of Figures 6

Summary 9

General Introduction 12

Chapter 1: Insight into the adaptive evolution of mitochondrial genomes in intertidal chitons

1.1 Introduction 15

1.2 Material and Methods

1.2.1 Sequence retrieval 17

1.2.2 Sequence alignment and phylogenetic analysis 17

1.2.3 Selection pressure analyses 20

1.2.4 Protein modelling and structural analysis 20

1.3 Results and Discussion

1.3.1 Phylogenetic analyses and diversity 21

1.3.2 Selection studies 26

1.3.3 Alterations of amino acid properties 31

1.3.4 Tertiary structure prediction of mitochondrial proteins 31

1.3.5 Mapping of positively selected sites in the predicted protein structures 32

Chapter 2: Comparison of evolutionary selection acting on the mitochondrial protein-coding genes between intertidal and deep sea gastropods

2.1 Introduction 37

2.2 Material and Methods

2.2.1 Sequence retrieval 38

2.2.2 Sequence alignment and model selection 38

2.2.3 Selection pressure analyses 40

2.3 Results and Discussion

2.3.1 Selection studies 41

Chapter 3: *In silico* analysis of two conserved residues in the structure of big defensin from *Crassostrea gigas*

3.1 Introduction 45

3.2 Material and Methods

3.2.1 Dataset collection 48

3.2.2 Multiple sequence alignment and identification of conserved residues 48

3.2.3 Mutagenesis and structure validation 48

3.2.4 Identification of amino acid network and functional residues in protein structure 50

3.2.5 Prediction of deleterious mutations 50

3.2.6 Prediction of destabilizing mutations 50

3.2.7 Aggregation propensity analysis 50

3.2.8 Molecular dynamics simulation 50

3.3 Results

3.3.1 Sequence comparison 51

3.3.2 Mutagenesis and structure validation 53

3.3.3 Identification of amino acid network and functional residues in protein structure 54

3.3.4 Prediction of deleterious and destabilizing mutations 55

3.3.5 Aggregation propensity analysis 56

3.3.6 Mutation induced conformational stability, flexibility and dynamic changes 56

3.4 Discussion 59

References 61

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Helena Fortunato and co-supervisor, Dr. Takeo Horiguchi for their continuous support and guidance throughout my PhD study. Besides, I would like to thank Dr. Horiguchi and Dr. Ryuichi Masuda for reviewing my thesis with their insightful comments and suggestions. My sincere thanks to Dr. Hiroshi Kajihara, Dr. Toru Katoh and Dr. Keiichi Kakui, from the Department of Natural History Sciences (Biodiversity Research Group I), for their kindness and support. Also I want to thank Dr. Ryouichi Tanaka and Dr. Hisashi Ito for giving me a great opportunity to work in their project. I actually want to thank everyone from Tanaka Lab of Hokkaido University for extending their help and support to me. My heartfelt thanks to Dr. Tanaka not only for his suggestions in academic matters but also for helping me to settle in this city.

I am thankful to the members of Graduate School Educational Affairs Section Science Administration Department and Office for International Academic Support Faculty of Science of Hokkaido University for helping me throughout the application process and journey of my PhD.

Furthermore, I am very grateful to Rotary Yoneyama Memorial Foundation, for granting me the Rotary Yoneyama Scholarship for two years of my PhD study (April 2019 – March 2021) and a single one way air fare to Japan, to sustain my living in Sapporo.

Most importantly, I would like to express my heartiest gratitude to my grandparents (Mrs. Madhuri Debnath & Mr. Krishnabandhu Debnath), parents (Mrs. Mili Dhar & Mr. Bachchu Dhar and Mrs. Sreeparna Dey & Mr. Narendranath Dey) and elder sister (Mrs. Pallabita Das) for their immeasurable and invaluable contributions to my life. My special thanks to Mrs. Dey for creating a mini India in my small Japanese apartment, by procuring all my necessities. I also want to thank my former lab mate and an elder sister, Dr. Sucharita Das from University of Calcutta, for sudden helps in my research with her excellent computational skills.

Apart from all, my words and emotions fall short for a person who is my friend, philosopher and guide rather my soulmate or may be the ‘scriptwriter of my life’. He is none other than Mr. Debayan Dey. He helps and guides me always to be strong and a true human being. Thank you Debayan for always being there!

Finally, I want to dedicate this popular quote to a person: “He's a silent guardian, a watchful protector, a Dark Knight.” from the famous movie ‘The Dark Knight’. In my life, that ‘Dark Knight’ is ‘the woman’, Dr. Soumalee Basu. I actually fall short of words to explain her contribution to my life and career. To sum it up, this PhD would have been impossible without her support and guidance.

List of Tables

Table 1: Details of the chiton species considered. 18

Table 2: Ecology of the chiton species considered. 18

Table 3: The partitioning scheme and best-fitting models selected for the dataset of 13 PCGs and 2 rRNAs, using PartitionFinder 2. 19

Table 4: Estimation of transition/transversion bias (*R*) under appropriate nucleotide substitution models for the 13 PCGs and the large variable region, performed in MEGA X. 25

Table 5: Log-likelihood values of the site-specific models (CODEML) for *atp8*. Sites under positive selection are marked with * at $P < 0.05$. 26

Table 6: Genes and sites under positive selection, detected by MEME in the HyPhy package. 27

Table 7: Parameter estimates and likelihood values for PCGs inferred using the branch-site model. Sites under positive selection are marked with * at $P < 0.05$ & ** at $P < 0.01$. 29

Table 8: Details of protein structure prediction by I-TASSER. 32

Table 9: Details of protein structure prediction by homology modelling. 32

Table 10: Details of the gastropod species considered. 38

Table 11: Ecology of the gastropod species considered. 39

Table 12: The best-fitting models selected for the 13 PCGs using MEGA X. 40

Table 13: Sites of different mitochondrial PCGs under positive selection, as detected by MEME & FEL in the HyPhy package. 42

Table 14: Details of the invertebrate and vertebrate species containing the big defensin and β-defensin protein, respectively. 49

Table 15: Prediction of deleterious mutation in the big defensin protein using PROVEAN web server. 56

Table 16: Prediction of destabilizing mutation in the big defensin using different computational programs. 56

Table 17: Prediction of aggregation propensities of wild-type and mutant forms of big defensin using the SODA web server. 56

List of Figures

Figure 1: Bayesian consensus tree based on the concatenated mitochondrial PCGs from seven chitons. Branch support values (posterior probabilities) are shown. *Stomatella planulata* was used as outgroup. Families of the polyplacophorans corresponding to each clade are indicated on the right side of the tree. Scale bar indicates substitutions per site. 22

Figure 2: Bayesian tree based on the concatenated sequences of two rRNAs from seven chiton species. *Stomatella planulata* was used as outgroup. Posterior probabilities for each branch are shown. Scale bar indicates substitutions per site. 23

Figure 3: Multiple sequence alignment of the large variable region from seven chiton species, as generated by MUSCLE. The nucleotide bases are indicated as follows: A (adenine), red; T (thymine), blue; C (cytosine), green; and G (guanine), yellow. Abbreviations used: *Ccaverna*, *Cyanoplax caverna*; *Ncalifornica*, *Nuttallina californica*; *Cstelleri*, *Cryptochiton stelleri*; *Capiculata*, *Chaetopleura apiculata*; *Ktunicata*, *Katharina tunicata*; *Spelliserpentis*, *Sypharochiton pelliserpentis*; and *Ssinclairi*, *Sypharochiton sinclairi*. 24

Figure 4: Transition/transversion bias for the 13 PCGs and the large variable region of the chiton mitogenomes. 25

Figure 5: Percentage of sites under purifying selection in PCGs encoding mitochondrial subunits of chitons, as detected by the FEL test. The subunits are colored according to their corresponding complexes: blue, complex I; red, complex III; green, complex IV; orange, complex V. 27

Figure 6: Schematic representation of the topology for analysing selection pressures on individual foreground branches, as based on the branch–site model in CODEML. The different branches of the phylogeny are labelled from A to K. Different colored shapes represent the 13 mitochondrial PCGs; the number within each shape indicates the number of positively selected sites. 28

Figure 7: Computationally predicted atomic-level structure of mitochondrial complex I (ND1–ND6 and ND4L) of *Chaetopleura apiculata*. Positively selected sites inferred by CODEML and MEME are shown in red; residues with amino acid substitutions, as revealed by TreeSAAP, are in pink; and sites identified by all these methods are displayed in orange. 33

Figure 8: Computationally predicted atomic-level structure of mitochondrial CYTB subunit of *Chaetopleura apiculata*. Positively selected sites inferred by CODEML and MEME are shown in red and residues with amino acid substitutions, as revealed by TreeSAAP, are in pink. 34

Figure 9: Computationally predicted atomic-level structure of mitochondrial complex IV subunit (COX1–COX3) of *Chaetopleura apiculata*. Positively selected sites inferred by CODEML and MEME are shown in red and residues with amino acid substitutions, as revealed by TreeSAAP, are in pink. 34

Figure 10: Computationally predicted atomic-level structure of mitochondrial subunits ATP6 (A) and ATP8 (B) of *Chaetopleura apiculata*. Positively selected sites inferred by CODEML and MEME are shown in red; residues with amino acid substitutions, as revealed by TreeSAAP, are in pink; and sites identified by all these methods are displayed in orange. 35

Figure 11: Distribution of sites, under positive selection and/or showing radical amino acid changes, between the transmembrane helices and the loop areas in individual mitochondrial subunits. 35

Figure 12: Percentage of sites under purifying selection in PCGs encoding mitochondrial subunits of gastropods, detected by FEL test. The subunits are colored according to their corresponding complexes: complex I (blue), complex III (red), complex IV (green) and complex V (orange). 42

Figure 13: Distribution of positively selected sites, obtained from the branch-site models of CodeML, in 13 mitochondrial PCGs of intertidal and deep sea gastropods. The blue and orange bars represent the number of positively selected sites in intertidal and deep sea gastropods, respectively. 43

Figure 14: (A) Structural domain organization of big defensins. Hydrophobic (green frame) and β-defensin like (orange frame) domains are indicated at the amino acid sequence alignment of certain mature big defensins. Cysteine pairing is indicated by gray lines. *Tt*-BigDef = *Tachypleus tridentatus* Big Defensin; *Bj*-BigDef = *Branchiostoma belcheri tsingtauense* Big Defensin; *Hc*-BigDef = *Hyriopsis cumingii* Big Defensin; *Cg*-BigDef1 = *Crassostrea gigas* Big Defensin1; *Mg*-BigDef1 = *Mytilus galloprovincialis* Big Defensin1; *Ap*-BigDef1 = *Argopecten purpuratus* Big Defensin1. (B) The 3D structure of canonical big 47

defensins showing two structural domains connected by a flexible linker. Source: Gerdol et al., 2020.

Figure 15: Multiple sequence alignment of 24 mature peptides sequences of big defensin from different mollusc species performed using MUSCLE integrated in MEGA X software. The residues with conservation score above 80% have been colored. The two conserved residues to be mutated *in silico*, have been marked with stars. 52

Figure 16: Multiple sequence alignment of the amino acid sequences pertaining to the C-terminal β-defensin-like domain of big defensin from invertebrate mollusc and β-defensin from vertebrate species, performed using MUSCLE integrated in MEGA X software. The residues with conservation score above 80% have been colored. The two conserved residues to be mutated *in silico*, have been marked with stars. 53

Figure 17: Ramachandran plot showing the dihedral angle values for (A) wild-type big defensin, (B) R64A mutant and (C) E71A mutant, performed using the PROCHECK server. 54

Figure 18: Evolutionary conservation of amino acid residues in the three-dimensional structure of big defensin by ConSurf analysis (A) and in the primary sequence of the same protein by ConSeq analysis (B). 'e' refers to an exposed residue according to the neural-network algorithm; 'b' refers to a buried residue according to the neural-network algorithm; 'f' refers to a predicted functional residue (highly conserved and exposed); 's' refers to a predicted structural residue (highly conserved and buried). The two conserved residues that have been mutated are marked with stars. 55

Figure 19: RMSD plot of the backbone atoms of wild-type and mutant big defensin proteins over the 100 ns trajectory of the MD production run. 57

Figure 20: Radius of gyration of wild-type and mutant big defensin proteins over the 100 ns trajectory of the MD production run. 58

Figure 21: RMSF plots of each residue of wild-type and mutant big defensin proteins over the 100 ns trajectory of the MD production run. 59

Summary

The phylum Mollusca constitutes an ubiquitous, heterogeneous, ecologically and economically important group of invertebrates that function as ecosystem engineers. Although the most conspicuous symptom of any ecosystem deterioration is the decline or disappearance of sensitive species, some organisms such as molluscs display an unusual resilience towards environmental changes. Marine molluscs that survive in the challenging environments of different oceanic zones, are ideal systems for studying stress adaptation. The understanding of the adaptive evolution of mitochondrial genomes in molluscs and the structural characterization of mollusc defensin protein which forms the essential component of their innate immunity is a major baseline for my PhD studies in Hokkaido University.

Mitochondria are known to be critical for energy homeostasis and changes in environmental factors result in their dysfunction and consequent injury to the organism. Mitochondrial proteins and mitochondria-derived stress signals regulate both oxidative phosphorylation and innate immune response. Maintenance of mitochondrial integrity and signaling are important for cellular homeostasis and survival. Evolutionary changes in the constituent residues of the mitochondrial proteins might have an impact on their functional domains, such as the regions lining the proton translocation channel or the subunit interacting sites, thereby allowing animals to adapt to challenging environments. Therefore, the aim of this study is to estimate selection pressures acting on mitochondrial proteins that could provide insights into the adaptive evolution of the mitochondrial genome.

Chapter 1 provides insight into the molecular mechanisms underlying the adaptive strategies of polyplacophorans to the intertidal habitat from a mitochondrial perspective. The intertidal zone is one of the most stressful environments, with extreme shifts in temperature, salinity, pH and oxygen concentration. Chitons (Polyplacophora) are the most primitive marine molluscs. They predominantly inhabit the intertidal zone and are still able to maintain their mitochondrial homeostasis regardless of the regular oscillations of immersion and emersion, as well as extreme changes in temperature, salinity, pH and hydrodynamic forces in their environment. Here, I used mitochondrial genetic components from seven chitons of the intertidal zone to infer phylogenetic relationships. Selection analyses on individual protein-coding genes (PCGs) were performed to identify and map potentially adaptive residues in the modelled structures of the mitochondrial respiratory chain complexes. The results showed significant amino acid changes in sites under diversifying selection for all the PCGs, indicating that the mitochondrial genome in chitons is undergoing adaptive evolution. Such sites were observed in the proton pump as well as in the translocation channel of the transmembrane helices and the surrounding loop regions, thus implying functional modification of the mitochondrial proteins essential for survival in the dynamic environment of the intertidal zone. This study represents the first thorough investigation of evolutionary selection acting on the mitochondrial PCGs of polyplacophorans.

Chapter 2 sheds light into the mitogenomic adaptations of intertidal and deep sea gastropods. Of all classes of the phylum Mollusca, gastropods are the only ones to radiate into marine, freshwater and terrestrial habitats, successfully adapting and thriving in a diverse array of environmental conditions. In order to withstand the constant fluctuations in temperature, salinity and shifts in oxygen concentration of the intertidal zone, the gastropods inhabiting here rely on a modified and adaptive energy metabolism. The same is applicable for gastropods living in the deep sea environment, which is characterized by high hydrostatic pressure, low oxygen concentrations and abundance of heavy metals. Therefore, survival of the organisms in these extreme conditions may be correlated to their adaptive mitochondrial genome which serves as the principal site for cellular energy metabolism. Here, I estimated selection pressure acting on the mitochondrial protein-coding genes of thirteen intertidal and two deep sea gastropods based on site and branch-site specific models. The results showed a higher number of sites under diversifying selection in the mitochondrial protein-coding genes of intertidal gastropods compared to deep sea species. This study focusses on the adaptive mitogenome evolution of marine gastropods for survival in dynamic environments such as the intertidal zone and the deep sea. The novelty of this research work is that it provides the first account of the comparison of mitogenomic adaptations between intertidal and deep sea gastropods.

Chapter 3 deals with the structural characterization of big defensin of Pacific oyster (*Crassostrea gigas*) through various bioinformatic tools and molecular dynamics simulation. Defensins are antimicrobial peptides (AMPs) consisting of three or four intramolecular disulphide bonds formed by six or eight cysteine residues, respectively, in a complex array of two or three antiparallel β-sheets with or without an α-helix structure. They are produced by a vast range of organisms being constitutively expressed or induced in various tissues in response to different stimuli like infection, injury or other inflammatory factors. Two classes of invertebrate defensin exist, namely CSαβ defensin and big defensin, the latter being predominantly present in molluscs. Interestingly, an invertebrate big defensin gene has been hypothesized as the most probable ancestor of vertebrate β-defensins. In this study, conserved residues were identified for both the big defensin and β-defensin. I performed *in silico* mutation on conserved amino acid positions of the β-defensin-like domain to understand the effects of mutation on the structure and function of the protein. Molecular dynamics simulations were performed on wild-type and two mutants for 100 ns to assess the structural stability and conformational dynamics of the protein structure of big defensin in its wild-type and mutated form. The aforementioned mutations (R64A and E71A) have been identified as deleterious as well as destabilizing the three-dimensional structure of big defensin, as revealed by bioinformatic analyses. Changes in amino acid network with interacting residues and aggregation propensity further support the structural basis of big defensin upon mutating these two conserved charged amino acids. 100 ns molecular dynamics simulations of wild-type, R64A and E71A structures revealed significant conformational changes in the case of

mutants. This study aims to unveil the detailed structural characteristics of a molluscan defensin through computational approaches. In conclusion, these results provide first in-depth understanding of destabilization and loss of conformational dynamics of big defensin from *Crassostrea gigas* mediated by R64A and E71A mutation. It will provide insightful knowledge of this antimicrobial peptide for application in therapeutics and other aspects of protein engineering. The results may be further exploited to understand the immune response of oysters against pathogenic invasion.

General Introduction

Mitochondria are the central organelles involved in energy and redox homeostasis, cellular signaling and survival. Animal mitochondria are sensitive to environmental stress and stress-induced changes in the mitochondrial integrity and function have major consequences for the survival of organisms (Monlun et al., 2017). Studies in model organisms such as terrestrial mammals and insects (rodents, *Drosophila* and mammalian cell lines) show that mitochondria are sensitive to environmental stress as well as a center for adaptive cellular response (Sokolova, 2018). Mitochondrial stress can impair the ATP supply of the cell leading to oxidative injury if the mitochondrial production of reactive oxygen species (ROS) exceeds the capacity of cellular antioxidants to curb the damage (Vakifahmetoglu-Norberg et al., 2017).

Exposure to environmental stressors such as heat or cold, osmotic shock, toxins, or lack of oxygen can shift the mitochondrial balance from adaptive response to cell death due to energy deficiency and cellular damage (Bohovych & Khalimonchuk, 2016). However, many eukaryotes (such as the inhabitants of intertidal zones) are capable of surviving and maintaining their mitochondrial homeostasis despite frequent and drastic fluctuations in environmental conditions. Intertidal zone with its alternating periods of immersion and emersion, extreme fluctuations of temperature and salinity is one of the most stressful environments for the organisms living here (Brian et al., 2002; Jensen & Denny, 2016). Temperature in the intertidal zone can change by 10 – 30^{o}C within a few hours reaching values in excess of 35 – 40^{o}C during summer low tides (Helmuth et al., 2016).

These changes pose major challenges to homeostatic mechanisms of intertidal organisms, so the mitochondrial function must be maintained in these organisms and restored after major physiological disruptions such as extreme temperatures or prolonged lack of oxygen (Sokolova, 2018). The mechanisms of the exceptional mitochondrial resilience to temperature, salinity and hypoxic stress are not yet fully understood in intertidal organisms. Therefore, it is expected that studies on mitochondrial genome evolution may hence uncover interesting aspects of adaptation of the organisms living in intertidal conditions. This is the primary reason that intrigued me to work on mitochondrial stress adaptation in intertidal organisms.

Molluscs, being extremely diverse and globally distributed, belong to the ecologically dominant organisms in the intertidal zone and have successfully adapted to harsh life between the tides, representing an excellent model system for environmental and climate change studies (Prendergast et al., 2013; Sokolova, 2018). The use of molluscs in climate change and ocean acidification research is already providing valuable information regarding changes in the global ocean system that will facilitate future predictions and conservation strategies (Fortunato, 2015). The phylum Mollusca constitutes an ubiquitous, heterogeneous and

economically important group of invertebrates that function as ecosystem engineers (Wanninger & Wollesen, 2018). Ecosystem engineers can directly or indirectly modulate the availability of resources to other species, by causing physical state changes in biotic or abiotic materials (Jones et al., 1994). They are also important in regulating the local microclimate and are thus beneficial in maintaining habitats for other species amidst climate change (Cavieres et al., 2002). Although the most conspicuous symptom of any ecosystem deterioration is the decline or disappearance of sensitive species, some organisms such as molluscs display an unusual resilience towards environmental changes. This may be due to adaptive strategies that allow them to cope with changes in their surroundings. My research on molluscs will impart knowledge regarding acclimatization of these organisms to the dynamic conditions of different oceanic zones. Therefore, selection analyses on individual mitochondrial PCGs from different molluscan classes were performed to identify and map potentially adaptive residues in the modelled structures of their mitochondrial respiratory chain complexes using various bioinformatic tools and modelling software.

Chitons are an ancient lineage of molluscs that are classified in the class Polyplacophora of phylum Mollusca (Okusu et al., 2003). Chitons are often important members of rocky intertidal communities and are able to maintain mitochondrial homeostasis regardless of regular oscillations of immersion and emersion, as well as extreme alternations of temperature, salinity, pH and hydrodynamic forces in their environment (Brian et al., 2002; Jensen & Denny, 2016). Therefore, my aim was to estimate the selection pressure acting on mitochondrial proteins that would provide insights into the adaptive evolution of the mitochondrial genome of chitons. Although research studies on cephalopods, terrestrial gastropods and bivalves have revealed significant evidence of positive selection in their OXPHOS subunits and have been linked to mitogenomic adaptation, the same with polyplacophorans are not done yet (Almeida et al., 2015; Romero et al., 2016; M. Yang et al., 2019). My study represents the first thorough investigation of evolutionary selection acting on the mitochondrial protein-coding genes of polyplacophorans.

Apart from the intertidal zone, deep sea environments are also very stressful which are characterized by extreme physicochemical conditions, such as high hydrostatic pressure, fluctuating temperatures, low levels of oxygen and high concentrations of heavy metals and other toxic chemicals (Van Dover et al., 2002). In the phylum Mollusca, gastropods represent the largest class with number of species ranging from 80,000 to 150,000 (Appeltans et al., 2012; Parkhaev, 2007). Apart from being the only terrestrial molluscs, gastropods live in every conceivable place on Earth with marine ones occupying habitats ranging from the deepest ocean basins to the intertidal zone, as well as freshwater and other inland water habitats. Since acclimation to extreme habitat conditions are often correlated to the evolution of mitochondrial protein-coding genes (Almeida et al., 2015; Li et al., 2018; Sun et al., 2018; Yang et al., 2019), detailed studies on gastropod mitogenome may uncover interesting aspects of molecular adaptations to living in intertidal and deep sea environment. My study

on mitochondrial genome of gastropods focusses on its adaptive evolution as well as provides the first account of the comparison of mitogenomic adaptations between intertidal and deep sea gastropods.

The interest of marine invertebrates as food resources provides a major interest to study molluscan immunity for better understanding of the host response to pathogens. Molluscs possess a natural immunity formed by anatomical and chemical protective barriers that prevent damage of the underlying tissues, body fluid losses and infections caused by pathogenic microorganisms and parasites (Gliński & Jarosz, 1997). Defensins represent an evolutionary ancient family of antimicrobial peptides (AMPs), which play an undeniably important role in host defense (Lv et al., 2020). Invertebrate defensins can primarily be categorized into two classes: (1) the invertebrate cysteine-stabilized α-helix/β-sheet motif defensin (CSαβ defensin), containing an α-helix linked to an antiparallel two-stranded/triple-stranded β-sheet by disulphide bridges and is found in molluscs, nematodes and arthropods (Carvalho & Gomes, 2009; Zhu, 2008); (2) the invertebrate big defensin having a disulphide array stabilized β-sheet structure (Kouno et al., 2008; Zhu & Gao, 2013) which exists mainly in molluscs (Saito et al., 1995) but also in arthropods (Teng et al., 2012). They are initiated by an interaction with the negatively charged membranes of pathogens, subsequently causing membrane disruption or altering metabolic processes and eventually leading to cell death (Zasloff, 2019). In recent times, antimicrobial peptides constitute a global research hotspot, but limitations in design and application are prevalent. Therefore, combination of bioinformatics and computer-aided molecular dynamics simulation can help develop prospective AMPs by studying their structure and molecular mechanism (Huan et al., 2020). My study on molluscan defensin aims to unveil its detailed structural characteristics through computational approaches. Moreover, the structural characteristics of the big defensin was carried out for *Crassostrea gigas* as this Pacific oyster makes a major economic contribution to the region due to their dominant role in aquaculture and also as ecosystem engineers (Harris, 2008; Ruesink et al., 2005).

My PhD study on molluscs will shed light into the adaptation of these organisms to the dynamic conditions of different oceanic zones from a mitochondrial perspective. Furthermore, my study on the big defensin of *Crassostrea gigas* will provide knowledge on structurally and functionally important residues of this protein for its application in protein engineering and therapeutics. This effort will impart the fundamental basis for preventing mass mortality of oysters in the aquaculture industry from pathogen invasion and environmental deterioration by redesigning natural antimicrobial peptides.

Chapter 1

Insight into the adaptive evolution of mitochondrial genomes in intertidal chitons

1.1 Introduction

Animal mitochondrial DNA (mtDNA) is generally about 15,000–20,000 bp in size (Boore, 1999; Kolesnikov & Gerasimov, 2012). However, much larger mitogenomes exist due to the presence of large amounts of duplicated genes or noncoding nucleotides (Smith & Snyder, 2007). MtDNA of bilaterian animals contains 13 protein-coding genes (PCGs), 2 ribosomal RNAs (rRNAs) of the mitochondrial ribosome and 22 transfer RNAs (tRNAs) essential for translating oxidative phosphorylation (OXPHOS) proteins. Furthermore, it is considered that noncoding regions in the mtDNA are probably the sites for controlling transcription and replication (Boore, 1999). The 13 mitochondrial PCGs form subunits of the respiratory chain complexes and these are cytochrome c oxidase subunits 1–3 (COX1–COX3), cytochrome b (CYTB), NADH dehydrogenase subunits 1–6 (ND1–ND6), NADH dehydrogenase subunit 4L (ND4L), ATPase F0 subunit 6 (ATP6) and ATPase subunit 8 (ATP8). These complexes are responsible for producing up to 95% of the energy of eukaryotic cells (Blier et al., 2001). Four (I, III, IV and V) of the five complexes (I–V) of the electron transport chain are composed of these subunits, *i.e.*, complex I (ND1–ND6 and ND4L), complex III (CYTB), complex IV (COX1–COX3) and complex V (ATP6 and ATP8) (Saraste, 1999).

Mitochondria are key organelles involved in essential cellular processes such as energy and redox homeostasis, cellular signaling and survival (Wallace, 2012). Therefore, various stress-induced changes in the mitochondrial integrity and function have major consequences on the performance and fitness of the organism (Sokolova, 2018). Mitochondrial stress can hinder the ATP supply of the cell leading to oxidative injury if the mitochondrial production of reactive oxygen species (ROS) exceeds the capacity of cellular antioxidants to curb the impairment (Vakifahmetoglu-Norberg et al., 2017). Mitochondrial proteins and mitochondria-derived stress signals regulate both oxidative phosphorylation and innate immune responses thus providing the molecular basis for the cross-talk between mitochondrial bioenergetics and stress defense (Lartigue & Faustin, 2013). Changes in the constituent residues might have an impact on the functional domains of mitochondrial proteins, such as the regions lining the proton translocation channel or subunit interacting sites, thereby allowing animals to adapt to challenging environments (Almeida et al., 2015; da Fonseca et al., 2008). Therefore, the estimation of selection pressures acting on mitochondrial proteins could provide insights into the adaptive evolution of the mitochondrial genome. Most of the studies on the adaptive evolution of mitochondrial genomes in vertebrates primarily involve mammals and fish (Foote et al., 2011; Garvin et al., 2011). In contrast, the evolution and molecular basis of adaptation of the mitochondrial

genome in invertebrates, particularly marine ones, remain poorly understood (Yuan et al., 2020).

Mollusca is the second largest phylum of invertebrate animals after Arthropoda where the number of valid species is currently estimated to be approximately 110,000, making 23% of all named marine organisms (Appeltans et al., 2012; Rosenberg, 2014). The phylum comprises the classes Bivalvia, Cephalopoda, Gastropoda, Monoplacophora, Polyplacophora, Scaphopoda, Caudofoveata and Solenogastres (Kocot et al., 2011; Vinther et al., 2012). Molluscs vary in size from giant squids and clams to small snails about a millimeter long. Despite their amazing diversity, all molluscs share some unique characteristics that define their body plan (Fedosov & Puillandre, 2012). The three most universal features defining modern molluscs are a mantle with a significant cavity used for breathing and excretion, the presence of a radula (except for bivalves) and the structure of the nervous system (Takeuchi, 2017). In addition, this phylum is a ubiquitous, heterogeneous and economically important group of invertebrates, with many taxa functioning as ecosystem engineers (Astorga, 2014; Gomes-dos-Santos et al., 2020; Sigwart et al., 2015). Although the most conspicuous symptom of ecosystem deterioration is the decline or disappearance of sensitive species, certain molluscs display an unusual resilience towards environmental changes (Parker et al., 2013). Molluscs can adapt to different environmental conditions (Sun et al., 2017), from cold or temperate to tropical, and live in both freshwater and terrestrial habitats (Sowa et al., 2019).

Chitons (Polyplacophora) are ancient marine molluscs (Sigwart & Sutton, 2007; Vendrasco et al., 2004) and predominantly inhabit the rocky intertidal zones. The intertidal zone is one of the most stressful environments, with extreme shifts in temperature, salinity, pH and oxygen concentration. Marine molluscs, particularly chitons belong to the ecologically dominant organisms in the intertidal zone and have successfully adapted to harsh life between the tides, representing an excellent model system to investigate mitochondrial adaptation to stress (Helmuth et al., 2006; Sokolova, 2018; Vinagre et al., 2019). Chitons are able to maintain mitochondrial homeostasis regardless of regular oscillations of immersion and emersion, as well as extreme alternations of temperature, salinity, pH and hydrodynamic forces in their environment (Brian et al., 2002; Jensen & Denny, 2016).

With *c*. 1,000 living species of chitons identified, there are relatively few species in this group in comparison to gastropods and bivalves (Irisarri et al., 2020; Puchalski et al., 2008). This is also reflected in the number of currently available sequenced and annotated mitochondrial genomes of chitons in GenBank. As of December 2019, complete mitochondrial genomes of only seven chiton species were available in GenBank (Boore & Brown, 1994; Guerra et al., 2018; Irisarri et al., 2014; Veale et al., 2016). Detailed studies of the polyplacophoran mitochondrial genome evolution may hence uncover interesting aspects of adaptations to living in intertidal conditions. I used these seven mitochondrial genomes, all of species of the order Chitonida, to detect signs of natural selection in the 13 PCGs. I performed site- and

branch–site-specific likelihood analyses to estimate the probability of positive selection on each site and branch of individual genes across the seven species. Subsequently, I assessed changes in amino acids based on their physicochemical properties and mapped positively selected sites in structure-based homology models to ascertain the effect of these changes in functionally important regions of the OXPHOS proteins. This work also reassessed the phylogenetic relationships based on the PCGs and rRNAs of chiton mtDNA (Dhar et al., 2021).

1.2 Material and Methods

1.2.1 Sequence retrieval

The nucleotide sequences of 13 mitochondrial PCGs, 2 rRNAs and a large variable region (Guerra et al., 2018) of seven polyplacophoran species were downloaded from the National Center for Biotechnology Information database (https://www.ncbi.nlm.nih.gov/nucleotide). *Chaetopleura apiculata*, *Sypharochiton pelliserpentis*, *Sypharochiton sinclairi*, *Nuttallina californica*, *Cyanoplax caverna*, *Cryptochiton stelleri* and *Katharina tunicata* were used for the study. The details and ecology of the species are provided in Table 1 and 2, respectively.

1.2.2 Sequence alignment and phylogenetic analysis

Three nucleotide sequence datasets were constructed with the mitogenomes obtained from the seven chiton species: (1) the 13 mitochondrial PCGs (*cox1–cox3*, *cytb*, *atp6*, *atp8*, *nd1–nd6* and *nd4l*), (2) the 2 rRNA genes and (3) the large variable region (Guerra et al., 2018). A vetigastropod, *Stomatella planulata* (GenBank acc. no. NC_031861.1), was chosen as outgroup. For each of the three datasets, multiple sequence alignments, including codon-based alignment for PCGs, were performed using MUSCLE v. 3.8.31 (Edgar, 2004) integrated in the MEGA X software (Kumar et al., 2018) selecting the invertebrate mitochondrial genetic code. Ambiguously aligned regions of PCGs and rRNAs were removed using the Gblocks server v. 0.91b (Talavera & Castresana, 2007) specifying ‘codon’ as sequence type and using “do not allow many contiguous nonconserved positions” under “options for a more stringent selection”. These trimmed alignments were concatenated in the PhyloSuite v. 1.2.2 package, using the ‘concatenate sequence’ tool (Zhang et al., 2020).

Table 1: Details of the chiton species considered.

Species	Family	Accession no.	Length of mitogenome (bp)	Reference
Chaetopleura apiculata	Chaetopleuridae	KY824658.1	15108	Guerra *et al.* (2018)
Sypharochiton pelliserpentis	Chitonidae	NC_024174.1	15048	Veale *et al.* (2016)
Sypharochiton sinclairi	Chitonidae	NC_024173.1	15028	Veale *et al.* (2016)
Nuttallina californica	Lepidochitonidae	NC_026849.1	15604	Irisarri *et al.* (2014)
Cyanoplax caverna	Lepidochitonidae	NC_026848.1	15141	Irisarri *et al.* (2014)
Cryptochiton stelleri	Mopaliidae	NC_026850.1	15082	Irisarri *et al.* (2014)
Katharina tunicata	Mopaliidae	NC_001636.1	15532	Boore & Brown (1994)

Table 2: Ecology of the chiton species considered.

Species	Habitat	Depth range	Geographical distribution
Chaetopleura apiculata	Intertidal (rocky shore)	0 – 37m	Western Atlantic: USA and Colombia
Sypharochiton pelliserpentis	Intertidal (top of rocks above mid tide mark)	NA	Southwest Pacific: New Zealand; Chatham Islands, coasts of Tasmania, Victoria and New South Wales in Australia
Sypharochiton sinclairi	Intertidal (under stones and in rock pools over the lower tidal and sub-tidal region)	NA	Southwest Pacific: New Zealand
Nuttallina californica	Intertidal (on rocks or in crevices, between mid to high tide mark)	NA	Northwest Atlantic and Eastern Pacific: USA and Mexico
Cyanoplax caverna	Intertidal (sea cave)	NA	California
Cryptochiton stelleri	Intertidal (lower intertidal and subtidal zones of rocky coastline)	NA	Northern Pacific Ocean from Central California to Alaska, across the Aleutian Islands to the Kamchatka Peninsula and south to Japan
Katharina tunicata	Intertidal (stony or rocky bottoms or on rocky shores with heavy wave action)	0 – 40m	Kamchatka, Russia through the Aleutian Islands to southern California; Alaska to Point Conception; North of Santa Barbara

Data partition schemes and best-fitting substitution models, provided in Table 3, were estimated using PartitionFinder 2 (Lanfear et al., 2012) for the PCGs and rRNAs. The PartitionFinder analyses were run using a 'greedy' search scheme and all models assessed under the Bayesian information criterion. Substitution saturation was assessed for each codon position in the dataset using the Xia et al. (2003) statistical test, executed in DAMBE7 (Xia, 2018). Subsequently, the transition/transversion bias (*R*) was calculated for the 13 PCGs and the large variable region using MEGA X.

For each dataset, Bayesian phylogenetic analyses were performed using MrBayes v. 3.2.6 in the PhyloSuite software package (Ronquist et al., 2012). The partitions suggested by PartitionFinder were imported into MrBayes and all model parameters were set as unlinked across partitions. Two independent Bayesian runs, each with four MCMC chains, were implemented for 5 million generations, with sampling every 100 generations. The first 25% of the trees were discarded as burn-in. A consensus tree and Bayesian posterior probabilities were estimated based on the remaining trees. All the phylogenies generated were visualized using Figtree v. 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) and iTOL v5 (Letunic & Bork, 2019).

Table 3: The partitioning scheme and best-fitting models selected for the dataset of 13 PCGs and 2 rRNAs, using PartitionFinder 2.

Subset	Subset Partitions	Best Models
1	*atp6*	GTR + G
2	*atp8*	HKY + I
3	*cox1*	GTR + I + G
4	*cox2*	HKY + G
5	*cox3*	GTR + I + G
6	*cytb*	GTR + I + G
7	*nd1*	GTR + I + G
8	*nd2*	GTR + I + G
9	*nd3*	HKY + G
10	*nd4*	GTR + I + G
11	*nd4l*	HKY + G
12	*nd5*	GTR + I + G
13	*nd6*	HKY + G
14	12S rRNA	GTR + I
15	16S rRNA	GTR + G

1.2.3 Selection pressure analyses

The estimation of ω, the ratio of nonsynonymous substitutions (d_N) to synonymous substitutions (d_S), indicates the nature of natural selection operating on PCGs. Under neutrality, the ratio should not differ significantly from 1 (ω = 1), whereas cases in which the ratio differs significantly from 1 may be interpreted as indicative of positive selection (ω > 1) or negative selection (ω < 1). The ω values were determined using the codon-based maximum likelihood (CODEML) algorithm implemented in EasyCodeML v. 1.21 (Gao et al., 2019; Yang, 2007). Both site and branch–site models were used to identify the variation in selective pressures on individual mitochondrial PCGs. Seven codon substitution models described as M0 (one-ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), M8 (beta and ω > 1) and M8a (beta and ω = 1) were investigated and four likelihood ratio tests (M0 *vs* M3, M1a *vs* M2a, M7 *vs* M8 and M8a *vs* M8) were performed to assess the likelihood of positive selection on each site for individual PCGs (Anisimova et al., 2001; Yang et al., 2000; Yang et al., 2005). On the other hand, branch–site models enable ω to differ among sites and across branches of the phylogenetic tree (Zhang et al., 2005). Thus, I tested for positive selection on each of the 11 branches of the phylogeny (ML tree without outgroup), considering a single foreground branch (branch of interest) at a time. Bayes Empirical Bayes (BEB) analysis was used to compute posterior probabilities in order to recognize sites under positive selection on the selected branches if the likelihood ratio test (LRT) *P*-values are significant (BEB > 95%). The requirement for running EasyCodeML is an alignment file and a phylogenetic tree. For each of the 13 PCGs, 13ML trees were generated with identical topologies using MEGA X.

The HyPhy v. 2.5 package (www.hyphy.org) was used to assess codons under selection pressure (Pond et al., 2020). Mixed effects model of evolution (MEME) and fixed effects likelihood (FEL) analyses were carried out to detect individual sites subjected to episodic diversifying selection and purifying selection, respectively (Pond & Frost, 2005; Murrell et al., 2012). For both analyses, *P*-values < 0.05 were considered to be significant. The impact of potential positive selection was detected using the software TreeSAAP v. 3.2, which detects changes in local physicochemical properties due to amino acid replacements (McClellan & Ellison, 2010; Woolley et al., 2003). A sliding window equal to 20 codons along with magnitude categories ≥ 6 was used to assess amino acid property changes; *P*-values < 0.001 (z-score > 3.09) were treated as being significant.

1.2.4 Protein modelling and structural analysis

In order to understand the influence of sites undergoing positive selection and/or physicochemical property changes due to amino acid replacements, I mapped the amino acids onto the 3D structure of mitochondrial proteins (products of PCGs). Owing to the lack of 3D structures of molluscan mitochondrial subunits in the Protein Data Bank (PDB; https://www.rcsb.org) (Berman et al., 2000), I modelled the products of 13 PCGs for the

chiton *Chaetopleura apiculata*. The gene order of *C. apiculata* proved to be unexpectedly primitive among Polyplacophora (Guerra et al., 2018), which prompted me to select this species for the purpose of protein modelling. The structures of the COX subunits and the CYTB protein were modelled using the SWISS-MODEL server (https://swissmodel.expasy.org), with 3ABM and 3H1I.1.C as templates, respectively (A. Waterhouse et al., 2018). The models for the remaining mitochondrial proteins, belonging to complexes I and V, were generated using the I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/), using default parameters (Yang & Zhang, 2015; Zhang, 2008). The structural accuracy of the models was improved by a two-step atomic-level potential energy minimization process, implemented in the ModRefiner server (https://zhanglab.ccmb.med.umich.edu/ModRefiner) (D. Xu & Zhang, 2011). The stereochemical qualities of the models were verified using the PROCHECK server (https://servicesn.mbi.ucla.edu/PROCHECK) (Laskowski et al., 1993). Complexes I and IV from the constituent proteins were generated by structurally aligning the components to the corresponding complex I and IV templates from the sheep *Ovis aries* (PDB ID: 5LNK) and cattle *Bos taurus* (PDB ID: 3ABM), respectively, using the software PyMOL v. 2 (DeLano, 2002). Additional information about the orientation of transmembrane domains of these proteins was obtained using the TMHMM server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM) (Krogh et al., 2001). PyMOL was used for graphic modification, visualization and depicting the final illustrations.

1.3 Results and Discussion

1.3.1 Phylogenetic analyses and diversity

The phylogeny derived from the PCGs of seven chitons displayed a topology similar to that of Guerra et al. (2018) (Figure 1), with families Mopaliidae and Lepidochitonidae as sister clades to Chaetopleuridae and Chitonidae. I chose to use a vetigastropod as the outgroup in comparison to Guerra et al. (2018), who used *Haliotis rubra* and *Solemya velum*. The phylogeny based on the concatenated rRNA sequences yielded the same topology (Figure 2). Next, I analyzed the highly polymorphic noncoding large variable region in the chiton mitochondrial genomes. This region is located between $tRNA_{Glu}$ and *cox3* in all species studied here, except for *Cyanoplax caverna* and *Nuttallina californica*, where the region is located between $tRNA_{Glu}$ and 12S rRNA (Guerra et al., 2018). The multiple sequence alignment (Figure 3) of the (AT-rich) noncoding region showed a high sequence variability, rendering it unsuitable to infer phylogenetic relationships.

The evolutionary models obtained for estimating the transition/transversion bias were: GTR + I + G for *nd1*, *nd2*, *nd4*, *nd5*, *cytb*, *cox1*, *cox3*; GTR + G for *atp6*; HKY + G for *nd3*, *nd4l*, *nd6*, *cox2*; HKY + I for *atp8*; GTR for the large variable region. Calculating transition/transversion bias is important for deciphering genome evolution since it allows

verifying the occurrence of nucleotide conversions (Yang & Yoder, 1999). Generally, nucleotide transition is favored over transversion (Stoltzfus & Norris, 2016). Here, all PCGs except *atp8* follow the same trend (Figure 4; Table 4). However, this transition bias is absent in the large variable region. This finding agrees with the observation that transversions in noncoding regions show larger effects on the alterations of the DNA backbone and regulatory element activities (changes in gene expression, transcription factor binding and gene regulation) (Guo et al., 2017).

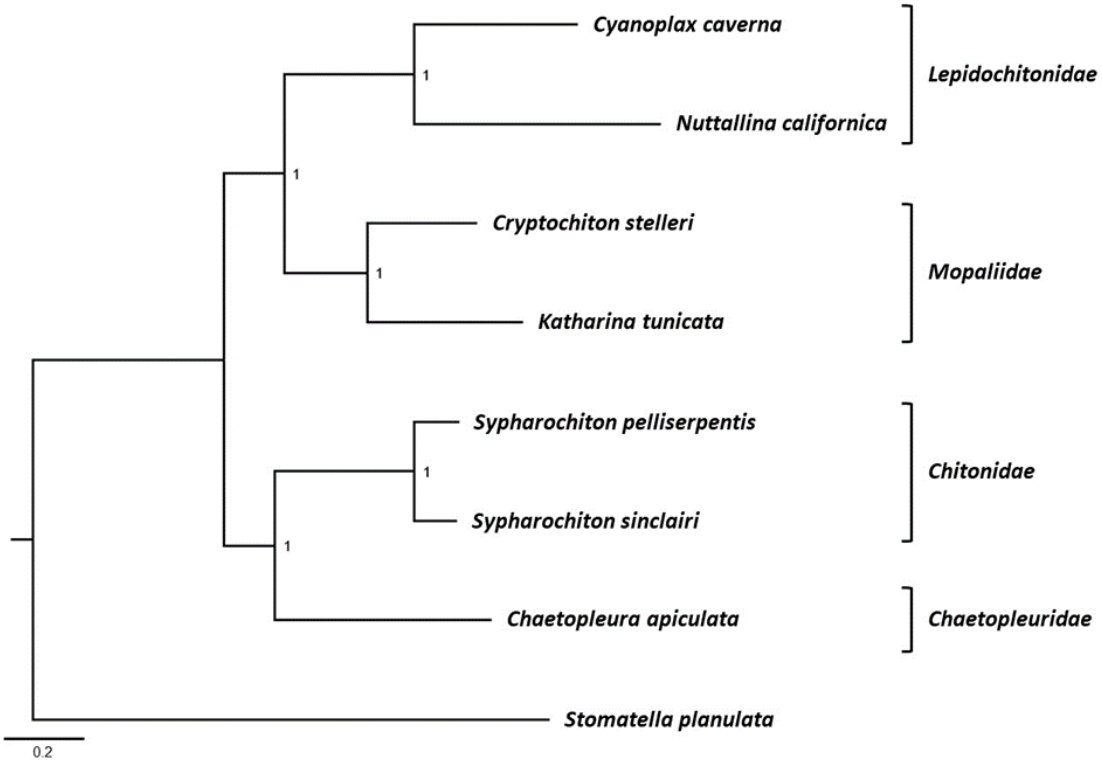


Figure 1: Bayesian consensus tree based on the concatenated mitochondrial PCGs from seven chitons. Branch support values (posterior probabilities) are shown. *Stomatella planulata* was used as outgroup. Families of the polyplacophorans corresponding to each clade are indicated on the right side of the tree. Scale bar indicates substitutions per site.

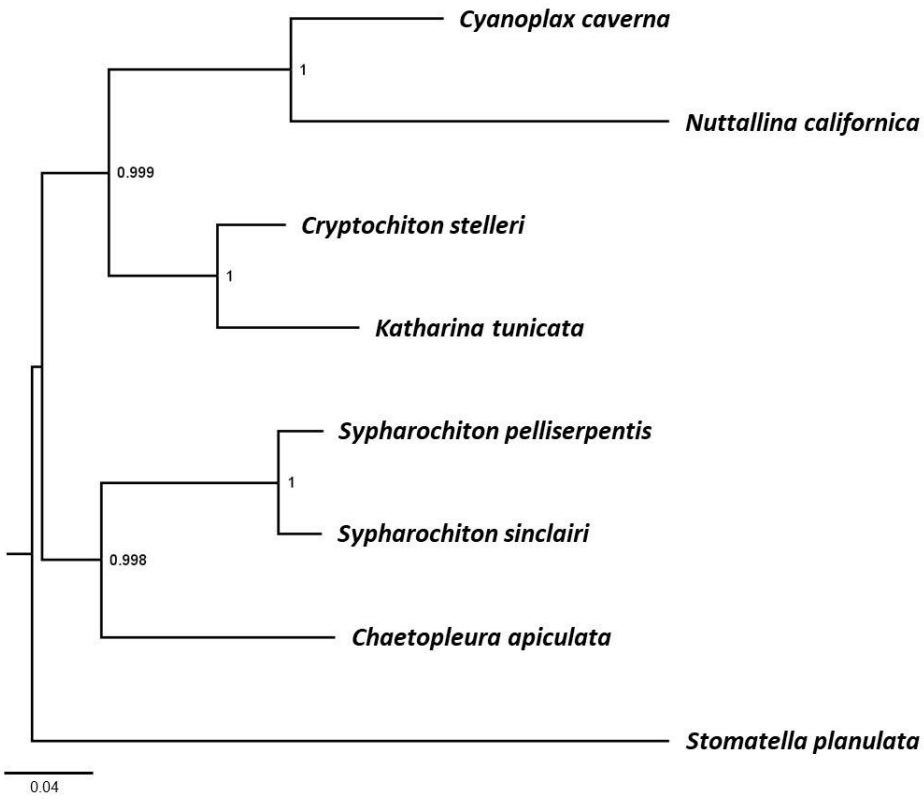


Figure 2: Bayesian tree based on the concatenated sequences of two rRNAs from seven chiton species. *Stomatella planulata* was used as outgroup. Posterior probabilities for each branch are shown. Scale bar indicates substitutions per site.

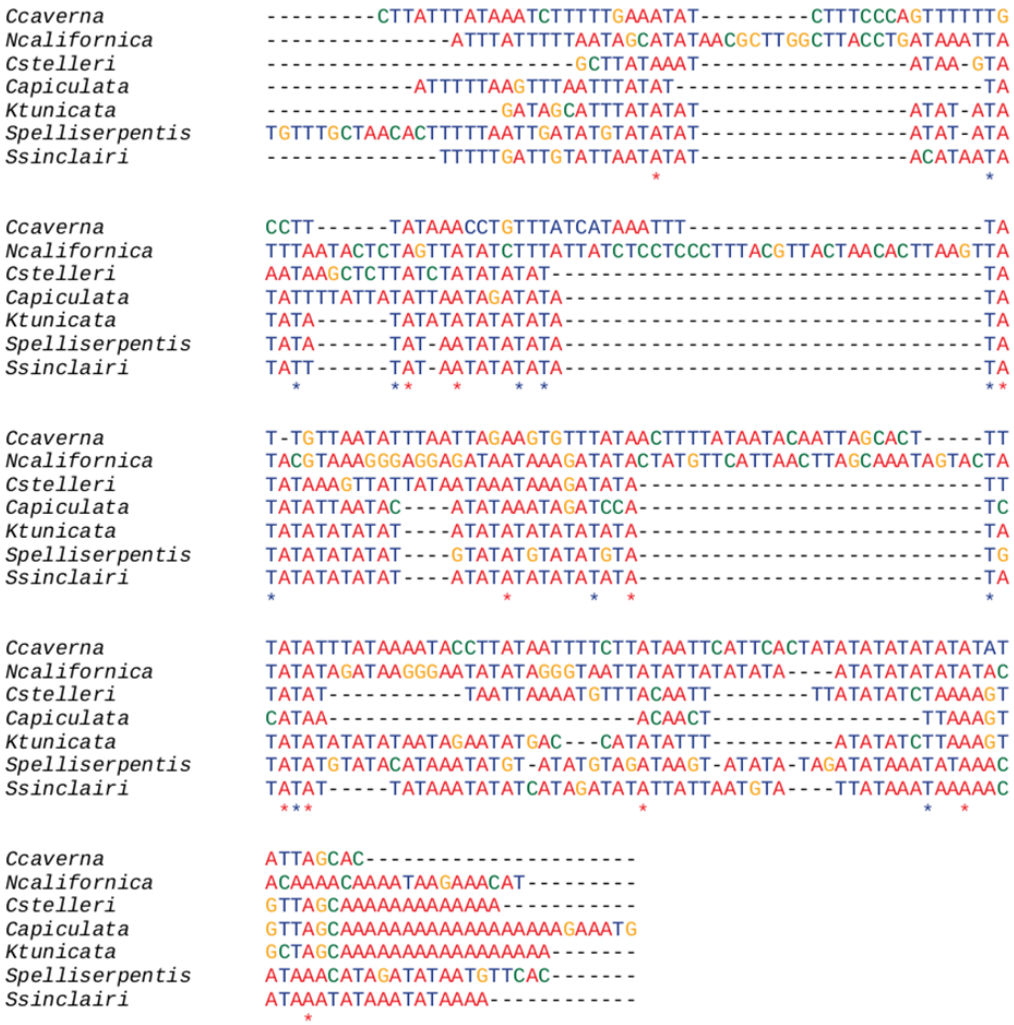


Figure 3: Multiple sequence alignment of the large variable region from seven chiton species, as generated by MUSCLE. The nucleotide bases are indicated as follows: A (adenine), red; T (thymine), blue; C (cytosine), green; and G (guanine), yellow. Abbreviations used: *Ccaverna*, *Cyanoplax caverna*; *Ncalifornica*, *Nuttallina californica*; *Cstelleri*, *Cryptochiton stelleri*; *Capiculata*, *Chaetopleura apiculata*; *Ktunicata*, *Katharina tunicata*; *Spelliserpentis*, *Sypharochiton pelliserpentis*; and *Ssinclairi*, *Sypharochiton sinclairi*.

Table 4: Estimation of transition/transversion bias (*R*) under appropriate nucleotide substitution models for the 13 PCGs and the large variable region, performed in MEGA X.

Gene	Model	Transition/ Transversion bias (*R*)
atp6	GTR + G	1.76
atp8	HKY + I	0.88
cox1	GTR + I + G	2.31
cox2	HKY + G	1.97
cox3	GTR + I + G	2.2
cytb	GTR + I + G	2.05
nd1	GTR + I + G	1.91
nd2	GTR + I + G	1.74
nd3	HKY + G	2.31
nd4	GTR + I + G	1.44
nd4l	HKY + G	2.16
nd5	GTR + I + G	1.52
nd6	HKY + G	1.61
large variable region	GTR	0.64

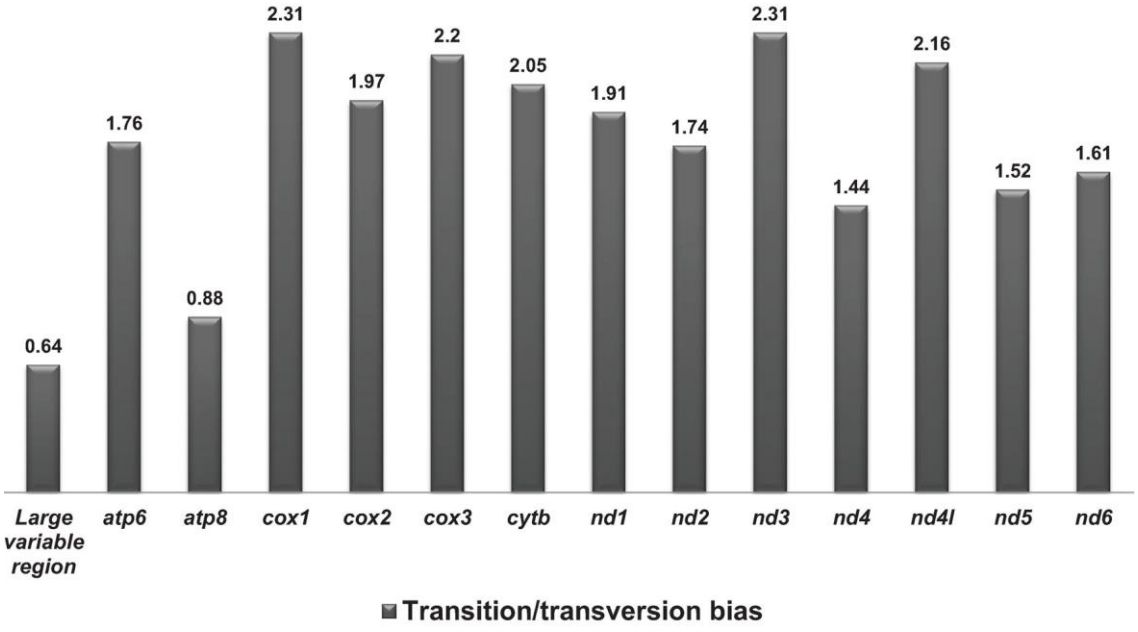


Figure 4: Transition/transversion bias for the 13 PCGs and the large variable region of the chiton mitogenomes.

1.3.2 Selection studies

Purifying selection is the prevalent force in the evolution of mitogenomes (Ballard & Whitlock, 2004). Since mitochondria serve as the major site for energy metabolism and production in the cell, weak and/or episodic positive selection may act on its PCGs allowing adaptation to environmental changes (Shen et al., 2010; Tomasco & Lessa, 2011). Intertidal marine molluscs are ecologically important organisms in the intertidal zone and have successfully adapted to this dynamic environment.

The results from the site-specific models of CODEML show that purifying selection dominated the evolution of all mitogenes except for *atp8* (Table 5). In other words, the likelihood ratio test of M7 *vs* M8 provided evidence of positive selection for *atp8* at codon position 44 (LRT: *P*-value = 0.000204865; BEB value = 0.967). This observation may be related to the small size of the ATP8 subunit (55 amino acids). The comparison of the LRTs based on site models (M0 *vs* M3, M1a *vs* M2a, M7 *vs* M8 and M8a *vs* M8) showed that the evolutionary rates of the remaining PCGs are under negative constraints. Furthermore, FEL analyses indicated that purifying selection was prevalent on all mitochondrial PCGs, with few sites evolving under neutrality ($P < 0.05$). According to FEL results, the PCGs (*cox1–cox3*) encoding complex IV subunits had the highest percentage of codons under negative selection (Figure 5). I also found that *cytb* or PCG encoding complex III displayed high levels of purifying selection, whereas genes encoding complexes I and V showed comparatively more relaxed selection, except for *nd4l*, which showed negative selection. In contrast, MEME-based analyses ($P < 0.05$), which employed a mixed effects ML approach to identify individual sites subjected to episodic diversifying selection, identified nine positively selected sites in six genes (*cox1*, *cytb*, *nd2*, *nd3*, *nd4* and *nd5*) (Table 6).

Table 5: Log-likelihood values of the site-specific models (CODEML) for *atp8*. Sites under positive selection are marked with * at $P < 0.05$.

Model	No. of parameters	lnL	Model compared	LRT *P*-value	Codon position	Amino acid	BEB values
M3	17	-837.143002	M0 *vs.* M3	0.000000000	-		
M0	13	-881.828094			Not allowed		
M2a	16	-853.115959	M1a *vs.* M2a	1.000000000	-		
M1a	14	-853.115959			Not allowed		
M8	16	-838.583856	M7 *vs.* M8	0.000204865	22	M	0.948
					44	C	0.967*
M7	14	-847.077014			Not allowed		
M8a	15	-842.279835	M8a *vs.* M8	0.006551613	Not allowed		

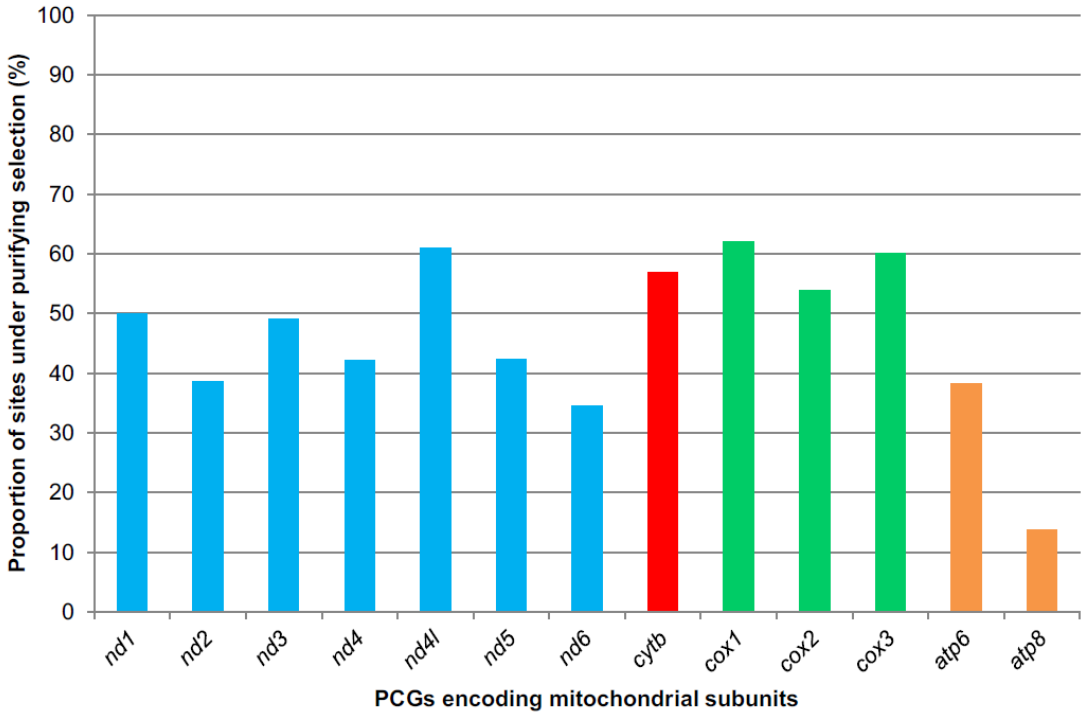


Figure 5: Percentage of sites under purifying selection in PCGs encoding mitochondrial subunits of chitons, as detected by the FEL test. The subunits are colored according to their corresponding complexes: blue, complex I; red, complex III; green, complex IV; orange, complex V.

Table 6: Genes and sites under positive selection, detected by MEME in the HyPhy package.

Gene	Sites under positive selection (*p*-value < 0.05)
cox1	507
cytb	14
nd2	74, 225
nd3	117
nd4	2, 300, 389
nd5	482

Since positive selection occurs at only a few positions in a particular lineage for most genes, detection of natural selection becomes possible only on the basis of site models. Therefore, the more stringent branch–site model was applied, which allows ω value to vary between branches and sites simultaneously, to measure divergent selective pressure on the specified foreground branch against the remaining background lineages. Identical topologies were obtained for all PCGs in 13 ML trees. A representative tree used as the working topology is shown in Figure 6. Using the BEB method, I detected 69 sites under potential positive selection in 12 PCGs with posterior probabilities ≥ 95% (Figure 6; Table 7). From the results, it is evident that the PCGs encoding complex I and V subunits possessed more positively selected sites per branch than PCGs encoding complex IV.

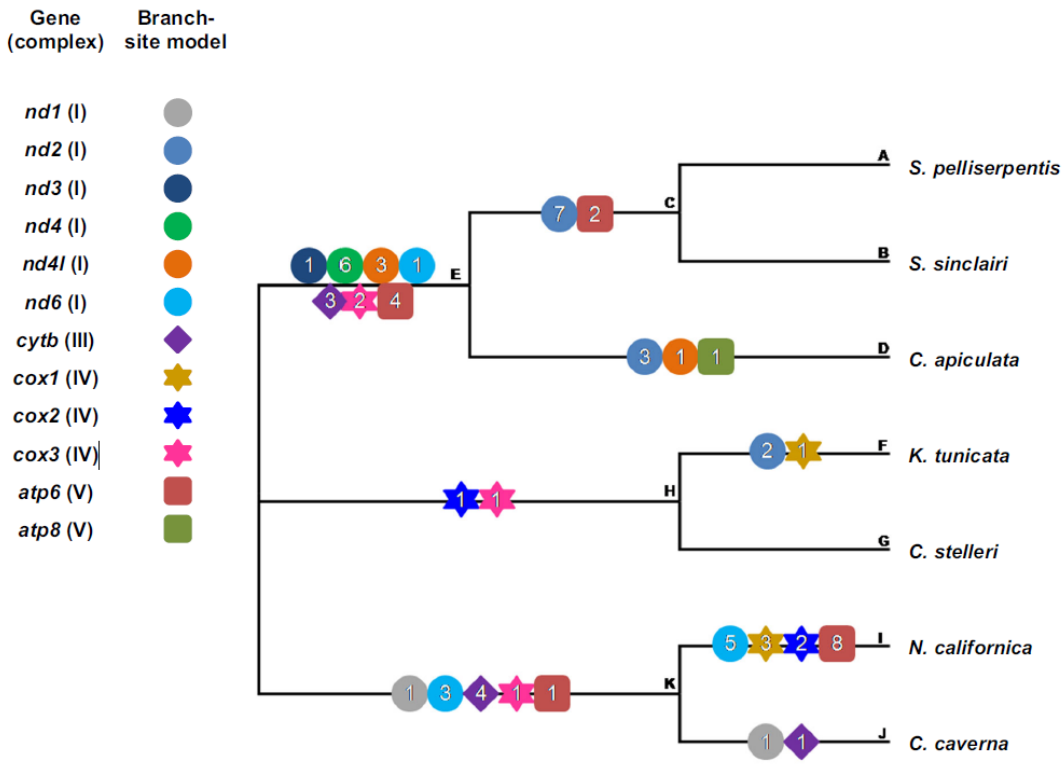


Figure 6: Schematic representation of the topology for analysing selection pressures on individual foreground branches, as based on the branch–site model in CODEML. The different branches of the phylogeny are labelled from A to K. Different colored shapes represent the 13 mitochondrial PCGs; the number within each shape indicates the number of positively selected sites.

Table 7: Parameter estimates and likelihood values for PCGs inferred using the branch-site model. Sites under positive selection are marked with * at $P < 0.05$ & ** at $P < 0.01$.

Gene	Branch	Models (no. of parameters)	InL	LRT *p*-value	Codon position	Amino acid	BEB values
atp6	C	Model A (16) Model A Null (15)	-3520.395410 -3520.806490	0.364549389	94 105	S L	0.998** 0.977*
	E	Model A (16) Model A Null (15)	-3515.87977 -3518.721861	0.017118518	8 33 50 124	Q Q S S	0.987* 0.968* 0.962* 0.965*
	I	Model A (16) Model A Null (15)	-3505.435324 -3508.751060	0.010019243	7 17 39 70 121 124 170 207	E P Y V S S G L	0.962* 0.980* 0.984* 0.953* 0.976* 0.955* 0.952* 0.978*
	K	Model A (16) Model A Null (15)	-3526.295771 -3526.295771	1.000000000	69	N	0.952*
atp8	D	Model A (16) Model A Null (15)	-850.602983 -850.602983	1.000000000	30	W	0.952*
cox1	F	Model A (16) Model A Null (15)	-5762.277719 -5764.268241	0.046015035	499	H	0.988*
	I	Model A (16) Model A Null (15)	-5751.066253 -5753.626005	0.023658378	327 404 491	R R D	0.971* 0.970* 0.976*
cox2	H	Model A (16) Model A Null (15)	-3103.062964 -3105.449412	0.028911154	49	K	0.998**
	I	Model A (16) Model A Null (15)	-3102.698834 -3102.736834	0.782793109	6 205	Q S	0.983* 0.982*
cox3	E	Model A (16) Model A Null (15)	-3415.190076 -3419.147264	0.004904349	106 217	T I	0.951* 0.997**
	H	Model A (16) Model A Null (15)	-3418.700076 -3419.807070	0.136764722	71	K	0.978*
	K	Model A (16) Model A Null (15)	-3415.056594 -3415.278312	0.505468690	97	A	0.968*
nd1	J	Model A (16) Model A Null (15)	-4411.035287 -4411.559978	0.305649461	287	S	0.985*
	K	Model A (16) Model A Null (15)	-4415.651735 -4415.651735	1.000000000	161	S	0.986*
nd2	C	Model A (16) Model A Null (15)	-5563.703249 -5563.782031	0.691409257	36 70 78 108	N Y L S	0.981* 0.974* 0.975* 0.954*

					114 135 316	C S W	0.953* 0.974* 0.952*
	D	Model A (16) Model A Null (15)	-5566.494831 -5566.494831	1.000000000	97 167 236	S S S	0.980* 0.979* 0.978*
	F	Model A (16) Model A Null (15)	-5566.472186 -5567.956701	0.084872732	121 215	W Q	0.954* 0.973*
nd3	E	Model A (16) Model A Null (15)	-1805.264409 -1808.912759	0.006908138	114	Y	0.980*
nd4	E	Model A (16) Model A Null (15)	-6950.584020 -6954.066838	0.008308989	46 60 267 294 322 373	S S K G S F	0.960* 0.983* 0.969* 0.984* 0.989* 0.994**
nd4l	D	Model A (16) Model A Null (15)	-1361.212224 -1361.264972	0.745331326	17	C	0.984*
	E	Model A (16) Model A Null (15)	-1358.749310 -1359.768974	0.153277350	27 38 58	S G S	0.998** 0.957* 0.988*
nd6	E	Model A (16) Model A Null (15)	-2936.774563 -2936.911784	0.600367486	78	S	0.967*
	I	Model A (16) Model A Null (15)	-2935.117597 -2935.117597	1.000000000	44 107 110 151 154	E K M K Y	0.951* 0.993** 0.982* 0.973* 0.980*
	K	Model A (16) Model A Null (15)	-2933.082458 -2934.482546	0.094253961	128 129 162	S L P	0.954* 0.994** 0.976*
cytb	E	Model A (16) Model A Null (15)	-5126.708470 -5130.743683	0.004499354	18 361 365	S F N	0.995** 0.974* 0.976*
	J	Model A (16) Model A Null (15)	-5138.463278 -5139.330524	0.187838243	367	Y	0.970*
	K	Model A (16) Model A Null (15)	-5122.220963 -5127.377406	0.001321045	8 76 301 342	I T H F	0.955* 0.965* 0.999** 0.976*

1.3.3 Alterations of amino acid properties

Values of $\omega > 1$ indicate that positive selection is possibly the driving force behind adaptive evolution of protein-coding DNA sequences. Even if several sites have undergone positive selection for the PCGs considered here, it is almost always the case that synonymous substitutions are favored in functionally conserved genes. For the nonsynonymous substitutions, changes in amino acid properties were identified using TreeSAAP, which compares random distributions of amino acid changes based on variable physicochemical properties with an expected distribution of changes deviating from neutral conditions.

Several codon sites (for $P < 0.001$ and categories 6–8) exhibiting deviations in amino acid properties were identified within 13 mitochondrial PCGs from seven chitons. Most of the PCGs, except for *atp8* and *nd6*, exhibited changes in the ionization properties for the majority of the amino acid substitutions keeping the hydrophobic or hydrophilic nature of the amino acids undisturbed. I observed changes in several measures for specific genes: changes in solvent accessible reduction ratio (R_a) for *atp6*, *cox1*, *nd1*, *nd2*, *nd4* and *nd5*; in buriedness (B_r) for *atp8*, *nd4* and *nd5*; in power to be at the C-terminal of the α-helix (α_c) for *cox2*, *nd4* and *nd6*; and in α-helical tendencies (P_α) for *cox3* and *nd6*. Furthermore, *nd3* and *nd4* showed changes in their power to be at the middle of α-helix (α_m) and surrounding hydrophobicity (H_p), respectively. It is noteworthy that *nd4l* and *cytb* exhibited changes only in the ionization properties of the amino acids. In order to check whether these changes affect the structural and functional organization of the mitochondrial proteins, I mapped the residues in the corresponding tertiary structures.

1.3.4 Tertiary structure prediction of mitochondrial proteins

The three-dimensional structures of polyplacophoran mitochondrial proteins were modelled in the absence of available structures in PDB. The tertiary structures of members from complexes I and V were generated using the protein threading approach implemented in I-TASSER, following lack of suitable templates (experimentally determined structures) for homology modelling. Subsequently, the most likely models for each protein were selected on the basis of highest C-score value, which is assessed according to the significance of threading template alignments and the convergence parameters of the structure assembly simulations (Table 8). The models of the remaining mitochondrial subunits, belonging to complexes III and IV, were built through homology modelling in the SWISS-MODEL server. The model details and evaluation parameters, provided by template-based modelling, are given in Table 9. Ramachandran plot analysis of the protein models, carried out in PROCHECK, revealed less than 1% of the residues as outliers.

Table 8: Details of protein structure prediction by I-TASSER.

Complex	Subunit	C-score	Estimated TM-score	Estimated RMSD (Å)
I	ND1	2	0.99 ± 0.04	2.2 ± 1.7
I	ND2	1.39	0.91 ± 0.06	3.5 ± 2.4
I	ND3	0.57	0.79 ± 0.09	3.2 ± 2.3
I	ND4	1.77	0.96 ± 0.05	3.4 ± 2.4
I	ND4L	0.89	0.83 ± 0.08	2.0 ± 1.6
I	ND5	2	0.99 ± 0.04	3.4 ± 2.3
I	ND6	-0.72	0.62 ± 0.14	6.5 ± 3.9
V	ATP6	0.29	0.75 ± 0.10	5.1 ± 3.3
V	ATP8	-2.30	0.44 ± 0.14	7.5 ± 4.3

Table 9: Details of protein structure prediction by homology modelling.

Complex	Subunit	Parent template	Species	Sequence identity	Sequence similarity	GMQE	RMSD with parent template (Å)
III	CYTB	3H1I.1.C	*Gallus gallus*	60.05	0.49	0.89	0.114
IV	COX1	3ABM.1.A	*Bos taurus*	76.34	0.54	0.97	0.080
IV	COX2	3ABM.1.B	*Bos taurus*	59.01	0.49	0.80	0.088
IV	COX3	3ABM.1.C	*Bos taurus*	67.97	0.52	0.88	0.080

Note: GMQE (Global Model Quality Estimation) is a quality estimation combining properties from the target-template alignment and the template search method.

1.3.5 Mapping of positively selected sites in the predicted protein structures

The putative positively selected sites, identified by CODEML and MEME, were mapped on the corresponding mitochondrial subunits, to determine their positions and significance in protein structure and function. Additionally, sites displaying changes in the physicochemical properties of amino acids, as inferred by TreeSAAP, were marked (Figures: 7–10). I used TMHMM to identify which of these sites are located in the transmembrane region or in the loop and found that a majority of the sites were within or in the vicinity of the domain. The subunits of complex I, except for ND6, show more instances of positive selection in the proton pump than the loop region (Figure 11). Complex IV is an exception with only one out of the three subunits having more positively selected sites in the transmembrane domain. Thus, complex III and ATP6 of complex V follow the trend of complex I. Furthermore, the

presence of positively selected residues mostly in the loop region of ND6 and ATP8 may provide a probable correlation with the restriction of amino acid changes in ionization properties, as obtained from TreeSAAP.

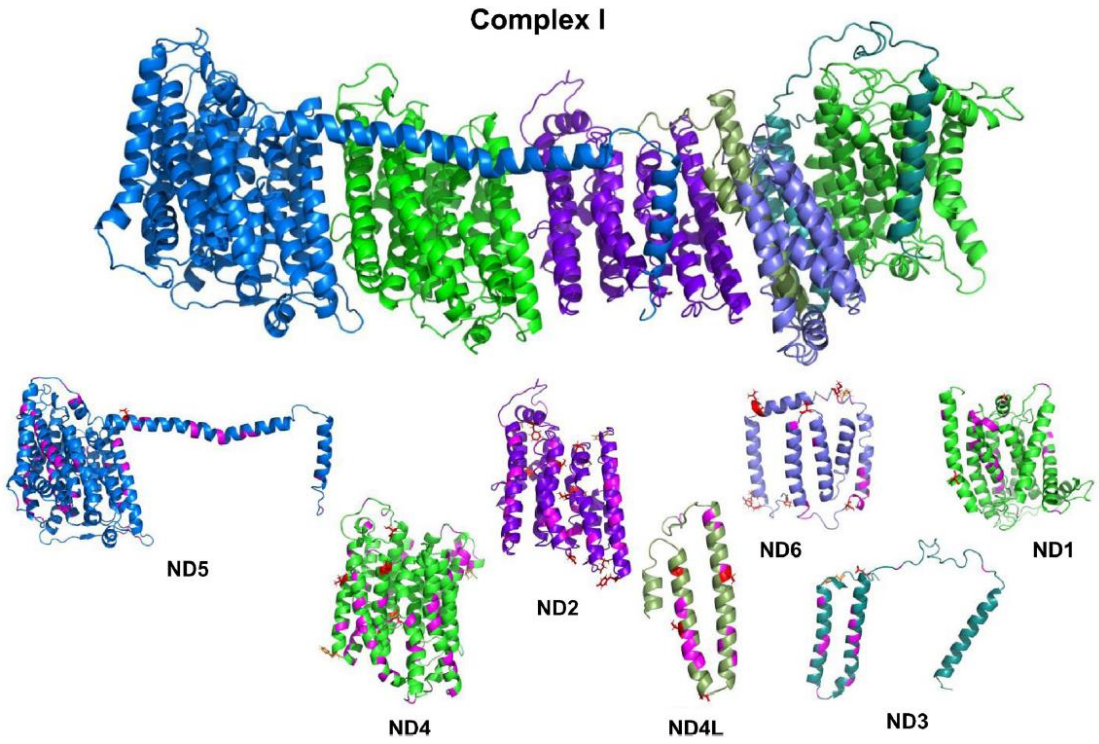


Figure 7: Computationally predicted atomic-level structure of mitochondrial complex I (ND1–ND6 and ND4L) of *Chaetopleura apiculata*. Positively selected sites inferred by CODEML and MEME are shown in red; residues with amino acid substitutions, as revealed by TreeSAAP, are in pink; and sites identified by all these methods are displayed in orange.

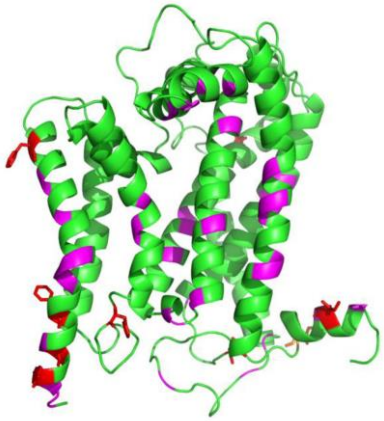


Figure 8: Computationally predicted atomic-level structure of mitochondrial CYTB subunit of *Chaetopleura apiculata*. Positively selected sites inferred by CODEML and MEME are shown in red and residues with amino acid substitutions, as revealed by TreeSAAP, are in pink.

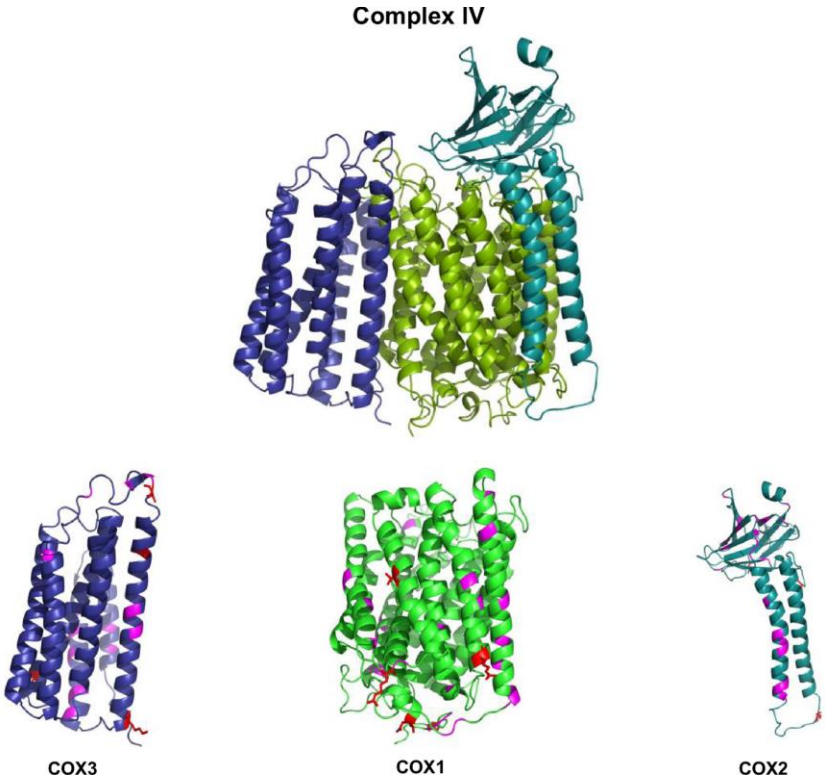


Figure 9: Computationally predicted atomic-level structure of mitochondrial complex IV subunit (COX1–COX3) of *Chaetopleura apiculata*. Positively selected sites inferred by CODEML and MEME are shown in red and residues with amino acid substitutions, as revealed by TreeSAAP, are in pink.

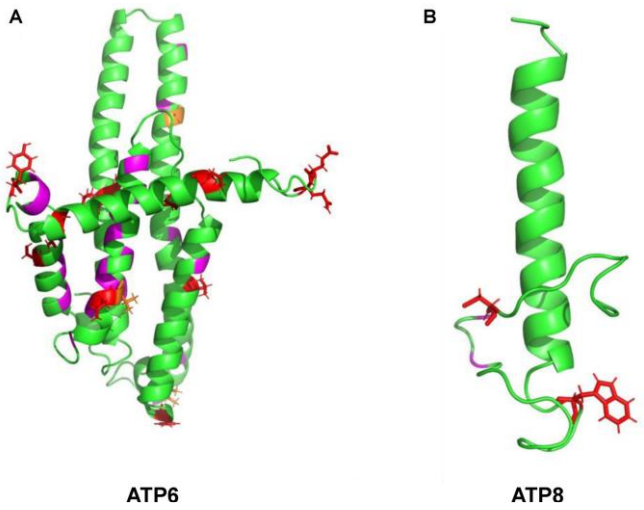


Figure 10: Computationally predicted atomic-level structure of mitochondrial subunits ATP6 (A) and ATP8 (B) of *Chaetopleura apiculata*. Positively selected sites inferred by CODEML and MEME are shown in red; residues with amino acid substitutions, as revealed by TreeSAAP, are in pink; and sites identified by all these methods are displayed in orange.

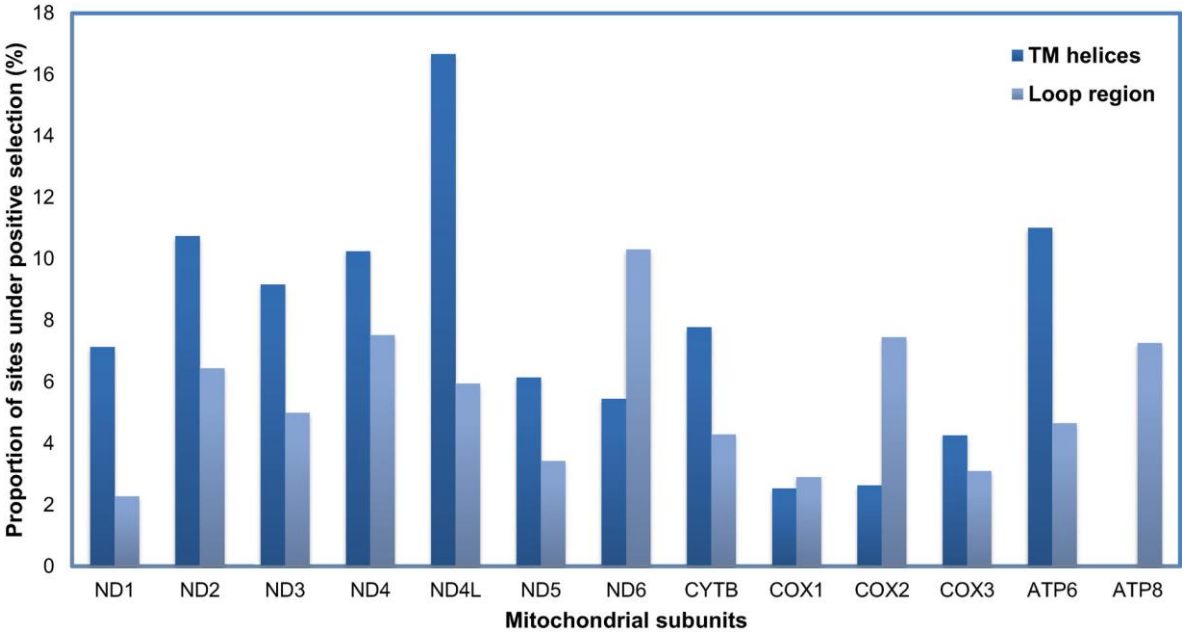


Figure 11: Distribution of sites, under positive selection and/or showing radical amino acid changes, between the transmembrane helices and the loop areas in individual mitochondrial subunits.

Complex I (proton-pumping NADH:ubiquinone oxidoreductase), the largest and most complicated enzyme set of the OXPHOS pathway, is responsible for NADH oxidation, ubiquinone reduction and proton pumping leading to ATP synthesis by complex V (Wirth et al., 2016). ND2, ND4 and ND5 are considered to be the pivotal members for maintaining the proton pump due to their sequence homology with a class of Na^+/H^+ antiporters (Mathiesen & Hägerhäll, 2002). Nonsynonymous substitutions in these subunits may either increase or decrease the efficacy of the proton pump. This could happen through biochemical changes that enhance/impair the proton translocation mechanism or improve/interrupt the redox-linked conformational changes (Brandt, 2006). I observed more sites with radical amino acid changes in ND proteins that can be attributed to their overall high mutation rate (da Fonseca et al., 2008).

CYTB of complex III in the electron transport chain plays a fundamental role in energy production by catalysing reversible electron transfer from ubiquinol to cytochrome *c* coupled to proton translocation (Trumpower, 1990). I observed instances of positive selection in the functionally significant domains of CYTB, which may be correlated with the adaptation of chitons to the fluctuating environment of the intertidal zone. In line with previous studies (Luo et al., 2008; Sun et al., 2018), I have substantial evidence of positive selection in the COX subunits (COX1–COX3), which is indicative of adaptive tolerance in intertidal organisms to intermittent hypoxia. Furthermore, sites under diversifying selection were identified in members of complex V (ATP synthase), the final enzyme complex of the respiratory chain that is involved in the electrochemical gradient across the inner mitochondrial membrane to drive ATP production (Hassanin et al., 2009; Zhang et al., 2017; Zhou et al., 2014). Therefore, mutations in *ATPase* genes have implications for the adaptive evolution of stress-tolerant intertidal chitons.

Similar studies on cephalopods, gastropods and bivalves have revealed significant evidence of positive selection in their OXPHOS subunits and these have been linked to mitogenomic adaptation (Almeida et al., 2015; Romero et al., 2016; Yang et al., 2019). Yet, unlike previous observations on other molluscan classes, I found several instances of diversifying selection acting differentially on all the members of mitochondrial complexes. Although the mechanisms of physiological adaptation in intertidal chitons remain elusive, my results clearly reflect the dynamic evolution of their mitochondrial genomes. The adaptive mitochondrial genomes may enable the chitons to thrive in the harsh environment of the intertidal zone. Intertidal chitons display elevated sensitivity to environmental stressors, such as thermal stress, changing ion concentration, fluctuation in osmolarity and reduced oxygen availability. Survival of these organisms depends on a modified and adaptive energy metabolism. This study provides insight into the molecular mechanisms underlying the adaptive strategies of polyplacophorans to the intertidal habitat from a mitochondrial perspective.

Chapter 2

Comparison of evolutionary selection acting on the mitochondrial protein-coding genes between intertidal and deep sea gastropods

2.1 Introduction

Animal mitochondrial DNA usually contains 13 protein-coding genes (PCGs), 2 rRNAs and 22 tRNAs essential for the translation of oxidative phosphorylation (OXPHOS) proteins. Cytochrome *c* oxidase subunit 1 - 3 (COX1 - COX3), cytochrome *b* (CYTB), NADH dehydrogenase subunit 1 - 6 (ND1 - ND6 & ND4L) and ATPase subunits (ATP6 & ATP8) are the subunits of the respiratory chain complexes, namely Complex I, III, IV and V, respectively (Blier et al., 2001). Mitochondrial genomes are an excellent source of information for phylogenetic and evolutionary studies but their application in marine invertebrates, particularly molluscs remain limited (Zhang et al., 2021).

Oxidative phosphorylation as well as innate immune response are regulated by mitochondrial PCGs (Lartigue & Faustin, 2013). Therefore, cellular homeostasis and survival relies largely on the maintenance of mitochondrial integrity and signaling (Sokolova, 2018). Determination of selection pressures acting on mitochondrial PCGs could shed light into the adaptive evolution of the mitochondrial genome. In the present study, I utilized gastropod mitogenomes to elucidate their environmental adaptation in intertidal and deep sea zones.

Gastropods represent the largest class of the phylum Mollusca, with number of species ranging from 80,000 to 150,000 (Appeltans et al., 2012; Parkhaev, 2007). Apart from being the only terrestrial molluscs, gastropods live in every conceivable place on Earth with marine ones occupying habitats ranging from the deepest ocean basins to the intertidal zone, as well as freshwater and other inland water habitats. The intertidal zone is the region between the highest and lowest tides, depicting the transition from ocean to terrestrial conditions. Organisms inhabiting this zone are able to maintain mitochondrial homeostasis amidst regular oscillations of immersion and emersion, along with tremendous variations of abiotic factors such as light, temperature, salinity and oxygen concentrations (Jensen & Denny, 2016). Consequently, gastropods living in this area need to withstand the dynamic conditions of the intertidal zone. On the other hand, gastropods are also an important component of the fauna in hydrothermal vents with regard to abundance and biomass (Chen et al., 2015). Deep sea hydrothermal vents are characterized by extreme physicochemical conditions, such as high hydrostatic pressure, fluctuating temperatures, low levels of oxygen and high concentrations of heavy metals and other toxic chemicals (Van Dover et al., 2002). Since acclimation to extreme habitat conditions are often correlated to the evolution of mitochondrial PCGs, detailed studies of gastropod mitogenome may uncover interesting aspects of molecular adaptations to living in intertidal and deep sea environment. In the

present study, fifteen gastropod mitochondrial genomes were considered to detect signs of natural selection and ascertain the adaptive evolution of the mitochondrial PCGs. I performed site and branch-site specific likelihood analyses to estimate the probability of positive selection on each site and branch of individual mitochondrial protein-coding genes across these species (Dhar & Dey, 2021).

2.2 Material and Methods

2.2.1 Sequence retrieval

The nucleotide sequences of 13 mitochondrial PCGs of 15 gastropods were downloaded from the National Center for Biotechnology Information (NCBI) database. Among the 15 gastropods, 13 species inhabiting the intertidal zone and 2 deep sea species were considered. The details and the ecology of the organisms are provided in Table 10 and 11, respectively.

2.2.2 Sequence alignment and model selection

The datasets of the 13 mitochondrial PCGs (*cox1–cox3*, *cytb*, *atp6*, *atp8*, *nd1–nd6* and *nd4l*) were constructed with the mitogenomes obtained from the 15 gastropod species. Codon-based alignment was performed using MUSCLE (Edgar, 2004) integrated in the MEGA X software (Kumar et al., 2018) selecting the invertebrate mitochondrial genetic code. The best-fitting substitution models for 13 mitochondrial PCGs were estimated using the option 'Find Best DNA/Protein Models' in the MEGA X software (Table 12).

Table 10: Details of the gastropod species considered.

Serial No.	Accession no.	Species	Family
1	AY588938.1	*Haliotis rubra*	Haliotidae
2	JN790612.1	*Fissurella volcano*	Fissurellidae
3	DQ862058.1	*Conus textile*	Conidae
4	NC_030536.1	*Conus striatus*	Conidae
5	NC_032377.1	*Californiconus californicus*	Conidae
6	KF728890.1	*Nerita versicolor*	Neritidae
7	NC_010090.1	*Reishia clavigera*	Muricidae
8	NC_017886.1	*Concholepas concholepas*	Muricidae
9	NC_012383.1	*Siphonaria pectinata*	Siphonariidae
10	NC_012376.1	*Onchidella celtica*	Onchidiidae
11	NC_016181.1	*Peronia peronii*	Onchidiidae
12	NC_031861.1	*Stomatella planulata*	Trochidae
13	NC_030595.1	*Littorina saxatilis*	Littorinidae
14	LC107880.1	*Lepetodrilus nux*	Lepetodrilidae
15	MK404176.1	*Pseudorimula sp.*	Lepetodrilidae

Table 11: Ecology of the gastropod species considered.

Species	Habitat	Depth range	Geographical distribution
Haliotis rubra	Intertidal (rocky surfaces, hidden in crevices, caves, fissures or vertical rock faces; frequent in estuaries, in natural beds on a shingle and mud bank)	0 – 50m	Indo-West Pacific: Endemic to Australia, from Fremantle, Western Australia to Angourie, New South Wales and south to Blubber Head, Tasmania
Fissurella volcano	Intertidal (underside of rocks)	NA	Eastern and Western Pacific Ocean
Conus textile	Intertidal	0 – 10m	Red Sea, the tropical Indo-Pacific, off Australia (New South Wales, Northern Territory, Queensland and Western Australia), New Zealand, the Indian Ocean from eastern Africa to Hawaii, and French Polynesia
Conus striatus	Intertidal	1 – 50m	Red Sea, Indian Ocean off the Aldabra Atoll, Madagascar, the Mascarene Basin, Mauritius and Tanzania; Pacific Ocean off the Philippines, Australia (Northern Territory, Queensland, Western Australia), New Zealand, New Caledonia and Thailand and Hawaiian islands.
Californiconus californicus	Intertidal (rocky and sandy bottoms)	0 – 45m	Eastern Pacific: USA to Mexico
Nerita versicolor	Intertidal (upper part of shores, often in crevices and pits of rock benches, or on branches of littoral trees overhanging the water)	0 – 1m	Indo-Pacific and Western Central Atlantic
Reishia clavigera	Intertidal (rocky shores)	NA	Western Pacific: Southeast and Northeast Asia
Concholepas concholepas	Intertidal (rocky substrates, tidepools and rock crevices)	0 – 40m	Eastern Pacific and Western Atlantic: From Lobos de Afuera Island, Peru to Cape Horn, Chile
Siphonaria pectinata	Intertidal (rocky shores)	NA	North Atlantic Ocean and Mediterranean Sea
Onchidella celtica	Intertidal (rocky shorelines)	NA	East Atlantic coastline, from western Scotland, western England and the Channel Islands, south to Spain and the Azores, including Great Britain and Coastline of Brittany

Peronia peronii	Intertidal (crevices and under rocks)	NA	Indo-West Pacific: Red Sea to Western Pacific
Stomatella planulata	Intertidal (stones and gravel)	0 – 30m	Western Pacific: Northern Marianas, Japan, Philippines and Taiwan
Littorina saxatilis	Intertidal (crevices of intertidal bedrock, in empty barnacle shells, and under rocks)	0 – 46m	Shores of the North Atlantic Ocean, including Hudson Bay, Baffin Island, Greenland, and the Barents Sea, south along the American East Coast to Chesapeake Bay, and along the European coast to the Straits of Gibraltar
Lepetodrilus nux	Deep sea (hydrothermal vents)	640 – 1580m	Northwest Pacific
Pseudorimula sp.	Deep sea (hydrothermal vents)	750 – 4200m	West Pacific, North Atlantic, and Indian oceans

Table 12: The best-fitting models selected for the 13 PCGs using MEGA X.

PCGs	Model selected
atp6	GTR + G
atp8	T92 + G
cox1	GTR + G
cox2	T92 + G
cox3	HKY + G
cytb	GTR + G + I
nd1	HKY + G + I
nd2	GTR + G
nd3	T92 + G
nd4	GTR + G + I
nd4l	T92 + G
nd5	GTR + G + I
nd6	GTR + G

2.2.3 Selection pressure analyses

The estimation of ω, *i.e.*, the ratio of non-synonymous substitutions (dN) to synonymous substitutions (dS) has been extensively used as an indicator of selection pressure on protein-coding genes. For protein-coding nucleotide sequences, the ratio of non-synonymous to synonymous substitutions (ω) differentiates neutrally evolving sequences ($\omega = 1$) from those subjected to negative ($\omega < 1$) or positive ($\omega > 1$) selection. The ω values were measured in EasyCodeML v1.21 that uses the codon-based maximum likelihood (CODEML) algorithm

(Gao et al., 2019; Yang, 2007). I performed both site and branch-site model analyses to identify the contrasting selection pressure acting on individual mitochondrial PCGs. Seven codon substitution models namely M0 (one-ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), M8 (beta and $\omega > 1$) and M8a (beta and $\omega = 1$) were considered and four likelihood ratio tests (M0 *vs.* M3, M1a *vs.* M2a, M7 *vs.* M8 and M8a *vs.* M8) were implemented to estimate the likelihood of positive selection on each site for individual PCGs (Anisimova et al., 2001; Yang et al., 2000; Yang et al., 2005). On the other hand, branch-site models enable ω to differ among sites and across branches of the phylogenetic tree (Zhang et al., 2005). Therefore, I evaluated selection on each site considering a single foreground branch (branch of interest) at a time. Posterior probabilities were computed using Bayes Empirical Bayes (BEB) analysis in order to recognize sites under positive selection on the selected branches if the Likelihood Ratio Test (LRT) *P*-values are significant (BEB > 95%). The requirement of running EasyCodeML is an alignment file and a phylogenetic tree. For each of the 13 PCGs, 13 Maximum Likelihood trees were generated using MEGA X software. In this study, only the leaf nodes were taken into consideration. HyPhy package v2.5 (available from www.hyphy.org) was used to assess codons under selection pressure (Pond et al., 2020). MEME (Mixed Effects Model of Evolution) and FEL (Fixed Effects Likelihood) analyses were performed to detect individual sites subjected to episodic diversifying selection and purifying selection, respectively (Pond & Frost, 2005; Murrell et al., 2012). Both analyses were done using a *P*-value ≤ 0.05 significance level.

2.3 Results and Discussion

2.3.1 Selection studies

Episodic positive selection acts on the mitochondrial PCGs that allow the organisms to adapt to extreme environmental changes (Shen et al., 2010; Tomasco & Lessa, 2011). The intertidal zone is one of the most stressful environments with extreme alterations in temperature, pH, salinity, and oxygen concentrations. The gastropods inhabiting this zone have successfully adapted to this dynamic environment. On the other hand, deep sea gastropods also face high hydrostatic pressure, variable temperatures and pH, and high levels of hydrogen sulphide, methane, and heavy metals. Considering the challenging conditions of both the zones, I investigated positive selection in the mitochondrial genes of intertidal and deep sea gastropods using EasyCodeML.

According to the results from the site-specific models of CODEML, negative selection dominated the evolution of all mitogenes. In other words, the comparison of the LRTs based on site models (M0 *vs.* M3, M1a *vs.* M2a, M7 *vs.* M8 and M8a *vs.* M8) revealed that the evolutionary rates of the PCGs are under negative constraints. Additionally, purifying selection was prevalent on all mitochondrial PCGs as revealed from the FEL analyses, with few sites evolving under neutrality (*P*-value < 0.05). From Figure 12, it is evident that the

PCGs (*cox1–cox3*) encoding complex IV subunits had the highest percentage of codons under negative selection. PCG-encoding complex III (*cytb*) also displayed high levels of purifying selection, whereas genes encoding complex I and V showed relatively more relaxed selection. In contrast, sixteen positively selected sites were identified in eight genes (*atp8*, *cytb*, *cox1*, *cox3*, *nd2*, *nd4*, *nd5* and *nd6*) using MEME and FEL (Table 13).

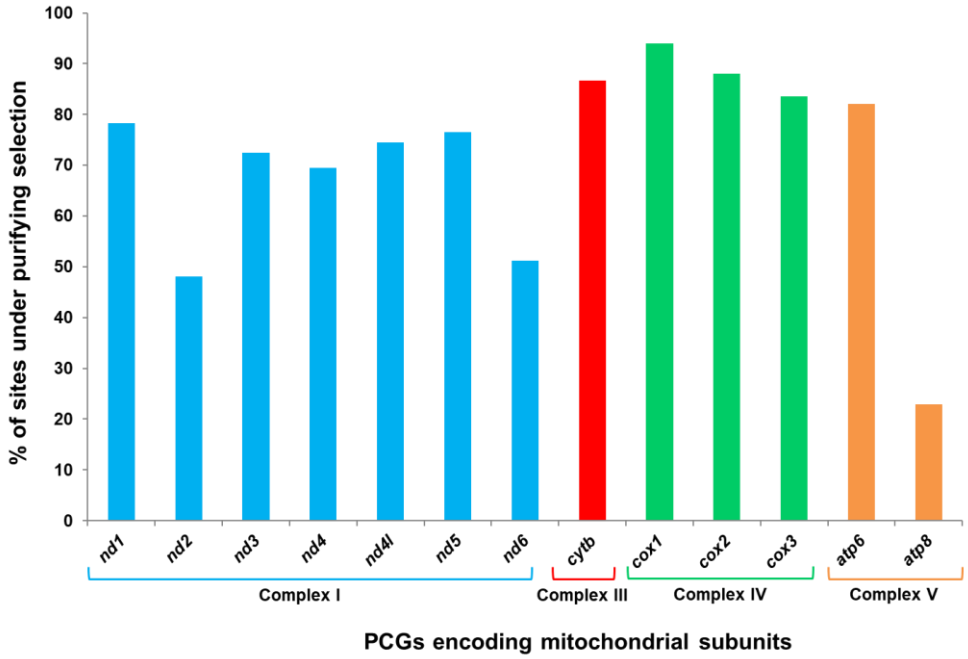
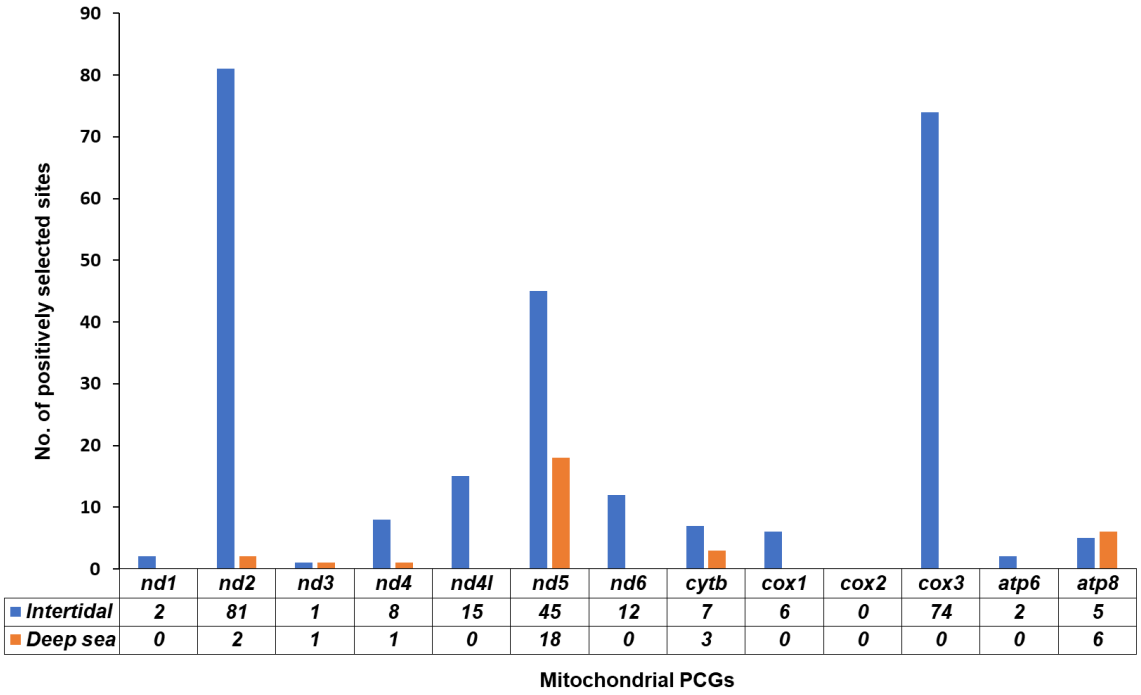


Figure 12: Percentage of sites under purifying selection in PCGs encoding mitochondrial subunits of gastropods, detected by FEL test. The subunits are colored according to their corresponding complexes: complex I (blue), complex III (red), complex IV (green) and complex V (orange).

Table 13: Sites of different mitochondrial PCGs under positive selection, as detected by MEME & FEL in the HyPhy package.

Genes	***Sites under positive selection***	
	MEME	**FEL**
atp8	46	-
cytb	8	-
cox1	17, 254	-
cox3	50	-
nd2	133	-
nd4	46, 198, 400	188
nd5	188, 455, 551	581
nd6	92, 173	173

As diversifying selection occurs at only few sites in a certain lineage for most protein-coding genes, identification of natural selection becomes restricted based on-site models only. Hence, I applied branch-site model, which allows ω value to differ between branches and sites at the same time, to measure varying selection pressure on the specified foreground branch against the remaining background lineages. I detected potential positively selected sites in 12 PCGs with posterior probabilities ≥ 95% using the Bayes Empirical Bayes method (Figure 13). Comparison of the mitochondrial PCGs of intertidal and deep sea gastropods revealed that the former lineage has experienced most diversifying selection. Only few sites located on the *nd2*, *nd3*, *nd4*, *nd5*, *cytb* and *atp8* gene were found to be positively selected along the branches of deep sea gastropods. Moreover, it is apparent from the results that for both intertidal and deep sea gastropods, PCGs encoding complex I subunits exhibited higher number of positively selected sites than PCGs encoding other complexes.



	nd1	nd2	nd3	nd4	nd4l	nd5	nd6	cytb	cox1	cox2	cox3	atp6	atp8
Intertidal	2	81	1	8	15	45	12	7	6	0	74	2	5
Deep sea	0	2	1	1	0	18	0	3	0	0	0	0	6

Figure 13: Distribution of positively selected sites, obtained from the branch-site models of CodeML, in 13 mitochondrial PCGs of intertidal and deep sea gastropods. The blue and orange bars represent the number of positively selected sites in intertidal and deep sea gastropods, respectively.

As a proton pump, the NADH dehydrogenase (Complex I) is the largest and most complicated enzyme complex of the respiratory chain. The efficacy of the proton pumping process can be influenced by mutation of this complex forming subunits, a feature crucial for

adaptive evolution (Ning et al., 2010; Yu et al., 2011). ND2 and ND5 are considered to be the central members for maintenance of the proton pump due to their sequence similarity with a class of Na^+/H^+ antiporters (Mathiesen & Hägerhäll, 2002). The effectiveness of the proton pump may either increase or decrease upon non-synonymous substitutions in these subunits. High number of positively selected sites are exhibited by these two ND protein-coding genes for both intertidal and deep sea gastropods that can be attributed to their high mutation rate (da Fonseca et al., 2008). CYTB of complex III plays a primary role in production of energy by catalyzing reversible electron transfer from ubiquinol to cytochrome *c* coupled to proton translocation (Trumpower, 1990). Evidence of positive selection was observed in *cytb*, which may be correlated with the adaptation of gastropods to their environment. In line with previous studies (Luo et al., 2008; Sun et al., 2018), signatures of positive selection were also observed in *cox1* and *cox3* along the branches of intertidal gastropods, which probably allows intertidal organisms to adapt to intermittent hypoxia. Additionally, few sites that were identified under diversifying selection in members of complex V (*atp6* & *atp8*), may be involved with the regulation of the electrochemical gradient across the inner mitochondrial membrane to drive ATP production (Zhang et al., 2017). Therefore, the non-synonymous substitutions in *ATPase* genes provide possible implications on adaptive evolution of the stress-tolerant intertidal gastropods and also may help deep sea gastropods to thrive under harsh deep sea conditions.

Unlike two deep sea dwelling gastropod species (*Lepetodrilus nux* and *Pseudorimula sp.*), I found maximal instances of diversifying selection acting differentially on all the members of mitochondrial complexes in the 13 intertidal gastropods considered here. This may be attributed to the fact that though deep sea vent systems are driven by several challenging abiotic conditions, *i.e.*, no sunlight and highly acidic and toxic vent fluids emerging at extreme temperatures, most vent species are restricted to their particular biogeographic territory (Rogers et al., 2012). On the other hand, intertidal gastropods have to persist in the extremely variable environment of the intertidal zone caused by periodic changes of temperature, salinity, pH and hydrodynamic forces (Jensen & Denny, 2016). Hence, survival of intertidal organisms depends on an extensively modified and adaptive energy metabolism. Although the mechanisms of physiological adaptation in intertidal gastropods remain elusive, the results clearly reflect the dynamic evolution of their mitochondrial genomes. Overall, this study provides insight into the mitogenomic adaptations of intertidal and deep sea gastropods. It is worth mentioning that if the sample size is increased, there might be some deviations in the conclusion. However, more species could not be considered in this case owing to the unavailability of the complete mitochondrial genome sequences (particularly PCGs) of gastropods, specifically those inhabiting the deep sea, in the database.

Chapter 3

In silico analysis of two conserved residues in the structure of big defensin from *Crassostrea gigas*

3.1 Introduction

Antimicrobial peptides (AMPs) are major constituents of the innate immune system in all living multicellular organisms (Bulet et al., 2004). The main families of AMPs found in vertebrates are cathelicidins and defensins (De Smet & Contreras, 2005). They play important roles in defending against bacterial/viral invasion by interfering with bacteria and virus replication (Avila, 2017). Additionally, they possess other functions such as immune modulation, apoptosis and wound healing (Lai & Gallo, 2009; Zasloff, 2002). Apart from vertebrates and arthropods, several classes of AMPs have been identified in bivalves, such as defensins, mytilins, myticins and mytimicin in mussels (Novoa et al., 2016; W. Xu & Faisal, 2010); defensins, big defensins, proline-rich AMPs and bactericidal/permeability-increasing (BPI) proteins in oysters (Gueguen et al., 2009; Rosa et al., 2011; Schmitt et al., 2010; Zhang et al., 2011); defensin, macin and BPI proteins in clams (Yang et al., 2019a, 2019b; Zhang et al., 2015).

Defensins are a diverse family of small, cysteine rich and cationic antimicrobial peptides, that consist of three or four intramolecular disulphide bonds formed by six or eight cysteine residues in a complex folded arrangement of two or three antiparallel β-sheets with or without an α-helix structure (Shafee et al., 2017; Tassanakajon et al., 2015). The most distinct molecular feature of defensins is their high pI value, which ranges from +6 to +12 in the monomeric form, manifested by abundant arginine and lysine residues in their sequences (Suresh & Verma, 2006). They kill microorganisms through permeabilization of the microbial membrane composed of negatively charged components such as phospholipids, teichoic acids and lipopolysaccharides (Jenssen et al., 2006). Vertebrate defensins possess six cysteine residues and can be divided into three distinct families: α, β and θ defensins of 29-35, 35-45 and 18 amino acid residues, respectively (Nguyen et al., 2003). On the other hand, invertebrate defensins can primarily be categorized into two classes: (1) the invertebrate cysteine-stabilized α-helix/β-sheet motif defensin (CSαβ defensin), containing an α-helix linked to an antiparallel two-stranded/triple-stranded β-sheet by disulphide bridges and is found in molluscs, nematodes and arthropods (Carvalho & Gomes, 2009; Zhu, 2008); (2) the invertebrate big defensin having a disulphide array stabilized β-sheet structure (Kouno et al., 2008; Zhu & Gao, 2013) which exists mainly in molluscs (Saito et al., 1995) but also present in arthropods (Teng et al., 2012).

Big defensin precursors are composed of a signal peptide followed by a prodomain holding a dibasic cleavage site, and a multi-domain polypeptide (mature big defensin) (Gerdol et al., 2020). Mature big defensins possess an N-terminal hydrophobic region and a C-terminal region consisting of six cysteine residues (Figure 14A). Till date, only the tertiary structures of big defensin from the horseshoe crab and the Pacific oyster have been determined (Figure 14B) (Kouno et al., 2008; Loth et al., 2019). The peptides are composed of two distinct globular domains connected by a flexible linker where the N-terminal hydrophobic domain adopts a β1-α1-α2-β2 fold while the cationic C-terminal domain shows the typical cysteine pairing of β-defensins ($Cys_{1-5}Cys_{2-4}Cys_{3-6}$). Interestingly, both structural and phylogenetic studies have provided compelling evidence about a shared evolutionary origin for vertebrate β-defensins and invertebrate big defensins. It is noteworthy that the N-terminal hydrophobic region is the hallmark of big defensins, a trait lost during the transition from basal chordates to their vertebrate relatives (Shafee et al., 2017; Zhu & Gao, 2013). Therefore, the elucidation of the structural characteristics of big defensins is of utmost significance.

In this study, I have analyzed 24 big defensin protein sequences obtained from different molluscan classes, out of which 11 unannotated sequences were identified using the tBLASTn tool. Multiple sequence alignment was carried out to estimate the conserved residues within the big defensins. Additionally, conserved residues common for both big defensins and β-defensins were identified. Furthermore, *in silico* mutation analysis was carried out by constructing the mutants of two conserved residues in the structure of big defensin1 from *Crassostrea gigas* to understand the effect of mutations on the structure and function of the protein. The structural and conformational changes of the wild-type and mutant proteins were studied at an atomic level using 100 ns molecular dynamics (MD) simulation.

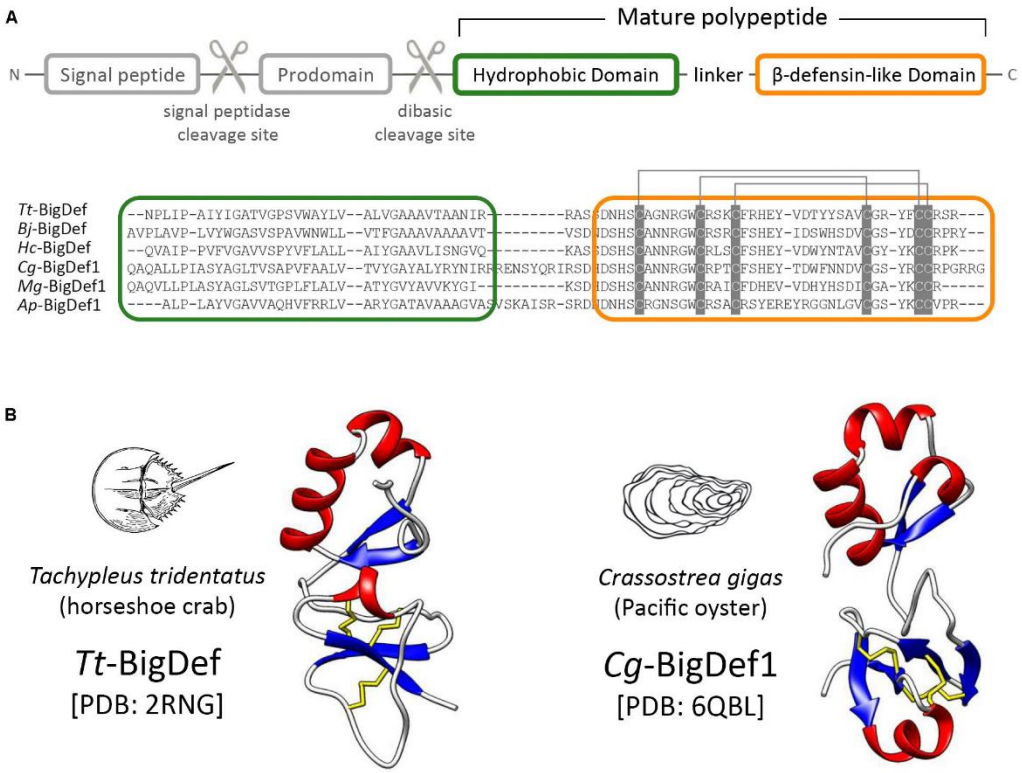


Figure 14: (A) Structural domain organization of big defensins. Hydrophobic (green frame) and β-defensin like (orange frame) domains are indicated at the amino acid sequence alignment of certain mature big defensins. Cysteine pairing is indicated by gray lines. *Tt*-BigDef = *Tachypleus tridentatus* Big Defensin; *Bj*-BigDef = *Branchiostoma belcheri tsingtauense* Big Defensin; *Hc*-BigDef = *Hyriopsis cumingii* Big Defensin; *Cg*-BigDef1 = *Crassostrea gigas* Big Defensin1; *Mg*-BigDef1 = *Mytilus galloprovincialis* Big Defensin1; *Ap*-BigDef1 = *Argopecten purpuratus* Big Defensin1. (B) The 3D structure of canonical big defensins showing two structural domains connected by a flexible linker. Source: Gerdol et al., 2020.

3.2 Material and Methods

3.2.1 Dataset collection

The amino acid sequences of big defensin from 11 mollusc species were downloaded from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/protein). Additionally, unannotated protein sequences of big defensins from 11 other molluscs were identified using the tBLASTn tool (Gertz et al., 2006). Furthermore, β-defensin protein sequences from 10 vertebrate species were retrieved from the NCBI database. The details of the species analyzed are provided in Table 14.

3.2.2 Multiple sequence alignment and identification of conserved residues

Multiple sequence alignments of (1) mature peptide sequences of big defensins from 22 mollusc species and (2) amino acid sequences pertaining to the C-terminal β-defensin-like domain of big defensin and β-defensin from 22 mollusc and 10 vertebrate species, respectively, were carried out using MUSCLE v. 3.8.31 (Edgar, 2004) integrated in the MEGA X software (Kumar et al., 2018) with default settings. Identification and marking of the conserved residues in the alignments were performed in Jalview v2.11.1.4 (Waterhouse et al., 2009).

3.2.3 Mutagenesis and structure validation

Two conserved, charged amino acid residues obtained from the multiple sequence alignment of the β-defensin-like domain of mollusc big defensin and vertebrate β-defensin protein sequences, were mutated. The atomic coordinates of big defensin1 protein from *Crassostrea gigas* were retrieved from the Protein Data Bank (PDB ID: 6QBL) (Loth et al., 2019). Two mutated monomeric structures of the aforementioned protein (R64A and E71A) were generated in PyMOL using the Mutagenesis wizard by selecting the most probable rotamers for each amino acid substitution (Delano, 2002). The stereochemical quality of the wild-type and two mutated protein structures after energy minimization in GROMACS 2020.1 (Abraham et al., 2015) were assessed through PROCHECK (Laskowski et al., 1993) in the Structural Analysis and Verification Server (SAVES) v. 5.0 server (https://servicesn.mbi.ucla.edu/SAVES/).

Table 14: Details of the invertebrate and vertebrate species containing the big defensin and β-defensin protein, respectively.

		Class	Species	Accession no.
Invertebrates	**Big defensin**	Bivalvia	*Mizuhopecten yessoensis*	XP_021358476.1
		Bivalvia	*Mimachlamys nobilis*	ANY95044.1
		Bivalvia	*Argopecten irradians*	ABC61319.1
		Bivalvia	*Argopecten purpuratus*	ANK58567.1
		Bivalvia	*Pecten maximus*	XP_033753528.1
		Bivalvia	*Crassostrea virginica*	XP_022319212.1
		Bivalvia	*Hyriopsis cumingii*	AEP26934.1
		Bivalvia	*Mytilus galloprovincialis*	AFC37168.1
		Bivalvia	*Crassostrea gigas*	AEV46761.1
				AEV46758.1
				AEV46760.1
		Bivalvia	*Potamilus streckersoni*	-
		Bivalvia	*Megalonaias nervosa*	-
		Bivalvia	*Venustaconcha ellipsiformis*	-
		Gastropoda	*Pomacea canaliculata*	XP_025098815.1
		Gastropoda	*Lanistes nyassanus*	-
		Gastropoda	*Marisa cornuarietis*	-
		Gastropoda	*Pomacea maculata*	-
		Cephalopoda	*Octopus sinensis*	XP_036356192.1
		Cephalopoda	*Nautilus pompilius*	-
		Cephalopoda	*Hapalochlaena maculosa*	-
		Cephalopoda	*Sepia pharaonis*	-
		Cephalopoda	*Architeuthis dux*	-
		Polyplacophora	*Acanthopleura granulata*	-
Vertebrates	**β-defensin**	Mammalia	*Homo sapiens*	AAQ09526.1
		Mammalia	*Mus musculus*	AAT67592.1
		Mammalia	*Pan troglodytes*	CAL68954.1
		Actinopterygii	*Oplegnathus fasciatus*	ADJ21805.1
		Actinopterygii	*Oryzias latipes*	ACG55699.1
		Actinopterygii	*Odonus niger*	QWB49506.1
		Actinopterygii	*Channa striata*	QJW82621.1
		Actinopterygii	*Oreochromis niloticus*	QBO59814.1
		Actinopterygii	*Oncorhynchus mykiss*	CBB12546.1
		Actinopterygii	*Epinephelus coioides*	AET25528.1

3.2.4 Identification of amino acid network and functional residues in protein structure

The webPSN v. 2 server (Felline et al., 2020) was utilized for the analysis of network of interacting amino acids wherein the wild-type and mutant proteins were considered. The ConSurf server (http://consurf.tau.ac.il/) was used to determine the functional regions in the protein structure of *Crassostrea gigas* big defensin. This tool analyzes the evolutionary dynamics of amino acid substitutions among homologous sequences and maps them onto the structure of the query protein (Ashkenazy et al., 2016). Additionally, ConSeq v. 1.1, integrated in the ConSurf server, was used to identify the functionally and structurally important residues in the amino acid sequence of big defensin (Berezin et al., 2004).

3.2.5 Prediction of deleterious mutations

PROVEAN (Protein Variation Effect Analyzer), a software tool which predicts whether an amino acid substitution has an impact on the biological function of a protein (Choi & Chan, 2015), was used to predict the deleterious or damaging effect of the mutations (R64A and E71A) in the protein structure of big defensin.

3.2.6 Prediction of destabilizing mutations

Protein stability is often represented by the change in the Gibbs free energy (ΔG) upon folding. In this study, I have utilized four protein stability prediction tools, DynaMut2 (Rodrigues et al., 2021), DUET (Pires et al., 2014b), mCSM (Pires et al., 2014a) and CUPSAT (Parthiban et al., 2006). DynaMut can be used to analyze and visualize protein dynamics by sampling conformations and assess the impact of mutations on protein dynamics and stability resulting from vibrational entropy changes. DUET is an integrated computational method used for calculating the impact of mutations on the stability of the protein. The mCSM algorithm establishes effects of mutations depending on the structural signature. CUPSAT uses protein structure environment-specific potential and torsion angle potential to find the differences in stability of wild-type and mutant proteins.

3.2.7 Aggregation propensity analysis

Solubility is an important feature required for the functioning of a protein. The solubility of the mutant forms of big defensin from *C. gigas* were predicted using the SODA web-server (http://protein.bio.unipd.it/soda). The program utilizes the propensity of amino acid sequence to aggregate, intrinsic disorder, secondary structure and hydrophobicity to calculate the differences in protein solubility (Paladin et al., 2017).

3.2.8 Molecular dynamics simulation

Molecular dynamics simulation was performed on the wild-type and two mutated monomers of the big defensin protein from *C. gigas* using the GROMACS simulation package 2020.1. OPLS-AA/L all-atom force field was used to model the intramolecular protein interactions

and the intermolecular interactions between the protein and solvent molecules (Kaminski et al., 2001). Initially the energy of each system was minimized using 500 steps of the steepest descent algorithm followed by 20,000 steps of the Polak-Ribiere conjugate gradient method to remove the strain in the initial structure. The relaxed structure was immersed in a cubic box of extended simple point charge (SPC/E) water molecules with periodic boundary conditions in all directions. A minimum distance between the protein and the wall of the cell was set to 1 nm. Prior to energy minimization with periodic boundary conditions, each solvated system was neutralized by the addition of sodium and chloride ions.

MD simulation consists of equilibration and production phases. In the first stage of equilibration, the solutes (protein, counter ions) were fixed, and the solvent (water molecules) was equilibrated for 100 ps of MD at 200 K using an integral time step of 0.001 ps. During the equilibration phase, velocity was assigned to the atoms using Maxwell distribution. The system was coupled to the heat bath and heated to 300 K in a short run of 100 ps (0.001 ps time step) in which the system was allowed to relax in the new condition. This was followed by another short simulation of 100 ps with pressure coupling at 1 atm. Finally, the production phase of MD simulation was run keeping the temperature, pressure and number of molecules of the ensemble invariant. Production phase was continued up to 100 ns using 0.002 ps time step for each of the wild-type and mutated big defensin proteins. The average structures for each monomer were obtained using the 100 ns trajectory of the MD production run. Subsequent analyses that include RMSD (Root Mean Square Deviation), RMSF (Root Mean Square Fluctuation) and radius of gyration (*Rg*) were performed using different programs of the GROMACS package on the 100 ns trajectory of the production run.

3.3 Results

This study sheds light into the structural and functional significance of two conserved residues in the β-defensin-like domain of big defensin from *Crassostrea gigas*. In addition, deleterious/destabilizing effect of mutations, aggregation behavior and conservation score were screened using advanced computational methods. Finally, the atomistic details of two amino acid mutations (R64A and E71A) have been analyzed through MD simulation approach.

3.3.1 Sequence comparison

Amino acid variability among 24 big defensin protein sequences from different classes of mollusc were observed from the analysis of the multiple sequence alignment. Figure 15 shows the alignment of the conserved residues both in the hydrophobic domain and β-defensin-like domain of the mature peptide sequences. Apart from the six conserved cysteines in the C-terminal region, two uncharged non-polar residue (glycine and tryptophan), two positively charged and one negatively charged amino acid (arginine and histidine; glutamic acid) were found to be conserved. On the other hand, owing to a shared ancestral

origin between vertebrate β-defensins and invertebrate big defensins, a multiple sequence alignment was carried out to identify the conserved residues therein. Two conserved charged amino acids were found (except six conserved cysteines) on aligning the amino acid sequences pertaining to the C-terminal β-defensin-like domain of big defensin from invertebrate mollusc and β-defensin from vertebrate species (Figure 16).

Figure 15: Multiple sequence alignment of 24 mature peptides sequences of big defensin from different mollusc species performed using MUSCLE integrated in MEGA X software. The residues with conservation score above 80% have been colored. The two conserved residues to be mutated *in silico*, have been marked with stars.

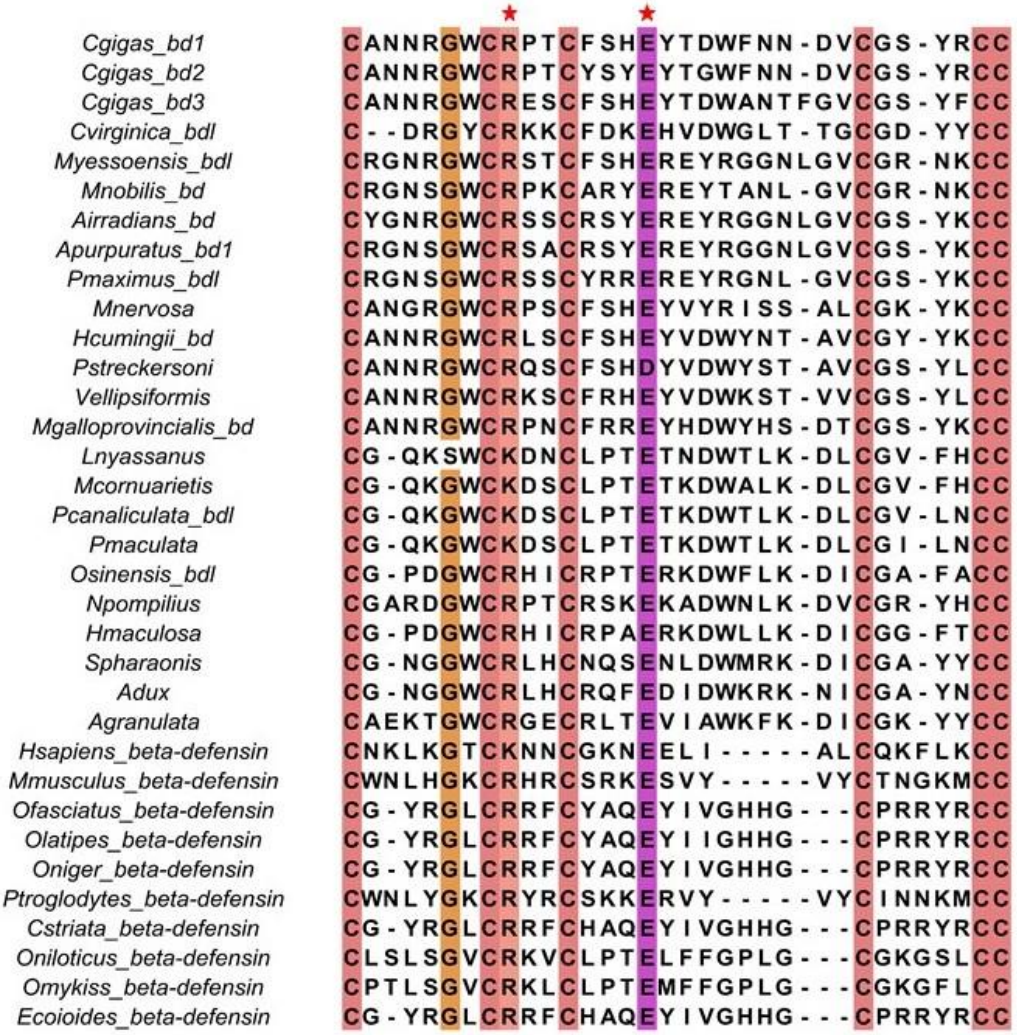


Figure 16: Multiple sequence alignment of the amino acid sequences pertaining to the C-terminal β-defensin-like domain of big defensin from invertebrate mollusc and β-defensin from vertebrate species, performed using MUSCLE integrated in MEGA X software. The residues with conservation score above 80% have been colored. The two conserved residues to be mutated *in silico*, have been marked with stars.

3.3.2 Mutagenesis and structure validation

Both structural and phylogenetic studies have provided compelling evidence that big defensins could be the missing link in vertebrate defensin evolution, as an invertebrate big defensin gene has been hypothesized as the most probable ancestor of β-defensins (Shafee et al., 2017). Both gene types share a phase I intron in a highly conserved position, at the 5' end of the region encoding the C-terminal cysteine rich region (Gerdol et al., 2020). Figure 16

also revealed the presence of two conserved charged amino acids in the C-terminal region of both big defensin and β-defensin sequences. So, two mutated monomeric structures (R64A and E71A) of big defensin from *Crassostrea gigas* were constructed in this study to understand their structural and functional significance. The wild-type and mutated monomers of big defensin were subjected to energy minimization and validated in PROCHECK, available online on the SAVES server, to assess the quality of the protein structures. Ramachandran plot for each structure demonstrated that they follow all the stereochemical properties with favourable phi (ϕ) and psi (ψ) values with no residues in the outlier region (Figure 17).

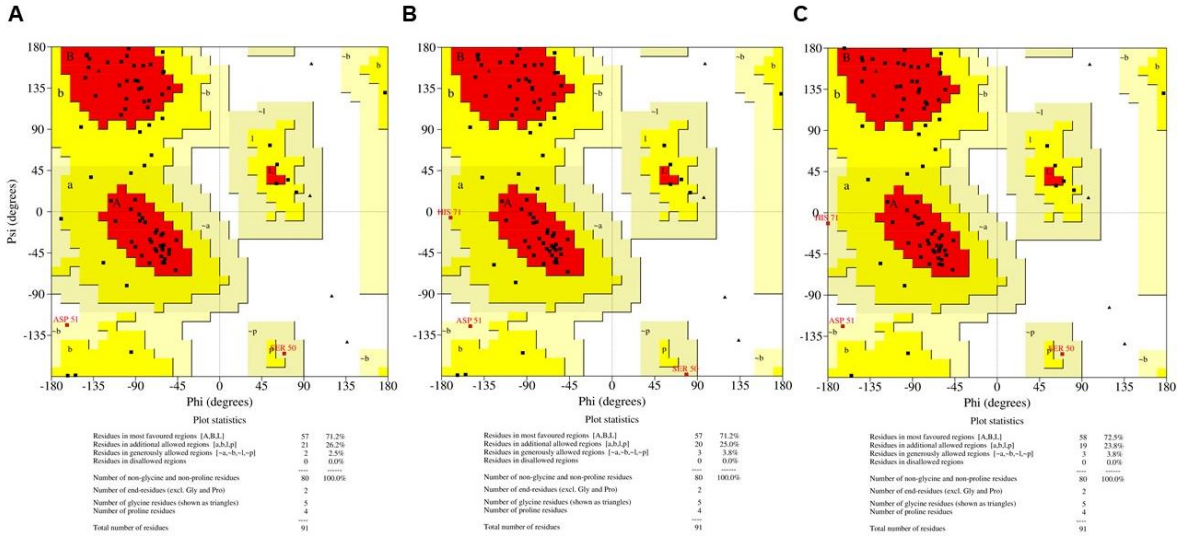


Figure 17: Ramachandran plot showing the dihedral angle values for (A) wild-type big defensin ,(B) R64A mutant and (C) E71A mutant, performed using the PROCHECK server.

3.3.3 Identification of amino acid network and functional residues in protein structure

Amino acid network analysis of the wild-type protein structure revealed R64 and E71 to be hub residues, *i.e.*, they are engaged in different kinds of intramolecular interaction with other residues for maintaining the native protein scaffold. At the same time, when the mutated structures (R64A & E71A) were considered, the respective amino acid position lost their role as hub residue.

Analysis of conserved amino acid residues in protein provides a better understanding of the significance of a particular amino acid residue and probably its localized evolution. Therefore, ConSurf was used to find the evolutionary conservation score of each amino acid residue in big defensin, where a score of less than 3 and more than 7 indicate variable and conserved residues, respectively (Figure 18A). The result of ConSeq analysis shows the degree of

conservation as well as the structurally and functionally important residues along the sequence of big defensin. Here, R64 and E71 (marked with a star above the residue) have been considered as functional residues (Figure 18B).

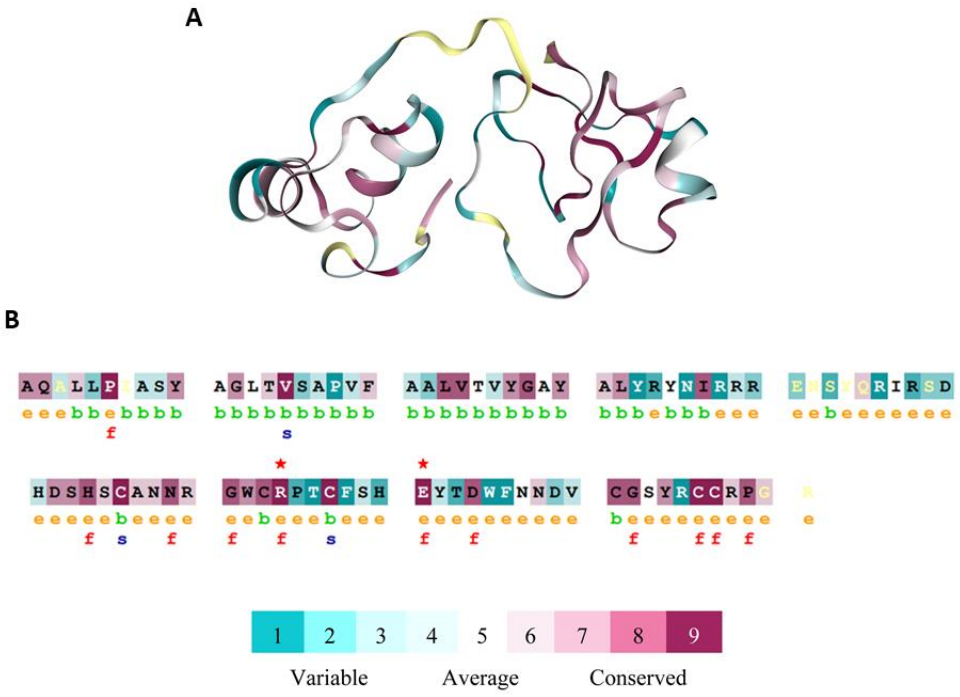


Figure 18: Evolutionary conservation of amino acid residues in the three-dimensional structure of big defensin by ConSurf analysis (A) and in the primary sequence of the same protein by ConSeq analysis (B). 'e' refers to an exposed residue according to the neural-network algorithm; 'b' refers to a buried residue according to the neural-network algorithm; 'f' refers to a predicted functional residue (highly conserved and exposed); 's' refers to a predicted structural residue (highly conserved and buried). The two conserved residues that have been mutated are marked with stars.

3.3.4 Prediction of deleterious and destabilizing mutations

R64A and E71A mutants depicted a deleterious effect on the function of big defensin when analyzed by PROVEAN (Table 15). PROVEAN analyzed the consequence of the two mutations on the function carried out by the big defensin. PROVEAN score ≤ -2.5 indicates that the amino acid variant has damaging effect whereas variant having score > -2.5 is considered to have neutral effect on the protein. The program predicted that these two mutations have a deleterious effect on the big defensin. Similarly, prediction of destabilizing

mutations employing DynaMut2, DUET, mCSM and CUPSAT showed destabilizing behavior of the protein by these two mutations (Table 16).

Table 15: Prediction of deleterious mutation in the big defensin protein using PROVEAN web server.

Variant	PROVEAN score	Prediction (cutoff = -2.5)
R64A	-6.000	Deleterious
E71A	-6.000	Deleterious

Table 16: Prediction of destabilizing mutation in the big defensin using different computational programs.

Computational method	Predicted Stability Change ($\Delta\Delta G^{Stability}$)	
	R64A	E71A
DynaMut2	Destabilizing	Destabilizing
DUET	Destabilizing	Destabilizing
mCSM	Destabilizing	Destabilizing
CUPSAT	Destabilizing	Destabilizing

3.3.5 Aggregation Propensity Analysis

Solubility of a protein plays a critical role in maintaining the conformation of the protein. Analysis of aggregation behavior using the SODA web server aids in molecular identification of a mutation to the conformational level. The results revealed that the mutation of R64 and E71 in the protein structure of big defensin will increase its solubility (Table 17).

Table 17: Prediction of aggregation propensities of wild-type and mutant forms of big defensin using the SODA web server.

Variants	Aggregation	Disorder	SODA	
			Score	Solubility
Wild-type	-4.652	0.049		
R64A	3.664	-0.025	0.513	More soluble
E71A	2.507	0.08	1.619	More soluble

3.3.6 Mutation induced conformational stability, flexibility and dynamic changes

The disruptive effects of R64A and E71A mutations on the conformational stability of big defensin from *Crassostrea gigas* were analyzed through 100 ns MD simulations. Overall changes caused by the mutations in the protein structure of big defensin were investigated by RMSD calculation. It shows the differences in flexibility of the backbone atoms of the wild-

type big defensin and its mutants. The RMSD plot of each protein form reflected convergence of the simulation, indicating the overall structural stability of all the monomers (Figure 19).

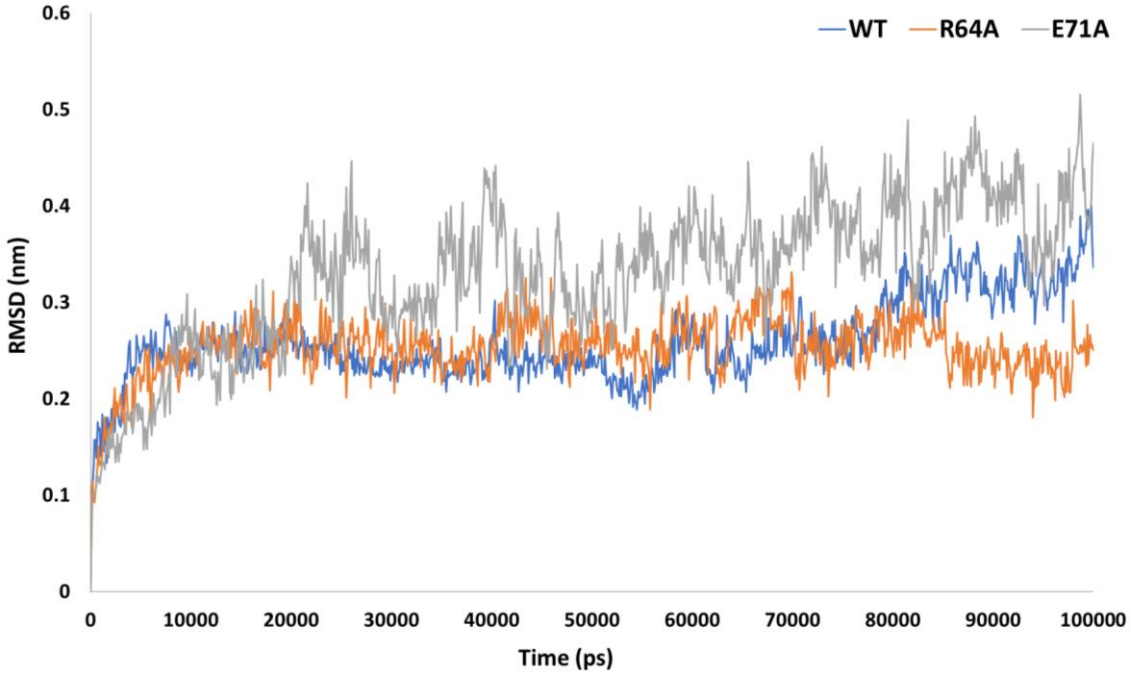


Figure 19: RMSD plot of the backbone atoms of wild-type and mutant big defensin proteins over the 100 ns trajectory of the MD production run.

I also calculated the radius of gyration (*Rg*) which is a measure of the compactness of a protein. Differences in *Rg* values between wild-type and mutated big defensin are shown in Figure 20. The mutation E71A shows a significant increase in *Rg*, suggesting a loss in compactness.

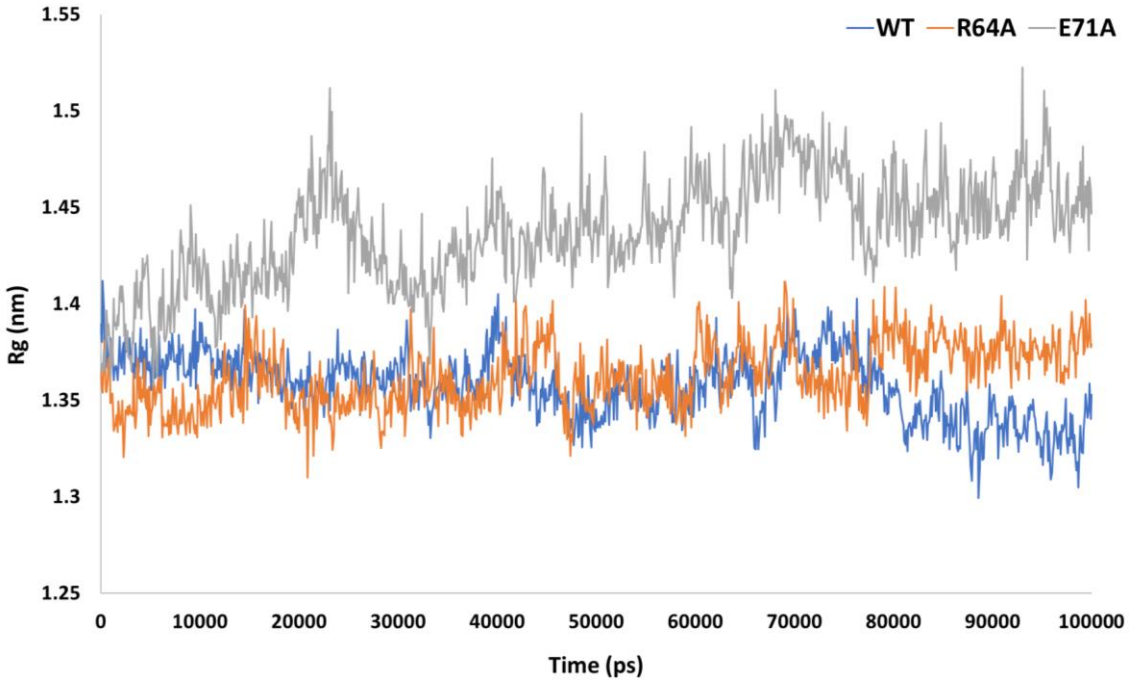


Figure 20: Radius of gyration of wild-type and mutant big defensin proteins over the 100 ns trajectory of the MD production run.

The RMSF for individual residues for wild-type big defensin and mutated monomers were computed to infer the residue specific flexibility, taking into account the respective input structure for MD production run as reference (Figure 21). It is evident that the R64A mutation reduces the fluctuations of the amino acid residue at the 64th position compared to the wild-type big defensin and E71A mutant. On the other hand, a significant increase in fluctuation of the residue at the 71st position was observed in case of E71A mutant.

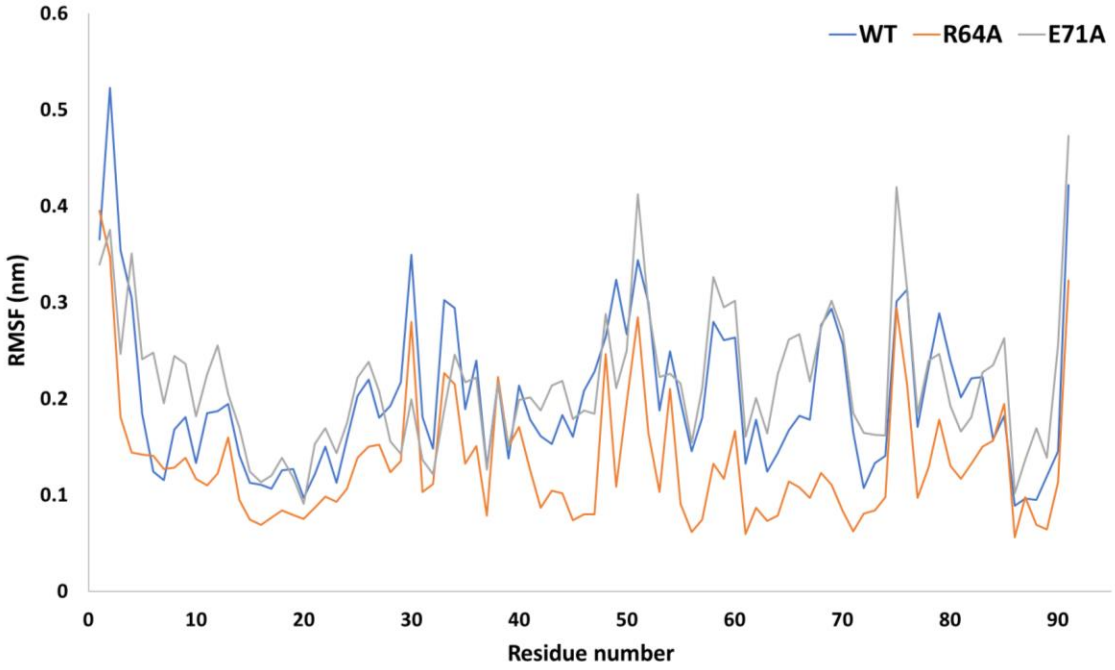


Figure 21: RMSF plots of each residue of wild-type and mutant big defensin proteins over the 100 ns trajectory of the MD production run.

3.4 Discussion

In this study, various computational approaches were used to predict the effect of mutating two conserved residues (R64 and E71) in the structure of big defensin from *Crassostrea gigas*. First, this work tested the significance of the two conserved residues using WebPSN and ConSeq where both R64 and E71 have been considered as hub and functional residues. Results from ConSurf analysis revealed that these residues are located in the evolutionarily conserved regions of the tertiary structure of big defensin.

Defensins are amphipathic molecules with a net positive charge possessing disulfide bonds which are important for their molecular function (Park et al., 2018). When disulfide linkages are affected, the resulting misfolded defensins could exhibit altered biological functions affecting its antimicrobial and chemotactic activity (Wang, 2012). The R64A and E71A mutants were considered as deleterious by PROVEAN. These two mutants were further analyzed by DynaMut2, DUET, mCSM and CUPSAT, showing that the mutations decreased the protein stability. These results indicate that R64A and E71A mutants may affect both the protein's structure and functions. They could be considered as potentially significant residues to be targeted for therapeutic purposes. Changes in backbone conformation, residual

flexibility, compactness and solvent accessibility probably indicate significant structural and functional losses in the big defensin protein due to the two analyzed mutations.

The prediction of the impact of these mutations on the structural organization of big defensin done using bioinformatics tools require further experimental analyses in order to validate their effects on the structure and function of big defensin. This effort will add to the knowledge in this area especially if investigations are done on other conserved residues in both the hydrophobic and the β-defensin-like domains of big defensin, that could not be completed in the thesis owing to the shortage of time.

In conclusion, these results provide the first in-depth understanding of the destabilization and loss of conformational dynamics of big defensin from *Crassostrea gigas* mediated by R64A and E71A mutations. These results should be further exploited in order to better understand the immune response of oysters to environmental stress. Such is especially important in view of an observed increase in anthropogenic impacts (*i.e.* temperature rise, nutrient overload and chemical pollution of near shore areas) that may result in toxic dinoflagellate blooms (Gonzalez-Romero et al., 2017). A good example of this is the recent red tide that affected the Pacific coast of Hokkaido resulting in losses of the order of million yens. A better knowledge of oysters' immune system and its components represent an important contribution in the search for ways to reinforce the immune responses of both oyster seedlings and adults so they can better withstand pathogens and other stress mediated by environmental changes predicted for the near future.

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/https://doi.org/10.1016/j.softx.2015.06.001

Almeida, D., Maldonado, E., Vasconcelos, V., & Antunes, A. (2015). Adaptation of the Mitochondrial Genome in Cephalopods: Enhancing Proton Translocation Channels and the Subunit Interactions. *PloS One*, *10*(8), e0135405. https://doi.org/10.1371/journal.pone.0135405

Anisimova, M., Bielawski, J. P., & Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution*, *18*(8), 1585–1592. https://doi.org/10.1093/oxfordjournals.molbev.a003945

Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A., Błażewicz-Paszkowycz, M., Bock, P., Boxshall, G., Boyko, C. B., Brandão, S. N., Bray, R. A., Bruce, N. L., Cairns, S. D., Chan, T.-Y., … Costello, M. J. (2012). The Magnitude of Global Marine Species Diversity. *Current Biology*, *22*(23), 2189–2202. https://doi.org/https://doi.org/10.1016/j.cub.2012.09.036

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(W1), W344–W350. https://doi.org/10.1093/nar/gkw408

Astorga, M. P. (2014). Genetic considerations for mollusk production in aquaculture: current state of knowledge. *Frontiers in Genetics*, *5*, 435. https://doi.org/10.3389/fgene.2014.00435

Avila, E. E. (2017). Functions of Antimicrobial Peptides in Vertebrates. *Current Protein & Peptide Science*, *18*(11), 1098–1119. https://doi.org/10.2174/1389203717666160813162629

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729–744. https://doi.org/10.1046/j.1365-294x.2003.02063.x

Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R., & Ben-Tal, N. (2004). ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics (Oxford, England)*, *20*(8), 1322–1324. https://doi.org/10.1093/bioinformatics/bth070

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Blier, P. U., Dufresne, F., & Burton, R. S. (2001). Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends in*

Genetics : TIG, *17*(7), 400–406. https://doi.org/10.1016/s0168-9525(01)02338-1

Bohovych, I., & Khalimonchuk, O. (2016). Sending Out an SOS: Mitochondria as a Signaling Hub. *Frontiers in Cell and Developmental Biology*, *4*, 109. https://doi.org/10.3389/fcell.2016.00109

Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, *27*(8), 1767–1780. https://doi.org/10.1093/nar/27.8.1767

Boore, J. L., & Brown, W. M. (1994). Complete DNA sequence of the mitochondrial genome of the black chiton, Katharina tunicata. *Genetics*, *138*(2), 423–443. https://doi.org/10.1093/genetics/138.2.423

Brandt, U. (2006). Energy converting NADH:quinone oxidoreductase (complex I). *Annual Review of Biochemistry*, *75*, 69–92. https://doi.org/10.1146/annurev.biochem.75.103004.142539

Brian, H., G., H. C. D., M., H. P., Michael, O., E., H. G., & A., B. C. (2002). Climate Change and Latitudinal Patterns of Intertidal Thermal Stress. *Science*, *298*(5595), 1015–1017. https://doi.org/10.1126/science.1076814

Bulet, P., Stöcklin, R., & Menin, L. (2004). Anti-microbial peptides: from invertebrates to vertebrates. *Immunological Reviews*, *198*, 169–184. https://doi.org/10.1111/j.0105-2896.2004.0124.x

Carvalho, A. de O., & Gomes, V. M. (2009). Plant defensins--prospects for the biological functions and biotechnological properties. *Peptides*, *30*(5), 1007–1020. https://doi.org/10.1016/j.peptides.2009.01.018

Cavieres, L., Arroyo, M. T. K., Peñaloza, A., Molina-Montenegro, M., & Torres, C. (2002). Nurse Effect of Bolax gummifera Cushion Plants in the Alpine Vegetation of the Chilean Patagonian Andes. *Journal of Vegetation Science*, *13*(4), 547–554. http://www.jstor.org/stable/3236739

Chen, C., Linse, K., Copley, J. T., & Rogers, A. D. (2015). The 'scaly-foot gastropod': a new genus and species of hydrothermal vent-endemic gastropod (Neomphalina: Peltospiridae) from the Indian Ocean. *Journal of Molluscan Studies*, *81*(3), 322–334. https://doi.org/10.1093/mollus/eyv013

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics (Oxford, England)*, *31*(16), 2745–2747. https://doi.org/10.1093/bioinformatics/btv195

da Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J., & Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, *9*, 119. https://doi.org/10.1186/1471-2164-9-119

De Smet, K., & Contreras, R. (2005). Human antimicrobial peptides: defensins, cathelicidins and histatins. *Biotechnology Letters*, *27*(18), 1337–1347. https://doi.org/10.1007/s10529-005-0936-5

Delano, W. L. (2002). The PyMOL Molecular Graphics System. *Http://Www.Pymol.Org*. https://ci.nii.ac.jp/naid/10020095229/en/

Dhar, D., & Dey, D. (2021). Comparison of Evolutionary Selection Acting on the Mitochondrial Protein-Coding Genes Between Intertidal and Deep-Sea Gastropods. *International Journal for Research in Applied Sciences and Biotechnology*, *8*(5 SE-Articles), 59–64. https://doi.org/10.31033/ijrasb.8.5.9

Dhar, D., Dey, D., Basu, S., & Fortunato, H. (2021). Insight into the adaptive evolution of mitochondrial genomes in intertidal chitons. *Journal of Molluscan Studies*, *87*(2). https://doi.org/10.1093/mollus/eyab018

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Fedosov, A. E., & Puillandre, N. (2012). Phylogeny and taxonomy of the Kermia–Pseudodaphnella (Mollusca: Gastropoda: Raphitomidae) genus complex: a remarkable radiation via diversification of larval development. *Systematics and Biodiversity*, *10*(4), 447–477. https://doi.org/10.1080/14772000.2012.753137

Felline, A., Seeber, M., & Fanelli, F. (2020). webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules. *Nucleic Acids Research*, *48*(W1), W94–W103. https://doi.org/10.1093/nar/gkaa397

Foote, A. D., Morin, P. A., Durban, J. W., Pitman, R. L., Wade, P., Willerslev, E., Gilbert, M. T. P., & da Fonseca, R. R. (2011). Positive selection on the killer whale mitogenome. *Biology Letters*, *7*(1), 116–118. https://doi.org/10.1098/rsbl.2010.0638

Fortunato, H. (2015). Mollusks: Tools in Environmental and Climate Research. *American Malacological Bulletin*, *33*(2), 310–324. https://doi.org/10.4003/006.033.0208

Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., & Ho, S. Y. W. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecology and Evolution*, *9*(7), 3891–3898. https://doi.org/10.1002/ece3.5015

Garvin, M. R., Bielawski, J. P., & Gharrett, A. J. (2011). Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. *PloS One*, *6*(9), e24127. https://doi.org/10.1371/journal.pone.0024127

Gerdol, M., Schmitt, P., Venier, P., Rocha, G., Rosa, R. D., & Destoumieux-Garzón, D. (2020). Functional Insights From the Evolutionary Diversification of Big Defensins . In *Frontiers in Immunology* (Vol. 11, p. 758). https://www.frontiersin.org/article/10.3389/fimmu.2020.00758

Gertz, E. M., Yu, Y.-K., Agarwala, R., Schäffer, A. A., & Altschul, S. F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biology*, *4*, 41. https://doi.org/10.1186/1741-7007-4-41

Gliński, Z., & Jarosz, J. (1997). Molluscan immune defenses. *Archivum Immunologiae et*

Therapiae Experimentalis, *45*(2–3), 149–155.

Gomes-dos-Santos, A., Lopes-Lima, M., Castro, L. F. C., & Froufe, E. (2020). Molluscan genomics: the road so far and the way forward. *Hydrobiologia*, *847*(7), 1705–1726. https://doi.org/10.1007/s10750-019-04111-1

Gonzalez-Romero, R., Suarez-Ulloa, V., Rodriguez-Casariego, J., Garcia-Souto, D., Diaz, G., Smith, A., Pasantes, J. J., Rand, G., & Eirin-Lopez, J. M. (2017). Effects of Florida Red Tides on histone variant expression and DNA methylation in the Eastern oyster Crassostrea virginica. *Aquatic Toxicology (Amsterdam, Netherlands)*, *186*, 196–204. https://doi.org/10.1016/j.aquatox.2017.03.006

Gueguen, Y., Bernard, R., Julie, F., Paulina, S., Delphine, D.-G., Franck, V., Philippe, B., & Evelyne, B. (2009). Oyster hemocytes express a proline-rich peptide displaying synergistic antimicrobial activity with a defensin. *Molecular Immunology*, *46*(4), 516–522. https://doi.org/10.1016/j.molimm.2008.07.021

Guerra, D., Bouvet, K., & Breton, S. (2018). Mitochondrial gene order evolution in Mollusca: Inference of the ancestral state from the mtDNA of Chaetopleura apiculata (Polyplacophora, Chaetopleuridae). *Molecular Phylogenetics and Evolution*, *120*, 233–239. https://doi.org/10.1016/j.ympev.2017.12.013

Guo, C., McDowell, I. C., Nodzenski, M., Scholtens, D. M., Allen, A. S., Lowe, W. L., & Reddy, T. E. (2017). Transversions have larger regulatory effects than transitions. *BMC Genomics*, *18*(1), 394. https://doi.org/10.1186/s12864-017-3785-4

Harris, J. (2008). Pacific oyster, Crassostrea gigas (thunberg, 1793). *Aquatic Invasive Species Profile. Aquatic Invasion Ecology*, 1–12.

Hassanin, A., Ropiquet, A., Couloux, A., & Cruaud, C. (2009). Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). *Journal of Molecular Evolution*, *68*(4), 293–310. https://doi.org/10.1007/s00239-009-9208-7

Helmuth, B., Choi, F., Matzelle, A., Torossian, J. L., Morello, S. L., Mislan, K. A. S., Yamane, L., Strickland, D., Szathmary, P. L., Gilman, S. E., Tockstein, A., Hilbish, T. J., Burrows, M. T., Power, A. M., Gosling, E., Mieszkowska, N., Harley, C. D. G., Nishizaki, M., Carrington, E., … Zardi, G. (2016). Long-term, high frequency in situ measurements of intertidal mussel bed temperatures using biomimetic sensors. *Scientific Data*, *3*, 160087. https://doi.org/10.1038/sdata.2016.87

Helmuth, B., Mieszkowska, N., Moore, P., & Hawkins, S. J. (2006). Living on the Edge of Two Changing Worlds: Forecasting the Responses of Rocky Intertidal Ecosystems to Climate Change. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 373–404. https://doi.org/10.1146/annurev.ecolsys.37.091305.110149

Huan, Y., Kong, Q., Mou, H., & Yi, H. (2020). Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields . In *Frontiers in Microbiology* (Vol. 11, p. 2559). https://www.frontiersin.org/article/10.3389/fmicb.2020.582779

Irisarri, I., Eernisse, D. J., & Zardoya, R. (2014). Molecular phylogeny of Acanthochitonina (Mollusca: Polyplacophora: Chitonida): three new mitochondrial genomes, rearranged gene orders and systematics. *Journal of Natural History*, *48*(45–48), 2825–2853. https://doi.org/10.1080/00222933.2014.963721

Irisarri, I., Uribe, J. E., Eernisse, D. J., & Zardoya, R. (2020). A mitogenomic phylogeny of chitons (Mollusca: Polyplacophora). *BMC Evolutionary Biology*, *20*(1), 22. https://doi.org/10.1186/s12862-019-1573-2

Jensen, M. M., & Denny, M. W. (2016). Life in an extreme environment: Characterizing wave-imposed forces in the rocky intertidal zone using high temporal resolution hydrodynamic measurements. *Limnology and Oceanography*, *61*(5), 1750–1761. https://doi.org/https://doi.org/10.1002/lno.10327

Jenssen, H., Hamill, P., & Hancock, R. E. W. (2006). Peptide antimicrobial agents. *Clinical Microbiology Reviews*, *19*(3), 491–511. https://doi.org/10.1128/CMR.00056-05

Jones, C. G., Lawton, J. H., & Shachak, M. (1994). Organisms as Ecosystem Engineers. *Oikos*, *69*(3), 373–386. https://doi.org/10.2307/3545850

Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., & Jorgensen, W. L. (2001). Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B*, *105*(28), 6474–6487. https://doi.org/10.1021/jp003919d

Kocot, K. M., Cannon, J. T., Todt, C., Citarella, M. R., Kohn, A. B., Meyer, A., Santos, S. R., Schander, C., Moroz, L. L., Lieb, B., & Halanych, K. M. (2011). Phylogenomics reveals deep molluscan relationships. *Nature*, *477*(7365), 452–456. https://doi.org/10.1038/nature10382

Kolesnikov, A. A., & Gerasimov, E. S. (2012). Diversity of mitochondrial genome organization. *Biochemistry. Biokhimiia*, *77*(13), 1424–1435. https://doi.org/10.1134/S0006297912130020

Kosakovsky Pond, S. L., & Frost, S. D. W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*(5), 1208–1222. https://doi.org/10.1093/molbev/msi105

Kosakovsky Pond, S. L., Poon, A. F. Y., Velazquez, R., Weaver, S., Hepler, N. L., Murrell, B., Shank, S. D., Magalis, B. R., Bouvier, D., Nekrutenko, A., Wisotsky, S., Spielman, S. J., Frost, S. D. W., & Muse, S. V. (2020). HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Molecular Biology and Evolution*, *37*(1), 295–299. https://doi.org/10.1093/molbev/msz197

Kouno, T., Fujitani, N., Mizuguchi, M., Osaki, T., Nishimura, S., Kawabata, S., Aizawa, T., Demura, M., Nitta, K., & Kawano, K. (2008). A novel beta-defensin structure: a potential strategy of big defensin for overcoming resistance by Gram-positive bacteria. *Biochemistry*, *47*(40), 10611–10619. https://doi.org/10.1021/bi800957n

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting

transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lai, Y., & Gallo, R. L. (2009). AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. *Trends in Immunology*, *30*(3), 131–141. https://doi.org/10.1016/j.it.2008.12.003

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*(6), 1695–1701. https://doi.org/10.1093/molbev/mss020

Lartigue, L., & Faustin, B. (2013). Mitochondria: metabolic regulators of innate immune responses to pathogens and cell stress. *The International Journal of Biochemistry & Cell Biology*, *45*(9), 2052–2056. https://doi.org/10.1016/j.biocel.2013.06.014

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*(2), 283–291. https://doi.org/https://doi.org/10.1107/S0021889892009944

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

Li, X.-D., Jiang, G.-F., Yan, L.-Y., Li, R., Mu, Y., & Deng, W.-A. (2018). Positive Selection Drove the Adaptation of Mitochondrial Genes to the Demands of Flight and High-Altitude Environments in Grasshoppers . In *Frontiers in Genetics* (Vol. 9, p. 605). https://www.frontiersin.org/article/10.3389/fgene.2018.00605

Loth, K., Vergnes, A., Barreto, C., Voisin, S. N., Meudal, H., Da Silva, J., Bressan, A., Belmadi, N., Bachère, E., Aucagne, V., Cazevielle, C., Marchandin, H., Rosa, R. D., Bulet, P., Touqui, L., Delmas, A. F., & Destoumieux-Garzón, D. (2019). The Ancestral N-Terminal Domain of Big Defensins Drives Bacterially Triggered Assembly into Antimicrobial Nanonets. *MBio*, *10*(5). https://doi.org/10.1128/mBio.01821-19

Luo, Y., Gao, W., Gao, Y., Tang, S., Huang, Q., Tan, X., Chen, J., & Huang, T. (2008). Mitochondrial genome analysis of Ochotona curzoniae and implication of cytochrome c oxidase in hypoxic adaptation. *Mitochondrion*, *8*(5–6), 352–357. https://doi.org/10.1016/j.mito.2008.07.005

Lv, C., Han, Y., Yang, D., Zhao, J., Wang, C., & Mu, C. (2020). Antibacterial activities and mechanisms of action of a defensin from manila clam Ruditapes philippinarum. *Fish & Shellfish Immunology*, *103*, 266–276. https://doi.org/https://doi.org/10.1016/j.fsi.2020.05.025

Mathiesen, C., & Hägerhäll, C. (2002). Transmembrane topology of the NuoL, M and N subunits of NADH:quinone oxidoreductase and their homologues among membrane-bound hydrogenases and bona fide antiporters. *Biochimica et Biophysica Acta*, *1556*(2–3), 121–132. https://doi.org/10.1016/s0005-2728(02)00343-2

McClellan, D. A., & Ellison, D. D. (2010). Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. *International Journal of Bioinformatics Research and Applications*, *6*(2), 120–133. https://doi.org/10.1504/IJBRA.2010.032116

Monlun, M., Hyernard, C., Blanco, P., Lartigue, L., & Faustin, B. (2017). Mitochondria as Molecular Platforms Integrating Multiple Innate Immune Signalings. *Journal of Molecular Biology*, *429*(1), 1–13. https://doi.org/10.1016/j.jmb.2016.10.028

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*(7), e1002764. https://doi.org/10.1371/journal.pgen.1002764

Nguyen, T. X., Cole, A. M., & Lehrer, R. I. (2003). Evolution of primate theta-defensins: a serpentine path to a sweet tooth. *Peptides*, *24*(11), 1647–1654. https://doi.org/10.1016/j.peptides.2003.07.023

Ning, T., Xiao, H., Li, J., Hua, S., & Zhang, Y. P. (2010). Adaptive evolution of the mitochondrial ND6 gene in the domestic horse. *Genetics and Molecular Research : GMR*, *9*(1), 144–150. https://doi.org/10.4238/vol9-1gmr705

Novoa, B., Romero, A., Álvarez, Á. L., Moreira, R., Pereiro, P., Costa, M. M., Dios, S., Estepa, A., Parra, F., & Figueras, A. (2016). Antiviral Activity of Myticin C Peptide from Mussel: an Ancient Defense against Herpesviruses. *Journal of Virology*, *90*(17), 7692–7702. https://doi.org/10.1128/JVI.00591-16

Okusu, A., Schwabe, E., Eernisse, D. J., & Giribet, G. (2003). Towards a phylogeny of chitons (Mollusca, Polyplacophora) based on combined analysis of five molecular loci. *Organisms Diversity & Evolution*, *3*(4), 281–302. https://doi.org/https://doi.org/10.1078/1439-6092-00085

Paladin, L., Piovesan, D., & Tosatto, S. C. E. (2017). SODA: prediction of protein solubility from disorder and aggregation propensity. *Nucleic Acids Research*, *45*(W1), W236–W240. https://doi.org/10.1093/nar/gkx412

Park, M. S., Kim, J. Il, Lee, I., Park, S., Bae, J.-Y., & Park, M.-S. (2018). Towards the Application of Human Defensins as Antivirals. *Biomolecules & Therapeutics*, *26*(3), 242–254. https://doi.org/10.4062/biomolther.2017.172

Parker, L. M., Ross, P. M., O'Connor, W. A., Pörtner, H. O., Scanes, E., & Wright, J. M. (2013). Predicting the response of molluscs to the impact of ocean acidification. *Biology*, *2*(2), 651–692. https://doi.org/10.3390/biology2020651

Parkhaev, P. (2007). The Cambrian 'basement' of gastropod evolution. *Geological Society London Special Publications*, *286*, 415–421. https://doi.org/10.1144/SP286.31

Parthiban, V., Gromiha, M. M., & Schomburg, D. (2006). CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Research*, *34*(suppl_2), W239–W242. https://doi.org/10.1093/nar/gkl190

Pires, D. E. V, Ascher, D. B., & Blundell, T. L. (2014a). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics (Oxford, England)*, *30*(3), 335–342. https://doi.org/10.1093/bioinformatics/btt691

Pires, D. E. V, Ascher, D. B., & Blundell, T. L. (2014b). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Research*, *42*(Web Server issue), W314–W319. https://doi.org/10.1093/nar/gku411

Prendergast, A. L., Azzopardi, M., O'Connell, T. C., Hunt, C., Barker, G., & Stevens, R. E. (2013). Oxygen isotopes from Phorcus (Osilinus) turbinatus shells as a proxy for sea surface temperature in the central Mediterranean: A case study from Malta. *Chemical Geology*, *345*, 77–86. https://doi.org/https://doi.org/10.1016/j.chemgeo.2013.02.026

Puchalski, S., Eernisse, D., & Johnson, C. (2008). The effect of sampling bias on the fossil record of chitons (Mollusca, Polyplacophora)*. *American Malacological Bulletin - AMER MALACOL BULL*, *25*. https://doi.org/10.4003/0740-2783-25.1.87

Rodrigues, C. H. M., Pires, D. E. V, & Ascher, D. B. (2021). DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Science*, *30*(1), 60–69. https://doi.org/https://doi.org/10.1002/pro.3942

Rogers, A. D., Tyler, P. A., Connelly, D. P., Copley, J. T., James, R., Larter, R. D., Linse, K., Mills, R. A., Garabato, A. N., Pancost, R. D., Pearce, D. A., Polunin, N. V. C., German, C. R., Shank, T., Boersch-Supan, P. H., Alker, B. J., Aquilina, A., Bennett, S. A., Clarke, A., … Zwirglmaier, K. (2012). The discovery of new deep-sea hydrothermal vent communities in the southern ocean and implications for biogeography. *PLoS Biology*, *10*(1), e1001234. https://doi.org/10.1371/journal.pbio.1001234

Romero, P. E., Weigand, A. M., & Pfenninger, M. (2016). Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life. *BMC Evolutionary Biology*, *16*(1), 164. https://doi.org/10.1186/s12862-016-0735-8

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542. https://doi.org/10.1093/sysbio/sys029

Rosa, R. D., Santini, A., Fievet, J., Bulet, P., Destoumieux-Garzón, D., & Bachère, E. (2011). Big defensins, a diverse family of antimicrobial peptides that follows different patterns of expression in hemocytes of the oyster Crassostrea gigas. *PloS One*, *6*(9), e25594. https://doi.org/10.1371/journal.pone.0025594

Rosenberg, G. (2014). A New Critical Estimate of Named Species-Level Diversity of the Recent Mollusca. *American Malacological Bulletin*, *32*(2), 308–322. https://doi.org/10.4003/006.032.0204

Ruesink, J. L., Lenihan, H. S., Trimble, A. C., Heiman, K. W., Micheli, F., Byers, J. E., & Kay, M. C. (2005). Introduction of Non-Native Oysters: Ecosystem Effects and Restoration Implications. *Annual Review of Ecology, Evolution, and Systematics*, *36*(1), 643–689. https://doi.org/10.1146/annurev.ecolsys.36.102003.152638

Saito, T., Kawabata, S., Shigenaga, T., Takayenoki, Y., Cho, J., Nakajima, H., Hirata, M., & Iwanaga, S. (1995). A novel big defensin identified in horseshoe crab hemocytes: isolation, amino acid sequence, and antibacterial activity. *Journal of Biochemistry*, *117*(5), 1131–1137. https://doi.org/10.1093/oxfordjournals.jbchem.a124818

Saraste, M. (1999). Oxidative phosphorylation at the fin de siècle. *Science (New York, N.Y.)*, *283*(5407), 1488–1493. https://doi.org/10.1126/science.283.5407.1488

Schmitt, P., Wilmes, M., Pugnière, M., Aumelas, A., Bachère, E., Sahl, H.-G., Schneider, T., & Destoumieux-Garzón, D. (2010). Insight into invertebrate defensin mechanism of action: oyster defensins inhibit peptidoglycan biosynthesis by binding to lipid II. *The Journal of Biological Chemistry*, *285*(38), 29208–29216. https://doi.org/10.1074/jbc.M110.143388

Shafee, T. M. A., Lay, F. T., Phan, T. K., Anderson, M. A., & Hulett, M. D. (2017). Convergent evolution of defensin sequence, structure and function. *Cellular and Molecular Life Sciences : CMLS*, *74*(4), 663–682. https://doi.org/10.1007/s00018-016-2344-5

Shen, Y.-Y., Liang, L., Zhu, Z.-H., Zhou, W.-P., Irwin, D. M., & Zhang, Y.-P. (2010). Adaptive evolution of energy metabolism genes and the origin of flight in bats. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19), 8666–8671. https://doi.org/10.1073/pnas.0912613107

Sigwart, J. D., Green, P. A., & Crofts, S. B. (2015). Functional morphology in chitons (Mollusca, Polyplacophora): influences of environment and ocean acidification. *Marine Biology*, *162*(11), 2257–2264. https://doi.org/10.1007/s00227-015-2761-2

Sigwart, J. D., & Sutton, M. D. (2007). Deep molluscan phylogeny: synthesis of palaeontological and neontological data. *Proceedings. Biological Sciences*, *274*(1624), 2413–2419. https://doi.org/10.1098/rspb.2007.0701

Smith, D. R., & Snyder, M. (2007). Complete mitochondrial DNA sequence of the scallop Placopecten magellanicus: evidence of transposition leading to an uncharacteristically large mitochondrial genome. *Journal of Molecular Evolution*, *65*(4), 380–391. https://doi.org/10.1007/s00239-007-9016-x

Sokolova, I. (2018). Mitochondrial Adaptations to Variable Environments and Their Role in Animals' Stress Tolerance. *Integrative and Comparative Biology*, *58*(3), 519–531. https://doi.org/10.1093/icb/icy017

Sowa, A., Krodkiewska, M., Halabowski, D., & Lewin, I. (2019). Response of the mollusc communities to environmental factors along an anthropogenic salinity gradient. *The Science of Nature*, *106*(11), 60. https://doi.org/10.1007/s00114-019-1655-4

Stoltzfus, A., & Norris, R. W. (2016). On the Causes of Evolutionary Transition:Transversion Bias. *Molecular Biology and Evolution*, *33*(3), 595–602. https://doi.org/10.1093/molbev/msv274

Sun, J., Zhang, Y., Xu, T., Zhang, Y., Mu, H., Zhang, Y., Lan, Y., Fields, C. J., Hui, J. H. L., Zhang, W., Li, R., Nong, W., Cheung, F. K. M., Qiu, J.-W., & Qian, P.-Y. (2017). Adaptation to deep-sea chemosynthetic environments as revealed by mussel genomes. *Nature Ecology & Evolution*, *1*(5), 121. https://doi.org/10.1038/s41559-017-0121

Sun, S., Hui, M., Wang, M., & Sha, Z. (2018). The complete mitochondrial genome of the alvinocaridid shrimp Shinkaicaris leurokolos (Decapoda, Caridea): Insight into the mitochondrial genetic basis of deep-sea hydrothermal vent adaptation in the shrimp. *Comparative Biochemistry and Physiology. Part D, Genomics & Proteomics*, *25*, 42–52. https://doi.org/10.1016/j.cbd.2017.11.002

Suresh, A., & Verma, C. (2006). Modelling study of dimerization in mammalian defensins. *BMC Bioinformatics*, *7 Suppl 5*(Suppl 5), S17. https://doi.org/10.1186/1471-2105-7-S5-S17

Takeuchi, T. (2017). Molluscan Genomics: Implications for Biology and Aquaculture. *Current Molecular Biology Reports*, *3*(4), 297–305. https://doi.org/10.1007/s40610-017-0077-3

Talavera, G., & Castresana, J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology*, *56*(4), 564–577. https://doi.org/10.1080/10635150701472164

Tassanakajon, A., Somboonwiwat, K., & Amparyup, P. (2015). Sequence diversity and evolution of antimicrobial peptides in invertebrates. *Developmental and Comparative Immunology*, *48*(2), 324–341. https://doi.org/10.1016/j.dci.2014.05.020

Teng, L., Gao, B., & Zhang, S. (2012). The first chordate big defensin: identification, expression and bioactivity. *Fish & Shellfish Immunology*, *32*(4), 572–577. https://doi.org/10.1016/j.fsi.2012.01.007

Tomasco, I. H., & Lessa, E. P. (2011). The evolution of mitochondrial genomes in subterranean caviomorph rodents: adaptation against a background of purifying selection. *Molecular Phylogenetics and Evolution*, *61*(1), 64–70. https://doi.org/10.1016/j.ympev.2011.06.014

Trumpower, B. L. (1990). The protonmotive Q cycle. Energy transduction by coupling of proton translocation to electron transfer by the cytochrome bc1 complex. *The Journal of Biological Chemistry*, *265*(20), 11409–11412.

Vakifahmetoglu-Norberg, H., Ouchida, A. T., & Norberg, E. (2017). The role of mitochondria in metabolism and cell death. *Biochemical and Biophysical Research Communications*, *482*(3), 426–431. https://doi.org/10.1016/j.bbrc.2016.11.088

Van Dover, C. L., German, C. R., Speer, K. G., Parson, L. M., & Vrijenhoek, R. C. (2002). Evolution and biogeography of deep-sea vent and seep invertebrates. *Science (New York,*

N.Y.), *295*(5558), 1253–1257. https://doi.org/10.1126/science.1067361

Veale, A. J., Williams, L., Tsai, P., Thakur, V., & Lavery, S. (2016). The complete mitochondrial genomes of two chiton species (Sypharochiton pelliserpentis and Sypharochiton sinclairi) obtained using Illumina next generation sequencing. *Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis*, *27*(1), 537–538. https://doi.org/10.3109/19401736.2014.905846

Vendrasco, M. J., Wood, T. E., & Runnegar, B. N. (2004). Articulated Palaeozoic fossil with 17 plates greatly expands disparity of early chitons. *Nature*, *429*(6989), 288–291. https://doi.org/10.1038/nature02548

Vinagre, C., Dias, M., Cereja, R., Abreu-Afonso, F., Flores, A. A. V, & Mendonça, V. (2019). Upper thermal limits and warming safety margins of coastal marine species – Indicator baseline for future reference. *Ecological Indicators*, *102*, 644–649. https://doi.org/https://doi.org/10.1016/j.ecolind.2019.03.030

VINTHER, J., JELL, P., KAMPOURIS, G., CARNEY, R., RACICOT, R. A., & BRIGGS, D. E. G. (2012). The origin of multiplacophorans – convergent evolution in Aculiferan molluscs. *Palaeontology*, *55*(5), 1007–1019. https://doi.org/https://doi.org/10.1111/j.1475-4983.2012.01180.x

Wallace, D. C. (2012). Mitochondria and cancer. *Nature Reviews. Cancer*, *12*(10), 685–698. https://doi.org/10.1038/nrc3365

Wang, G. (2012). Post-translational Modifications of Natural Antimicrobial Peptides and Strategies for Peptide Engineering. *Current Biotechnology*, *1*(1), 72–79. https://doi.org/10.2174/2211550111201010072

Wanninger, A., & Wollesen, T. (2018). The evolution of molluscs. *Biological Reviews of the Cambridge Philosophical Society*, *94*(1), 102–115. https://doi.org/10.1111/brv.12439

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky427

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*(9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Wirth, C., Brandt, U., Hunte, C., & Zickermann, V. (2016). Structure and function of mitochondrial complex I. *Biochimica et Biophysica Acta*, *1857*(7), 902–914. https://doi.org/10.1016/j.bbabio.2016.02.013

Woolley, S., Johnson, J., Smith, M. J., Crandall, K. A., & McClellan, D. A. (2003). TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics (Oxford, England)*, *19*(5), 671–672. https://doi.org/10.1093/bioinformatics/btg043

Xia, X. (2018). DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution. *Molecular Biology and Evolution*, *35*(6), 1550–1552.

https://doi.org/10.1093/molbev/msy073

Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, *26*(1), 1–7. https://doi.org/10.1016/s1055-7903(02)00326-3

Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, *101*(10), 2525–2534. https://doi.org/10.1016/j.bpj.2011.10.024

Xu, W., & Faisal, M. (2010). Defensin of the zebra mussel (Dreissena polymorpha): molecular structure, in vitro expression, antimicrobial activity, and potential functions. *Molecular Immunology*, *47*(11–12), 2138–2147. https://doi.org/10.1016/j.molimm.2010.01.025

Yang, D., Han, Y., Chen, L., Cao, R., Wang, Q., Dong, Z., Liu, H., Zhang, X., Zhang, Q., & Zhao, J. (2019a). A macin identified from Venerupis philippinarum: Investigation on antibacterial activities and action mode. *Fish & Shellfish Immunology*, *92*, 897–904. https://doi.org/10.1016/j.fsi.2019.07.031

Yang, D., Han, Y., Chen, L., Cao, R., Wang, Q., Dong, Z., Liu, H., Zhang, X., Zhang, Q., & Zhao, J. (2019b). A bactericidal permeability-increasing protein (BPI) from manila clam Ruditapes philippinarum: Investigation on the antibacterial activities and antibacterial action mode. *Fish & Shellfish Immunology*, *93*, 841–850. https://doi.org/10.1016/j.fsi.2019.08.050

Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv342

Yang, M., Gong, L., Sui, J., & Li, X. (2019). The complete mitochondrial genome of Calyptogena marissinica (Heterodonta: Veneroida: Vesicomyidae): Insight into the deep-sea adaptive evolution of vesicomyids. *PloS One*, *14*(9), e0217952. https://doi.org/10.1371/journal.pone.0217952

Yang, Z, Nielsen, R., Goldman, N., & Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, *155*(1), 431–449.

Yang, Z, & Yoder, A. D. (1999). Estimation of the transition/transversion rate bias and species sampling. *Journal of Molecular Evolution*, *48*(3), 274–283. https://doi.org/10.1007/pl00006470

Yang, Ziheng. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, Ziheng, Wong, W. S. W., & Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, *22*(4), 1107–1118. https://doi.org/10.1093/molbev/msi097

Yu, L., Wang, X., Ting, N., & Zhang, Y. (2011). Mitogenomic analysis of Chinese snub-nosed monkeys: Evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. *Mitochondrion*, *11*(3), 497–503.

https://doi.org/10.1016/j.mito.2011.01.004

Yuan, M.-L., Zhang, L.-J., Zhang, Q.-L., Zhang, L., Li, M., Wang, X.-T., Feng, R.-Q., & Tang, P.-A. (2020). Mitogenome evolution in ladybirds: Potential association with dietary adaptation. *Ecology and Evolution*, *10*(2), 1042–1053. https://doi.org/10.1002/ece3.5971

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, *415*(6870), 389–395. https://doi.org/10.1038/415389a

Zasloff, M. (2019). Antimicrobial Peptides of Multicellular Organisms: My Perspective. *Advances in Experimental Medicine and Biology*, *1117*, 3–6. https://doi.org/10.1007/978-981-13-3588-4_1

Zhang, B., Zhang, Y.-H., Wang, X., Zhang, H.-X., & Lin, Q. (2017). The mitochondrial genome of a sea anemone Bolocera sp. exhibits novel genetic structures potentially involved in adaptation to the deep-sea environment. *Ecology and Evolution*, *7*(13), 4951–4962. https://doi.org/10.1002/ece3.3067

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X., & Wang, G. T. (2020). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources*, *20*(1), 348–355. https://doi.org/https://doi.org/10.1111/1755-0998.13096

Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, *22*(12), 2472–2479. https://doi.org/10.1093/molbev/msi237

Zhang, K., Sun, J., Xu, T., Qiu, J.-W., & Qian, P.-Y. (2021). Phylogenetic Relationships and Adaptation in Deep-Sea Mussels: Insights from Mitochondrial Genomes. *International Journal of Molecular Sciences*, *22*(4), 1900. https://doi.org/10.3390/ijms22041900

Zhang, L., Yang, D., Wang, Q., Yuan, Z., Wu, H., Pei, D., Cong, M., Li, F., Ji, C., & Zhao, J. (2015). A defensin from clam Venerupis philippinarum: Molecular characterization, localization, antibacterial activity, and mechanism of action. *Developmental and Comparative Immunology*, *51*(1), 29–38. https://doi.org/10.1016/j.dci.2015.02.009

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*, 40. https://doi.org/10.1186/1471-2105-9-40

Zhang, Y., He, X., Li, X., Fu, D., Chen, J., & Yu, Z. (2011). The second bactericidal permeability increasing protein (BPI) and its revelation of the gene duplication in the Pacific oyster, Crassostrea gigas. *Fish & Shellfish Immunology*, *30*(3), 954–963. https://doi.org/10.1016/j.fsi.2011.01.031

Zhou, T., Shen, X., Irwin, D. M., Shen, Y., & Zhang, Y. (2014). Mitogenomic analyses propose positive selection in mitochondrial genes for high-altitude adaptation in galliform birds. *Mitochondrion*, *18*, 70–75. https://doi.org/10.1016/j.mito.2014.07.012

Zhu, S. (2008). Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CSalphabeta defensins. *Molecular Immunology*, *45*(3),

828–838. https://doi.org/10.1016/j.molimm.2007.06.354

Zhu, S., & Gao, B. (2013). Evolutionary origin of β-defensins. *Developmental and Comparative Immunology*, *39*(1–2), 79–84. https://doi.org/10.1016/j.dci.2012.02.011