DOMINANT VIRUSES AND MITOCHONDRIAL GENOME OF *Cladocora caespitosa*

by

İdil Çağla Ertaş

B.Sc. in Molecular Biology and Genetics, Boğaziçi University, 2020

Submitted to the Institute of Environmental Sciences

in partial fulfillment of the requirements for the degree of

Master of Science

in

Environmental Sciences

Boğaziçi University

2024

DOMINANT VIRUSES AND MITOCHONDRIAL GENOME OF *Cladocora caespitosa*

APPROVED BY:

Prof. Dr. İ. Raşit Bilgin . . . . . .
Thesis Advisor

Assoc. Prof. Dr. Ulaş Tezel . . . . . .

Assist. Prof. Dr. Emrah Çoraman . . . . . .

DATE OF APPROVAL: 11/01/2024

ACKNOWLEDGEMENTS

Foremost, I would like to express my gratitude to my thesis advisor Prof. Raşit Bilgin for his both professional and emotional support during this journey, even in the hardest times. I could not have done it without his guidance.

I also would like to thank my thesis committee, Assoc. Prof. Dr. Ulaş Tezel and Assist. Prof. Dr. Emrah Çoraman for their generous time and evaluation of this study.

I would like to thank Ünsal Karhan and Ziya Çaylarbaşı for their efforts in sampling coral colonies and making the field trip fun. I would like to thank my project partner Nehir Sevim for helping me in the lab and in handling coral colonies, he has been both a mentor and a friend throughout the project. I also would like to thank Elizabeth Somsen for her assistance in the project and for brightening up the lab during her internship in Bilgin Lab.

I would like to thank Bilgin Lab members: Merve Sıvacı, Maha Salameh, Anıl Örün, Onur Doğan, Ayşe Mergenci and Aygün Karaçay for their friendship and tolerating me whenever I complain.

Special thanks to another Bilgin Lab member Merjan Koşucu, my dearest friend, for her wisdom, friendship, and great support. She helped me stay sane during the toughest times, and I cannot thank her enough.

I would like to thank my dear partner Deniz Onat Öztürk, for his extreme emotional support during this journey, and listening to me talking about scientific topics he has no idea about.

Last but not least, I would like to thank my father Yusuf Ertaş, my mother Şükran Ertaş, and my sister Güneş Ertaş for their endless supports and being there for me all the time, even when I insisted on not searching for a paid job and keep on studying.

The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). This study was supported by a Boğaziçi University Research Fund grant (Project no: 20Y00M6) to RB.

ABSTRACT

DOMINANT VIRUSES AND MITOCHONDRIAL GENOME OF *Cladocora caespitosa*

Global warming leads to increased coral bleaching due to higher sea temperatures and increased pathogenic activity affecting coral communities. Together with rising temperatures, the pathogenic activity affecting coral communities also increases. However, the role of viruses in coral bleaching has been underestimated. Recent studies demonstrated the presence of virus-like particles (VLP) on the surface of bleached corals, indicating the need for further analysis to determine the impact of viruses on the bleaching phenomena. In this context, our objective was to detect the dominant viruses on *Cladocora caespitosa*, a coral species endemic to the Mediterranean basin, when exposed to a bacterial pathogen, namely *Vibrio coralliilyticus*, and heat stress. Thermally stressed and both thermally and pathogenically stressed colonies were bleached during the experiment, representing the detrimental impact of the mentioned stressors on coral health. The viral load of the control groups significantly differed from each other prior to the introduction of any stress factors, indicating the complex interactions within coral microbiota. The increased abundance of viruses in thermally and both thermally and pathogenically stressed corals demonstrates a positive correlation between the viral activity and stress. Interestingly, while the increased abundance of viruses was attributed to bacteriophages in thermally stressed corals, the abundance of eukaryotic viruses increased in both thermally and pathogenically stressed corals, presumably indicating a difference in the bleaching mechanisms observed in those tanks. In addition, using the sequenced reads of *C. caespitosa*, we conducted mitochondrial genome assembly of the species and undertook phylogeographic comparisons.

ÖZET

Cladocora caespitosa'NIN BASKIN VİRÜSLERİ VE MİTOKONDRİYAL GENOMU

Küresel ısınma, daha yüksek deniz sıcaklıkları ve mercan topluluklarını etkileyen patojenik aktivitenin artması nedeniyle mercan ağarmasının artmasına neden olmaktadır. Artan sıcaklıklarla birlikte mercan topluluklarını etkileyen patojenik aktivite de artmaktadır. Bununla birlikte, virüslerin mercan ağarmasındaki rolü üzerine yapılan çalışmalar yetersiz kalmaktadır. Yakın zamanda yapılan çalışmalar, ağarmış mercanların yüzeyinde virüs benzeri partiküllerin (VLP) varlığını göstermiş ve virüslerin ağarma fenomeni üzerindeki etkisini belirlemek için daha fazla analiz yapılması gerektiğini ortaya koymuştur. Bu bağlamda, amacımız Akdeniz havzasına endemik bir mercan türü olan *Cladocora caespitosa*'da, bakteriyel bir patojen olan *Vibrio coralliilyticus* ve ısı stresine maruz kaldığında ortaya çıkan baskın virüsleri tespit etmektir. Sıcaklık stresine maruz kalan ve hem sıcaklık hem de patojen stresine maruz kalan koloniler deney sırasında ağarmış ve bu da söz konusu stres faktörlerinin mercan sağlığı üzerindeki zararlı etkisini göstermiştir. Kontrol gruplarının viral yükü herhangi bir stres faktöründen bağımsız olarak birbirinden önemli ölçüde farklı çıkması mercan mikrobiyotası içindeki karmaşık etkileşimleri temsil etmektedir. Sıcaklık ve hem sıcaklık hem de patojen stresi altındaki mercanlarda virüs bolluğunun artması, viral aktivite ile stres arasında pozitif bir korelasyon olduğunu göstermektedir. İlginç bir şekilde, sıcaklık stresi altındaki mercanlarda artan virüs bolluğu bakteriyofajlara atfedilirken, ökaryotik virüslerin bolluğu hem sıcaklık hem de patojen stresi altındaki mercanlarda artmıştır, bu da muhtemelen bu tanklarda gözlemlenen ağarma mekanizmalarındaki farklılığa işaret etmektedir. Buna ek olarak, *C. caespitosa*'nın dizilenmiş okumalarını kullanılarak, türün mitokondriyal genom montajını gerçekleştirilmiş ve filocoğrafik karşılaştırmalar yapılmıştır.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....iii
ABSTRACT.....iv
ÖZET.....v
TABLE OF CONTENTS.....vi
LIST OF FIGURES.....viii
LIST OF TABLES.....xii
LIST OF SYMBOLS/ABBREVIATIONS.....xiv
1. INTRODUCTION.....1
1.1. The Effect of Global Warming on Marine Environments.....1
1.2. The Microbial Environment of Coral Reefs.....1
1.2.1. Coral Bleaching.....2
1.2.2. The Role of Viruses in Coral Microbiome.....3
1.3. Main Aim of The Study.....3
2. LITERATURE REVIEW.....5
2.1. Effect of Elevated Sea Temperatures on Coral Bleaching.....5
2.2. Importance of Microorganisms.....5
2.2.1. Coral Bleaching Caused by Pathogens.....6
2.2.2. Evidence for the Viral Activity.....7
2.3. The Habitat, Morphology and Ecology of *Cladocora caespitosa*.....8
2.3.1. The Distribution and Significance of *Cladocora caespitosa* in Turkey.....9
2.3.2. The Microbiome of *C. caespitosa*.....10
2.4. High-throughput Sequencing and Mitochondrial Genome Assembly.....10
3. MATERIALS AND METHODS.....12
3.1. Collection of Coral Colonies.....12
3.2. Experimental Set-up and Sampling.....14
3.3. Isolation of DNA.....15
3.4. DNA Fragmentation by Sonication.....17
3.5. Library Preparation and Sequencing.....18
3.6. Bioinformatics Analyses.....18
3.6.1. Read Normalization.....21
3.7. Statistical Analyses.....21
3.8. Mitochondrial Genome Assembly.....22

3.8.1. Haplotype Network ……..22
4. RESULTS.....24
4.1. Gel Electrophoresis of the Extracted DNA……24
4.2. Sonication for DNA Fragmentation.....……....25
4.3. Genome Alignment and Contig Assembly……..27
4.3.1. Quality Control..……..27
4.3.2. Removal of the Host Genome…...……..31
4.3.3. Contig Assembly…….……33
4.4. Taxonomic Assignment……..35
4.4.1. Application of 80% Similarity Threshold……..……..52
4.4.2. Nucleocytoplasmic Large DNA Viruses (NCLDVs)……..……58
4.5. Read Normalization …….....59
4.6. Statistical Analyses ……63
4.7. Viral-Host Search……...……67
4.8. Mitochondrial Genome Assembly……..68
4.8.1. Haplotype Network ……..……...73
5. DISCUSSION.....74
5.1. Comparison of Viral Loads……74
5.2. Detection of the Dominant Viral Families…….……76
5.3. Similarities Among Samples Based on Statistical Findings……79
5.4. Viral-Host Search……...80
5.5. Mitochondrial Genome Assembly and Haplotype Network…….……..81
6. LIMITATIONS OF THE STUDY AND RECOMMENDATIONS……83
7. CONCLUSIONS…….....85
REFERENCES……...86
APPENDIX A: THE DRAFT MITOCHONDRIAL GENOME OF *Cladocora caespitosa*……....96

LIST OF FIGURES

Figure 2.1. An image of *Cladocora caespitosa* ……..…….....9

Figure 3.1. The sampling site of healthy *C. caespitosa* in Mordoğan, İzmir.....12

Figure 3.2. The collected *C. caespitosa* colonies from Mordoğan placed in plastic buckets, avoiding their contact with air.…….……..……13

Figure 3.3. The experimental set-up consisting of four tanks, each containing a total of three replicas……..……..14

Figure 3.4. The timeline for the experiment. 10 days of acclimation period is followed by two samplings (Day 1 and Day 12)…….15

Figure 3.5. Steps followed during the bioinformatic analyses.……...19

Figure 4.1. The gel electrophoresis image of extracted DNA samples (left), and the gel electrophoresis image of the DNA isolates after treated with Clean & Concentrator Kit (right).……...……..24

Figure 4.2. The gel image of the field samples (MD10, MD19, MD20, MD23, MD25, respectively) from the first trial after sonication.……..…….26

Figure 4.3. The gel electrophoresis image after sonication of DNA samples (522, 822, 825, 525, 523, 323, 402) from the second trial…….……..26

Figure 4.4. The electropherogram plots of the four different sonicated DNA samples……...27

Figure 4.5. The sequence length distributions a) before and b) after the trimming step for the forward read of the sample A2_Nostr_T1…….31

Figure 4.6. Comparison between first and last samplings at realm level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold……..37

Figure 4.7. The viral composition representing the relative proportion of viral realms between samples……38

Figure 4.8. Comparison between first and last samplings at kingdom level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold……..39

Figure 4.9. The viral composition representing the relative proportion of viral kingdoms between samples……40

Figure 4.10. Comparison between first and last samplings at phylum level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold……41

Figure 4.11. The viral composition representing the relative proportion of viral phyla between samples……..42

Figure 4.12. Comparison between first and last samplings at class level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold……....43

Figure 4.13. The viral composition representing the relative proportion of viral classes between samples……..44

Figure 4.14. Comparison between first and last samplings at order level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold……45

Figure 4.15. The viral composition representing the relative proportion of viral orders between samples……..46

Figure 4.16. Comparison between first and last samplings at family level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold……48

Figure 4.17. The viral composition representing the relative proportion of viral families between samples……..49

Figure 4.18. The viral composition representing the relative proportion of viral genera between samples……..51

Figure 4.19. Comparison between first and last samplings at family level within each tank for the number of viral contigs that matched to the viral protein database with a stricter threshold of 80% similarity score……..53

Figure 4.20. Viral composition distribution after application of 80% threshold and data filtration at family level……54

Figure 4.21. The viral composition representing the relative proportion of viral genera between samples after the application of 80% threshold……..55

Figure 4.22. PCoA plots showing the clustering of viral families (a) before the data filtration and (b) after the data filtration process……...56

Figure 4.23. PCoA plots showing the clustering of viral genera (a) before the data filtration and (b) after the data filtration process……...57

Figure 4.24. The viral contig abundance of the members of Nucleocytoplasmic Large DNA viruses (NCLDV) in each sample……..58

Figure 4.25. Comparison between first and last samplings at family level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold after the read normalization step……61

Figure 4.26. The comparison of viral compositions at family level between samples after the read normalization process……62

Figure 4.27. Shannon diversity indexes for each sample at family level……63

Figure 4.28. Alpha diversities calculated for each sample at genus level based on Jaccard distances……....64

Figure 4.29. Beta-diversity heatmap that represents the similarities/dissimilarities between samples based on Jaccard distances...……..65

Figure 4.30. The dendrogram displaying the clustering of the samples based on the calculated average distances between the samples, using the Jaccard distance method……66

Figure 4.31. The PCoA plot showing the clustering of samples based on the dominant viral families that are constantly found on the surface of coral colonies……..……...67

Figure 4.32. Annotated mitogenome of *C. caespitosa* in sample 306 (A1_Con_T0)……..69

Figure 4.33. Annotated mitogenome of Favites abdita……...69

Figure 4.34. Circularized annotated mitogenome of *C. caespitosa* in sample 306 (A1_Con_T0)…..71

Figure 4.35. The haplotype network of the COX1 gene region of the assembled *C. caespitosa* mitogenomes (this study) and COX1 sequences of *C. caespitosa* available from RefSeq……73

LIST OF TABLES

Table 3.1. The concentrations of DNA isolates from the field samples in the first trial of sonication.17

Table 3.2. The list of experimental samples that are used in the second attempt of sonication……18

Table 4.1. The DNA concentrations of the samples…..……..25

Table 4.2. FastQC report outcomes. Total number of sequences, total base pairs, sequence lengths and GC% of the forward and reverse reads of all samples are presented……...……28

Table 4.3. FastQC report after the trimming and quality filtering step with fqtrim.……....30

Table 4.4. Number of reads after each step of contaminant genome removal……...……...32

Table 4.5. Alignment rates of seven samples after every step of genome removal calculated in bowtie2…….33

Table 4.6. The detailed output of contigs from MegaHit.……..…..……34

Table 4.7. Alignment rates of the assembled contigs to the quality trimmed reads…….34

Table 4.8. The total number of assembled contigs, assembled contigs that matched to the viral protein database and the viral contigs that matched to the viral protein database with 75% similarity threshold in all samples…….....35

Table 4.9. The detailed output of contigs from MegaHit after the read normalization step…….59

Table 4.10. The alignment rates calculated from the mapping of contigs after the read normalization to their trimmed reads they originated from...…….……60

Table 4.11. Comparison of number of contigs, number of viral contigs and number of viral contigs that has similarity value higher than 75%, between samples after data normalization.…60

Table 4.12. The abundance of viruses in each sample based on the type of organisms they infect…..68

Table 4.13. The presence-absence of annotated genes in *Cladocora caespitosa* and *Favites abdita*...70

Table 4.14. The read numbers and mapping statistics of mitogenome mapping process……72

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation
µl	Microliter
Cl	Chlorine
pH	Potential of Hydrogen
℃	Celsius
V	Volt

Abbreviation	Explanation
A	Adenine
ANOSIM	Analysis Of Similarities
bp	Base Pair
C	Cytosine
COX1	Cytochrome C Oxidase Subunit I
DNA	Deoxyribonucleic Acid
dsDNA	Double-Stranded Deoxyribonucleic Acid
EDTA	Ethylenediamine Tetraacetic Acid
E-value	Expect Value
FU	Fluorescence Units
G	Guanine
GC	Guanine-Cytosine
HPC	High Performance Computing
HS	High Sensitivity
HTS	High Throughput Screening
ICTV	International Committee on Taxonomy of Viruses
IUCN	International Union for Conservation of Nature
Gbp	Giga base pair
Gr	Gram
MEGA	Molecular Evolutionary Genetics Analysis

mM	Millimolar
ng	Nanogram
NCLDV	Nucleocytoplasmic Large DNA Virus
OGDRAW	Organellar Genome DRAW
ORF	Open Reading Frame
PCoA	Principal Coordinate Analysis
RNA	Ribonucleic Acid
SCUBA	Self-Contained Underwater Breathing Apparatus
T	Thymine
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
TEM	Transmission Electron Microscopy
Tris	Tris(hydroxymethyl)aminomethane
VBNC	Viable But Not Culturable
VLP	Virus Like Particle
WS	White Syndrome

1. INTRODUCTION

1.1. The Effect of Global Warming on Marine Environments

Global climate change causes severe disturbances on most natural ecosystems, especially when combined with anthropogenic pressures. The fluctuations within temperature induce alterations in timing of the reproductive activities, as well as geographic ranges of species (Pratchett et al., 2013). Relative abundance of species has significantly been altered due to these environmental pressures, and some species face the risk of extinction.

The delicate and intricate relationships that make up marine ecosystems are threatened by the released greenhouse gas that is primarily absorbed by the ocean (Hoegh-Guldberg and John F. Bruno, 2010). Apart from the rising sea levels seen in recent years, the oceans also serve as a sink for carbon dioxide emitted by human activity, which causes the ocean to become more acidic (Doney, 2009). The metabolism of marine organisms is directly impacted by the observed increases in water temperature, which controls population expansion. The primary production of the oceans has gradually decreased as a result of the changes in phytoplankton population abundance brought on by the warmer seas (Gregg et al., 2003). The entities that rely on the primary producers for nourishment have also been impacted by the fluctuations in the microbial community. For instance, coral reef ecosystems are quite prone to the fluctuations in temperature, such that even 1.0℃ increase in temperature significantly disrupts the symbiotic relationship between corals and algae, leading to the phenomena called coral bleaching (Pratchett et al., 2013).

1.2. The Microbial Environment of Coral Reefs

Coral reefs provide one of the most biodiverse ecosystems to numerous organisms in marine habitats. Many fish species inhabit or feed on coral reefs, some of these polyp-feeders are claimed to sustain 80% of their feeding on coral colonies, hence making them dependent on coral colonies (Ambalavanan et al., 2021). The rich biodiversity of coral reefs is not limited to eukaryotic multicellular organisms; in fact, the diverse microbiota of the coral colonies are the main sources of organic and inorganic compounds that attract various other marine species (Raina et al., 2010; Thurber et al., 2017). However, the elevated sea temperatures, mainly resulting from anthropogenic activities, have been shown to have an adverse impact on the coral ecosystem (Hughes et al., 2017).

The coral reefs have become more susceptible to these environmental perturbations, threatening both the health of the coral colonies and the organisms that are dependent on the coral reefs.

The primary components of the coral microbiota are the photosynthetic algae, which are responsible for supplying nutrients to their coral host. Therefore, the sustainability of this symbiotic relationship between corals and the photosynthetic algae is vital for coral health. The preservation of this relationship is thought to be influenced by the interactions of other microorganisms, such as bacteria. Numerous studies on the bacteriome characterization of coral colonies have indicated a shift in the bacterial composition when exposed to environmental stressors (Ziegler et al, 2017; Ainsworth et al., 2010; Vega Thurber et al., 2009). On the other hand, it has been suggested that changes in the composition of the coral microbiome could serve as an indicator of an adaptation strategy. Another possible role of microorganisms on the surface of coral reefs is the production of antimicrobial substances against pathogens, increasing the endurance of the coral colonies to environmental perturbations (ElAhwany et al., 2013).

1.2.1. Coral Bleaching

The primary cause of coral mortalities is the coral bleaching phenomena, which is the disruption of the symbiotic relationship between the coral and the photosynthetic algae on the surface of the coral (Goreau & Hayes, 1994; Morrow et al., 2018; Vezzulli et al., 2010). While elevated sea temperatures have been argued to be the main underlying cause behind this condition, the microbiome and virome analyses from the coral ecosystems have highlighted the complex interactions in coral microbiota that require further investigation in terms of their relation to the coral bleaching. Coral reefs are in a symbiotic relationship with algae, which provide the reefs with their beautiful colors (Ainsworth et al., 2010). When exposed to environmental stress factors, such as fluctuations within temperature or acidity, these algae leave the reef surface resulting in white coloration of the corals (Hughes et al., 2017). The loss of symbiotic algae from the surface of the corals is associated with lower nutrient uptake for the reef communities as well as being more susceptible to pathogens (Goreau & Hayes, 1994; Morrow et al., 2018; Vezzulli et al., 2010). It is a highly common condition recorded for many coral reef communities. In addition to its aesthetic appearance, the coral reef communities provide a living space for a wide variety of marine organisms from microorganisms to fish. The conservation of coral reef communities, especially the endangered ones, is therefore significant for both their ecological and economic contributions to the environment and humanity.

1.2.2. The Role of Viruses in Coral Microbiome

Although there have been various studies underlying the effect of the temperature increases in the coral bleaching phenomena, the effect of microorganisms is not completely understood in the coral bleaching process. However, the viruses are thought to be significant components of the coral microbiota. Banin et al. (2000) claimed that the challenge in determining the effect of microorganisms in coral bleaching process stems from the non-culturable characteristic of some pathogens, the state called as viable but not culturable (VBNC) (Banin et al., 2000). Evolutionarily dynamic characteristics of microorganisms, on the other hand, bear a significant potential in determining the origin of the disease as well as providing possible solutions (Ainsworth et al., 2010). Despite the unexplored role of viruses within the coral microbiome, the presence of viruses or Virus Like Particles (VLPs) have also been detected in coral ecosystems in several studies. One of the possible roles of viruses is their participation in nutrient cycles of marine ecosystems via viral shunt, in which viruses release organic and inorganic compounds that are not readily available for other microorganisms, through lytic activity (Ambalavanan et al., 2021).

1.3. Main Aim of the Study

The correlation between the coral bleaching incidents and increased temperatures clearly indicates the effect of global warming on marine ecosystems. The elevated sea temperatures not only affect the future of the coral reefs, but the entire ecosystem where marine organisms are dependent on these coral reefs. The economic and environmental contributions of these coral reef communities are remarkable, and conservation of these ecosystems is a necessity, considering the increased number of incidents related to coral bleaching or other coral diseases. To make provisions for the protection of the coral reefs, we need to first understand the causes and outcomes of coral bleaching comprehensively. Despite its obvious influence on coral bleaching, the 1992 incident of mass bleaching mentioned above demonstrates that the temperature by itself fails to provide a comprehensive explanation for the condition (Goreau & Hayes, 1994). Although the susceptibility of the coral ecosystem to fluctuations of temperature is undoubted, further investigations are needed in defining the root of the issue. The algal communities on the surface of corals and the interactions between these algae and other microorganisms possess an important potential in understanding the coral bleaching phenomena. The studies concerned with the microbial components of coral bleaching demonstrate that the microbial communities on the surface of the bleached corals dynamically change

in response to the environmental stress factors. Therefore, the studies focusing on the microbial perspective in the coral bleaching process would be much more precise.

Vibrio species occupy an important place in determining the pathogenicity related causes of the coral bleaching disease. Several studies demonstrated the causative agents of certain coral species as these pathogenic *Vibrio* species (Banin et al., 2000; Ben-Haim et al., 2003; Ben-Haim & Rosenberg, 2002; Morrow et al., 2018). Among these *Vibrio* species, *Vibrio coralliilyticus* has a noticeable characteristic of being temperature dependent (Ben-Haim et al., 2003; Ben-Haim & Rosenberg, 2002). Dependence of *Vibrio coralliilyticus* to the temperature enables the use of temperature as a controllable parameter for investigations on the issue. Therefore, the pathogen has a noteworthy potential in conducting experiments aiming to understand the actual mechanisms of coral bleaching. It also should be taken into consideration that the *Vibrio coralliilyticus* induced coral bleaching is contagious and this feature increases the importance of the pathogen for induction of bleaching.

The presence of VLPs on the bleached corals is an interesting finding, which requires further investigation. Weynberg's study (2015) indicates that the pathogenicity of especially *Vibrio coralliilyticus* may stem from a viral infection. Their findings represent a similarity between the *Vibrio cholerae* and *V. coralliilyticus*, claiming that in both cases the pathogenicity of the bacteria is induced by a viral infection. Their hypothesis is supported by the presence of similar genome sequences found in *V. coralliilyticus* to that of the bacteriophage infecting the *V. cholera* (Weynberg et al., 2015). Therefore, further analyses with *V. coralliilyticus* might be helpful for a better understanding of the causes of the coral bleaching.

The aim of this study is to contribute to a better understanding of the causalities and consequences of the coral bleaching disease that has been observed within several coral reef communities. Considering the significance of the microbial community in the process of coral bleaching, the pathogenic activity and diversity will be monitored for *Cladocora caespitosa*. The vulnerability of the species to heat stress and the presence of *Vibrio* species found on the bleached coral make this species an appropriate candidate for our study. The effect of increased temperature and exposure to coral pathogen of *Vibrio coralliilyticus* will be observed and the viral load of the exposed corals were aimed to be detected.

2. LITERATURE REVIEW

Since the coral reef communities are in a symbiotic relationship with microorganisms, it is important to investigate the coral bleaching process from a microbial perspective. Therefore, studies related with the microbial activities on coral bleaching process are required to be investigated, trying to correlate this microbial perspective with possible viral infections.

2.1. Effect of Elevated Sea Temperatures on Coral Bleaching

Several mass coral bleaching events have been recorded over the past 40 years. The incidence of these major bleaching cases coincides with the recorded elevated sea temperatures as a result of the global climate change since 1980s (Glynn, 1993; Hughes et al., 2017). Goreau and Hayes (1994) showed that, on a local or a larger scale, the locations of these bleaching events are in concordance with the rises in sea temperature, based on data provided from the satellite systems. However, the 1992 incident of mass coral bleaching has taken place in a period where the temperatures were below seasonal normals, questioning the temperatures being the major cause of the coral bleaching phenomena (Goreau & Hayes, 1994). In 2016, 60% of the coral reefs in Great Reef Barrier in Australia had experienced bleaching (Hughes et al., 2017). Coral reefs are highly sensitive to fluctuations within temperature. Although the increased sea temperatures are correlated with most of the coral bleaching incidents, the occurrence of bleaching cannot be explained completely with the elevated sea temperatures. Hoegh-Guldberg (1999) describes the study conducted by Goenaga et al. (1988) where the bleaching intensities differ in terms of their locations within the same coral reef colonies. Since the temperature is not expected to differ between the upper and lower parts of the same coral colony, elevated temperatures are not sufficient to determine the cause of the coral bleaching events (Hoegh-Guldberg, 1999).

2.2. Importance of Microorganisms

The symbiotic relationship between the corals and algae, referred to as the holobiont, and the interactions between the holobiont and microorganisms is essential in sustaining the reef ecosystem. The symbiotic relationship between corals and algae enables the nutrient uptake for the corals, therefore possible disruptions to this relationship threatens the health of the corals. Ainsworth et al. (2009) indicated that the evolutionarily and environmentally dynamic structure of microorganisms

might provide a variety of advantages for coral reefs (Ainsworth et al., 2010; Morrow et al., 2018). It has been argued that the microbial communities in the coral ecosystem could protect the corals by producing antibiotics (Ainsworth et al., 2010; Morrow et al., 2018).

Environmental stress factors might affect microbial communities. For instance, it has been shown that different compositions of microbial communities on the coral surface have been recorded in the different seasons (Morrow et al., 2018). The microbial communities that are related with a certain coral colony have also been shown to be highly specific to corals (Ainsworth et al., 2010). Hence, there is a unique interaction between the corals and their microbial associations, almost like a fingerprint (Ainsworth et al., 2010; Bourne et al., 2008). Ziegler et al. (2017) showed that the microbiome significantly differs between the heat-tolerant and heat-sensitive coral reef communities, hypothesizing that the shift observed in bacterial communities in certain coral species favors the adaptation of the species to the changing environment. The difference between the microbial communities indicates the potential of microbial communities in shedding light on the causalities of the coral bleaching phenomenon.

2.2.1. Coral Bleaching Caused by Pathogens

The effect of the environmental stress factors on the microbial communities eventually results in the increased vulnerability of the coral reefs against pathogens because virulence and the growth rates of the pathogens increases with thermal stress (Morrow et al., 2018; Vezzulli et al., 2010). It has been shown that the antimicrobial features of bleached corals decrease significantly, leaving the corals more susceptible to pathogens (Morrow et al., 2018). Kushmaro et al. (Banin et al., 2000; Morrow et al., 2018) demonstrated that the coral bleaching event that has been observed in *Oculina patagonica* is caused by a pathogen named *Vibrio shiloi*. The penetration of the bacteria into the coral epidermis induces the intoxication of the symbiotic algae on the surface, inhibiting the photosynthesis mechanism (Banin et al., 2000; Morrow et al., 2018). Ben Haim et al. (2002, 2003) identified a novel pathogen affecting the *Pocillopora damicornis* and this pathogen has been named as *Vibrio coralliilyticus*. Their study demonstrated that the infection of the *P. damicornis* by this novel bacterium was dependent on the temperature and the pathogen showed an increased lysogenic activity directly proportional to the temperature (Ben-Haim et al., 2003; Ben-Haim & Rosenberg, 2002). Thermal dependence of the *Vibrio* species draws attention to the significance of pathogens in the process of coral bleaching in the face of global warming (Bally & Garrabou, 2007). The data also indicated that the coral bleaching induced by *Vibrio coralliilyticus* was contagious among corals (Ben-Haim & Rosenberg, 2002).

2.2.2. Evidence for the Viral Activity

Although the significance of the microbial communities in coral bleaching process is clear, studies investigating the role of the viruses in coral bleaching are quite limited. However, several studies demonstrated that there was evidence for viral infections to be a cause for the coral bleaching phenomena. In a study conducted by Correa et al. (2016), transmission electron microscopy analyses indicated the presence of virus-like particles (VLP) on the surface of bleached corals from Heron Island, and the bleaching event was due to the heavy rainfalls (Correa et al., 2016). Similar results were obtained for *Porites compressa* and *Acropora hyacinthus*, where the presence of VLP was observed in bleached corals but not in healthy corals (Lawrence et al., 2014; Pollock et al., 2014).

The susceptibility of coral colonies against pathogens has also been suggested to be related to viral activity. Weynberg et al. (2015) hypothesized that the pathogenicity of the *Vibrio* species might be induced through phage activity, claiming that toxigenic activity of the pathogens may be activated by viral infections of these pathogens (Weynberg et al., 2015). Together with environmental perturbations, the importance of viruses on coral health has become prominent. For instance, Nguyen-Kim et.al. (2015) demonstrated that the heat-stress (+4°C) introduced coral colonies were bleached in 11 days, and the microbial communities measured from the mucus of the corals revealed an increase diversity of the bacterial communities together with the increased abundance of viruses. The mucus of the stressed corals was demonstrated to be rich in protein, lipid content and antimicrobial activity, possibly as a defending mechanism against pathogens (Wright et al., 2019). Nguyen-Kim et al. argue that the increase observed in viral load might be due to the increase in mucosal activity in coral colonies, resulting in an increase in the nutrients in the environment that possibly attract microorganisms. This finding represents the complexity of the interactions within the microbiome of coral colonies.

Messyasz et al. (2020) demonstrated that the observed VLPs on the surface of the corals or their symbionts are morphologically similar to the Nucleocytoplasmic Large DNA Viruses (NCLDV), based on their sizes (Messyasz et al., 2020). The study further focused on the characterization of these VLPs with the help of metagenomic analyses and custom bioinformatics. Assembling and sorting the contigs obtained from the sequencing, significant shift between the viral load of bleached and non-bleached corals has been elucidated. The microbiome of the non-bleached corals mainly consisted of phages, whereas the microbiome of the bleached corals was dominated by eukaryotic viruses. On the other hand, this remarkable difference between the viral load of bleached and non-bleached corals were difficult to interpret due to the limited amount of information that exist for both the corals and

the viruses living on them (Messyasz et al., 2020). Therefore, further studies are required to clarify the role of viruses in the process of coral bleaching.

2.3. The Habitat, Morphology and Ecology of *Cladocora caespitosa*

C. caespitosa, depicted in Figure 2.1, is an endemic species to the Mediterranean Sea, and a member of the Faviidae family. It is found in locations with calm water or strong currents, on stony bottoms, and down to about 40 m below the surface, where there is sufficient light for the symbiotic zooxanthellae to perform photosynthesis (Kružić and Benković, 2008). Shallow colonies that resemble bushes either occur alone or in groups to form "banks" or "beds," which are substantial formations that are several square meters in size and more than one meter high (Aguirre and Jiménez, 1997). *Cladocora caespitosa* is a zooxanthellate, colonial coral with the ability to construct large formations. Its morphology and physiology are similar to those of other corals that build tropical reefs. Since these varieties of *C. caespitosa* do not reach the water's surface, Schuhmacher (1985) classified them as constructional and formally ahermatypic, non-reef-building.

The seasonal temperature fluctuations observed in the Mediterranean Sea, varying from 8 °C in the winter to 28 °C in the summer, have been well-adapted to by *Cladocora caespitosa* (Rudolfo-Metalpa et al., 2006). Nonetheless, during positive thermal anomalies, several death events, including *C. caespitosa*, happened in the Adriatic and Ligurian Seas (Kružić and Požar-Domac, 2003). The exact environmental component that impacts skeletal growth is still unknown. However, Montagna et al. (2007) have reported seasonal fluctuations in *C. caespitosa* linear growth rates, correlated with light, temperature and nutrient availability.



Figure 2.1. An image of *Cladocora caespitosa.*

2.3.1. The Distribution and Significance of *Cladocora caespitosa* in Turkey

In Turkey, the *C. caespitosa* populations have been recorded in Aegean Sea at Gökçeada island (Öztürk, 2004), in Edremit Bay (Çınar, 2003), and in the Marmara Sea (Özalp and Alparslan, 2011). It is currently included in IUCN Red List as an endangered species (Chefaoui et al., 2017). C. caespitosa is considered as the remnant of the ancient tropical reefs that became extinct after the Messinian Salinity Crisis (Chefaoui et al., 2017; Kersting and Linares, 2012). Fluctuations in temperature have been shown to be linked to the recurrent mortality events in *C. caespitosa* (Rubio-Portillo et al., 2018). However, thermal anomalies themselves do not provide a satisfactory explanation for the bleaching incidence observed in this species. For instance, Rubio-Portillo et al. (2018) demonstrated that the coral bacterial associated community differs between healthy and bleached corals, pointing out the possible role of the microbial community in the process of coral

bleaching (Rubio-Portillo et al., 2018). The presence of the coral pathogen *Vibrio coralliilyticus* has also been observed in the bleached corals (Rubio-Portillo et al., 2018). These studies show that the bleaching incidences of *C. caespitosa* might be the result of a combined stress factor induced by thermal anomalies and pathogenic activities.

2.3.2. The Microbiome of *C. caespitosa*

The coral microbiome is thought to consist of three primary components, according to Hernandez-Agreda et al. (2017): (i) core microbiome; (ii) a dynamic site and/or species-specific community; and (iii) a highly changeable community formed in response to biotic and abiotic oscillations. Rubio-Portillo et al. (2018) demonstrated that the core microbiome of *C. caespitosa* includes the coral pathogen *Vibrio coralliilyticus*, which is thought to participate in nutrient cycling in the coral holobiont. The tissue necrosis in *C. caespitosa* was observed in colonies that are found in places with high environmental perturbations, and the pathobiome of these colonies consisted of *V. coralliilyticus* and *V. mediterranei*. *C. caespitosa* tissue necrosis is likely to have a complex etiology, with rising sea temperatures being the primary cause (Kersting et al., 2013). Therefore, the coral microbial community's contribution to necrosis events may also be strongly influenced by regional environmental factors and human activity.

2.4. High-throughput Sequencing and Mitochondrial Genome Assembly

The intricate dynamics within microbial communities have been revealed by high-resolution analyses with the development in High Throughput Screening (HTS) techniques (Pérez-Losada et al., 2020). These discoveries have advanced the study of the coral microbiome and advanced our understanding of the relationships between microorganisms and corals. HTS has a number of benefits that are essential for monitoring microbiome changes. Comparing it to conventional sequencing methods, it generates a fraction of the cost while producing an enormous amount of data. This increases statistical power and lowers the possibility of mistakes by enabling researchers to replicate their studies with a larger number of samples. One common HTS technique is amplicon sequencing, which amplifies particular gene sequences, such as the 16S rRNA gene, to facilitate accurate phylogenetic analysis (Callahan et al., 2019). Numerous HTS systems, including Miseq and NovaSeq from Illumina, allow for flexibility in read length and depth to meet the unique needs of each study. The integration of HTS methods into viral ecology field is particularly useful, due to the highly

dynamic and infectious characteristics of viruses. Most viruses are not culturable in lab environment as well. Hence, the application of HTS techniques makes it easier to study the field of virology.

In molecular ecology, population genetics, evolutionary biology, and phylogeographic and phylogenetic studies of animals, mitochondrial DNA is a traditionally valuable and preferred marker. Mitochondrial DNA is frequently the molecular marker of choice due to its rapid rate of mutation, absence of recombination, maternal inheritance, high copy number, and relatively easy accessibility (Hahn et al., 2013). Initially, only relatively short portions of the mitochondria were targeted, but as technique improved, sequencing entire mitochondrial genomes became increasingly routine, especially when examining challenging templates like ancient DNA. When trying to find answers to long-standing puzzles about human evolution, or when delving into the evolutionary histories of mysterious animals like polar bears, mammoths, aurochs, or Tasmanian tigers, whole mitochondrial genomes are very helpful (Hahn et al., 2013).

3. MATERIALS AND METHODS

3.1. Collection of Coral Colonies

Healthy *Cladocora caespitosa* samples were collected from Mordoğan/Izmir (Figure 3.1) in Turkey by SCUBA diving from the depth range of 3 to 5 meters. Several different coral colonies were selected from the same region, and they were placed in white plastic buckets under the water as they were collected. The buckets containing the coral colonies were then placed carefully into a cooler tank filled with seawater and aquarium gravel on the bottom (Figure 3.2). The coral colonies were then immediately transferred into the aquarium set-up prepared prior to the arrival of the coral samples.

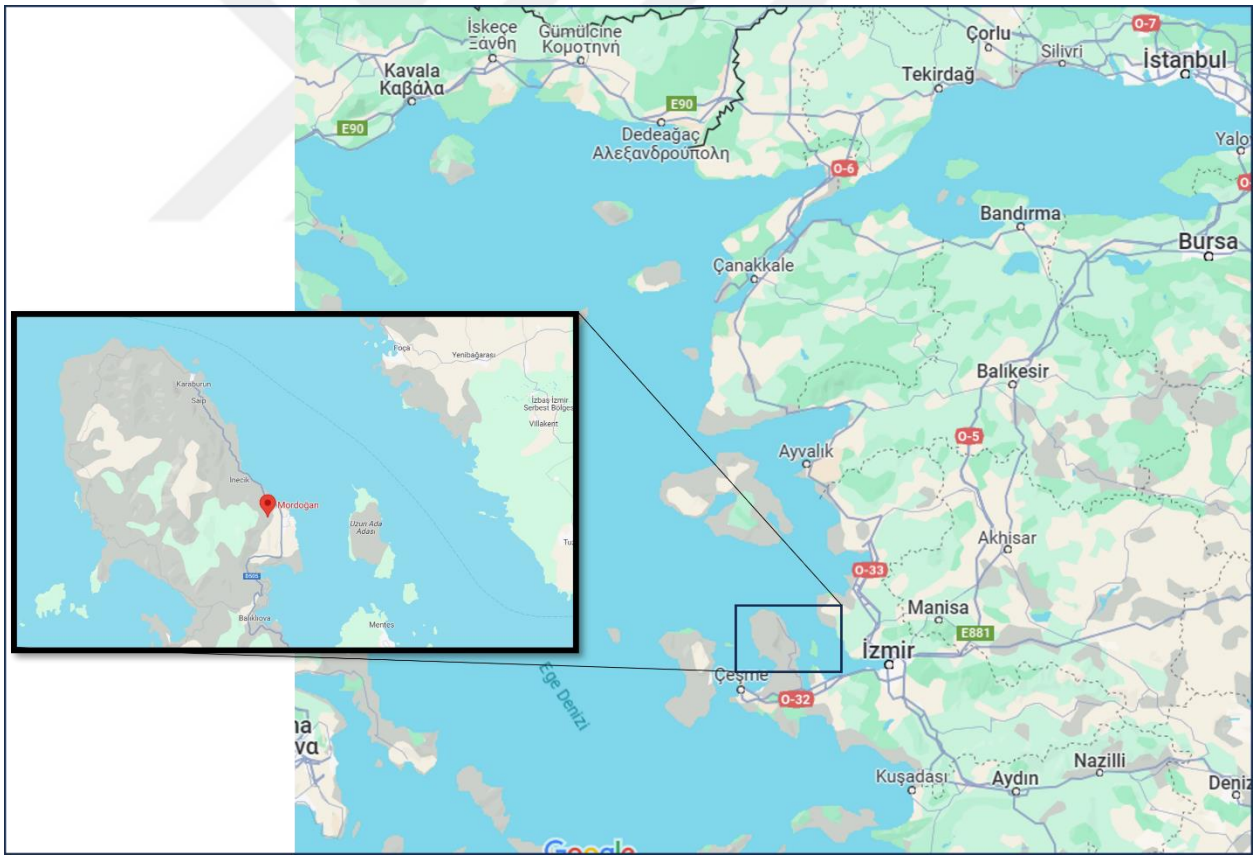


Figure 3.1. The sampling site of healthy *C. caespitosa* in Mordoğan, İzmir.

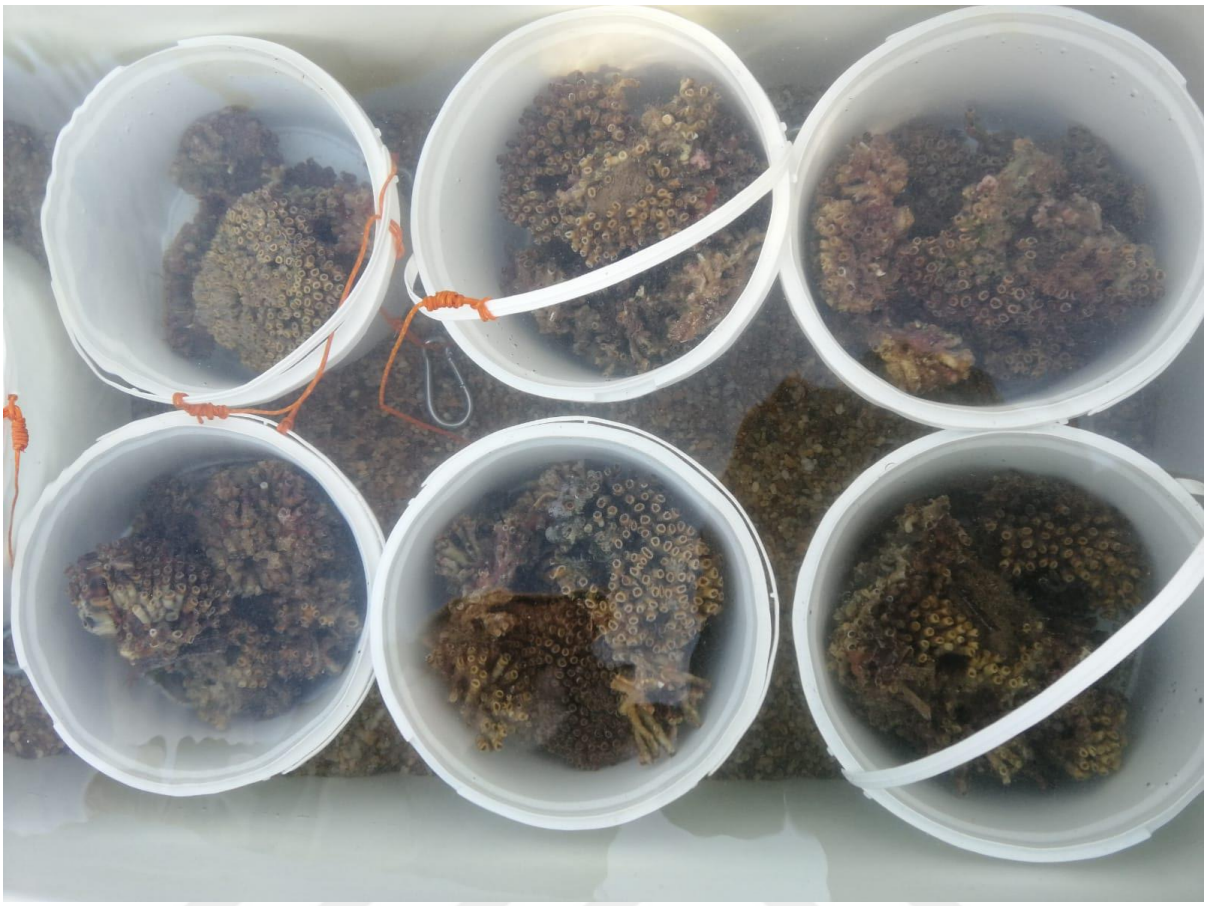


Figure 3.2. The collected *C. caespitosa* colonies from Mordoğan placed in plastic buckets, avoiding their contact with air.

Multiple polyp branches were separated from the colonies and placed in 15 ml falcon tubes, and the tubes were filled with RNA-Later to preserve the coral tissue. The separated field samples were placed in a cool pack and immediately transferred to -80°C for storage, after arrival to the laboratory.

3.2. Experimental Set-up and Sampling

The aquarium setup comprises four tanks, each representing distinct environmental stress conditions. Within each tank, there were three smaller aquariums serving as replicates, as illustrated in Figure 3.3.

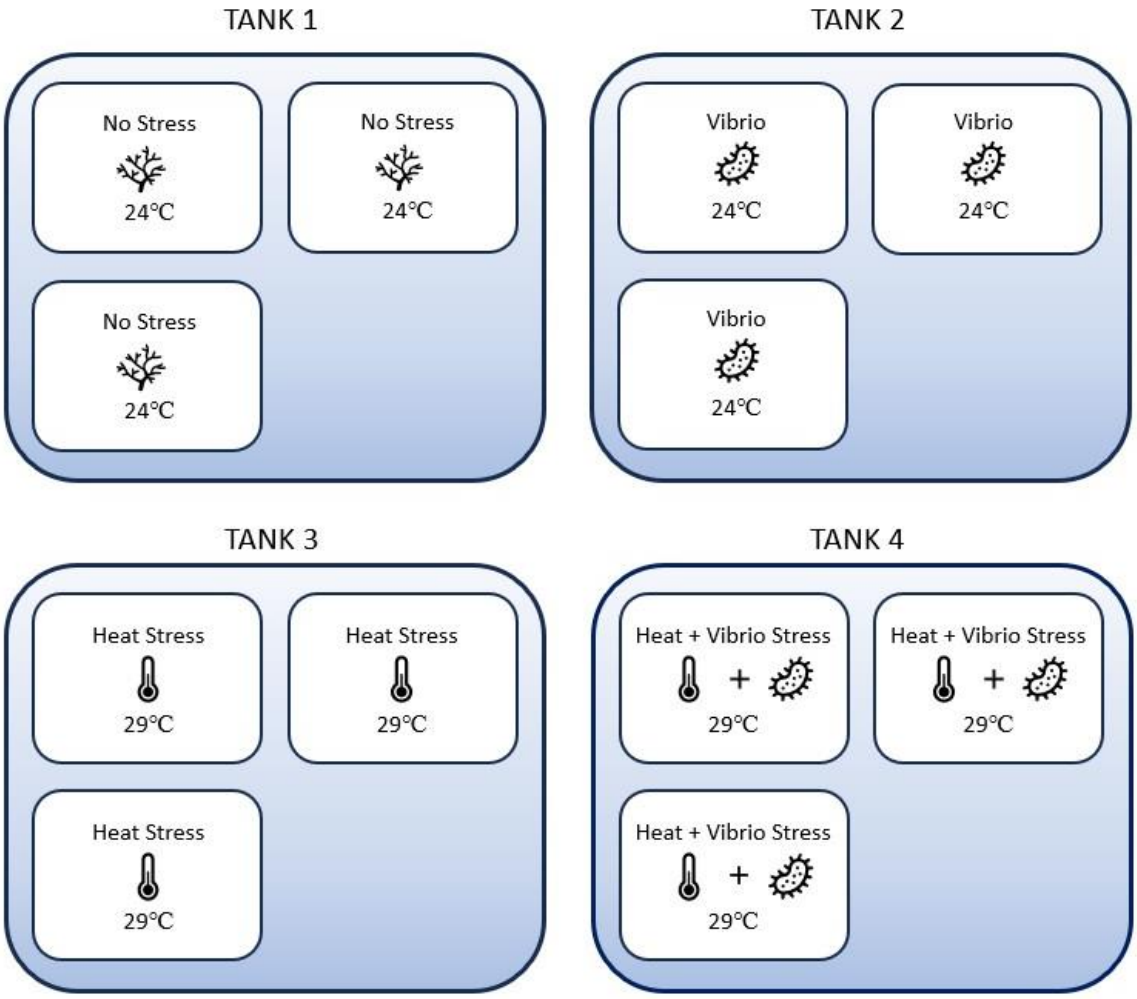


Figure 3.3. The experimental set-up consisting of four tanks, each containing a total of three replicas. The coral colonies in Tank 1 represent no stress condition at 24°C. The corals in Tank 2 were introduced to *V. coralliilyticus* at 24°C. Tank 3 colonies were subjected to heat stress at 29°C and colonies in Tank 4 were exposed to both heat stress at 29°C and *V. coralliilyticus*.

In these aquariums, individual coral colonies were carefully planted, and an acclimation phase lasting 10 days at 24°C was initiated, as depicted in Figure 3.4. Following this acclimation period, the first round of sampling was conducted, involving the collection of 10 polyps from each of the 12 colonies, establishing a baseline assessment.

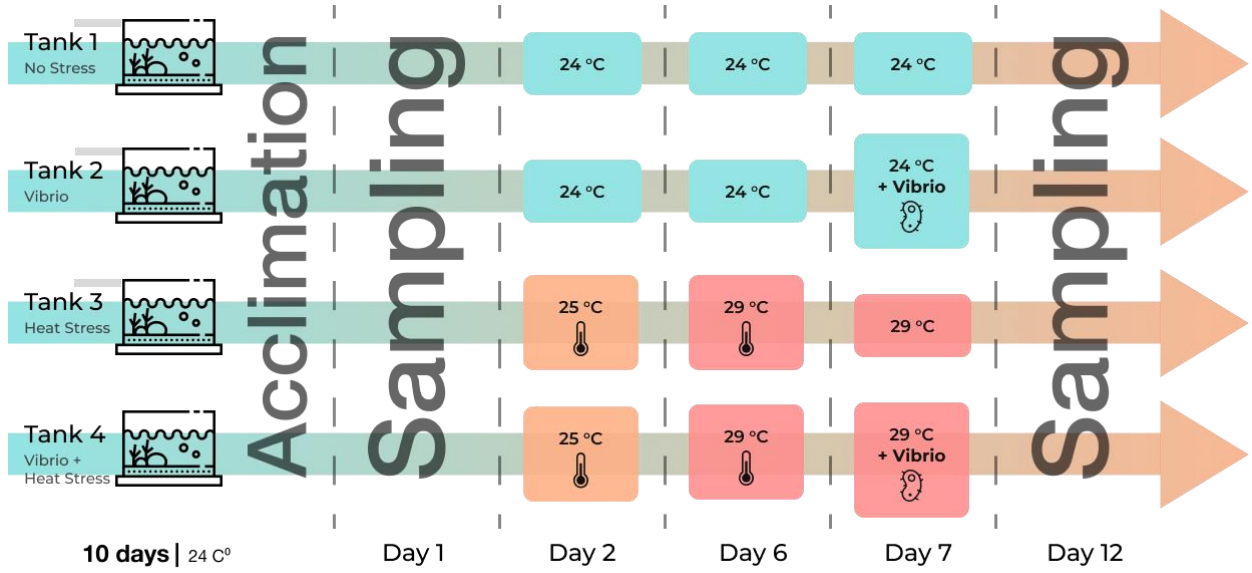


Figure 3.4. The timeline for the experiment. 10 days of acclimation period is followed by two samplings (Day 1 and Day 12). The temperature gradually increased from 24°C to 29°C in Tank 3 and Tank 4. *Vibrio coralliilyticus* was introduced to Tank 2 and Tank 4 at Day 7.

Subsequently, the introduction of environmental stress conditions commenced. In Tank 1, no additional environmental stressors were imposed on the coral colonies throughout the 12-day duration of the experiment. In Tank 2, on the other hand, the colonies were exposed exclusively to the coral pathogen *Vibrio coralliilyticus* while maintaining a constant temperature of 24°C. The coral colonies in Tank 3 were subjected to heat stress, with the temperature gradually increasing each day until reaching 29°C. Tank 4 presented a combination of both vibrio pathogen and heat stress. The introduction of heat stress in Tank 3 and Tank 4 commenced on the second day of the experiment, with the temperature rising by 1°C each day for the subsequent four days until it reached 29°C. When the temperature reached 29°C in Tank 3 and Tank 4, *Vibrio* was introduced to both Tank 2 and Tank 4. The final sampling occurred five days after the introduction of *Vibrio*, on Day 12. During this sampling, the collected polyps were carefully placed in Eppendorf tubes containing RNA-later and were promptly transferred to -80°C for storage. A total of 240 polyps were collected.

3.3. Isolation of DNA

For DNA extractions, 72 polyps, comprising three polyps from each aquarium for each of the two sampling points, were utilized. Prior to the DNA extraction, the coral polyps within the Eppendorf tubes were subjected to a tissue lysing process. Initially, the polyps were crushed with a plier and a tweezer inside the tubes. Both the plier and the tweezer were sterilized with 70% ethanol and flame, before proceeding to the next sample. Subsequently, the metal beads were added to the tubes and the

tubes were placed in TissueLyser LT (Qiagen, Hilden, Germany) for further tissue lysing with oscillation for 2 minutes.

For the lysis of the coral tissue, 37.5 µl Proteinase K (Zymo Research, Irvine, CA, USA) and 75 µl Proteinase Digestion Buffer (Zymo Research, Irvine, CA, USA) were added to each tube with a total volume of 700 µl. The tubes were then incubated for 1 hour. Following the incubation period, genomic content of each lysed sample was isolated using the Quick-DNA/RNA™ Miniprep Plus Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. During the elution step of the DNA extraction protocol, Tris-EDTA (TE) Buffer (8 mM Tris-Cl, 0.8 mM EDTA, pH 8.15) was used.

The quality of the extracted DNA was monitored with gel electrophoresis. 3 µl volume of DNA was mixed with 3 µl Loading Dye, and run on 1% agarose gel. For the preparation of the agarose gel, 0.4 gr of agarose and 40 ml TAE Buffer was mixed and heated for 2 minutes in the Microwave. 4 ml of RedSafe (iNtRON Biotechnology, Burlington, USA) dye was added to the mixture and the gel was poured into the cassette with comb, until the gel was solidified. The gel tank was connected to the power supplier at 120V for 30 minutes. The resulting gel images were photographed under blue light.

The DNA isolates were later treated with DNA Clean & Concentrator®-5 Kit (Zymo Research, Irvine, CA, USA) for additional purification and concentration of the isolated DNA, following the manufacturer's instructions. Purified DNA isolates were stored at -20°C. 1 µl of DNA from samples were again run on 1% agarose gel for quality monitoring purposes.

The concentrations of the DNA samples were measured with Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA) was utilized during measurements, according to the manufacturer's instructions. The DNA concentrations were recorded. However, the concentrations of some samples were above the concentration range of Qubit Fluorometer. These samples were diluted and measured again. Dilution was performed by mixing 8 µl of water with 4 µl of DNA isolate for each of the samples. Consequently, 10 µl of DNA isolates in Eppendorf tubes labeled with sample codes were prepared for commercial DNA library preparation and Illumina sequencing by Gen-Era (İstanbul, Türkiye), for each of the 72 samples. The prepared samples were stored in -20°C until shipment.

3.4. DNA Fragmentation by Sonication

Our initial plan involved using the ThruPLEX DNA-Seq Kit (Takara Bio, USA) to generate a genomic library from the isolated DNA samples. In accordance with the kit manual's instructions, we aimed to achieve initial DNA fragments with lengths falling within the range of 300-500 base pairs. To achieve this target range, we attempted sonication as the fragmentation method for the isolated DNA samples. Each sample was prepared in a volume of 30 µl.

The first trial of sonication was performed with non-experimental field samples. These samples consisted of branches from a coral colony, each containing several polyps. They were carefully placed in Falcon tubes filled with RNA-shield solution, awaiting further processing. To optimize our sonication procedure, we tested samples with varying concentrations and exposure times. Specifically, the MD19, MD20, and MD25 samples, as indicated in Table 3.1, were subjected to sonication for a total duration of 44 minutes. On the other hand, the MD10 and MD23 samples were subjected to sonication for a total of 64 minutes. The efficiency of the sonication was visualized by gel electrophoresis by running the fragmented DNA isolates on 1% agarose gel.

Table 3.1. The concentrations of DNA isolates from the field samples in the first trial of sonication.

Sample Code	Sample Type	DNA conc. (ng/µl)
MD 10	Field sample	29.1
MD 19	Field sample	26
MD 20	Field sample	15.7
MD 23	Field sample	8
MD 25	Field sample	13.1

Since the desired target DNA size was not achieved during the initial sonication run, a second round of sonication was carried out using the experimental samples. To ensure consistency, all the samples were carefully diluted to a concentration of 2 ng/µl, effectively controlling the concentration parameter. These samples, as listed in Table 3.2, were prepared with a total volume of 52.5 µl. The sonication process for these samples lasted for a total of 44 minutes.

Table 3.2. The list of experimental samples that are used in the second attempt of sonication.

Sample Code	Sample Type	DNA conc. (ng/µl)
522	Experimental sample	2
822	Experimental sample	2
825	Experimental sample	2
525	Experimental sample	2
523	Experimental sample	2
323	Experimental sample	2
402	Experimental sample	2

It was taken into consideration that samples with low concentrations may not be easily detectable on a gel. Subsequently, Agilent BioAnalyzer service (SEM, Istanbul, Turkey) was purchased for detailed analysis of the quality of the genomic content within samples listed in Table 3.2. The resulting electropherograms were investigated.

3.5. Library Preparation and Sequencing

Following the unsuccessful optimization attempts for DNA fragmentation via sonication, the decision was made to send the DNA isolates directly to GenEra (İstanbul, Turkey) for the purpose of library preparation and subsequent Illumina sequencing. For each sample a total of 20M 150 bp paired end reads were requested, by sequencing on the Illumina Nova-Seq platform.

3.6. Bioinformatic Analyses

Prior to the bioinformatic analyses described below, the quality and the basic statistics of the sequencing performance were controlled for the raw reads using FastQC tool (Andrews, 2010). The bioinformatic workflow that is followed is depicted in the scheme below, Figure 3.5. The raw paired-end DNA reads were trimmed for removing adaptors and indexes. Once the quality of the reads was assured, the possible contaminant genomes were discarded from the trimmed reads. The reads were next assembled into contigs, and the Open Reading Frames (ORFs) were annotated. The resulting ORFs were finally utilized for homology search.

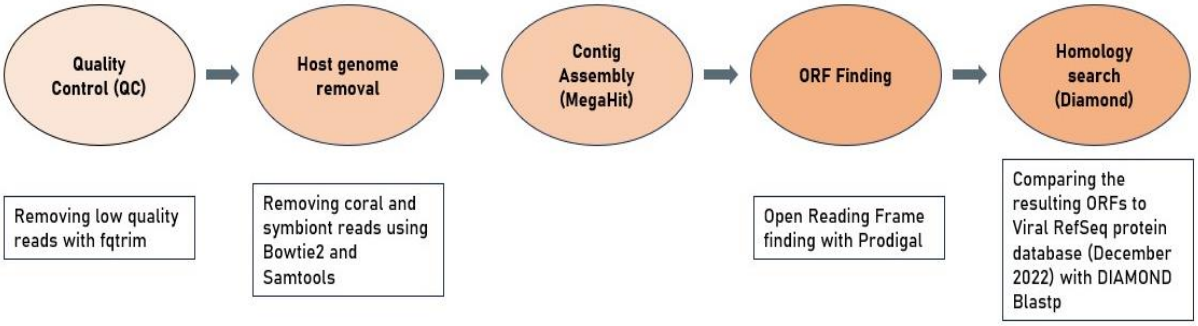


Figure 3.5. Steps followed during the bioinformatic analyses.

Bioinformatic analyses were conducted on eight samples, each denoted as follows: A1_Con_T0, A1_Nostr_T1, A2_Con_T0, A2_Vib_T1, A3_Con_T0, A3_Temp_T1, A4_Con_T0, and A4_ViTe_T1. All analyses were performed within the TRUBA (Turkish National e-Science e-Infrastructure) platform. Quality filtering, encompassing adapter and index trimming, was executed using the fqtrim program (Geo Pertea, 2018) with the following parameter settings: -R – pid5 – pid3 -D -A -q20 -w -m -l 50 -P33. The significance of each parameter employed during the analysis is outlined below:

- R: Look for terminal alignments with the reverse complement of the adapter sequence(s).
- pid5: Minimum percent identity for an adapter match at the 5' end (default 96.0).
- pid3: Minimum percent identity for an adapter match at the 3' end (default 94.0).
- D: Pass reads through a low-complexity (dust) filter and discard any reads with over 50% of their length masked as low complexity.
- A: Disable polyA/T trimming (enabled by default).
- q: Trim read ends where the quality value drops below 20.
- w: For -q, sliding window size for calculating the average quality (default 6).
- m: Maximum percentage of Ns allowed in a read after trimming (default 5).
- l: Minimum read length after trimming; reads shorter than this threshold will be discarded (default: 16).
- P33: Input is phred64/phred33 (use -P64 or -P33).

The quality of the trimmed reads was once again examined via FastQC (Andrews, 2010) to verify the efficiency of the trimming process. Subsequently, the trimmed forward and reverse DNA reads

were analyzed with bowtie2 (Langmead & Salzberg, 2012) to eliminate contaminant reads originating from host and symbiont genomes, employing the following parameters:

- --mp 4: Max penalty for mismatch.
- -X 1000: Maximum fragment length.
- -p 8: use up to 8 parallel processors

Due to the absence of a congeneric reference genome, two reference coral genomes were selected based on the evolutionary tree described in Cuif et al (2003). Additionally, two reference genomes for the symbiont genomes were chosen from the GenBank database, following the methodology used in the paper by Messyasz et al. (2020). In total, four reference genomes were aligned with our DNA reads, namely those of *Platygyra sinensis* (GCA_019787425.1), *Pocillopora verrucosa* (GCA_014529365.1), Symbiodinium sp. clade A Y106 (GCA_003297005.1), and Symbiodinium sp. clade C Y103 (GCA_003297045.1). The retrieved four reads were initially indexed using "build" option in bowtie2. Next, the previously trimmed paired-end reads were mapped to the indexed reference genomes. Alignment rates were recorded at each mapping step. Reads that did not map to these reference genomes were isolated using the samtools program (Danecek et al., 2021) using "view", "sort" and "fastq" options, and read counts and the alignment rates were recorded after each step of contaminant genome removal.

The remaining reads were employed for contig assembly using the software MegaHit (Li et al., 2015) with default parameters. The number of contigs assembled for each sample was recorded. The assembled contigs were mapped back to the trimmed reads that were filtered from contaminant genomes, using bowtie2 to verify the efficiency of the contig assembly process. The calculated alignment rates were recorded.

The assembled contigs were subsequently analyzed using the Open Reading Frame (ORF) finder program Prodigal (Hyatt et al., 2010) for gene annotation. The resulting ORFs were compared to the Viral RefSeq Protein database (from December 2022) via blastp in DIAMOND software (Buchfink et al., 2021), with a maximum e-value for protein comparison set to 1.7e-17, as suggested by Messyasz et al (2020). The number of viral contigs that matched to the Viral RefSeq protein database was counted and recorded for each sample.

Viral queries displaying at least 75% similarity to viral protein databases were selected, and taxonomic identifications for these contigs were performed using Viral RefSeq databases. The blast

hit counts of the identified viruses and the relative composition of viruses within each sample were presented in bar plots at every taxonomic level from genus to kingdom, using ggplot2 (Wickham, 2016) with R (Version 4.2.3).

To eliminate the false positive data, a stricter threshold of 80% is used instead of 75% for the viral families and genera that were only matched to one of the four viral protein databases. Additional bar plots based on filtered data are generated at family and genus levels, using ggplot2 (Wickham, 2016) package in R (Version 4.2.3).

The similarities/dissimilarities between the samples were visualized with Principal Coordinate Analysis (PCoA) in R (Version 4.2.3.), using ‘tidyverse’ (Wickham, Averick, et al., 2019), ‘vegan’ (Oksanen et al., 2022) and ‘ggplot2’ packages. The impact of the application of a stricter threshold was investigated from the PCoA plot generated with the data before and after the filtration at family and genus levels.

The contig abundance of the Nucleocytoplasmic Large DNA Viruses (NLDVs) were counted and depicted in bar plots generated with ggplot2 package in R (Version 4.2.3)

3.6.1. Read Normalization

As a next step, we aimed to investigate whether there existed a correlation between the number of reads and the viral profile obtained for each sample. To assess this, we implemented a read number filtering process, wherein we excluded the initial "n" reads from each sample that were previously cleaned from possible contaminant genomes. The aim was to standardize the read numbers to match the lowest read count observed in sample A2_Nostr_T1, which totaled 68,239,977 reads. Following the read normalization step, the obtained reads were analyzed in the same manner as the non-normalized reads, as mentioned above.

3.7. Statistical Analyses

The Shannon diversity indexes were calculated using family-level data for each sample with the 'dplyr' (Wickham, François, et al., 2019) and 'vegan' packages, and the results were visualized in bar plots using the 'ggplot2' package in R (Version 4.2.3).

Alpha and Beta diversities are computed with TaxonTableTools (Macher et al., 2021), based on Jaccard distances calculated from the presence/absence data at the genus level. The Analysis of Similarities (ANOSIM) test was applied to generate the beta diversity heatmap to compare the similarities and dissimilarities between samples, based on the stress conditions. The beta diversity dendrogram was generated from Jaccard distances using average linkage as the clustering method.

3.8. Mitochondrial Genome Assembly

The mitogenome assembly was performed using the NOVOPlasty tool (Dierckxsens et al., 2017) with previously trimmed reads. The *Cladocora caespitosa* voucher Cc718 cytochrome c oxidase subunit I (COX1) gene (MW032522.1) was obtained from GenBank and used as a seed during the mitogenome assembly process. The resulting circularized mitogenomes for seven samples were further annotated using the MITOS Webserver (Bernt et al., 2013). The annotated genes in all samples were compared with each other, and sample 306 was chosen for further mapping purposes. In comparison to the annotated mitogenome of the *C. caespitosa*, an accessible mitogenome of *Favites abdita* (NC_035879.1) was retrieved from GenBank and was also annotated with MITOS Webserver. The similarities among annotated mitogenomes of two species were investigated.

The trimmed reads of the eight samples were mapped to assembled mitogenome of 306 using bowtie2 (Langmead & Salzberg, 2012). Mapping statistics were computed with samtools (Danecek et al., 2021), using “index”, “view”, “sort”, “flagstat” and “coverage options.

Assembled mitogenomes of the seven samples were aligned altogether using MEGA 11 (Tamura et al., 2021). The differing bases between the assembled mitogenomes were investigated. The gene regions that span these divergent sites were recorded.

The circularized mitochondrial genome of the selected sample, A1_Con_T0 was annotated with GeSeq (Tillich et al., 2017) and OGDRAW (Greiner et al., 2019), using the RefSeq mitochondrial genome database of Cnidaria.

3.8.1. Haplotype Network

Using the assembled seven mitogenomes described above, the phylogeographic distribution of the *Cladocora caespitosa* species was aimed to be characterized by comparing the COX1 region of the assembled mitogenomes to the available COX1 gene sequence of *C. caespitosa* in the GenBank

database. A total of 26 COX1 gene sequences were retrieved from the database. The retrieved sequences were separated based on the spatial location they were isolated from. Two of the retrieved COX1 sequences were from Spain (Accession numbers: JQ611396.1, KR297263.1); total of 16 sequences were from Cyprus (Accession numbers: MW032500.1-MW032515.1); and eight of the COX1 sequences were from Greece (Accession numbers: MW032516.1-MW032523.1)

The assembled mitogenomes and the retrieved COX1 regions were aligned altogether in MEGA11 (Tamura et al., 2021), to locate the COX1 gene region in the assembled mitogenomes. The aligned COX1 regions were then utilized for generating the haplotype network among all 26 samples, using the PoPART software (Leigh & Bryant, 2015).

4. RESULTS

The coral colonies within Tank 3 and Tank 4 displayed a sign of bleaching in certain regions of the colonies on the seventh day of the experiment, a day after the temperature of these tanks reached 29°C. On the other hand, coral bleaching was not recorded for the colonies planted in Tank 1 and Tank 2.

4.1. Gel Electrophoresis of the Extracted DNA

The DNA isolates were run in a gel (Figure 4.1) to assess their quality, before subjecting them to treatment with the Zymo Clean & Concentrator Kit (Irvine, CA), see below. Although sample 706 (A3_Con_T0) exhibited a slight smear pattern on the gel, all other samples displayed a distinct, concentrated band, providing assurance that the genomic material in our samples remained structurally intact. Following the treatment with the Zymo Clean & Concentrator Kit, the DNA isolates were subjected to gel electrophoresis once more, as depicted in Figure 4.1 (right). The pattern we observed closely resembled the gel image of the isolates before the treatment.

The key distinction between the two sets of results lay in the intensity of the bands. This difference was primarily because, in Figure 4.1, we loaded a reduced amount of genomic content onto the gel. Despite this variation in band intensity, the overall similarity in the gel patterns indicated that the DNA isolates maintained their structural integrity after the Zymo Clean & Concentrator Kit treatment.

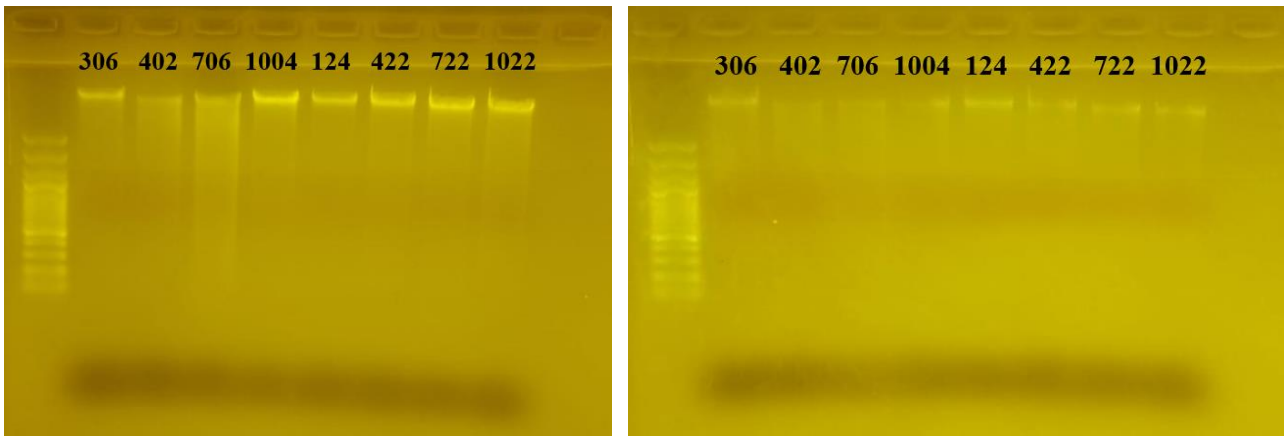


Figure 4.1. The gel electrophoresis image of extracted DNA samples (left), and the gel electrophoresis image of the DNA isolates after treated with Clean & Concentrator Kit (right).

The DNA concentration of the samples, measured with Qubit 3, are listed in Table 4.1. The lowest DNA concentration was recorded as 62 ng/µl for sample A1_Con_T0, whereas the highest DNA concentration is measured as 116 ng/µl for sample A3_Con_T0.

Table 4.1. The DNA concentrations of the samples.

Sample	Concentration (ng/µl)
A1_Con_T0	62
A1_Nostr_T1	73.6
A2_Con_T0	120
A2_Vib_T1	74.2
A3_Con_T0	116
A3_Temp_T1	78.8
A4_Con_T0	114
A4_ViTe_T1	93

4.2. Sonication for DNA Fragmentation

As depicted in Figure 4.2 below, we present the gel electrophoresis image of the sonicated field samples. Encouragingly, we observed a smeared pattern in all the samples, indicating a partially successful application of sonication.

Interestingly, the samples exposed to 44 minutes of sonication (MD19, MD20, and MD23) displayed reduced migration in the gel. In contrast, those subjected to 64 minutes of sonication (MD10 and MD23) traveled further through the gel. This observation suggests that 64 minutes of sonication was more effective in fragmenting the genomic content of our samples compared to the 44-minute duration.

Notably, MD19, which exhibited the least fragmentation in Figure 4.2, also had the highest concentration among the samples subjected to 44 minutes of sonication. This highlights the significance of sample concentration as a crucial parameter to consider during the sonication optimization process. In summary, our initial conclusion from the first sonication trial underscores the inefficiency of shorter sonication times for highly concentrated samples.

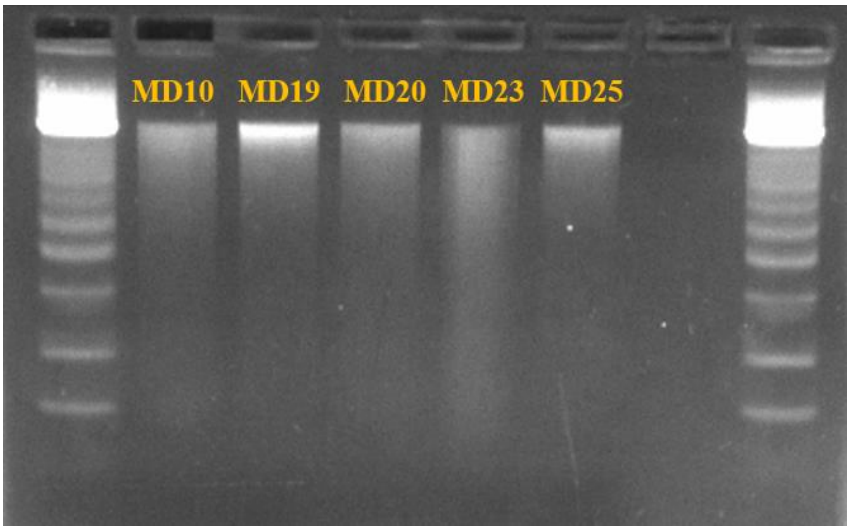


Figure 4.2. The gel image of the field samples (MD10, MD19, MD20, MD23, MD25, respectively) from the first trial after sonication.

However, it's important to note that this extremely low concentration posed a challenge when attempting to visualize the samples on the gel, as illustrated in Figure 4.3.



Figure 4.3. The gel electrophoresis image after sonication of DNA samples (522, 822, 825, 525, 523, 323, 402) from the second trial.

The outcomes of the Agilent analysis are presented in Figure 4.4, displaying the results obtained from four distinct samples. Despite sharing identical concentrations and undergoing the same sonication duration, the samples exhibited notably distinct fragmentation patterns when compared to each other, as depicted in Figure 4.4. For instance, sample 825 exhibited no noticeable fragmentation of its genomic content. On the other hand, sample 525 appeared to have undergone separation into shorter fragments falling outside the expected range of 300-500 base pairs.

A particularly intriguing observation was made with sample 323, as evidenced by its electropherogram plot. This sample displayed two distinct peaks, suggesting that the genomic material was fragmented into two separate base pair ranges. Such variations in fragmentation patterns underscore the complexity and diversity inherent in our experimental samples.

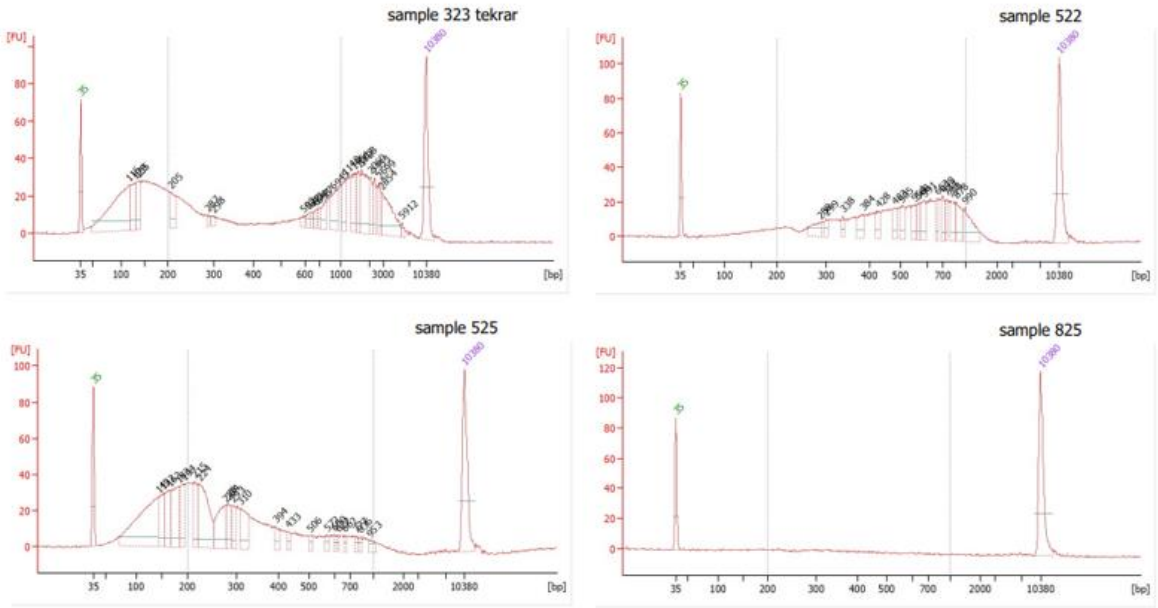


Figure 4.4. The electropherogram plots of the four different sonicated DNA samples.

As the optimization process for DNA fragmentation via sonication proved unsuccessful, it was decided that the samples would be sent directly to Gen-Era (İstanbul, Türkiye) for library preparation and subsequent Illumina sequencing.

4.3. Genome Alignment and Contig Assembly

4.3.1. Quality Control

The quality report on the basic statistics of the sequencing performance is presented in Table 4.2 for the raw paired end reads. The forward and reverse reads of the same samples resulted in the exact same number of total sequences as expected. The highest number of total sequences was obtained as 126597300 for the sample A2_Vib_T1. In contrast, the lowest number of total sequences was recorded as 70610033 for the reads belonging to the sample A3_Con_T0. While the GC% value

ranged between 30 to 53 among samples, this value did not vary greatly between the forward and reverse reads of the same sample. Overall, the GC content across all bases is lower than AT content in most of the samples with exceptions of samples A2_Con_T0 and A3_Con_T0. The 'Per Sequence GC Content' module, which represents the GC content across the whole length of each sequence in a file in FastQC report, failed for the forward and reverse reads of the following samples: A2_Con_T0, A3_Con_T0 and A3_Temp_T1. This failure was due to the sum of the deviations from normal distribution representing more than 30% of the reads, which might indicate a contaminated library or a systematic bias.

Table 4.2. FastQC report outcomes. Total number of sequences, total base pairs, sequence lengths and GC% of the forward and reverse reads of all samples are presented.

Sample	Read	Total Sequences	Total Bases (Gbp)	Sequence length	%GC
A1_Con_T0	Forward	112232867	16.6	35-151	40
	Reverse	112232867	16.9	151	41
A1_Nostr_T1	Forward	71873064	10.6	35-151	39
	Reverse	71873064	10.8	151	40
A2_Con_T0	Forward	108961761	16.1	35-151	53
	Reverse	108961761	16.4	151	53
A2_Vib_T1	Forward	126597300	18.7	35-151	39
	Reverse	126597300	19.1	151	39
A3_Con_T0	Forward	70610033	10.4	35-151	53
	Reverse	70610033	10.6	151	53
A3_Temp_T1	Forward	90311790	13.4	35-151	46
	Reverse	90311790	13.6	151	46
A4_Con_T0	Forward	84438245	12.5	35-151	39
	Reverse	84438245	12.7	151	39
A4_ViTe_T1	Forward	90720837	13.4	35-151	39
	Reverse	90720837	13.6	151	40

In all the reads, the 'Per Base Sequence Content' module, representing the proportion of each base, failed. This failure is caused by the difference between A and T, or G and C being greater than 20%. While an evenly distributed proportion of bases is expected (Figure 4.5-b), an imbalanced

distribution of bases at the start of sequenced read was observed, as shown in Figure 4.5-a for the forward read of the sample A1_Nostr_T1. The disrupted pattern observed in the starting sequence of forward read of the sample A1_Nostr_T1 persisted across all sequenced reads. This failure is attributed to sequencing bias and does not impact the absolute sequence.

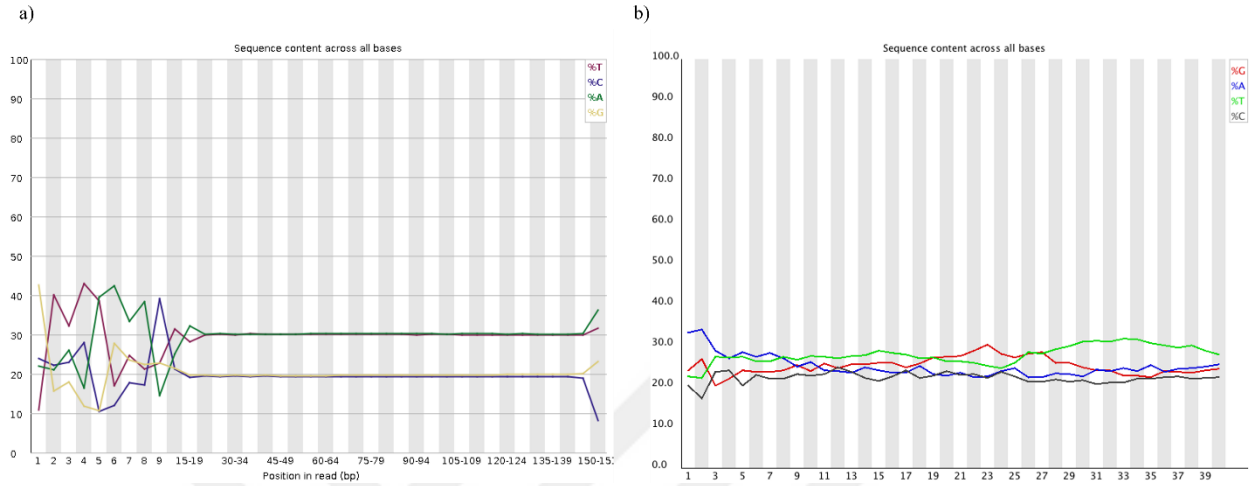


Figure 4.5. Per base sequence content of a) forward read of sample A1_Nostr_T1 as compared to the b) expected proportion of bases in a sequenced sample.

Performing the trimming and quality filtering step with fqtrim, reads that were shorter than 50 base pairs were eliminated (Table 4.3). While the GC% stayed the same for most of the samples, the GC% of forward and reverse reads of sample A1_Con_T0 was equalized after the trimming step. The overrepresented sequences were eliminated after the trimming step in reverse read of sample A1_Nostr_T1, sample A2_Vib_T1 and sample A4_ViTe_T0. The 'Per Base Sequence Content' module continued to fail for all the sequenced reads. The failure observed in the GC count per read module remained unchanged after the trimming step, thereby failing for the forward and reverse reads of the samples A2_Con_T0, A3_Con_T0 and A3_Temp_T1.

Table 4.3. FastQC report after the trimming and quality filtering step with fqtrim. Total number of sequences, total base pairs, sequence lengths and GC% of the forward and reverse reads of all samples after trimming step are presented.

Sample	Read	Total Sequences	Total Bases (Gbp)	Sequence length	%GC
A1_Con_T0	Forward	111228239	16.4	50-151	40
	Reverse	111436207	16.7	50-151	40
A1_Nostr_T1	Forward	71238576	10.5	50-151	39
	Reverse	71316282	10.6	50-151	40
A2_Con_T0	Forward	108077643	16	50-151	53
	Reverse	108296728	16.2	50-151	53
A2_Vib_T1	Forward	125440020	18.5	50-151	39
	Reverse	125645093	18.8	50-151	39
A3_Con_T0	Forward	69593530	10.2	50-151	53
	Reverse	69993314	10.5	50-151	53
A3_Temp_T1	Forward	89581286	13.2	50-151	46
	Reverse	89720335	13.4	50-151	46
A4_Con_T0	Forward	83742114	12.4	50-151	39
	Reverse	83797634	12.5	50-151	39
A4_ViTe_T1	Forward	89929463	13.3	50-151	39
	Reverse	90058654	13.5	50-151	40

One of the changes that was observed after the trimming step was the distribution of sequence length over all sequences. The ‘Sequence Length Distribution’ module delivered a warning in all reads after the application of the trimming step. Figure 4.5 below exhibits the sequence length distribution before (a) and after (b) the trimming and quality filtering step on forward reads of the sample A1_Nostr_T1. The elimination of low-quality reads and reads that are shorter than 50 bp resulted in a shifted distribution of sequence length over all sequences in the given sample, as shown in Figure 4.5 a-b.

a) b)

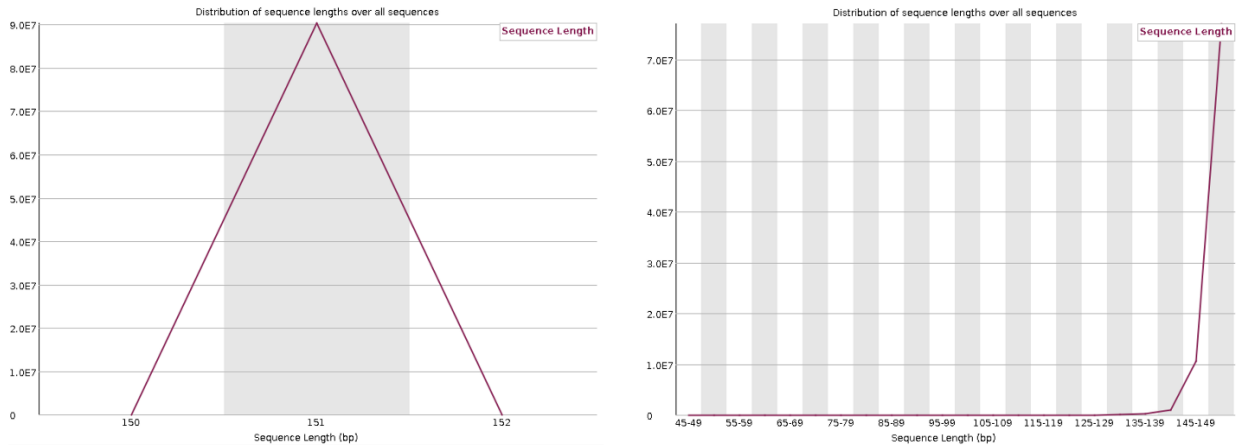


Figure 4.5. The sequence length distributions a) before and b) after the trimming step for the forward read of the sample A2_Nostr_T1.

4.3.2. Removal of the Host Genome

Initially, the host genome removal process was performed for the sample A4_Con_T0. 84,438,245 reads were obtained at the end of the trimming step (Table 4.4). After the elimination of a potential coral host *Platygyra sinensis* genome, the remaining read number was 75,860,941. Read numbers were identical for both forward and reverse DNA read data throughout the analyses.

Subsequently, the genome of *Pocillopora verrucosa* was discarded from our reads, as a result the read number was reduced to 75,499,252. During the removal of *Pocillopora verrucosa* from our data, the overall alignment rate was recorded as 0.32% in bowtie2 (Table 4.5), confirming that most of the coral host genome was removed with the elimination of *Platygyra sinensis*.

Symbiodinium sp. clade A Y106 genome was discarded next, and the resulting read number was recorded as 75,488,782. The same number was reduced to 75,476,956 after the removal of Symbiodinium sp. clade C Y103.

After all the host and symbiont genomes were removed from the sample A4_Con_T0, the same trimming and genome filtering steps were performed for the rest of the seven samples. The recorded read numbers at each genome removal process are depicted in Table 4.4. While the highest read number (yellow) was recorded as 119,448,738 for the sample A2_Vib_T1, the lowest read number (blue) was counted as 68,239,977 for the sample A3_Con_T0.

However, after the elimination of all the contaminant genomes in our dataset, the lowest number of reads no longer belonged to sample A3_Con_T0, but to sample A1_Nostr_T1. This suggests that, comparing the read numbers between A3_Con_T0 and A1_Nostr_T1, a greater number of reads were discarded after the elimination of *Platygyra sinensis* genome.

Table 4.4. Number of reads after each step of contaminant genome removal. The selected contaminant genomes are *Platygyra sinensis*, *Pocillopora verrucosa*, Symbiodinium sp. clade A Y106 and Symbiodinium sp. clade C Y103. The highest number of reads depicted in yellow, and the lowest number of reads depicted in blue.

Sample	**Read number after trimming**	***Platygyra sinensis* removal**	***Pocillopora verrucosa* removal**	**Symbiodinium sp. clade A Y106 removal**	**Symbiodinium sp. clade C Y103**
A1_Con_T0	112,232,867	107,237,828	106,830,928	106,785,872	106,752,478
A1_Nostr_T1	71,873,064	68,538,454	68,272,570	68,256,484	68,239,977
A2_Con_T0	108,961,761	108,780,268	108,772,489	108,707,404	108,692,095
A2_Vib_T1	126,597,300	120,025,111	119,503,460	119,479,651	119,448,738
A3_Con_T0	70,610,033	70,182,151	70,151,205	70,074,173	70,054,652
A3_Temp_T1	90,311,790	87,545,793	87,320,498	87,273,020	87,247,831
A4_Con_T0	84,438,245	75,860,941	75,499,252	75,488,782	75,476,956
A4_ViTe_T1	90,720,837	86,356,212	86,000,970	85,974,255	85,947,034

Table 4.5 represents the alignment rates calculated in bowtie2, aligning our reads with contaminant host and symbiont genomes. The highest alignment rate for *P. sinensis* was observed for the sample A2_Vib_T1 with 3.73%, and same alignment rate was the lowest for the sample A2_Con_T0 with 0.10%. The alignment rates decreased drastically when aligned with *P. verrucosa*, confirming the removal of *P. sinensis* at the previous step of genome removal.

However, it's worth highlighting that the alignment rate for sample A2_Con_T0 was notably lower. Following the removal of the *P. sinensis* genome, the alignment rate to the second host genome reached 0. This suggests the intriguing possibility of either the absence or a very low abundance of coral host genome reads within this particular sample.

A similar scenario appears to be valid for sample A3_Con_T0 as well. Although the alignment rates to the coral host genomes were not as minuscule as in A2_Con_T0, they were still notably lower compared to the rest of the samples. This observation suggests, once again, the presence of very few coral host-genome reads within this sample.

Furthermore, the highest alignment rate with Symbiodinium sp. clade C, hence presumably the highest content of the species, was recorded as 0.07% for the sample A3_Con_T0. In contrast, the lowest alignment rate was recorded for sample A2_Vib_T1 and A4_Con_T0 with 0.01%, representing the lowest abundance of the Symbiodinium species in this sample.

Table 4.5. Alignment rates of seven samples after every step of genome removal calculated in bowtie2.

Sample	*P. sinensis*	*P. verrucosa*	Symbiodinium sp. clade A Y106	Symbiodinium sp. clade C Y103
A1_Con_T0	3.17%	0.26%	0.02%	0.02%
A1_Nostr_T1	3.14%	0.26%	0.03%	0.02%
A2_Con_T0	0.10%	0.00%	0.04%	0.01%
A2_Vib_T1	3.73%	0.30%	0.01%	0.02%
A3_Con_T0	0.38%	0.03%	0.07%	0.02%
A3_Temp_T1	2.15%	0.18%	0.04%	0.02%
A4_Con_T0	3.85%	0.32%	0.01%	0.01%
A4_ViTe_T1	3.36%	0.28%	0.02%	0.02%

4.3.3. Contig Assembly

After the host removal step, we conducted contig assembly using MegaHit, and a comprehensive overview of the assembled contigs can be found in Table 4.6, provided for each of the samples. Notably, the sample A3_Con_T0 yielded the highest number of contigs (depicted in yellow), whereas the lowest number of contigs was documented for A2_Vib_T1 (depicted in blue).

Furthermore, for sample A2_Con_T0, which had the second-highest number of contigs, we observed that it contained the greatest total base pairs (yellow) as well as the longest individual contig

(yellow). The average base pair length of the assembled contigs fell within a range of 625 to 1087 (Table 4.6).

Table 4.6. The detailed output of contigs from MegaHit. The number of contigs, total base pairs, length of the assembled contigs, average length of the contigs and the N50 lengths are presented. The highest values are depicted in yellow, and the lowest values are depicted in blue.

Sample	Contigs	Total bp	Min bp	Max bp	Avg bp	N50 bp
A1_Con_T0	1041883	944531189	200	121484	906	1927
A1_Nostr_T1	968457	868362635	200	90552	896	2235
A2_Con_T0	2515457	1969230585	200	960479	782	919
A2_Vib_T1	592298	644362752	200	113573	1087	3772
A3_Con_T0	3083736	1930076033	200	212182	625	707
A3_Temp_T1	1920528	1606417647	200	320488	836	1207
A4_Con_T0	649813	661527447	200	158260	1018	3578
A4_ViTe_T1	680593	685960317	200	178110	1007	3406

The alignment rates for the assembled contigs to the trimmed reads for each sample are provided below in Table 4.7. The highest alignment rate, a remarkable 94.66%, was achieved for A2_Vib_T1, indicating the precision of the contig assembly process. Surprisingly, when we aligned the assembled contigs from A2_Con_T0 with the trimmed reads of the same sample, the alignment rate was only 63.86%, notably lower than the alignment rates observed for the other samples.

Table 4.7. Alignment rates of the assembled contigs to the quality trimmed reads.

Sample	Alignment rate
A1_Con_T0	88.67 %
A1_Nostr_T1	93.83 %
A2_Con_T0	63.86 %
A2_Vib_T1	94.66 %
A3_Con_T0	84.73 %
A3_Temp_T1	83.42 %
A4_Con_T0	95.55%
A4_ViTe_T1	91.27 %

4.4. Taxonomic Assignment

Among the assembled contigs, the number of viral contigs that matched the viral protein database is counted and presented in Table 4.8. Although the highest number of assembled contigs was observed for sample A3_Con_T0, the greatest number of viral contigs, reaching 9746, was recorded for sample A2_Con_T0. Additionally, this sample also had the highest number of viral contigs that matched the viral protein database with a minimum threshold of 75%. Conversely, the lowest number of viral contigs and those with a similarity percentage higher than 75% were recorded as 728 and 60, respectively, for sample A4_Con_T0.

Comparing the initial and final viral loads in each tank in Table 4.8, the number of contigs, number of viral contigs and the number of contigs having more than 75% similarity score all showed a decreasing pattern in Tank 1 (A1_Con_T0 and A1_Nostr_T1) and Tank 2 (A2_Con_T0 and A2_Vib_T1). On the other hand, while we observed a decrease in both the number of contigs and number of viral contigs in Tank 3 (A3_Con_T0 and A3_Temp_T1), the number of viral contigs having more than 75% similarity score has slightly increased in the same tank. Interestingly, there is a general increase in all types of contigs in Tank 4 (A4_Con_T0 and A4_ViTe_T1).

Table 4.8. The total number of assembled contigs, assembled contigs that matched to the viral protein database and the viral contigs that matched to the viral protein database with 75% similarity threshold in all samples.

Sample	# of Contigs	# of Viral Contigs	> 75%
A1_Con_T0	1041883	3201	141
A1_Nostr_T1	968457	1360	64
A2_Con_T0	2515457	9746	365
A2_Vib_T1	592298	938	68
A3_Con_T0	3083736	6632	163
A3_Temp_T1	1920528	6230	168
A4_Con_T0	649813	728	59
A4_ViTe_T1	680593	1753	90

Based on the taxonomic assignments resulted with DIAMOND, the number of contigs that matched to the viral protein database with similarity percentage higher than 75% is represented as bar plots for each of the taxonomic level in Figure 4.6-4.21. Comparing data from the first and last sampling, similar patterns were observed at each taxonomic level, except at order level. It is important

to note that most viruses lacked order level identification in the database, therefore severe decrease in BLAST hit counts was observed at order level.

In general, there is no similarity within control groups at any taxonomic level in terms of both viral contig abundance and viral composition (Figure 4.6-4.21). At all taxonomic levels, the highest viral contig abundance and viral diversity is observed in sample A2_Con_T0.

Furthermore, our analysis revealed that Duplodnaviria is the most dominant realm in every tank and the second most abundant realm is Varidnaviria (Figure 4.6). The viral contig abundance of Duplodnaviria realm exhibited fluctuations among samples, while the Varidnaviria realm appeared to maintain greater consistency in abundance across different samples. Approximately 30 viral contigs hit Varidnaviria member viruses in most samples, with two exceptions: A2_Con_T0 and A4_Con_T0, in which the number of viral contigs matched with Varidnaviria members was 50 and 4, respectively. On the other hand, the number of contigs that matched with Duplodnaviria member viruses displayed very little fluctuation in abundance between the first and last samplings in Tank 3 and Tank 4. However, the members of the Duplodnaviria realm in Tank 1 and Tank 2 were almost depleted between the first and last samplings. Blast hit counts that matched with Duplodnaviria members decreased from 96 to 20 in Tank 1, and from 296 to 32 in Tank 2. Members of the Riboviria realm were recorded in very few contig abundances in all samples, except in A1_Nostr_T1 and A4_Con_T0. Finally, two of the contigs in sample A2_Con_T0 matched with Monodnaviria realm members, and this realm only appeared in this sample.

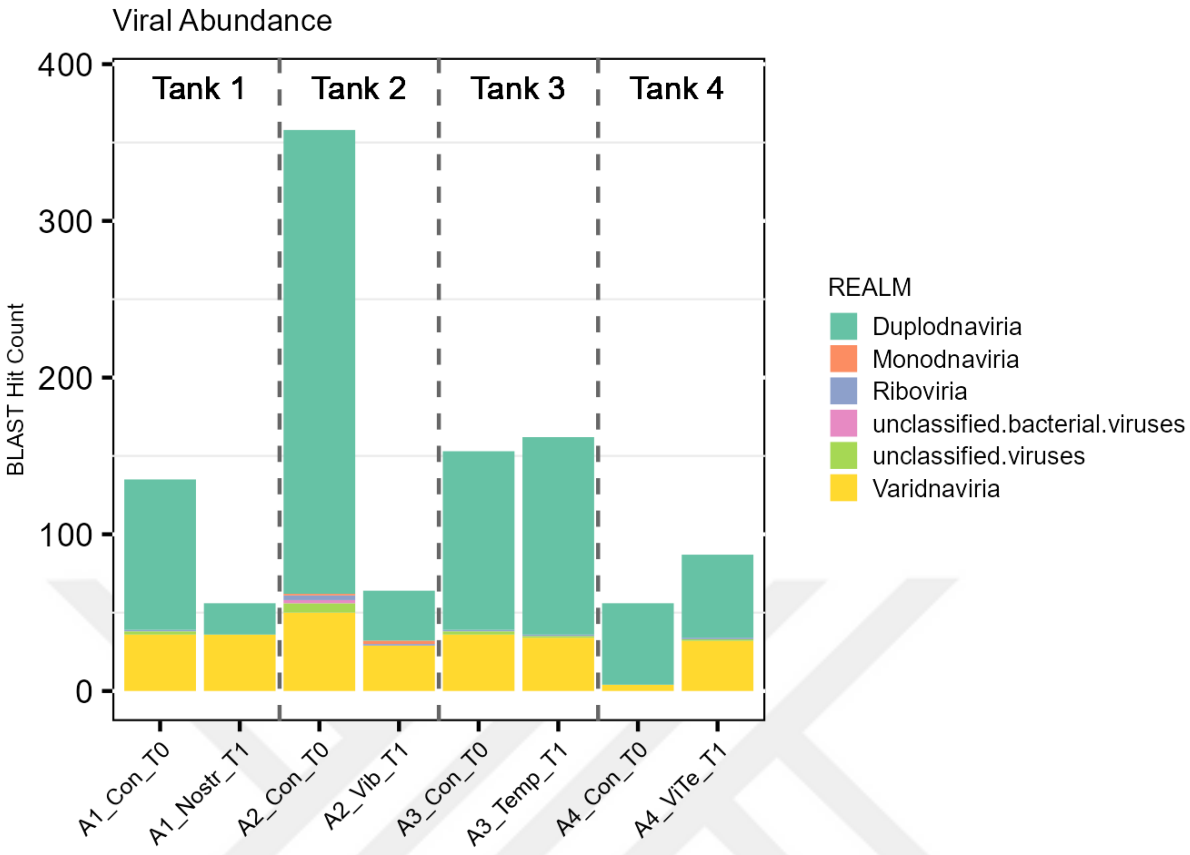


Figure 4.6. Comparison between first and last samplings at realm level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold.

The viral composition of the samples is illustrated in Figure 4.7. In the control groups, the dominant viral presence is attributed to the Duplodnaviria realm. This pattern holds true for the remaining samples, except for A1_Nostr_T1. Notably, in Tank 1, Tank 2, and Tank 4, there is an increased prevalence of Varidnaviria in the viral composition between first and last samplings. On the other hand, the viral compositions in Tank 3 remained almost the same in time. The composition of Varidnaviria realm members is quite similar between control groups of Tank 1, Tank 2 and Tank 3. However, the composition of Varidnaviria members in control group of Tank 4 is relatively lower than the rest of the samples.

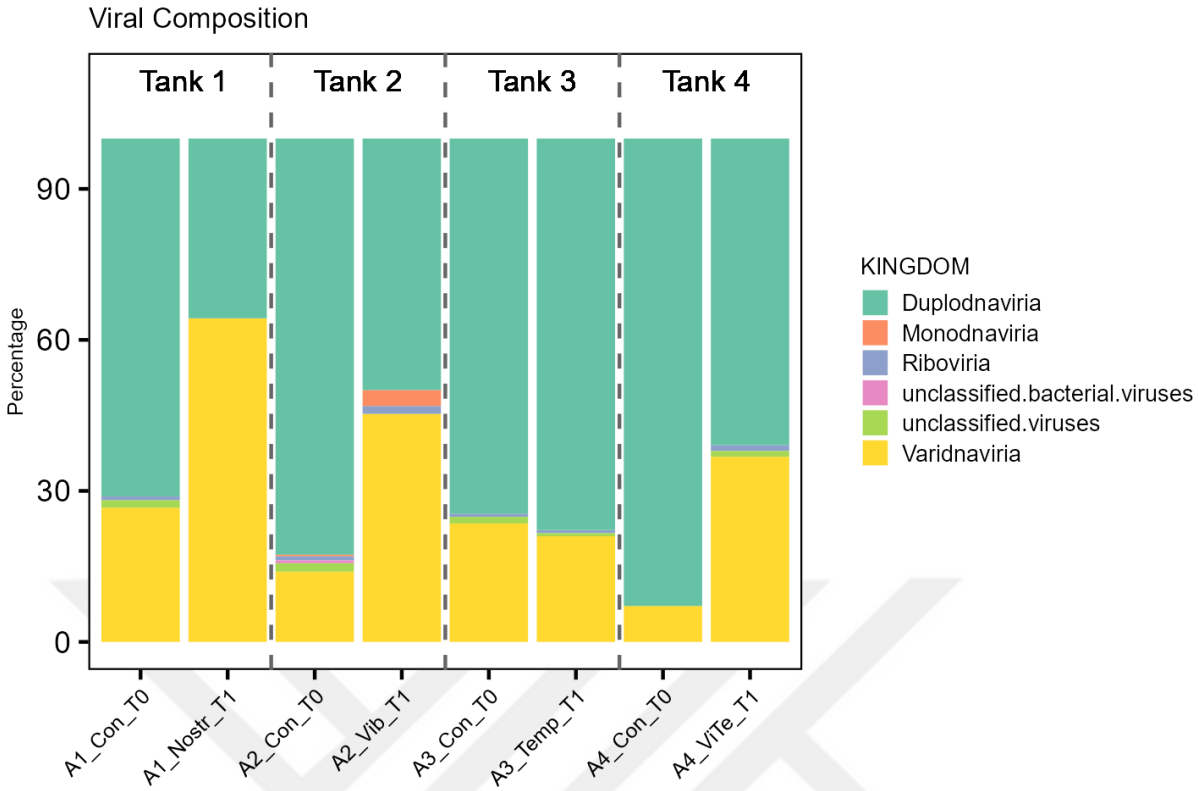


Figure 4.7. The viral composition representing the relative proportion of viral realms between samples.

The viral kingdoms within our samples are dominated by Heunggongvirae and Bamfordvirae (Figure 4.8), We observe the same pattern that we did for the distributions of the number of viral contigs at realm level, simply because these two kingdoms are the sub-category of Duplodnaviria and Varidnaviria, respectively. The contig abundance of Heunggongvirae kingdom members, which is equal to the number of contigs that matched with Duplodnaviria realm members, decreased in Tank 1 and Tank 2. The viral contig abundance of Bamfordvirae kingdom members is the same as its higher-level taxa, namely Varidnaviria.

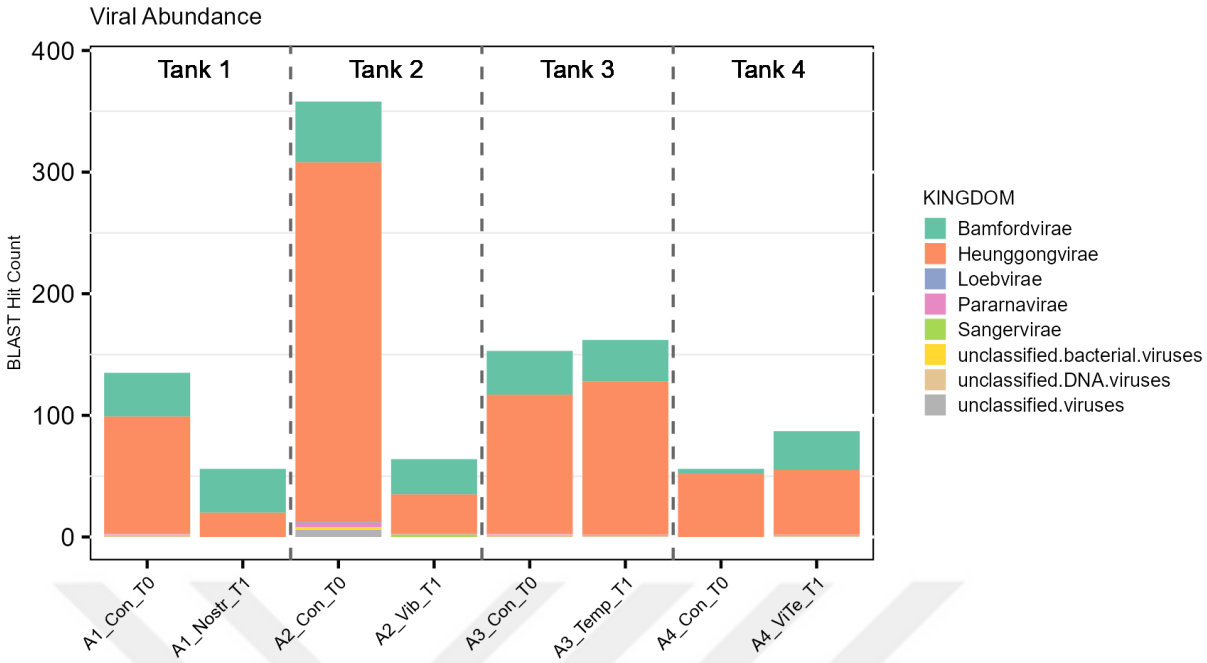


Figure 4.8. Comparison between first and last samplings at kingdom level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold.

The viral compositions at kingdom and phylum levels share a similar trend as the viral composition distribution displayed at realm level. The ratio of the Heunggongvirae kingdom within samples is the highest, and this ratio decreases at t=1 in Tank 1, Tank 2 and Tank 4 (Figure 4.9).

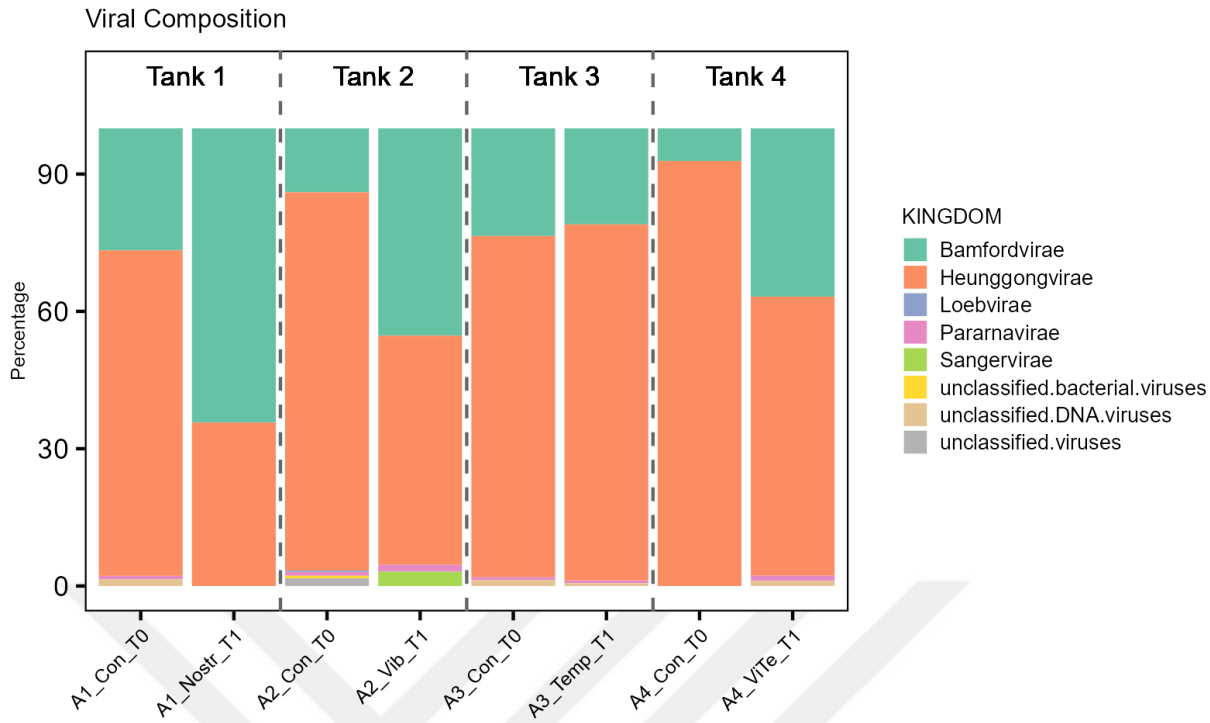


Figure 4.9. The viral composition representing the relative proportion of viral kingdoms between samples.

A similar condition was observed for the viral phyla where Uroviricota and Nucleocytoviricota, subcategories of Heunggongvirae and Bamfordvirae, are the most abundant among samples (Figure 4.10). The number of contigs that matched with Nucleocytoviricota phylum members are once again equal to the number of contigs observed in its higher-level taxa of Bamfordvirae and Varidnaviria. The Heunggongvirae kingdom members were split into Uroviricota and Peploviricota kingdoms, in which Uroviricota kingdom members are higher in abundance of viral contigs. Phixviricota phylum members were only observed in A2_Vib_T1 with a very low contig abundance of two.

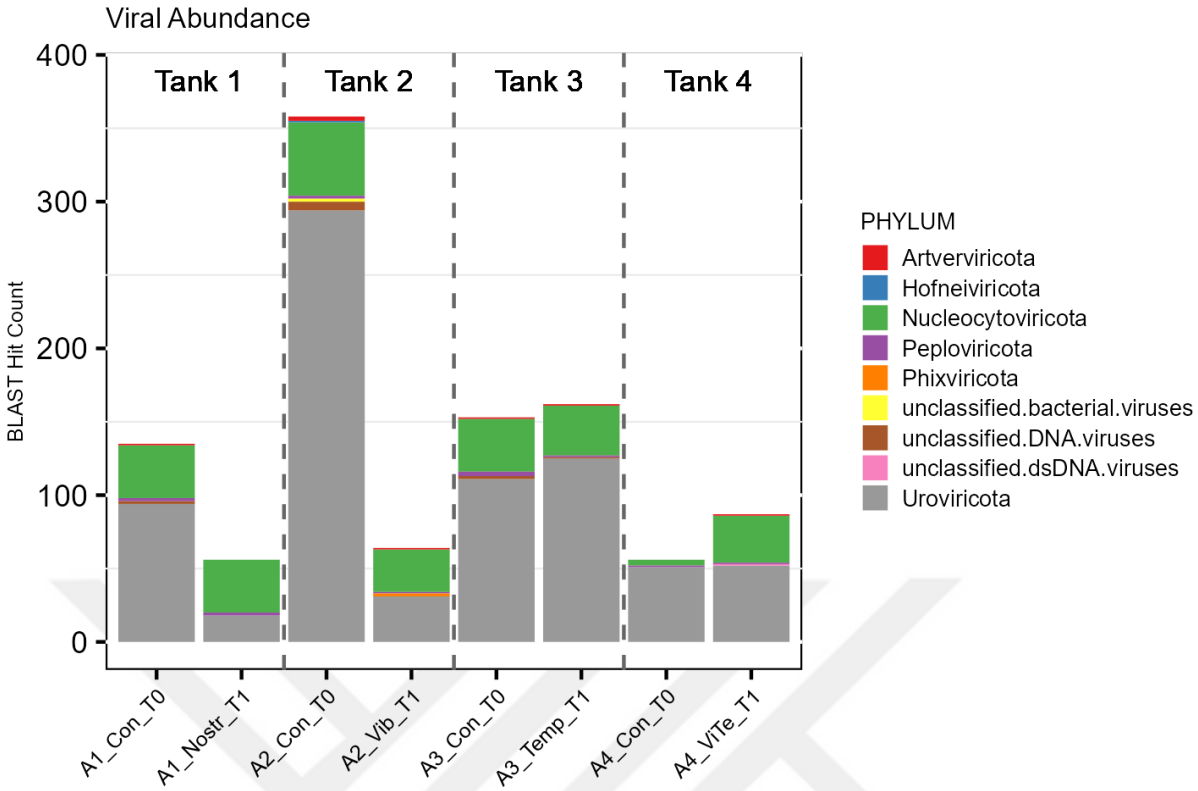


Figure 4.10. Comparison between first and last samplings at phylum level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold.

Similarly, the relative composition of the viral phylum is dominated by Uroviricota in all samples, however the composition of Uroviricota decreases at t=1 in Tank 1, Tank 2 and Tank 4 (Figure 4.11). The viral compositions of A3_Con_T0 and A3_Temp_T1 in Tank 3 remained very similar to each other, despite the introduction of heat stress in this tank.

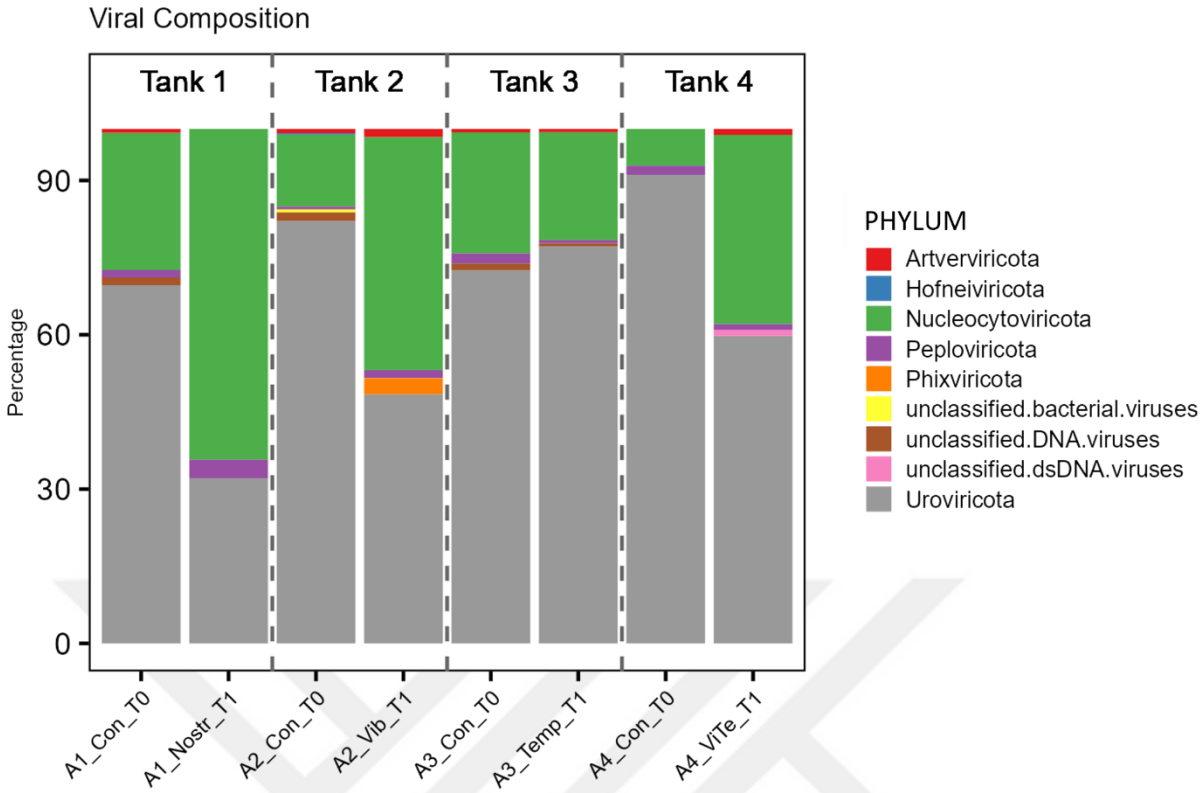


Figure 4.11. The viral composition representing the relative proportion of viral phyla between samples.

The only exception to this common pattern observed at three different taxonomic levels was sample A4_Con_T0. The abundance of Varidnaviria, Bamfordvirae and Nucleocytoviricota members appears to be the lowest when compared to the rest of the samples. After the introduction of heat stress and pathogen, the number of viruses in these taxonomic levels seems to have increased drastically, when compared to the increase observed for the Duplodnaviria realm members for the same tank, Tank 4.

The most abundant viral class is Caudoviricetes, which is the subcategory of the phylum Uroviricota (Figure 4.12). On the other hand, we see that the second most abundant phylum, Nucleocytoviricota, is dominated by two classes, Pokkesviricetes and Megaviricetes. The contig abundance of these two classes are very similar in number. Out of 257 viral contigs that matched to Nucleocytoviricota phylum members, 132 of them were matched with Megaviricetes class and 125 of them were matched with Pokkesviricetes class members in total for all the samples. A similar pattern was observed for the distribution of the two classes in each of the samples as well. All the

Caudoviricetes class members belong to the Uroviricota phylum members, therefore the abundance of contigs in these two taxa are the same.

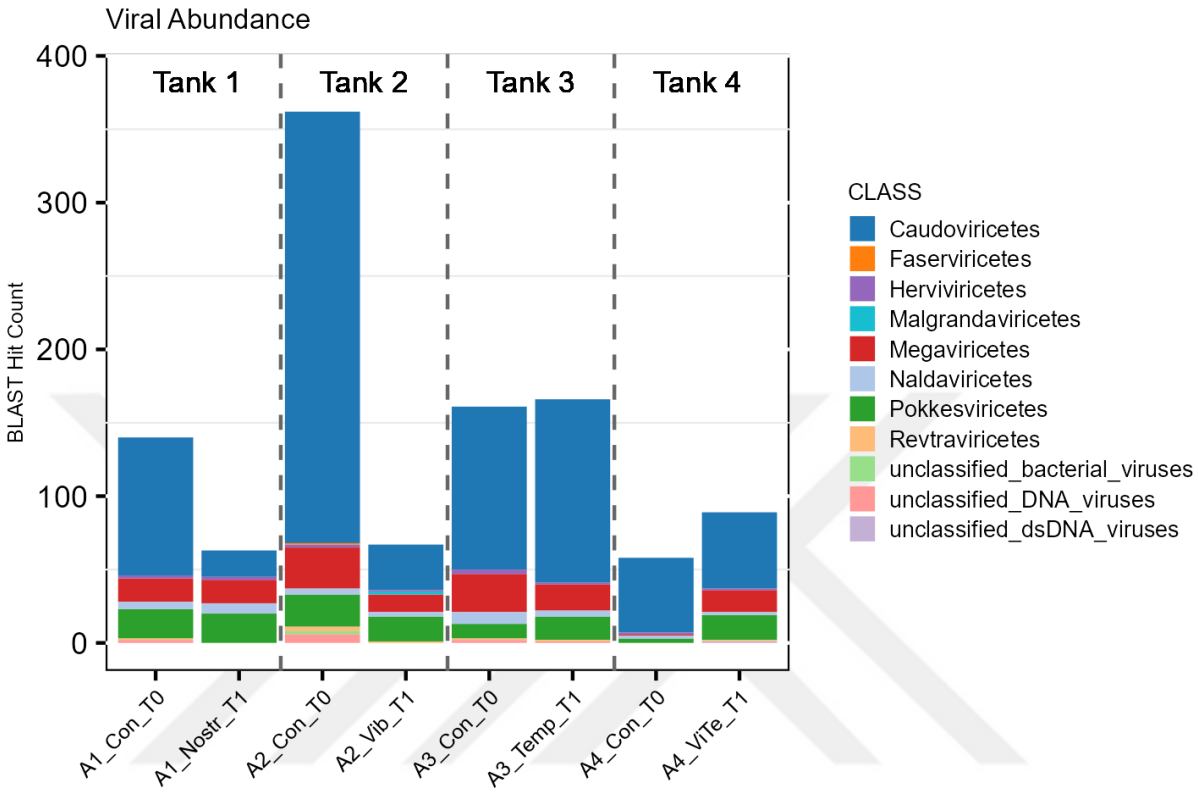


Figure 4.12. Comparison between first and last samplings at class level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold.

The highest ratio of viral class is observed for Caudoviricetes, a subcategory of the phylum Uroviricota, in all the samples, except A1_Nostr_T1, as depicted in Figure 4.13. The ratio of Caudoviricetes class members to the rest of the classes displayed a decrease in Tank 1, Tank 2, and Tank 4. On the other hand, an increase in the ratio of Megaviricetes was observed in the same tanks. Finally, the ratio of Pokkesviricetes within the viral composition of samples increased in all the tanks.

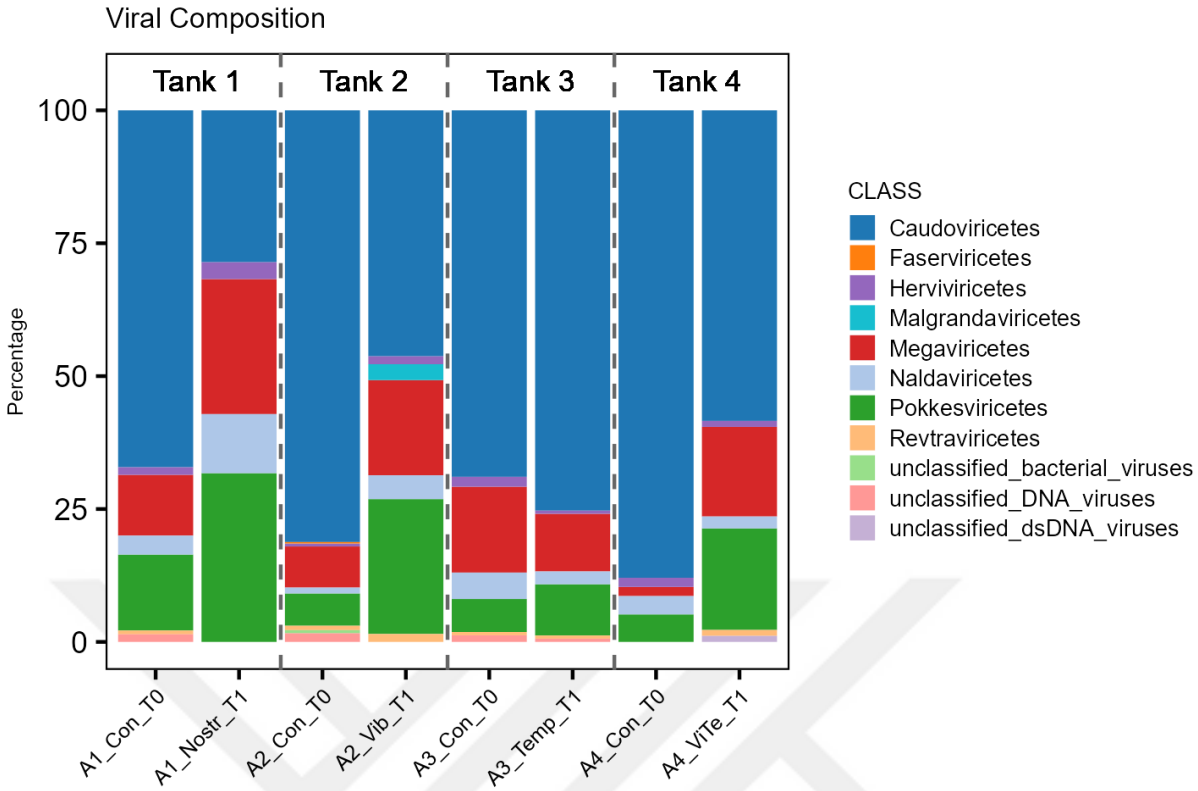


Figure 4.13. The viral composition representing the relative proportion of viral classes between samples.

However, we observed a different pattern between the samples at order level (Figure 4.14), due to the lack of information within the database. Despite the data loss at the order level, we see that the most abundant two orders are Chitovirales and Algavirales. The Ortervirales order completely disappeared in Tank 1 and emerged in Tank 4 at t=1. In Tank 2, Crassvirales order disappeared at t=1, on the other hand Petrivirales order has emerged.

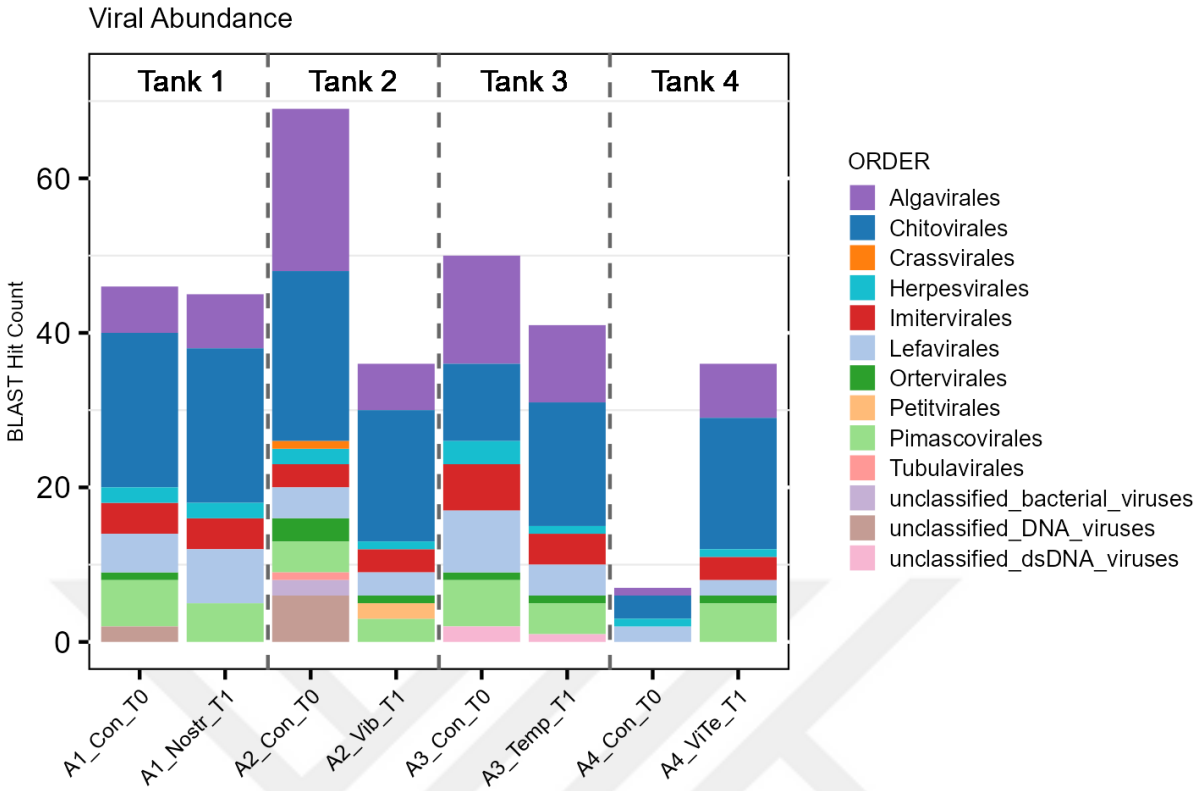


Figure 4.14. Comparison between first and last samplings at order level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold.

The viral composition in Tank 1 (A1_Con_T0 and A1_Nostr_T1) at order level was relatively stable, as compared to the changes observed for the rest of tanks (Figure 4.15). A similar observation was made for Tank 3, where the viral composition did not differ significantly between A3_Con_T0 and A3_Temp_T1. However, the sample A3_Temp_T1 became slightly dominated by Chitovirales order at t=1. Despite the severe data loss at the order level, it was possible to monitor the drastic change in the viral composition of Tank 2 (A2_Con_T0 and A2_Vib_T1) over time. Finally, the ratio of Lefavirales and Herpesvirales orders in Tank 4 decreased.

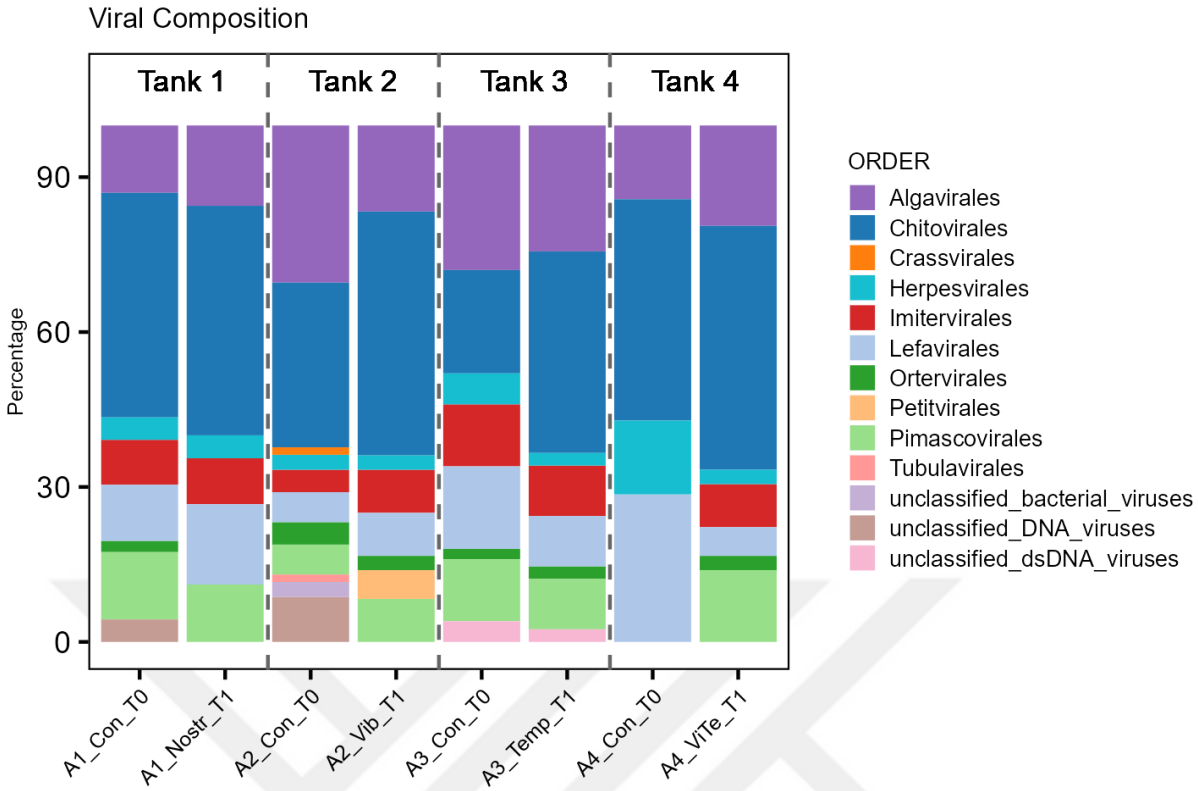


Figure 4.15. The viral composition representing the relative proportion of viral orders between samples.

The viral diversity increased at family level as shown in Figure 4.16. The highest diversity of families is observed in A2_Con_T0 sample, in which total of 32 families was recorded. Introducing *V. coralliilyticus* to Tank 2, the number of different families observed in A2_Vib_T1 decreased to 14. Similarly, the number of viral families observed in Tank 1 decreased from 21 to 13 between the first (A1_Con_T0) and last (A1_Nostr_T1) samplings. Conversely, we observed an increase in the number of unique families observed in Tank 2 and Tank 3. Although the total contig abundance in Tank 2 remained relatively similar between the first (A3_Con_T0) and last (A3_Temp_T1) samplings, the number of unique viral families increased from 18 to 23 in time. In Tank 4, the number of viral families reached 19 after the introduction of heat stress and *V. coralliilyticus*, and this value was recorded as 12 in the control group. Therefore, while the viral family diversities decreased in "No Stress" (Tank 1) and "Vibrio" (Tank 2) tanks, the viral family diversities increased in "Heat Stress" (Tank 3) and "Vibrio and Heat Stress" (Tank 4) tanks.

The number of viral families observed in all the samples is recorded as six, and these families are Kyanoviridae, Poxviridae, Phycodnaviridae, Baculoviridae, Herpesviridae, and

Polydnaviriformidae. In addition to these viral families, the Peduoviridae, Myoviridae, Autographiviridae, and Siphoviridae families were all observed in the control groups. However, no viral family was recorded as being unique to the control groups only.

The total number of contigs presented in each sample was also counted for comparison. The contig abundance is the highest in the sample A2_Con_T0, which was recorded as 237. This value dropped to 57 viral contigs for the sample A2_Vib_T1 in the same tank after the introduction of *V. coralliilyticus.* Although not as drastic as the decrease observed in Tank 2, the contig abundance at family level also displayed a decrease in Tank 1, going from 107 to 55. Despite the increased viral family diversity observed in Tank 3 between the sample A3_Con_T0 and A3_Temp_T1, the viral contig abundance corresponding to these viral families remained relative the same, exhibiting a slight increase from 90 to 93. This finding indicates that some of the viral families has matched with a higher number of contigs. Notably, the viral contigs that were acquired by the viral families in Tank 4 displayed an increase in abundance, comparing the contig abundances between A4_Con_T0 and A4_ViTe_T1.

The contig abundance of Kyanoviridae and Poxviridae is the highest among all other families, and the highest abundance of these viral contigs at family level were recorded as 48 and 22, respectively, for the sample A2_Con_T0. Both families exhibited a decrease in contig abundance after the introduction of heat stress in Tank 2; the contig abundance of Kyanoviridae family is reduced to 15 and the contig abundance of Poxviridae family decreased to 17. On the other hand, these two families displayed an increase in Tank 3 and Tank 4 after the introduction of corresponding environmental stress factors: heat stress and vibrio with heat stress, respectively. The initial contig abundance of the Kyanoviridae family was 10 in Tank 3 and 14 in Tank 4. With the introduction of corresponding environmental stress factors, these values increased to 14 in Tank 3 and 16 in Tank 4. Similarly, the contig abundance of the Poxviridae family increased from 10 to 16 in Tank 3 and from 3 to 17 in Tank 4, where the most drastic increase was observed in the contig abundance of this viral family. Interestingly, Chaseviridae family members were only observed in Tank 4, where the Chaseviridae family members completely disappeared after the introduction of *V. coralliilyticus* to the coral colonies in Tank 2. The initial contig abundance of the Chaseviridae family was recorded as 42 in A2_Con_T0 and 18, which decreased to 4 after the introduction of *V. coralliilyticus* with heat stress, in A4_Con_T0.

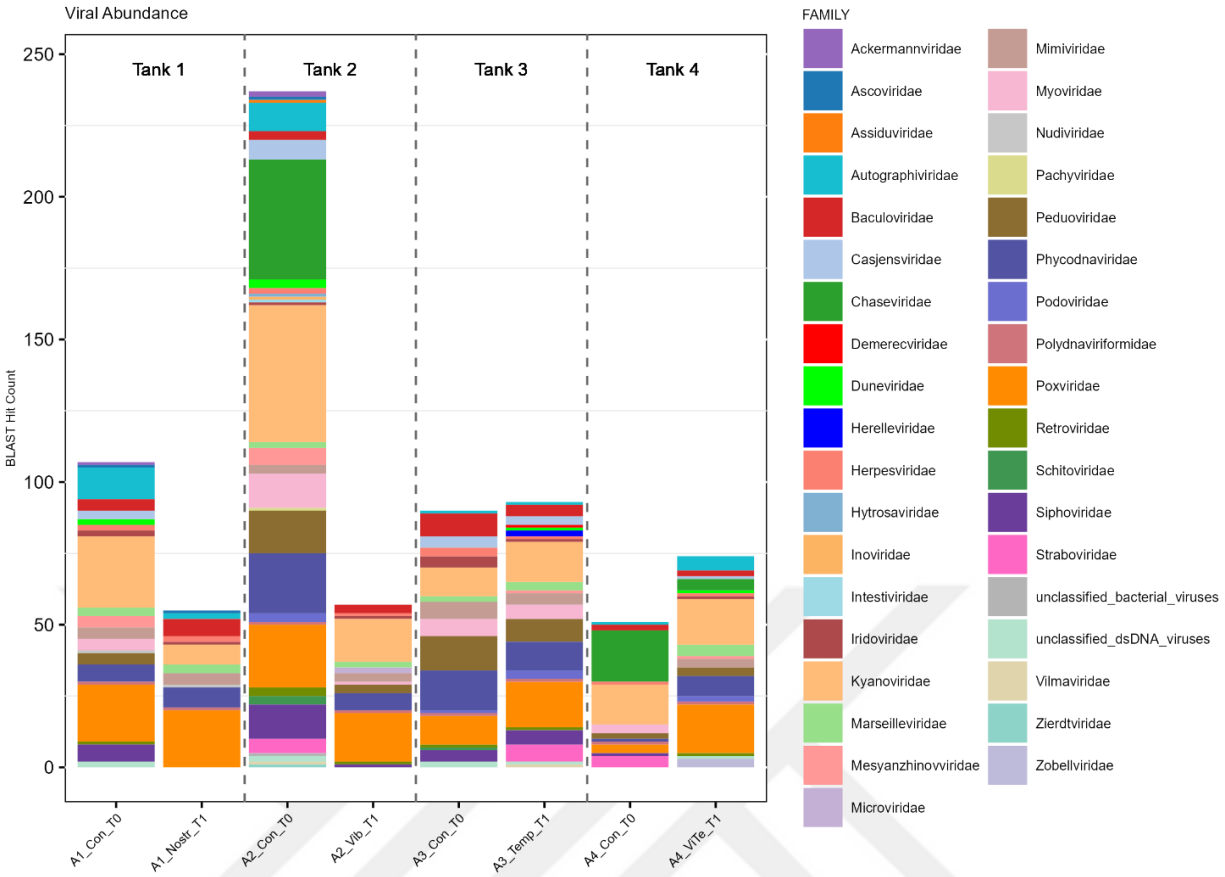


Figure 4.16. Comparison between first and last samplings at family level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold.

The ratio of the viral family Poxviridae increased in all tanks after the introduction of corresponding environmental stress factors (Figure 4.16 & 4.17). The viral compositions were quite similar to each other in Tank 3 between A3_Con_T0 and A3_Temp_T1. Once again, we failed to observe a similarity in the viral compositions of the control groups.

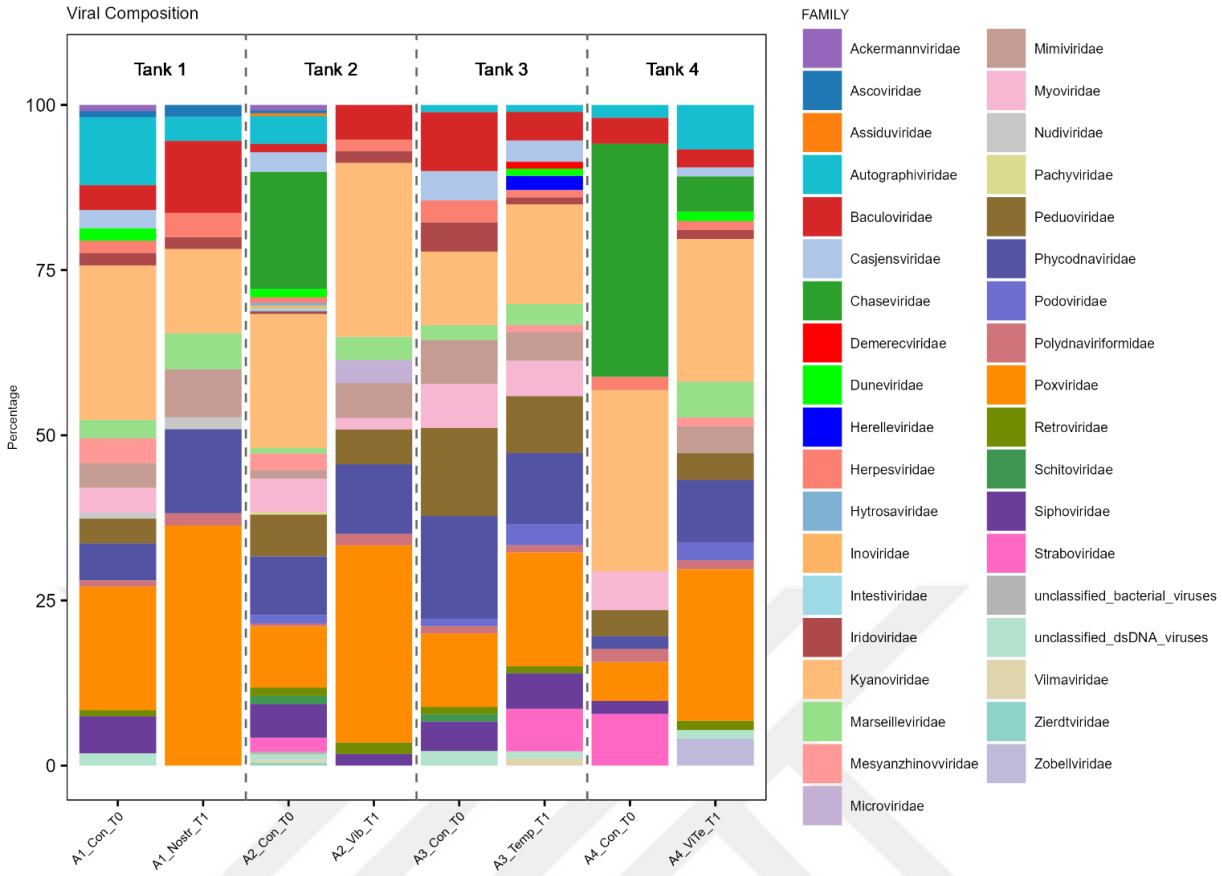


Figure 4.17. The viral composition representing the relative proportion of viral families between samples.

Finally, at genus level, the highest number of unique viral genera was attributed to the sample A2_Con_T0, as depicted in Figure 4.18, as we have also observed at higher taxonomic levels. The number of different viral genera decreased from 88 to 36 in Tank 1. However, the most drastic decrease in the viral diversity at genera level was observed in Tank 2, where the number of viral genera decreased from 173 to 41. Similar to the observations at the family level, the number of viral genera increased from 95 to 98 in Tank 3, and from 32 to 56 in Tank 4.

We see that the most abundant genera are Orthopoxvirus, Alphabaculovirus and Prasinovirus in all tanks. While no significant changes are observed in the contig abundance of Orthopoxvirus for Tank 1 and Tank 2, the contig abundance of Orthopoxvirus increased in Tank 3 and Tank 4 (Figure 4.18). Orthopoxvirus, Prasinovirus, Alphabaculovirus, Lowelvirus, Haifavirus, Shapirovirus, Rhadinovirus and Bracoviriform were the viral genera present in all the samples. Additionally, Jouyvirus and the viral genera listed above were found in all the control groups of the four tanks.

Comparing the contig abundances between the first and last samplings in all tanks, the most drastic change is observed in Tank 2, where the total number of contigs attributed to viral genera decreased from 362 to 68. Tank 1 also displayed a severe decrease in contig abundance from 139 to 64. In contrast, the total number of viral contigs at this taxonomic level increased drastically in Tank 4, from 58 to 90. However, the increase observed in viral contigs for Tank 3 was not as drastic, going from 160 to 165.

In general, while the highest viral diversity and the contig abundance is attributed to the sample A2_Con_T0, the lowest viral diversity and the contig abundance is recorded for the sample A4_Con_T0. This observation underlines the significance of the dissimilarities between the control samples. The most drastic increase in both viral diversity and viral contig abundance was observed in Tank 4 between the samples A4_Con_T0 and A4_ViTe_T1 at all taxonomic levels. On the other hand, the most severe decrease in viral diversity and viral contig abundance was attributed to Tank 2, between A2_Con_T0 and A2_Vib_T1. However, the general trend that we observed was that the viral diversity and the viral contig abundances display a decline in Tank 1 and Tank 2, whereas these values remain relatively unchanged in Tank 3 and increase in Tank 4. It is worth noting that Tank 1 and Tank 2 displayed a similar pattern in terms of the changes observed in the viral diversities and contig abundances over time.

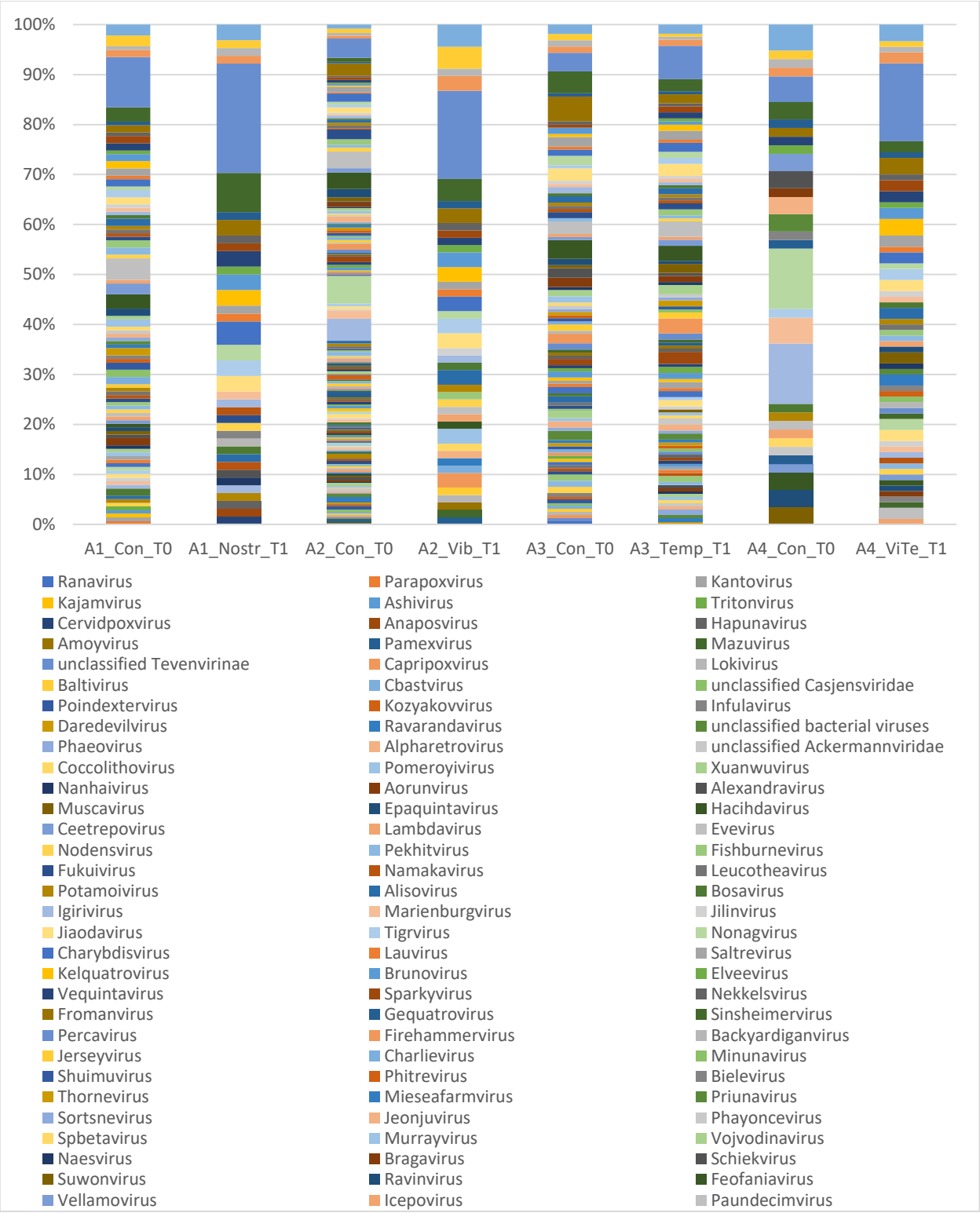


Figure 4.18. The viral composition representing the relative proportion of viral genera between samples.

4.4.1. Application of 80% Similarity Threshold

The lack of similarity within the control groups in the non-filtered data set led us to question the presence of false positive data. We encountered the appearance and disappearance of certain taxa between samples, and the very low abundance of such taxa might simply stem from a bioinformatics related or sequencing library error. Therefore, as mentioned above, in our next step, we aimed to detect and investigate the anomalous data and tried to obtain a coherent pattern among all samples.

Our strategy in detecting and eliminating the false positive data was to increase the similarity threshold that was initially set to 75%. While some taxa appeared in all four of the viral protein databases with similarity percentage higher than 75%, some were only observed in a single viral database. Therefore, we targeted the viral taxa that only appeared in one or two viral protein databases, and we set a new similarity threshold for such taxa as 80%.

The resulting bar plots post-filtration can be found in Figure 4.19-4.20, at family and genus levels. 37 different families were recorded prior to the data filtration process, and this value decreased to 27 when we applied a stricter threshold of 80%. However, elimination of the families that are very low in abundance resulted in a very similar pattern between the plots obtained with filtered and non-filtered data, despite a reduction in the total family counts (Figure 4.19). The most drastic decrease observed in the number of different viral families in samples was for A2_Con_T0, where 11 viral families were eliminated with the data filtration process. In the rest of the samples, a maximum of five viral families were eliminated. Therefore, the abundance of the potential false positive data was the highest for the sample A2_Con_T0.

A total of 16 viral contigs at family level were eliminated from the sample A2_Con_T0 with the data elimination. Although the number of viral families that were eliminated from the sample A3_Temp_T1 was only five with the application of a stricter threshold, a total of 17 viral contigs were eliminated after the data filtration, which is the highest number of eliminated viral contigs among all samples.

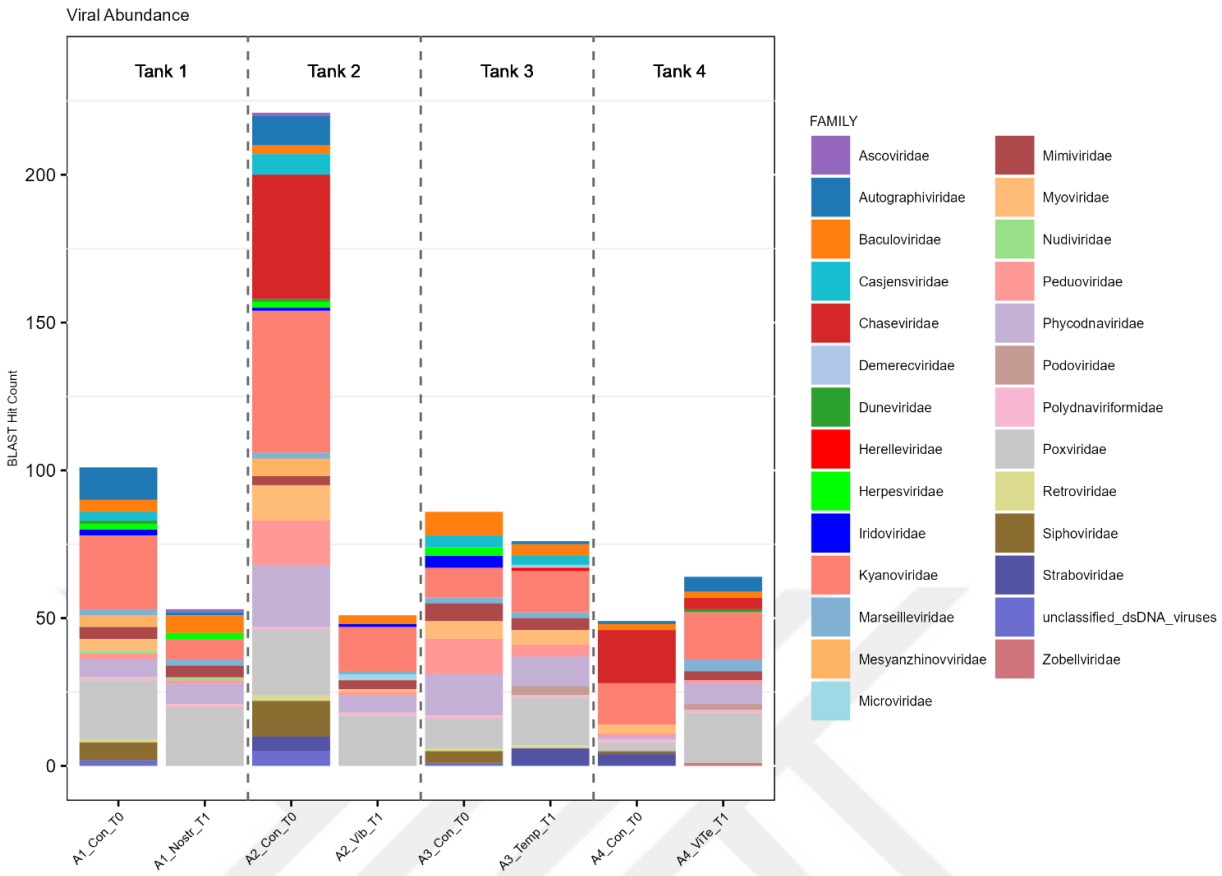


Figure 4.19. Comparison between first and last samplings at family level within each tank for the number of viral contigs that matched to the viral protein database with a stricter threshold of 80% similarity score.

The viral compositions in Tank 3 did not change drastically between A3_Con_T0 and A3_Temp_T1, as shown in Figure 4.20. A similar pattern was observed for Tank 1 as well, where the viral compositions are roughly the same between A1_Con_T0 and A1_Nostr_T1. However, the viral composition of A1_Nostr_T0 was dominated by Poxviridae family members, as opposed to the viral composition of A1_Con_T0 being dominated mostly the family Kyanoviridae. Notably, the most severe change in the viral composition was observed for Tank 2 and Tank 4.

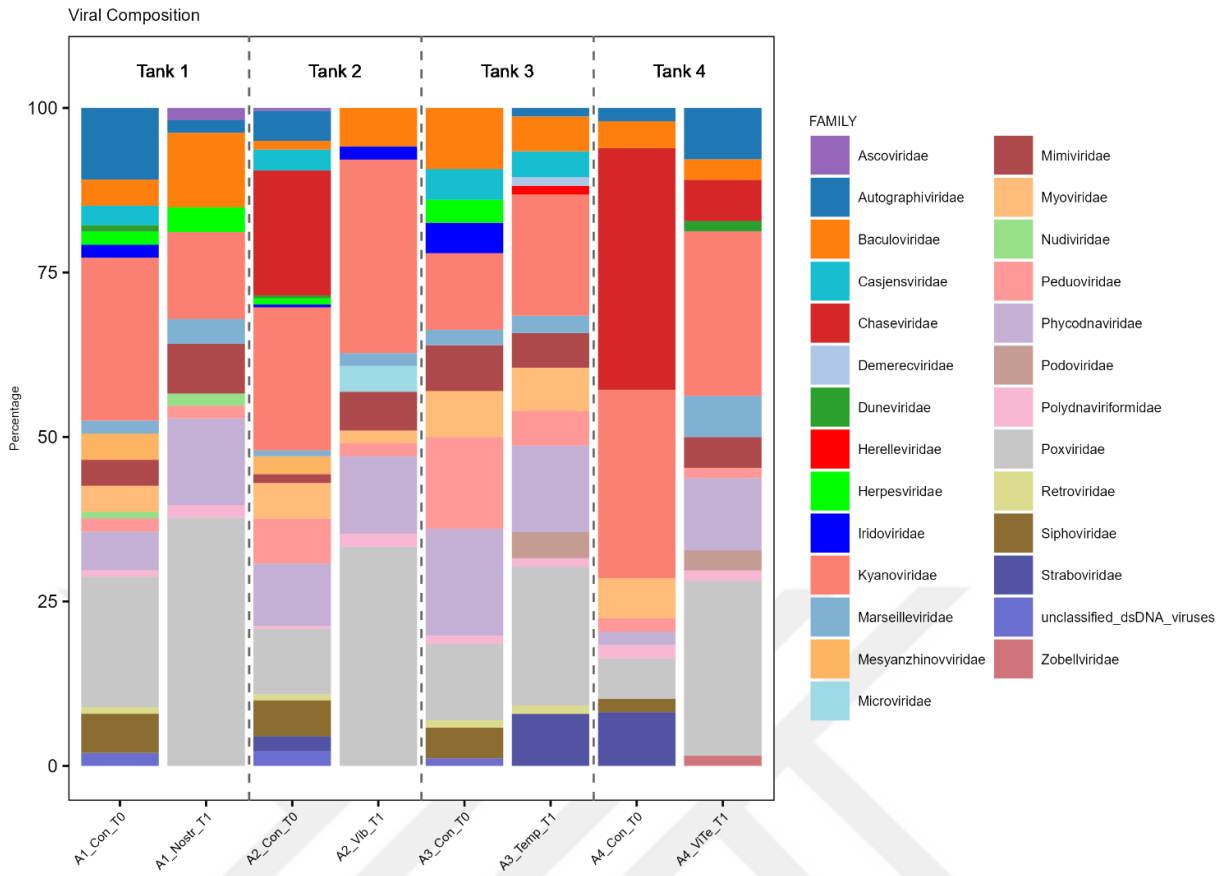


Figure 4.20. Viral composition distribution after application of 80% threshold and data filtration at family level.

Moving to the genus level in Figure 4.21, 233 different viral genera were recorded before the application of 80% threshold and data filtration, whereas 215 viral genera were obtained after the filtration process. The general trend between these viral genera did not significantly change before and after the data filtration process. The highest and total of 21 contigs were eliminated from A2_Con_T0 with the data filtration process. The number of viral genera eliminated from the same sample was also the highest, a total of 14 viral contigs were eliminated with a stricter threshold.

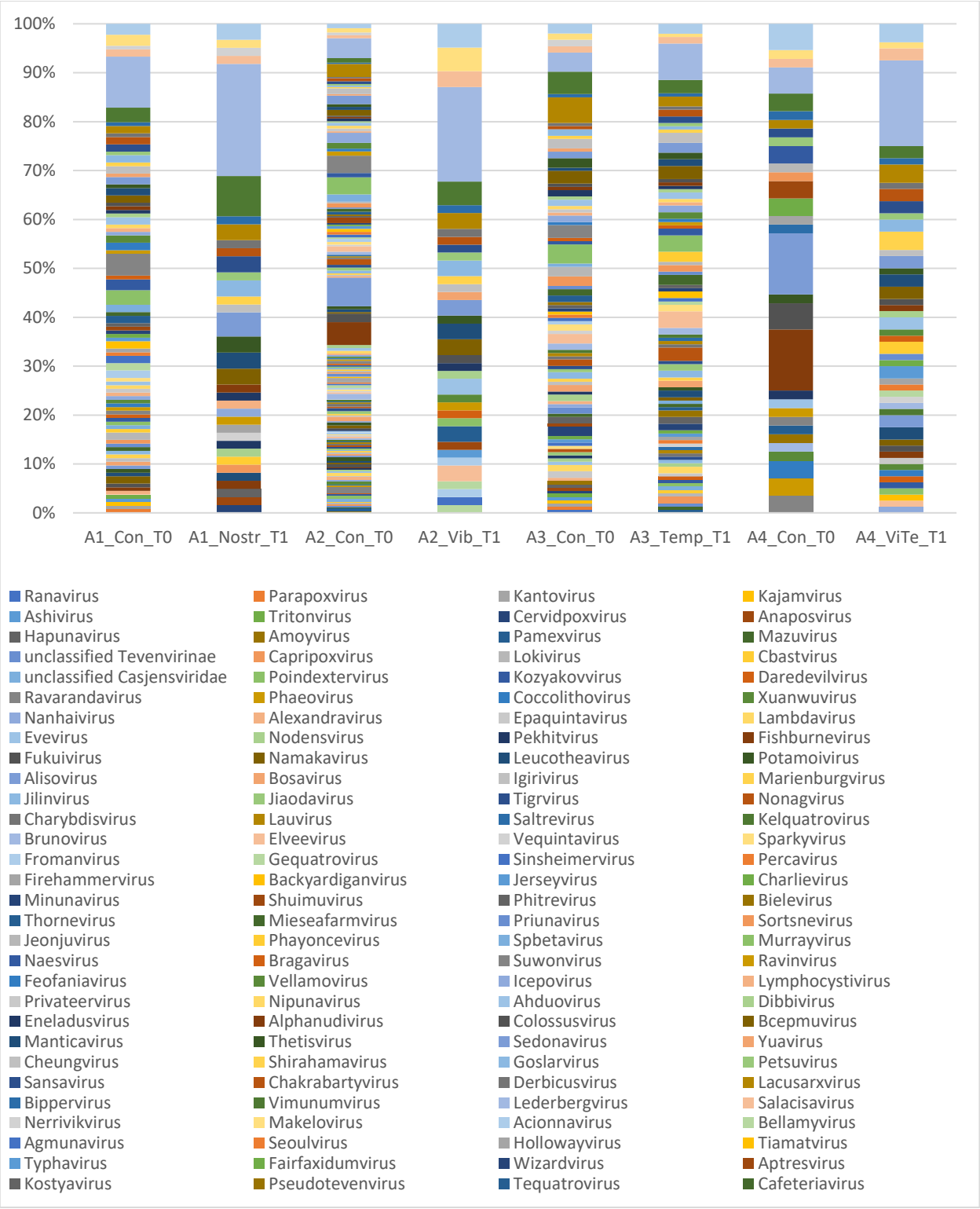


Figure 4.21. The viral composition representing the relative proportion of viral genera between samples after the application of 80% threshold.

To visualize the similarities and dissimilarities among the samples, Principal Coordinate Analysis (PCoA) is performed with filtered and unfiltered data at family and genus levels (Figure 4.22 & 4.23, respectively). Although we did not observe a pattern that we expected after the data filtration, the total percentage of variation explained slightly increased after the data elimination at family level in Figure 4.22 (b). We also see the clustering of the samples at t=1, however we again failed to see the clustering of control groups as the sample A4_Con_T0 distantly located from other control samples within the graph (Figure 4.22 (b)).

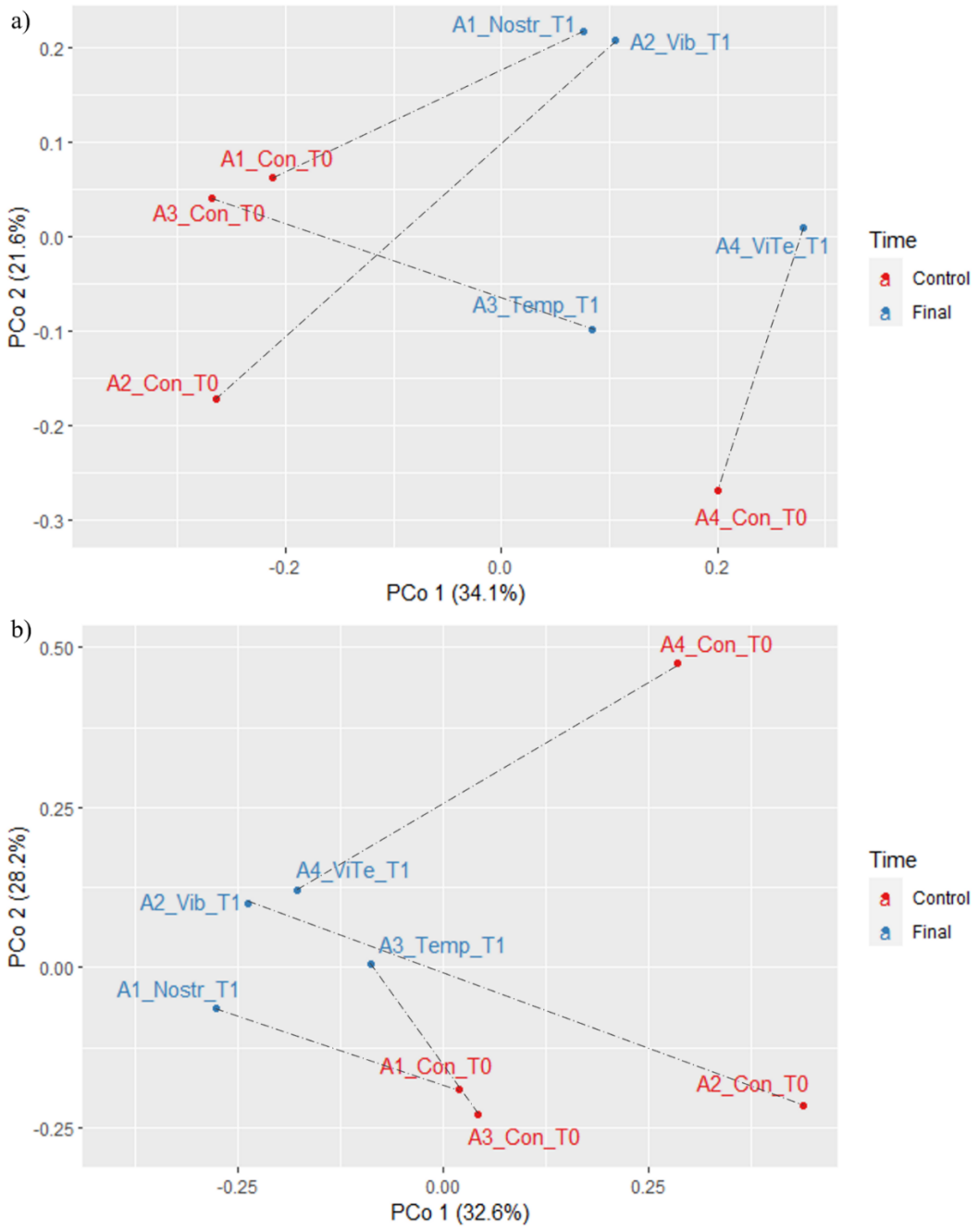


Figure 4.22. PCoA plots showing the clustering of viral families (a) before the data filtration and (b) after the data filtration process.

The PCoA plots did not represent a coherent pattern at genus level as seen in Figure 4.23. We fail to obtain a proximity between the control samples in PCoA graphs performed with data before or after the filtration. The total percentage of variation explained does not seem to change after the data filtration as well. Since the taxa branches and diversifies as we move towards to the lowest taxonomic level, the increased diversity presumably counteracts with the data filtration process. Overall, the PCoA graph at the family level after the data elimination process, presented above, delivers more promising results when compared to the genus level.

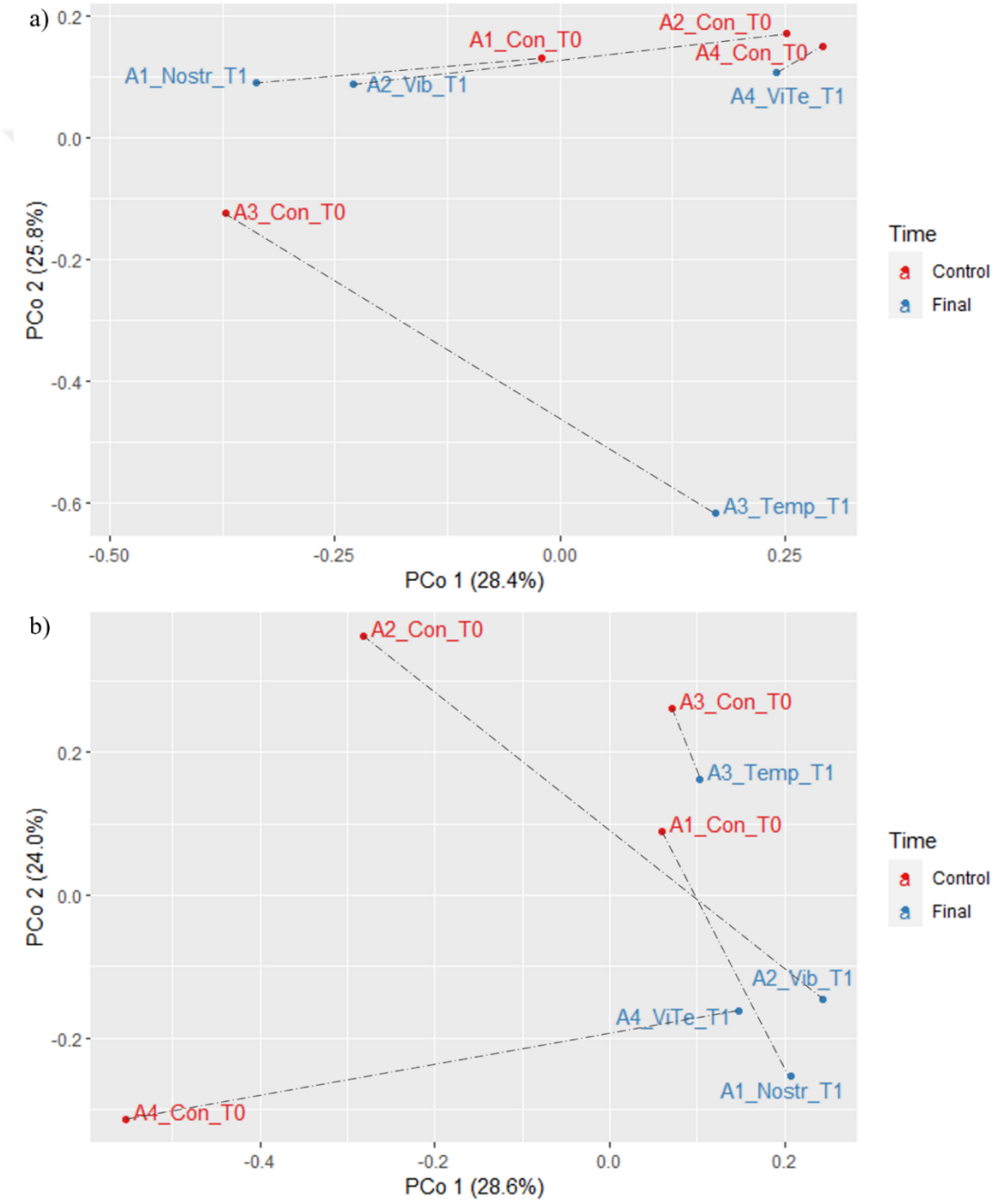


Figure 4.23. PCoA plots showing the clustering of viral genera (a) before the data filtration and (b) after the data filtration process.

4.4.2. Nucleocytoplasmic Large DNA Viruses (NCLDVs)

The abundance of Nucleocytoplasmic Large DNA Viruses (NCLDV) has also been recorded for each sample. In general, the number of NCLDV members decreased in Tank 1, Tank 2 and Tank 3 while they increased in Tank 4 as depicted in Figure 4.24. In Tank 1, the total number of NCLDV viruses decreased from 38 to 36. Similarly, the number of NLDV viruses decreased from 38 to 35 in Tank 3. However, the most drastic decrease was recorded for Tank 2 where the number of NCLDV family members decreased from 56 to 29. The contig abundance of the Poxviridae family increased in Tank 3 and Tank 4; conversely, the contig abundance of members from the same family displayed a decreasing trend in Tank 2 (Figure 4.24). On the other hand, the abundance of the Poxviridae family remained the same in Tank 1. The Ascoviridae family is observed only in Tank 1 and Tank 2; it disappeared from Tank 2 at t=1 and did not change in Tank 1. Meanwhile, Pandoravirus disappeared from Tank 1 and Tank 2, but emerged in Tank 4. Notably, the most drastic increase in NCLDV members was recorded for A4_ViTe_T1 in Tank 4, where the number of NCLDV viruses increased from 4 to 33 after the introduction of heat stress and coral pathogen. Interestingly, some members of the NCLDV viruses emerged from zero in Tank 4, specifically the families; Mimiviridae, Iridoviridae, Marseilleviridae and Pandoravirus. On the other hand, the disappearance of some NCLDV members was also recorded in Tank 1 and Tank 2. While Panodravirus was not observed in Tank 1 and Tank 2, Ascoviridae family members only disappeared in Tank 2.

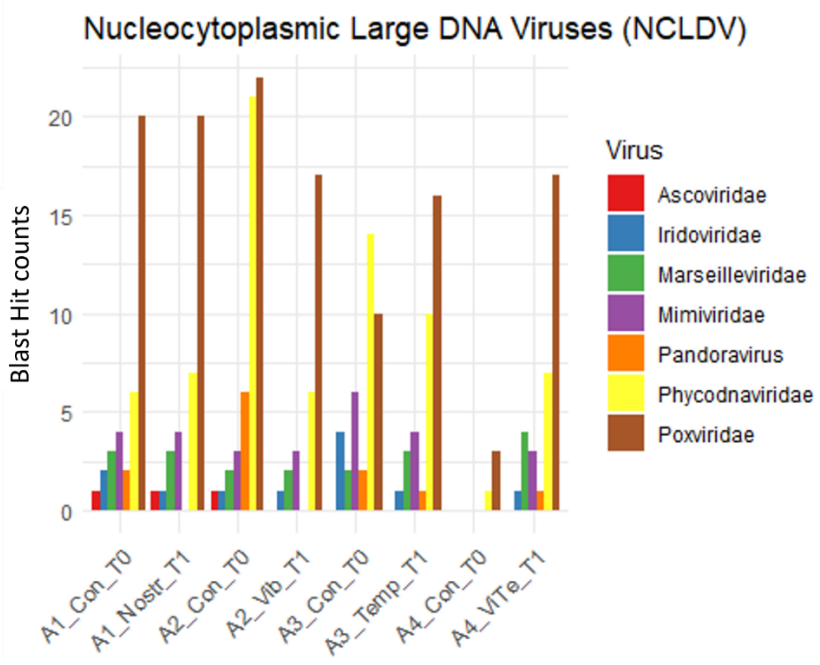


Figure 4.24. The viral contig abundance of the members of Nucleocytoplasmic Large DNA viruses (NCLDV) in each sample.

4.5. Read Normalization

A detailed report on the contig assembly process computed with MegaHit is presented in Table 4.6. All values, from number of contigs to N50, seem to increase during contig assembly for all the samples when certain number of reads were eliminated in each of the sample. The highest total base pair (in yellow) was recorded for sample A3_Con_T0, which originally belonged to sample A2_Con_T0 before the read normalization process. The lowest total base pair (in blue), on the other hand, was still associated with sample A2_Vib_T1.

Table 4.9. The detailed output of contigs from MegaHit after the read normalization step. The number of contigs, total base pairs, length of the assembled contigs, average length of the contigs and the N50 lengths are presented. The highest values are depicted in yellow, and the lowest values are depicted in blue.

Samples	Contigs	Total bp	Min bp	Max bp	Avg bp	N50 bp
A1_Con_T0	834666	809758874	200	162120	970	2222
A1_Nostr_T1	968457	868362635	200	90552	896	2235
A2_Con_T0	1734174	1380107086	200	593838	795	970
A2_Vib_T1	507715	605532234	200	103950	1192	3896
A3_Con_T0	3017951	1894212788	200	217545	627	709
A3_Temp_T1	1615534	1399758332	200	217500	866	1275
A4_Con_T0	632039	653665270	200	175518	1034	3599
A4_ViTe_T1	622154	658912476	200	177416	1059	3503

The alignment rates of the contig mapping process is listed in Table 4.10. While most of the values are above 80%, the alignment rate recorded for the contig mapping step of A2_Con_T0 is as low as 59.93%. This finding verifies once again the ineffectivity of the contig assembly step for the sample A2_Con_T0 before and after the read normalization step.

Table 4.10. The alignment rates calculated from the mapping of contigs after the read normalization to their trimmed reads they originated from.

Sample	Alignment rate
A1_Con_T0	87.75 %
A1_Nostr_T1	93.83 %
A2_Con_T0	59.93 %
A2_Vib_T1	94.35 %
A3_Con_T0	84.53 %
A3_Temp_T1	82.11 %
A4_Con_T0	95.52 %
A4_ViTe_T1	91.09 %

As a result of the decrease observed in the contig numbers in each sample after the read normalization step, the number of viral contigs and the number of viral contigs that has a percent similarity value higher than 75 also decreased with the reduced number of reads in each sample (see Table 4.8 and Table 4.11). Since the most abundant viral contigs were observed in the sample A2_Con_T0 prior to the read normalization step, the most drastic decrease in the number of viral contigs and the number of viral contigs with similarity value higher than 75% was unexpectedly observed for the same sample. In general, the read normalization step resulted in increased proximity between first and last sampling samples in Tank 1, Tank 2 and Tank 4, in terms of the abundance of the viral contigs having similarity value higher than 75%.

Table 4.11. Comparison of number of contigs, number of viral contigs and number of viral contigs that has similarity value higher than 75%, between samples after data normalization.

Sample	# of Contigs	# of Viral Contigs	> 75%
A1_Con_T0	834666	2263	88
A1_Nostr_T1	968457	1360	64
A2_Con_T0	1734174	7424	211
A2_Vib_T1	507715	683	36
A3_Con_T0	3017951	5883	128
A3_Temp_T1	1615534	5097	117
A4_Con_T0	632039	688	57
A4_ViTe_T1	622154	1457	52

The abundance of the viral families recorded for each sample slightly changed after data normalization. The control groups in Tank 1, Tank 3, and Tank 4 have become more similar to each

other in terms of their contig abundances. On the other hand, viral diversity and contig abundance were still the highest in sample A2_Con_T0 (Figure 4.25). Interestingly, the contig abundance of the viral families decreased from 82 to 67 in Tank 4 between the samples A4_Con_T0 and A4_ViTe_T1, in contrast to the increasing pattern, from 51 to 70 viral contigs, observed for the same tank prior to the read normalisation step. On the other hand, we continue to obtain a decrease in the number of contigs at family level in Tank 1 and Tank 2, whereas a slight increase in Tank 3 after the read normalization process. A total of 11 viral families were found present in all of the samples after the read normalization step. In addition to the six viral families, namely Kyanoviridae, Poxviridae, Phycodnaviridae, Baculoviridae, Herpesviridae and Polydnaformiviridae that were observed in all of the samples with pre-normalized data, the viral families of Mimiviridae, Iridoviridae, Autographiviridae, Peduoviridae and Marseilleviridae were also observed in all of the samples in the post-normalized data. On the other hand, Retroviridae and Straboviridae families together with the viral families listed above were present in all of the control samples.While the number of viral families observed in all of the tanks was recorded as 37, this value decreased to 33 after normalisation. However, the highest contig abundance of the Kyanoviridae and Poxviridae families within samples remained unchanged.

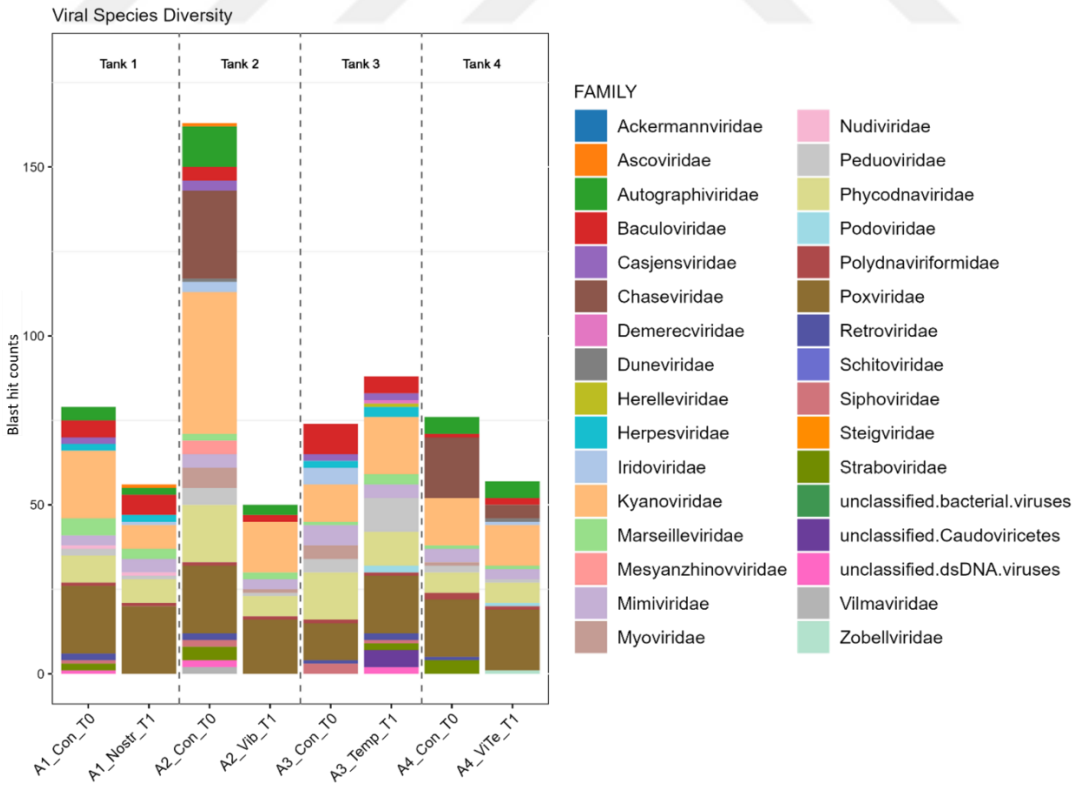


Figure 4.25. Comparison between first and last samplings at family level within each tank for the number of viral contigs that matched to the viral protein database with higher than 75% similarity threshold after the read normalization step.

The viral composition of the families were similar to those before the read normalisation, however the most drastic changes was observed again for Tank 4. Especially, the increase observed in Poxviridae composition in Figure 4.17 has become less apparent in Tank 4 after read normalisation as depicted in Figure 4.26.

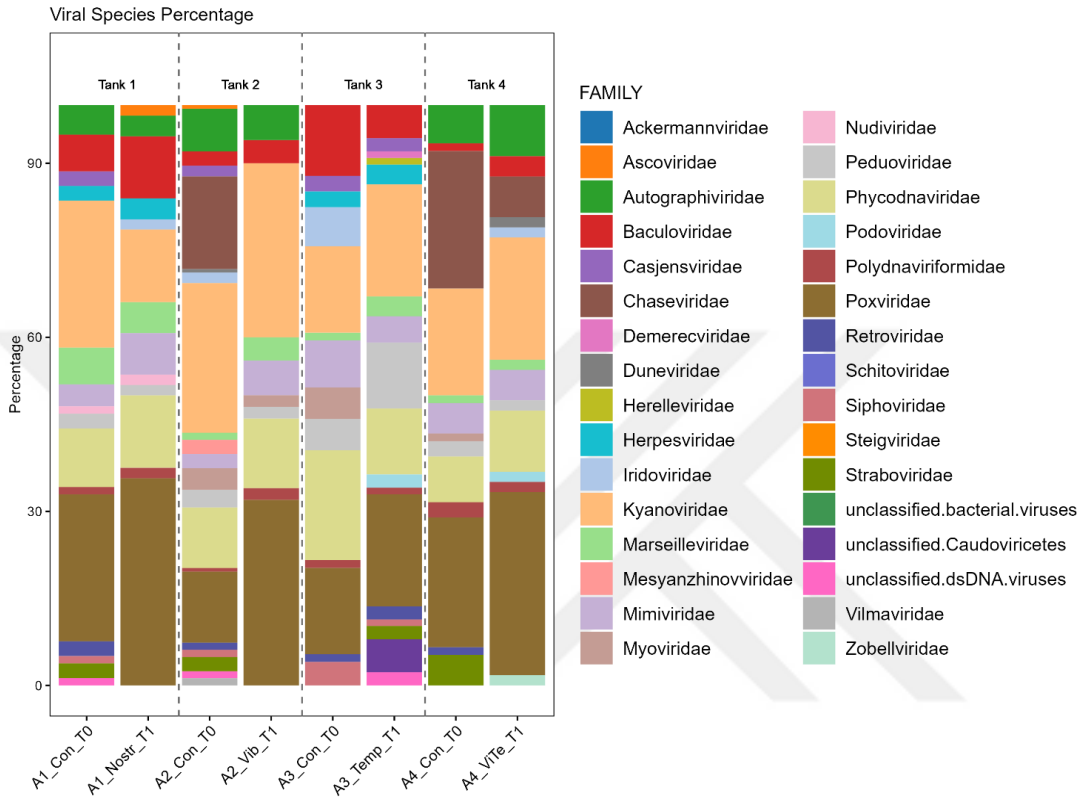


Figure 4.26. The comparison of viral compositions at family level between samples after the read normalization process.

4.6. Statistical Analyses

Figure 4.27 represents the Shannon diversity indexes for each sample at family level. While sample A2_Con_T0 has the highest Shannon diversity index with 2.91, the lowest Shannon diversity index belongs to the sample A2_Vib_T1 which is recorded as 1.74.

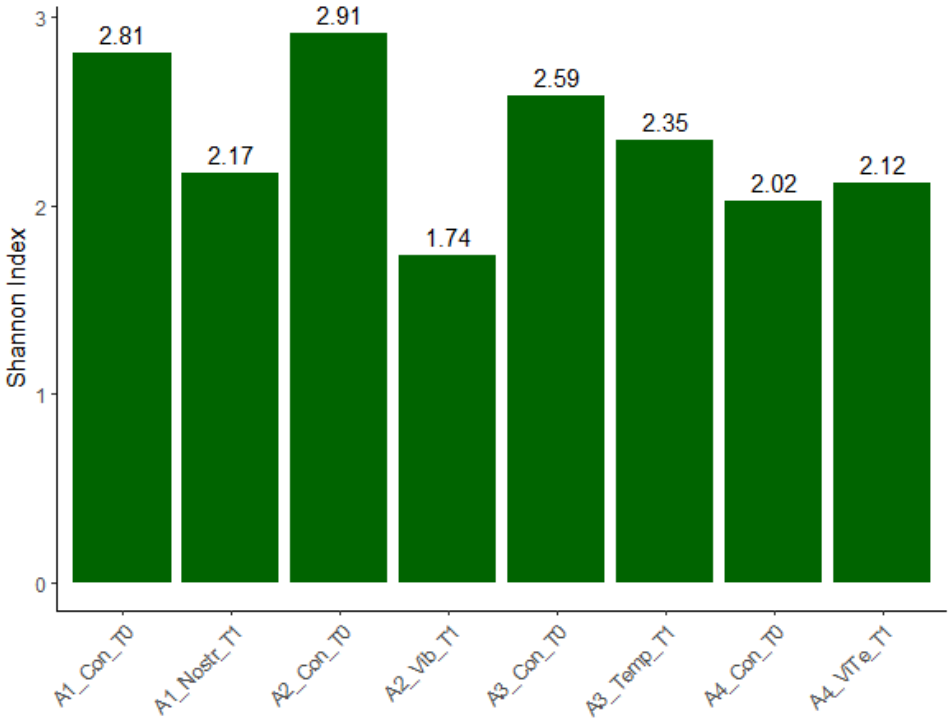


Figure 4.27. Shannon diversity indexes for each sample at family level.

The computed alpha diversities for all samples at genus level are displayed as plots, based on the calculated Jaccard distances among samples, in Figure 4.28. The difference in the viral diversities of control samples is quite drastic, comparing the viral diversities between the first and last samplings. While A4_Con_T0 has the lowest alpha diversity index, the highest alpha diversity is attributed to the sample A2_Con_T0. The alpha diversities of A1_Con_T0, A3_Con_T0 and A3_Temp_T1 resulted in similar values. The alpha diversity indexes of the samples A4_Con_T0, A1_Nostr_T1, A4_ViTe_T1 and A2_Vib_T1 also lined closer to each other in the plot (Figure 4.28).

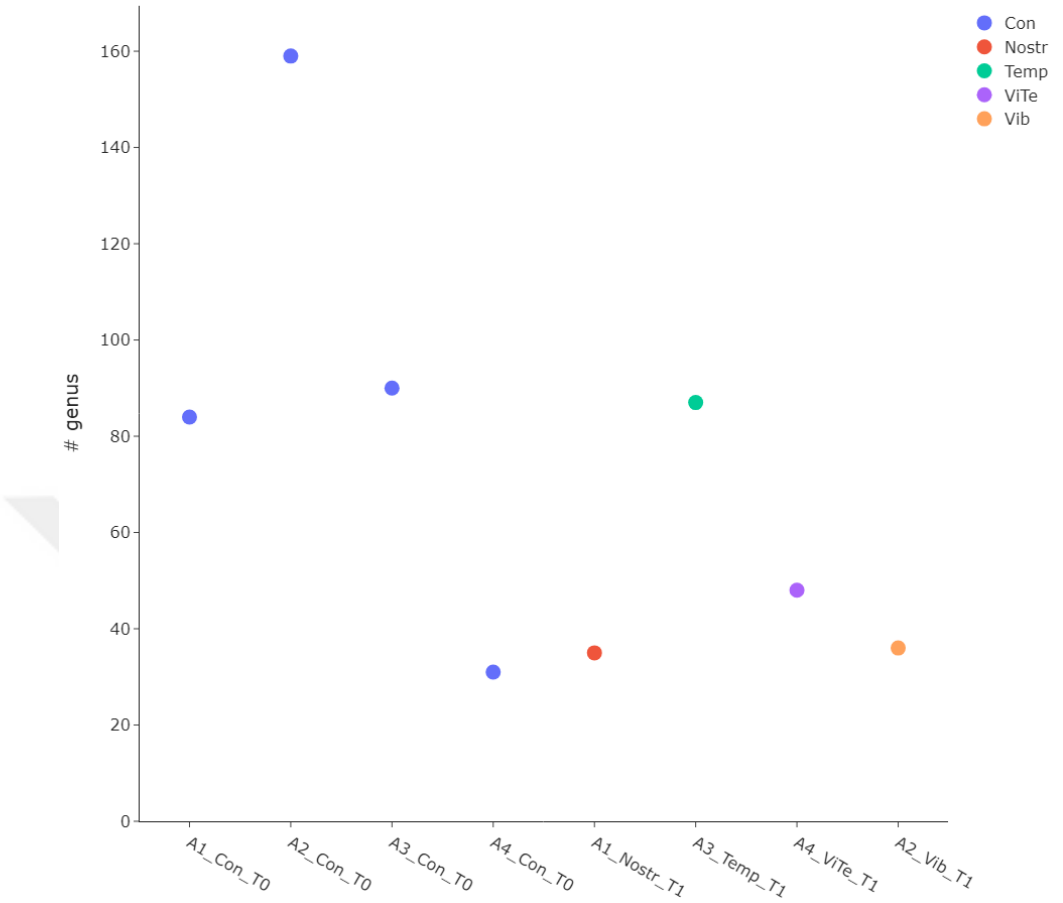


Figure 4.28. Alpha diversities calculated for each sample at genus level based on Jaccard distances.

Based on the calculated Jaccard distances between samples, a heatmap of the dissimilarity matrix is generated in Figure 4.29. The negative R value of -0.10606 indicates that the dissimilarty among control samples is higher than the dissimilarities between the first and last samplings. The similarity between the samples is the highest between A3_Con_T0 and A3_Temp_T1. On the other hand, the dissimilarity is the highest between A3_Con_T0 and A4_Con_T0. The p value is calculated as 0.557, which is noninformative in terms of statistical significance.

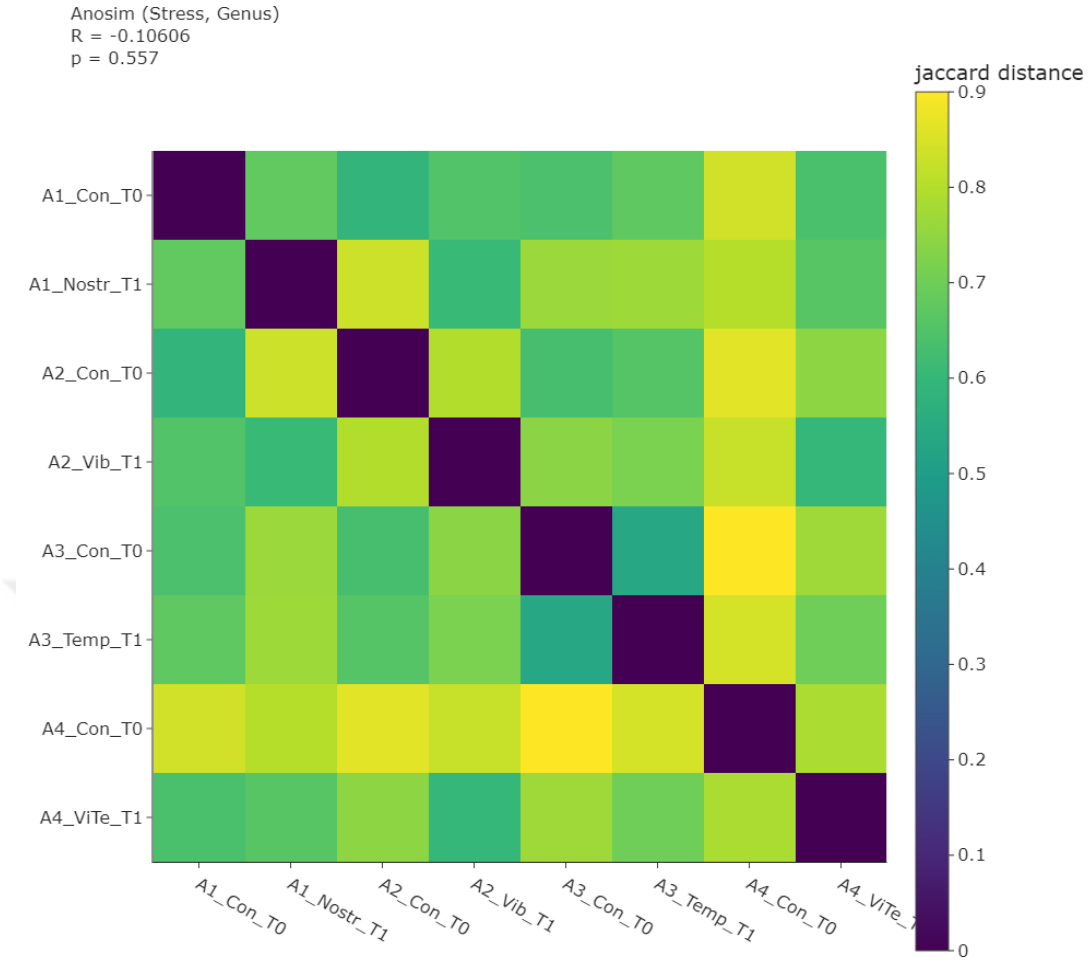


Figure 4.29. Beta-diversity heatmap that represents the similarities/dissimilarities between samples based on Jaccard distances. The ANOSIM test was performed utilizing stress as the selected factor at genus level.

The clustering of the samples based on the Jaccard similarities and dissimilarities is presented in a dendrogram in Figure 4.30. We see the clustering of the samples A3_Con_T0 and A3_Temp_T1 once again. Addidionally, the control groups of Tank 1 and Tank 2 also clustered together in the dendrogram, and the samples A2_Vib_T1 and A4_ViTe_T1 clustered together as well. The sample A4_Con_T0 is acting as an outlier in the dendrogram shown below.

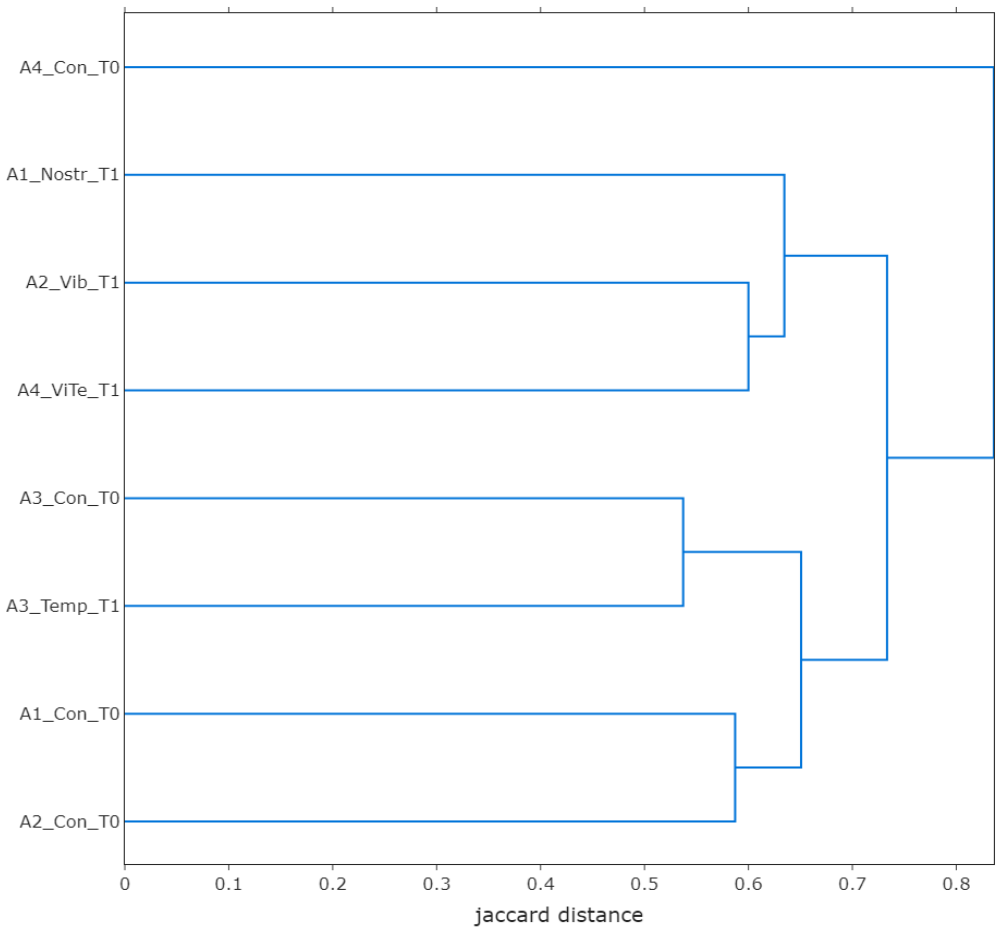


Figure 4.30. The dendrogram displaying the clustering of the samples based on the calculated average distances between the samples, using the Jaccard distance method.

The PCoA plot depicted in Figure 4.31 was generated with the dominant viral families obtained in our data and the coral families that are constantly found on the surface of coral colonies in the literature. These viral families consisted of; Phycodnaviridae, Poxviridae, Polydnaviridae, Baculoviridae, Herpesviridae, Kyanoviridae, Microviridae, Podoviridae, Retroviridae, Siphoviridae, Iridoviridae, Myoviridae, and Mimiviridae. The control groups once again appeared distant from each other, A4_Con_T0 acting as an outlier. The first and last samplings of Tank 3 clustered close to each other.

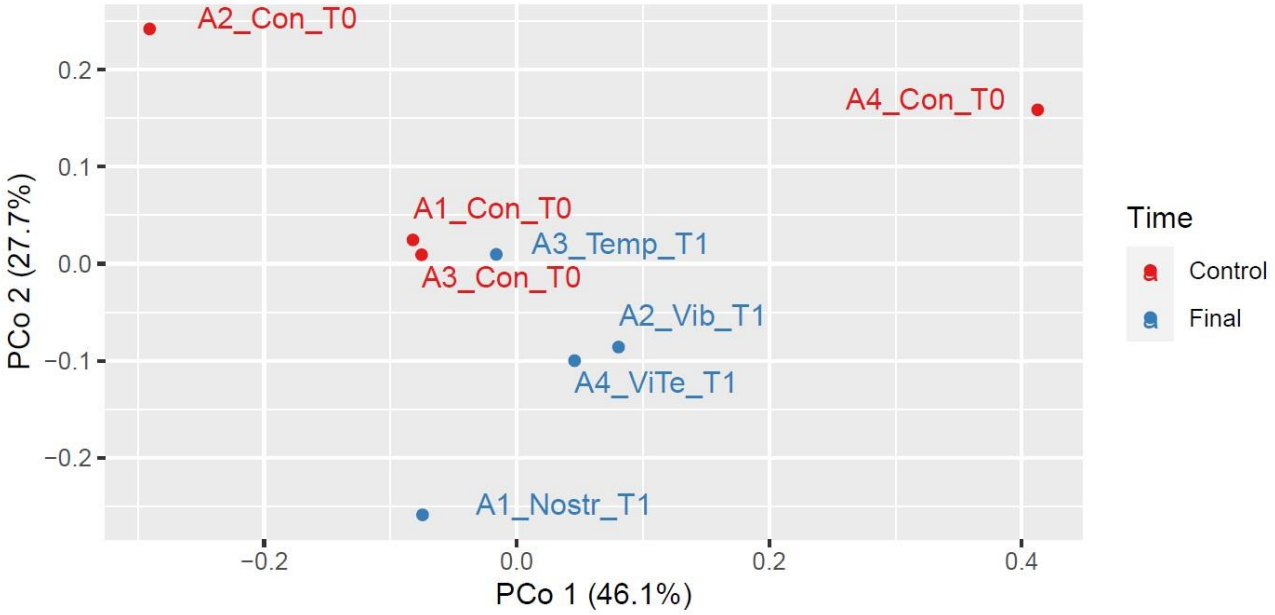


Figure 4.31. The PCoA plot showing the clustering of samples based on the dominant viral families that are constantly found on the surface of coral colonies. Control (red) represents the first sampling, Final (blue) represents the last sampling samples.

4.7. Viral-Host Search

To have a better understanding about the virome structures of each sample, the possible host of each virus was examined through Virus-Host Database (Mihara et al., 2016). Initially, the changes in the abundance of eukaryotic and bacterial viruses (phages) were examined for each sample. In Tank 1 and Tank 2 both eukaryotic viruses and phages displayed a decrease as depicted in Table 4.12 below. The abundance of the eukaryotic viruses decreased in Tank 3, however the abundance of these viruses increased in Tank 4. On the other hand, while the abundance of phages increased in Tank 3, the number of phages observed in Tank 4 remained relatively the same between the first and last samplings.

In addition to the comparison between eukaryotic viruses and phages, the changes in cyanobacteria infecting viruses have also been examined. The number of viruses targeting cyanobacteria, an important component of the coral ecosystem, increased in Tank 3 and Tank 4, while decreasing in Tank 1 and Tank 2, as depicted in Table 4.12. It is worth highlighting that, although the number of bacteria infecting viruses remained the same in Tank 4, the cyanobacteria infecting viruses increased.

Table 4.12. The abundance of viruses in each sample based on the type of organisms they infect.

Samples	Bacteria	Eukaryote	Cyanobacteria	Gammaproteo-bacteria	Alphaproteo-bacteria
A1_Con_T0	93	46	42	28	9
A1_Nostr_T1	30	40	8	5	1
A2_Con_T0	276	61	62	142	26
A2_Vib_T1	32	34	17	10	3
A3_Con_T0	109	47	16	50	10
A3_Temp_T1	124	40	19	52	11
A4_Con_T0	50	7	14	35	1
A4_ViTe_T1	51	36	24	14	5

4.8. Mitochondrial Genome Assembly

The mitogenome annotation step was only performed for seven of the eight samples, as the mitogenome assembly failed for sample A2_Con_T0. This sample was unable to use the *C. caespitosa* COX1 gene as a seed during the assembly, suggesting once again the absence or the very low abundance of the genomic content related with the coral. All annotated mitogenomes contained the same set of genes. A total of 19 genes were successfully annotated as depicted for sample A1_Con_T0 in Figure 4.32. These genes consist of; *nad6, atp6, trnG(ggc), trnL2(tta), nad4, rrnS, cox3, cox2, nad4l, nad3, nad5-0_b, trnW(tga), atp8, cox1, trnM(atg), rrnL, nad5-0_a, nad1, cob* and *nad2*. On the other hand, the following genes were not found within the assembled mitogenomes; *trnaA, trnC, trnE, trnF, trnH, trnI, trnK, trnL1, trnN, trnP, trnQ, trnR, trnS1, trnS2, trntT, trnY* and *trnV*. Additionally, the nad5 gene was duplicated into nad5-0_and nad5-0_b within the mitogenome.

Name	Start	Stop	Strand	Length	Structure
nad6	23	571	+	549	
atp6	760	1248	+	489	
trnG(ggc)	1217	1261	-	45	svg ps
trnL2(tta)	1321	1377	+	57	svg ps
nad4	1470	2564	+	1095	
rrnS	2697	3605	+	909	svg ps
cox3	3599	4375	+	777	
cox2	4411	5055	+	645	
nad4l	5112	5363	+	252	
nad3	5396	5707	+	312	
nad5-0_b	5771	6868	+	1098	
trnW(tga)	6873	6942	+	70	svg ps
atp8	6946	7140	+	195	
cox1	7158	8681	+	1524	
trnM(atg)	8714	8785	+	72	svg ps
rrnL	8925	10626	+	1702	svg ps
nad5-0_a	10841	11362	+	522	
nad1	11498	12418	+	921	
cob	12427	13557	+	1131	
nad2	13787	14689	+	903	

tRNA gene rRNA gene protein coding gene



Figure 4.32. Annotated mitogenome of *C. caespitosa* in sample 306 (A1_Con_T0).

The annotated mitogenome of *F. abdita* is presented in Figure 4.33. Compared to the annotated genes in A1_Con_T0 in Figure 4.32, 17 genes were represented in annotated mitogenome of *F. abdita* as shown in Figure 4.33. The genes that were annotated in *F. abdita* mitogenome were as follows; *trnM(atg), rrnL, nad5_a, nad1, cob, and-2_0, nad6, atp6, nad4, nad2-1, rrnS, cox3-0, cox1-1, cox3-1, cox2, nad4l, nad3, nad5_b, trnW(tga), atp8, cox1-0* and *rrnS*. Additional to the genes that were not found in the A1_Con_T0 mitogenome, genes; *trnG* and *trnL2* were also not found in *F. abdita* mitogenome. However, the number of duplicated genes was higher in *F. abdita* mitogenome. While *nad5* gene was duplicated in both mitogenomes, *cox1, cox3, nad2* and *rrnS* genes were found to be duplicated in only *F. abdita* mitogenome.

Name	Start	Stop	Strand	Length	Structure
trnM(atg)	1	72	+	72	svg ps
rrnL	300	1996	+	1697	svg ps
nad5_a	2049	2759	+	711	
nad1	2890	3810	+	921	
cob	3819	4949	+	1131	
nad2-0	5167	6081	+	915	
nad6	6272	6820	+	549	
atp6	7009	7497	+	489	
nad4	7719	8876	+	1158	
nad2-1	8975	9034	+	60	
rrnS	9085	9995	+	911	svg ps
cox3-0	10114	10890	+	777	
cox1-1	11262	11315	+	54	
cox3-1	12057	12158	+	102	
cox2	12402	13046	+	645	
nad4l	13079	13354	+	276	
nad3	13387	13698	+	312	
nad5_b	13753	14859	+	1107	
trnW(tga)	14864	14933	+	70	svg ps
atp8	14937	15131	+	195	
cox1-0	15152	16672	+	1521	
rrnS	16988	17769	-	782	svg ps

tRNA gene rRNA gene protein coding gene



Figure 4.33. Annotated mitogenome of *Favites abdita*.

The comparison of presence and absence of the annotated genes in two stony coral, *C caespitosa* and *F. abdita* is given in Table 4.13. The number of genes annotated in *C. caespitosa* mitogenome is higher than the number of annotated genes in *F. abdita* mitogenome. The transfer RNA genes of *trnG(ggc)* and *trnL2(tta)* were absent in *F. abdita* mitogenome.

Table 4.13. The presence-absence of annotated genes in *Cladocora caespitosa* and *Favites abdita*.

Gene	***C. caespitosa***	***F. abdita***
Nad6	✓	✓
Atp6	✓	✓
trnG(ggc)	✓	×
trnL2(tta)	✓	×
Nad4	✓	✓
rrnS	✓	✓
Cox3	✓	✓
Cox2	✓	✓
Nad4l	✓	✓
Nad3	✓	✓
Nad5	✓	✓
trnW(tga)	✓	✓
Atp8	✓	✓
Cox1	✓	✓
trnM(atg)	✓	✓
rrnL	✓	✓
Nad1	✓	✓
Cob	✓	✓
Nad2	✓	✓

The circular annotated mitochondrial genome of *C. caespitosa* is presented in Figure 4.34, with the data obtained from the sample A1_Con_T0. The annotated mitogenome consists of 14,856 bp. The order of the genes are quite similar to the linearly annotated mitogenome of the same sample in Figure 4.32.

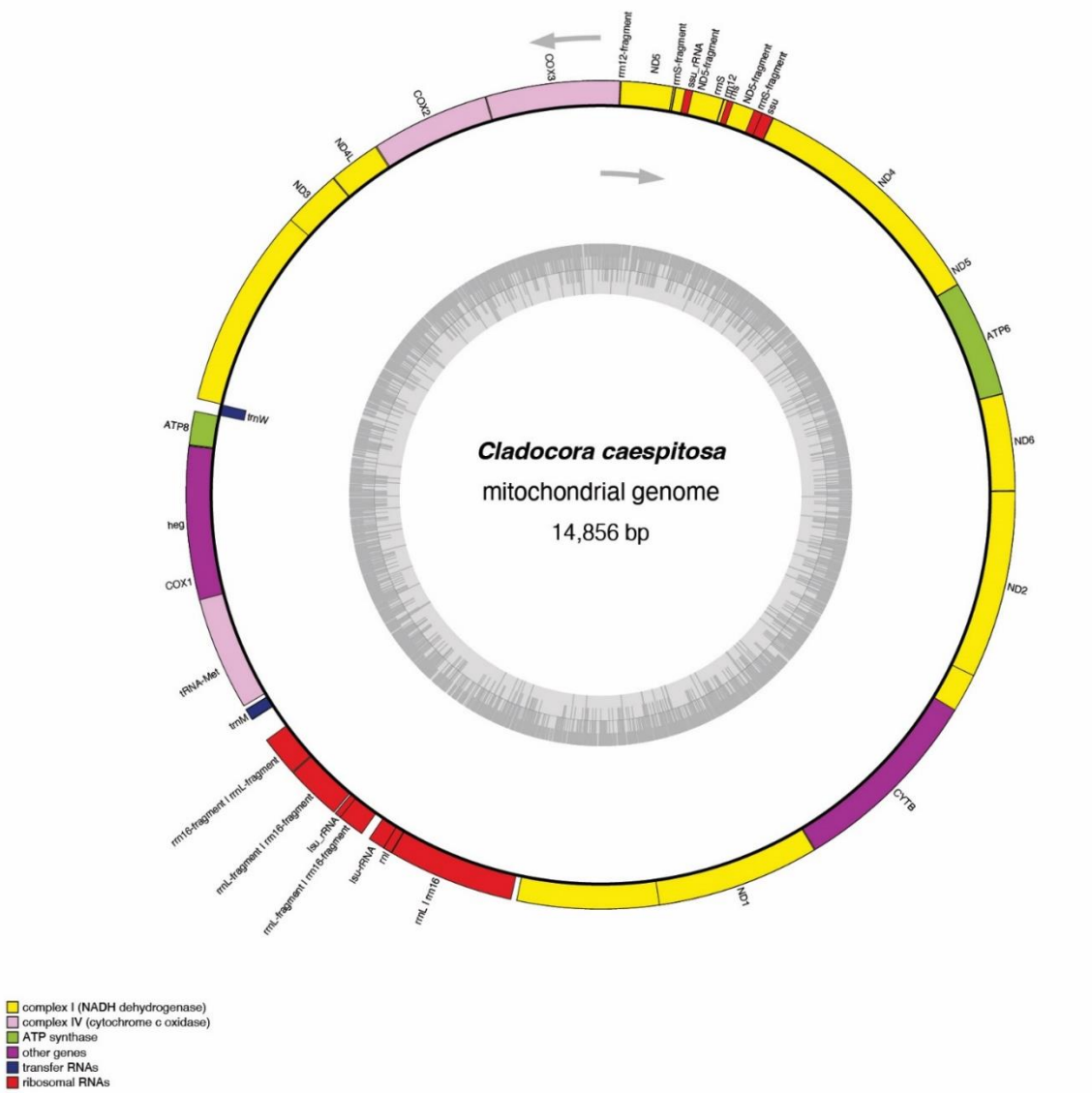


Figure 4.34. Circularized annotated mitogenome of *C. caespitosa* in sample 306 (A1_Con_T0).

The read numbers and the basic statistics on the mapping of trimmed reads to the assembled mitogenome of A1_Con_T0 is depicted in Table 4.14. The highest number of mapped reads to the assembled mitogenome was recorded as 212,807 for sample A2_Vib_T1, whereas the lowest number of mapped reads were recorded as 111 for sample A2_Con_T0. The very low abundance of the mapping reads of sample A2_Con_T0 to the assembled mitogenome verifies the unsuccessful trial of mitogenome assembly for the same sample. All the reads mapped to the assembled mitogenome of A1_Con_T0 resulted in 100% coverage, again except A2_Con_T0 which had a coverage of 54.29%.

Table 4.14. The read numbers and mapping statistics of mitogenome mapping process.

Sample	Read Number	Read Number After Trimming	Number of Mapped Reads	Mapping percentage	Meandepth	Coverage
A1_Con_T0	112,232,867	111,228,239	55,357	0.03	2147.2	100.00
A1_Nostr_T1	71,873,064	71,238,576	129,179	0.09	1288.64	100.00
A2_Vib_T1	126,597,300	125,440,020	212,807	0.08	2119.84	100.00
A2_Con_T0	108,961,761	108,077,643	111	0.00	1.09498	54.2878
A3_Con_T0	70,610,033	69,593,530	7,929	0.01	79.1492	100.00
A3_Temp_T1	90,311,790	89,581,286	92,703	0.05	924.466	100.00
A4_Con_T0	84,438,245	83,742,114	126,864	0.08	1265.36	100.00
A4_ViTe_T1	90,720,837	89,929,463	155,106	0.09	1546.56	100.00

The assembled mitogenomes of seven samples were compared to each other and the diverse regions among samples have been identified. In total, there are four informative sites that differ in base composition between samples. The first site that differs among samples is located within the atp6 gene region. In this location, while samples A3_Con_T0 and A3_Temp_T1 share the same bases of Adenine, the rest of the five samples have Thymine in the same location. The second variant site is spanned by the gene region of COX1, where A3_Con_T0 and A3_Temp_T1 has Cytosine base in this site, rest of the samples have Guanine instead. The third variant site is located within the cytochrome b gene; A3_Con_T0 and A3_Temp_T1 shares the same base of Cytosine in this site, and rest of samples have Thymine instead. Finally, the last divergent site was spanned by the gene region of nad2. In this region, the samples A1_Con_T0, A4_Con_T0 and A4_ViTe_T1 have Guanine in this position, whereas rest of the samples have Adenine. Notably, the samples A3_Con_T0 and A3_Temp_T1 are collected from the same colonies at different times. Therefore, having same bases in the divergent sites verifies the proximity of these samples. Similarly, samples A2_Con_T0 and A2_Vib_T1 are collected from the same coral colony in different timelines. This holds true for the samples A4_Con_T0 and A4_ViTe_T1. Conversely, the samples A1_Con_T0 and A1_Nostr_T1 are collected from different colonies.

4.8.1. Haplotype Network

The haplotype network based on the COX1 gene of the assembled and retrieved mitochondiral genomes of *C. caespitosa* are depicted in Figure 4.35. Within the limitations of the available COX1 gene data in the database, 480 base-pair long COX1 gene region was used for building the haplotype network. The number of variant sites between the COX1 region of the total 33 samples is recorded as one. While 26 of the samples have Guanine (Haplotype 2) base in this informative variant site, rest of the 7 samples have Cytosine (Haplotype 1) instead, resulting in two haplotypes among all samples used in the network.

Among 7 samples from Turkey, two of them, A3_Con_T0 and A3_Temp_T1, have Cytosine base in the variant site. Although the majority of the samples had Guanine in this site, both of the haplotypes were present in almost all regions, except Spain. The two COX1 gene sequences retrieved from Spain consisted of th single haplotype of Haplotype 1 (Figure 4.35). While the total 16 COX1 genes from Cyprus consisted of both haplotypes, only a single sample had the H1 haplotype. Finally, out of 8 COX gene regions from Greece, two of them has H1, whereas six of them had H2. It can be seen in the network that there is almost no differentiation of the samples among the four countries that sequences were obtained from.

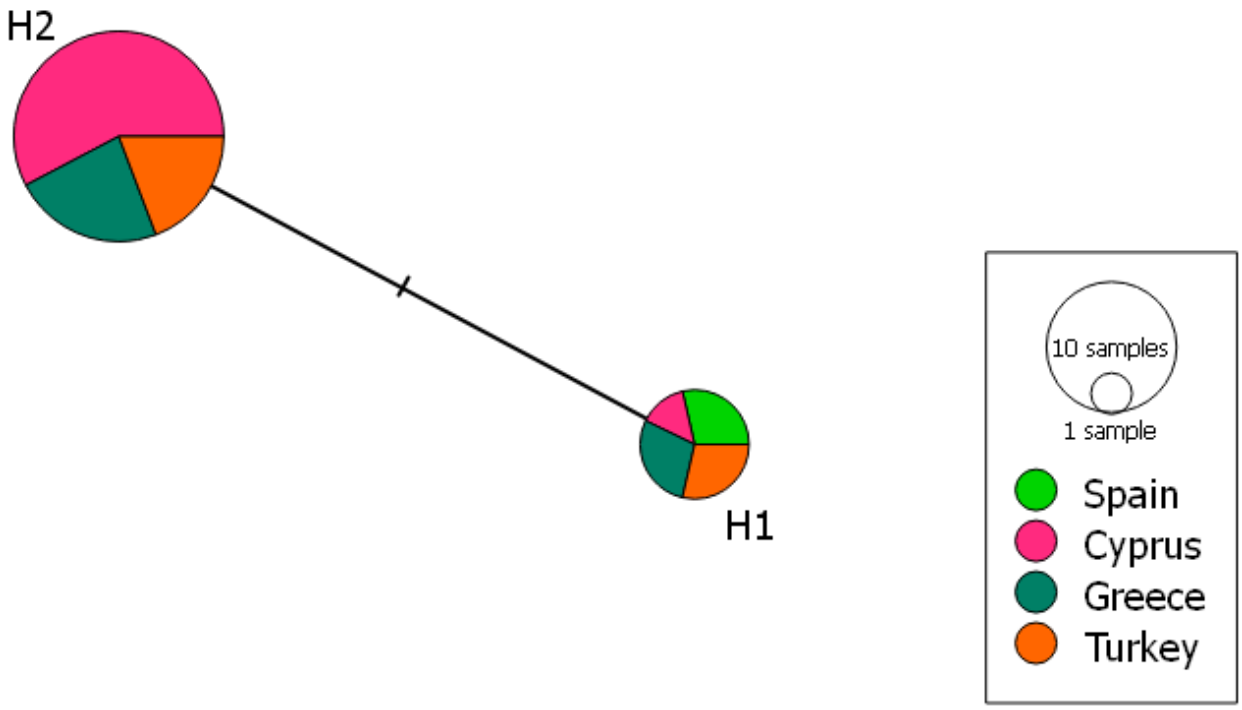


Figure 4.35. The haplotype network of the COX1 gene region of the assembled *C. caespitosa* mitogenomes (this study) and COX1 sequences of *C. caespitosa* available from RefSeq.

5. DISCUSSION

Despite the great efforts on coral bacteriome characterization, relatively few numbers of studies focusing on the detection of viruses on the surface of coral colonies have been performed (Rubio-Portillo et al., 2018). Nevertheless, here we present the first detection of dominant viral families of *Cladocora caespitosa*, an endangered coral species endemic to the Mediterranean basin. The objective of this study was to monitor the possible shifts in viral load on coral colonies corresponding to the applied environmental stressors. To that end, two of the well-studied causes of coral bleaching has been utilized in our study: heat stress and a coral pathogen known to contribute to coral bleaching, namely *Vibrio coralliilyticus*. The effect of these stress factors was experimentally designed to investigate the impact of each of the applied stressors by themselves and in combination, comparing to the no-stress applied coral colonies. Considering the increased abundance of virus like particles (VLPs) on the surface of bleached corals indicated in several studies (Ben-Haim et al., 2003; Correa et al., 2016; Nguyen-Kim et al., 2015; Weston et al., 2012), we anticipated to observe an increase in the viral load in stressed corals. However, our initial findings showed that the abundance of assigned viral contigs differed greatly between samples, particularly the control groups, regardless of the applied environmental stress factors.

5.1. Comparison of Viral Loads

Although initially unexpected, the dissimilarities among the control groups serve as vital indicators of the complex interactions inherent within the coral microbiome. It's important to note that the colonies were originally collected from the wild and subsequently transplanted into the aquarium setup. Therefore, the microbial compositions of the coral colonies in different tanks likely differed prior to the start of our experiment. Consequently, we evaluated it to be more valuable to investigate the changing patterns in each tank, rather than attempting to equate the control samples for baseline assessment. Examining the dissimilarities among control groups, the sample A2_Con_T0 was prominent, having the highest viral diversity and viral contig abundance. The lowest viral diversity and contig abundance, on the other hand was recorded for the sample A4_Con.T0. The extensive difference between these two control samples constrains us from forming a correlation between the four tanks representing four environmental conditions based on the relationship between the control groups. Therefore, we initially targeted the similarities and dissimilarities between the first and last samplings of each tank.

In general, the viral diversity and viral abundances decreased in Tank 1 and Tank 2, whereas these values displayed an increase in Tank 3 and Tank 4. The initial findings on the changes observed in viral diversity and viral contig abundances represent the similarity between the reactions to the applied environmental stressors in Tank 1 and Tank 2, in which both viral diversity and contig abundance declined over time. Notably, the *Vibrio* pathogen is activated within the temperature range of 27-29°C (Kimes et al., 2012). As such, we did not anticipate pathogenic activity in Tank 2, which was exposed to *Vibrio* at a lower temperature of 24°C. Considering that the corals are placed into a relatively closed system as compared to their natural habitats, the decline observed in the viral load in non-stressed corals of Tank 1 and Tank 2 might also be associated with the limited nutrients within the aquariums.

Despite the increase in viral contig abundance we observed in Tank 3 and Tank 4, it is worth noting that the increase recorded for Tank 4 was sharper than that observed in Tank 3. Comparing the environmental stress conditions, the heat stress in Tank 3 and *Vibrio* with heat stress in Tank 4 is expected to demonstrate the highest detrimental impact on the coral colonies. Accordingly, the observed incidence of coral bleaching in these tanks aligns with our expectations. On the other hand, despite the observed bleaching in Tank 3 and Tank 4, it is possible that these two tanks reacted to the applied stressors in different ways, such that the viral load in Tank 3 was relatively more stable as compared to the drastic increase in viral load and diversity observed in Tank 4. Several studies indicated a positive correlation between the bleaching incidence and the increase in viral load (Ben-Haim et al., 2003; Correa et al., 2016; Nguyen-Kim et al., 2015; Weston et al., 2012). In a study conducted by Pollock et al. (2014), a 65% higher abundance of virus-like particles (VLPs) was documented on the surface of white syndrome (WS) infected corals compared to their healthy counterparts. Therefore, the increase in viral diversity and abundance in Tanks 3 and 4 suggests a potential relationship between coral bleaching and the increased presence of viruses. Overall, it will be important to investigate the changes in specific viral groups to gain a deeper insight into the correlation between viral load and coral bleaching.

The lack of replicas in our experiments was one of the main limitations in analyzing the virome of each sample. However, we put extensive effort into the elimination of possible false positive data during bioinformatics analyses. Our initial strategy in elimination of the false positive data was to filter with a stricter threshold for the percent similarity to the viral protein database. To compare the filtered and non-filtered data, we performed a Principal Coordinate Analysis (PCoA). At the genus level, the PCoA plots obtained from filtered and non-filtered data do not reveal a coherent pattern. We encountered challenges in achieving proximity among the control samples in both the PCoA

graphs, whether conducted with data before or after the filtration process. Additionally, the total percentage of variation explained appeared to remain relatively stable after the data filtration. This observation could be attributed to the fact that as we move towards the lowest taxonomic level, the taxa tend to branch out and diversify. Consequently, the increased diversity at this level may counteract the effects of the data filtration process. Overall, the PCoA graph at the family level in , following the data elimination process, yields more promising and discernible results compared to the genus level. Therefore, comparing the samples based on the observed viral family abundances was deemed to be more effective from an ecological perspective.

A second approach on investigating the false positive data was the read-normalization step, which was based on the equation of the read counts across all samples prior to the downstream analyses. The taxonomic assignment with the non-normalized data resulted in a decrease in the number of viral contigs with similarity score higher than 75% for Tank 1 and Tank 2, and an increase in Tank 3 and Tank 4. While we failed to observe a significant change in pattern between samples after the normalization step, the number of contigs that had a similarity score higher than 75% decreased in Tank 3 and Tank 4, instead of the increase observed prior to the normalization. This finding brings the adverse impact of the lost data on the taxonomic assignment step into question. The normalization was performed by subtracting the first ‘n’ number of reads from each of the samples until the read numbers of all samples were equal. However, this step might have introduced a bias in terms of elimination of certain reads. Therefore, to not disrupt the integrity of the data in hand, we decided to further investigate the non-normalized data.

5.2. Detection of the Dominant Viral Families

Despite the ongoing debate on accepting viruses as a living matter, viruses occupy a noteworthy space in many ecosystems. The extremely divergent characteristic of viruses hinders their classification (Simmonds, 2015; Simmonds & Aiewsakun, 2018). This characteristic is claimed to stem from viruses having more than one origin (Nasir et al., 2012). The virus-first hypothesis claims that the unicellular organisms that evolved into modern day organisms originated from viruses, the hypothesis supported by the presence of virus-specific sequences. The reduction hypothesis, on the other hand, states that viruses are the reduced or primitive versions of the parasitic organisms that they originated from. The last hypothesis is the escape hypothesis which suggests that the viruses were once part of a host cell and escaped the host, evolving into modern day viruses. The complexity in the characterization of viruses obscures the classification of these biological entities higher than the family taxonomic level (Kuhn et al., 2013; Simmonds, 2015). Consequently, it is important to

note that the detection of dominant viruses performed in this study is limited by the availability of the assigned viral proteins in databases.

The virome of all samples were dominated by two realms: Duplodnaviria and Varidnaviria. Both realms encompass double-stranded DNA (ds-DNA) viruses capable of infecting a wide array of hosts, spanning all three domains – eukaryotes, bacteria, and archaea (Woo et al., 2021). Moving to the phylum taxonomic level, the distribution of the Nucleocytoviricota between samples is worth investigating, since this phylum represents all the Nucleocytoplasmic Large DNA Viruses (NCLDVs). These NCLDVs, which are known to infect eukaryotic organisms, are important due to their constant presence on the coral surfaces as indicated in various studies (Messyazs et. al., 2020; Correa et. al., 2016). The drastic increase in abundance of Nucleocytoviricota phylum members in Tank 4 might be representing a relationship between these viruses and *Vibrio*-induced environmental stress in corals. To unravel the underlying causes behind the abundance fluctuations within higher taxonomic levels, a deeper exploration into their subcategories is required.

Categorizing the viruses into groups to be able to emphasize certain differences, such as genome size and host range between the samples, the detection of dominant viruses was performed at the family taxonomic level. The viral families that constitute the core virome, found in all samples, were as follows: Phycodnaviridae, Poxviridae, Polydnaviridae, Baculoviridae, Herpesviridae and Kyanoviridae. All these viral families, except Kyanoviridae family, have been identified on coral ecosystem in several studies (Correa et al., 2016; Thurber et al., 2017; Cárdenas et al., 2020; Ambalavanan et al., 2021) with different methodologies from Transmission Electron Microscopy (TEM) to metagenomic analyses. Among these double-stranded DNA (dsDNA) virus families that constitute the core virome of coral colonies in this study, Phycodnaviridae and Poxviridae are the members of Nucleocytoplasmic Large DNA Viruses (NCLDVs) (Mönttinen et al., 2021). While Phycodnaviridae family members are known to infect phytoplankton and algae (Dunigan et al., 2006), the viruses included in Poxviridae family are known to infect various vertebrates from cattle to humans (Hughes et al., 2009). Polydnaviridae and Baculoviridae viruses, on the other hand, are known to infect insects (Strand and Burke, 2014; Herniou et al., 2004). Subsequently, we conducted further investigations into the viral families in terms of their relationships with coral colonies. A review of the existing literature on coral virome studies reveals that several families, including Microviridae, Podoviridae, Retroviridae, Siphoviridae, Iridoviridae, Myoviridae, Mimiviridae, Phycodnaviridae, Poxviridae, and Baculoviridae, have been consistently identified across various coral species (Mahmoud and Jose, 2017; Weynberg et al., 2017; Laffy et al., 2018), and all these families are detected in various abundances among samples in our study as well. These studies also

highlight the high abundance of members belonging to the Caudoviricetes class (Weynberg et al., 2017), a finding that aligns with our own dataset.

The eukaryotic viral family of Herpesviridae was hypothesized to be infecting the coral host in stressed coral in a study conducted by Vega Thurber et al. (2008). Similar to the observation by Laffy et al. (2018), the abundance of the Herpesviridae family appears to vary among samples, with higher abundance observed in the control groups within our dataset. On the other hand, families like Ascoviridae and Marseilleviridae, two of the NCLDV members (Mönttinen et al., 2021), are reported to be rare on coral surfaces (Weynberg et al., 2017); however, our data indicates a moderate abundance of Marseilleviridae family members. The remaining families found within our data do not exhibit specific relationships with coral colonies in the existing literature. These include Demerecviridae, Herelleviridae, Zobellviridae, Mesyanzhinovviridae, Nudiviridae, Duneviridae, Chaseviridae, Straboviridae, Casjensviridae, Autographiviridae, Peduoviridae, and Kyanoviridae. It is worth highlighting that the Kyanoviridae family stands out as one of the most abundant families in our dataset. Notably, the dynamic structure of the virus taxonomy results in the frequent changes in the classification or nomenclature of the viruses. The Kyanoviridae family was created in 2022 by the International Committee on Taxonomy of Viruses (ICTV), explaining the lack of this viral family in the existing literature (Lefkowitz et al., 2018). On the other hand, the Kyanoviridae family is part of the Caudoviricetes class, and the members of this viral family are known to infect cyanobacteria (Rong et al., 2022). Therefore, despite its absence in the literature, here we present a proximate relation between *C. caespitosa* and the viral family Kyanoviridae.

The fluctuations in viral abundance and diversity provide valuable clues for our data analyses. To extract meaningful insights from this dataset, it is essential to investigate the changing patterns within specific viral groups or families. For instance, the Kyanoviridae family has a well-established association with cyanobacteria, a significant component of the coral microbiome (Rong et al., 2022). Investigating the members of the family in our dataset, the main species appeared to consist of Synechococcus phages and Cyanophages. Notably, the abundance of Kyanoviridae family members appears to exhibit an intriguing decrease in Tanks 1 and 2, whereas Tanks 3 and 4 exhibit an increase. The increased abundance of these cyanobacteria infecting viruses is potentially related with the disruption of coral symbiont relationship, resulting in coral bleaching that has been recorded for Tank 3 and Tank 4.

Another group of viruses that has been linked to the bleaching phenomena are the Nucleocytoplasmic Large DNA Viruses (NCLDV) (Messyasz et al., 2020). The members of this

group consist of the viral families: Phycodnaviridae, Mimiviridae, Ascoviridae, Iridoviridae, Marseilleviridae, Asfarviridae, and Poxviridae (Mönttinen et al., 2021). The following viruses are also listed as the unclassified NCLDVs: Faustovirus, Kaumoebavirus, Klosneuvirus, Medusavirus, Mollivirus, Pithovirus and Pandoravirus (Mönttinen et al., 2021). Messyasz et al. (2020) demonstrated an increasing pattern of the members of NCLDV particularly in bleached corals and not in healthy corals. Similarly, in our study, an increase of NCLDV members were observed for the coral colonies planted in Tank 4, and not in coral colonies in Tank 1, Tank 2 and Tank 3. The presence of NCLDV represents the activity of eukaryotic viruses, rather than bacteriophages. This also shows that NCLDV members are potentially infecting the eukaryotic symbionts on the surface of corals, such as symbiotic algae, or the coral cells themselves, each of which eventually might lead coral bleaching. Despite the bleaching condition observed in both Tank 3 and Tank 4, the increase in NCLDVs was only observed in Tank 4. This finding once again suggests that the potential bleaching pathways might be different in thermally stressed and pathogenically stressed corals, in addition the different changes observed in viral load in Tank 3 and Tank 4.

Despite the lack of virome characterization of *C. caespitosa* in the existing literature, there is a single study by Rubio-Portillo et al. (2018) on the bacteriome characterization of this species, comparing the bacterial composition of healthy and necrosed corals at different locations. Rubio-Portillo et al. demonstrated that the bacterial communities of necrosed corals were more similar to each other than those of healthy corals. This finding suggests that the environmental stressors cause a shift in microbial communities in certain direction. Investigating the shared families between samples, the highest similarity was accounted for the stressed samples. Out of 23 viral families in A3_Temp_T2 and out of 19 viral families in A4_ViTe_T1, 17 of them were present in both samples. Consequently, the shift described in bacteriome study of *C. caespitosa* possibly underpins the shift in the viral families in stressed corals.

5.3. Similarities Among Samples Based on Statistical Findings

The low alignment rate between the assembled contigs and the trimmed reads in sample A2_Con_T0 highlighted the possible inefficiency of the contig assembly process. Therefore, restraining from speculative comments on the abundance of viral families or genera, the presence/absence data was decided to be integrated into the statistical methods (Jaccard distance method). The highest diversity was unsurprisingly attributed to the sample A2_Con_T0 in alpha diversity metrices. However, while the sample A2_Con_T0 seemed to be the outlier within all samples, the beta-diversity heatmap and dendrogram indicated that the dissimilarity between the

sample A4_Con_T0 and the rest of the samples was in fact the highest. Interestingly, despite the highest viral diversity recorded for the sample A2_Con_T0, the control groups of Tank 1 and Tank 2 (samples A1_Con_T0 and A2_Con_402) clustered together in the dendrogram. Similarly, the T1 samples (A1_Nostr_T1 and A2_Vib_T1) from the same tanks clustered close to each other. This is expected since *Vibrio* is not activated at 24°C, the coral colonies in these tanks (Tank 1 and Tank 2) should not have been stressed. The relatively high distance between the last samplings of the Tank 3 and Tank 4 (A3_Temp_T1 and A4_ViTe_T1) again potentially represents the difference between the coral bleaching pathways, as mentioned above. On the other hand, the clustering of A2_Vib_T1 and A4_ViTe_T1 samples might suggest a possible shift in the viral diversity due to the addition of *Vibrio* to these tanks. Although these statistical findings are noteworthy, an additional investigation into the virus-host interactions is required.

5.4. Viral-Host Search

Although we discussed the possibility of the difference in bleaching mechanisms in Tank 3 and Tank 4, the comparison between the abundance of phages and eukaryotic viruses bears the potential of unraveling the underlying causes behind these bleaching incidences. The bleaching observed in Tank 3 might be caused by viral infections that eliminate the beneficial bacteria from the coral surface. Conversely, the bleaching observed in Tank 4 might stem from the infection of the coral tissue itself or the eukaryotic algae that provides the necessary nutrients for the coral tissue. Alternatively, the shift in phages and eukaryotic viruses might be the result of the shift in bacterial composition of corals due to the applied stress factors, leading to bleaching. The heat stress introduced in Tank 3 might have led to a shift in the microbiome in favor of the bacteria rather than eukaryotic organisms. Similarly, the thermally activated *Vibrio* applied in Tank 4 might have outcompeted the other bacteria in the environment, resulting in the increase of the eukaryotic viruses.

Finally, the number of viruses that are known to infect Gammaproteobacteria was investigated. The *V. coralliilyticus* introduced to Tank 2 and Tank 4 is taxonomically in the Gammaproteobacteria class. Therefore, possible shifts in the abundance of Gammaproteobacteria infecting viruses were anticipated to be discovered. However, the abundance of such viruses seemed to decrease in both tanks. However, when we narrowed down the taxonomic level to Vibrionales order, to which the *Vibrio coralliilyticus* belongs, the pattern appears to change between samples. The number of viruses known to infect Vibrionales was recorded as two in Tank 2 and these viruses completely disappeared in the last sampling. The number of Vibrionales infecting viruses in Tank 4, on the other hand, was

recorded as zero in the first sampling and one in the last sampling. The small sample size of the found Vibrionales infecting viruses restrains us from commenting on the observed changes.

The bacterial community characterization of coral mucus study by Lee et al. demonstrated a shift in the bacteriome in thermally stressed (29°C) coral of *Acropora muricata*. The study focused on the sugar content of the mucus obtained from the coral colonies. The obtained shift in bacterial community of the thermally stressed corals was from Gammaproteobacteria to Alphaproteobacteria, together with increased abundance of Cyanobacteria. The decline in the abundance of both Gammaproteobacteria and Alphaproteobacteria infecting viruses in Tank 1 and Tank 2 might therefore be due to the decrease in the mucosal content of coral colonies in the absence of a stressor. On the other hand, although the colonies in Tank 3 were thermally stressed, the abundance of both Gammaproteobacteria and Alphaproteobacteria infecting viruses displayed a decrease over time. However, the abundance of Cyanobacteria infecting viruses increased in both thermally stressed (Tank 3) and *Vibrio* with thermally stressed (Tank 4) corals. While the viral host search gives us an idea about the structure of the bacterial community, it is important to note that the presence of the viruses does not necessarily indicate the presence of the host. Therefore, a proper comparison between the bacteriome and virome characterizations will be valuable as a follow-up to this study.

5.5. Mitochondrial Genome Assembly and Haplotype Network

Struggling to find a congeneric reference genome in the host-genome removal step, we anticipated filling the gap by performing a mitochondrial genome assembly of *C. caespitosa* with the sequenced data in hand. The assembled mitochondrial genomes of the seven samples provided quite similar results, with only four variant sites. The samples collected from the same colonies, such as A3_Con_T0 and A3_Temp_T1, shared the same bases in these divergent sites, validating the efficiency of the mitochondrial genome assembly process. Conversely, we failed to generate the mitochondrial genome of A2_Con_T0, possibly due to the lack or very few abundances of coral genome within the sample. This finding is also supported by the very low alignment rate between this sample and the reference coral genome of *P. sinensis* and *P. verrucosa* utilized during the host-genome removal step. The lack of coral genome within this sample might be related to the errors made during the sampling or library preparation. Comparing the annotated mitogenome of A1_Con_T0 and the readily available mitogenome of *F. abdita* (NC_035879.1), the genes that are present in *F. abdita* were all present in the assembled mitogenome of C. caespitosa, verifying once again the efficiency of the assembly.

Utilizing the assembled mitogenome in hand, the haplotype networks among the available regional data were aimed to be generated. The variations within the COX1 gene of the assembled mitogenomes and the available sequences on databases were non-informative in terms of explaining the phylogeographic relationship among samples. There were only two different haplotypes across all the samples, and these haplotypes were present in almost all the regions. Therefore, the distribution of *C. caespitosa* across the Mediterranean is relatively uniform. A recent study by Repullés et al. (2022), indicated the lack of significant association between genetic differentiation and geographic distance of *C. caespitosa* colonies, using microsatellite loci. However, the study also showed the low genetic connectivity among populations in the western and eastern Mediterranean. Considering that our haplotype network mainly comprises *C. caespitosa* samples from eastern Mediterranean, the uniformity of the distribution of the species is in accordance with the previous study. On the other hand, while the two samples obtained from Spain shared the same haplotype, they did not exhibit a distinct haplotype as compared to the rest of the samples from eastern Mediterranean. Nonetheless, the genetic similarity detected between western (Spain) and eastern Mediterranean (Greece, Cyprus and Turkey) might potentially be due the slower evolution rate of the COX1 gene used in this study than the microsatellites used by Repullés et al. (2022).

6. LIMITATIONS OF THE STUDY AND RECOMMENDATIONS

The bioinformatic workflow that was applied in this study does not allow the elimination of most contaminant genomes that are possible present within the samples. The host-genome removal process only contained the elimination of two coral species and two Symbiodinium clades, leaving the contaminant reads that possibly belong to other organisms, such as bacteria, fungi, amoeba, fish, and other organisms. Although, during the homology search step the remaining reads were mapped to the viral protein database, it is possible that some reads, not belonging to viruses, matched to the viral proteins, enabling the generation of false positive data. Therefore, an additional viral enrichment step can be performed during DNA extraction, in order to create a dataset consisting mostly of the DNA sequences that belong to viruses. Such viral enrichment practices may involve elimination of genomic content of organisms that are larger than viruses through membrane filtration or precipitation prior to the DNA isolation process. Similar approaches can be employed during the bioinformatics workflow, allowing the removal of all contaminant genomes through mapping practices. However, this strategy relies on the assumption of the presence of certain organisms, excluding the remaining organisms that are not expected to be observed on the surface of coral. Additionally, the assembled contigs should be investigated individually in order to examine the efficiency of homology search between the assembled contigs and the viral protein database. On the other hand, the challenges encountered during the taxonomic assignment process are also caused by the limited information on viruses that can be obtained from the databases. Therefore, further studies are required to contribute to the inclusiveness of the viral databases. Finally, the visualization of viruses or Virus Like Particles (VLPs) under microscopy is another approach in confirming the presence of certain viruses based on their phenotypes.

The dissimilarity between control groups is another limitation in our study that hinders the proper statistical interpretations. One possible solution for overcoming this challenge would be to reproduce newly formed colonies in the aquariums that sustain a favorable and sterile environment for the colonies, reducing the risk of contamination as much as possible. In this way, the starting point of all colonies could have been equalized, and the changing patterns in viral loads for each coral colony would be more apparent. Another approach in designing the experiment would be the in situ virome characterization of coral colonies, in order to obtain a deeper perspective on the interactions within coral microbiota in their natural habitats. This method would require for instance, the transplantation of coral colonies in different sites representing different environmental conditions; such an

experiment would be effective in investigating the impact of heat stress, though not pathogens, on viral composition of coral colonies.

One of the main challenges in conducting this study was the lack of replicas due to funding limitations. The absence of replicas restrained us from verifying the presence of the viruses in a sample by comparing them with replicas. In addition to the need for replicas, a further approach would be increasing the number of samplings and monitoring the changes or possible fluctuations in viral loads over shorter time periods.

The parasitic nature of viruses represents the dependency of these biological entities on certain organisms that they infect. Therefore, in order to understand the complex interactions within the coral microbiome, the virus-host interactions should be examined as well. The characterization of viruses on the surface of coral colonies eventually depends on the characterization of other microorganisms on the coral surfaces. Consequently, a bacteriome or algae characterization of this species is required to compare the overall microbial communities and have a better understanding of the possible interactions between viruses and bacteria or other microorganisms.

Finally, the contrast in phylogeographic distribution of *C. caespitosa* between two different genetic markers, namely microsatellite loci and COX1 gene, reveals the need for further phylogeographic studies. In this context, the range of sampling can be expanded, monitoring the genetic connectivity of colonies at a wider range. Due to different characteristics of different genetic markers, the population dynamics generated with several genetic markers can be compared to each other in the future.

7. CONCLUSIONS

In this study, we present the first detection of the dominant viruses found on the surface of *C. caespitosa*, a Mediterranean endemic coral species which is currently listed as an endangered species by IUCN. The dissimilarities between the viral loads of control groups highlights the complexity and the dynamic structure of the microbial communities in coral colonies. The fact that these colonies were immediately transferred to an aquarium set-up from the wild poses a challenge in comparing the viral load of the control groups.

The detected dominant viral families of *C. caespitosa* showed resemblance to the formerly characterized virome of different coral species. The core virome of *C. caespitosa* did not consist of any viruses previously not associated with coral colonies. The only exception to this was the presence of Kyanoviridae family, which is a viral family that was established in 2022. The well-known relationship between Kyanoviridae family member viruses and cyanobacteria corroborates the presence of these viruses in coral colonies.

The comparison between the possible host of viruses gives insightful information about the possible differences in the bleaching mechanisms of coral colonies. The increased abundance of eukaryotic viruses in pathogenically stressed corals can be associated with the infection of coral tissue or eukaryotic algae on the surface of the corals. The increased abundance of bacteriophages in thermally stressed corals, on the other hand, may be representing the viral infections affecting the cyanobacteria community, a finding supported by the increased abundance of cyanobacteria infecting viruses. However, as mentioned in the discussion section the presence of a virus does not necessarily indicate the presence of the host.

Finally, we aimed to generate the mitochondrial genome of *C. caespitosa*, due to the lack of a congeneric reference genome in databases. Furthermore, the haplotype networks were generated using the assembled mitogenome and retrieved COX1 genes from the database, and the distribution of *C. caespitosa* across the Mediterranean was relatively uniform. We concluded that the genetic similarity between western and eastern Mediterranean samples may be due to the slow evolution rate of COX1 gene as compared to the microsatellite loci.

REFERENCES

Aguirre, J., & Jiménez, A. P. (1997). Census assemblages in hard-bottom coastal communities: A case study from the Plio-Pleistocene Mediterranean. Palaios, 12(6). https://doi.org/10.2307/3515415

Ainsworth, T. D., Thurber, R. V., & Gates, R. D. (2010). The future of coral reefs: a microbial perspective. In Trends in Ecology and Evolution (Vol. 25, Issue 4, pp. 233–240). https://doi.org/10.1016/j.tree.2009.11.001

Ambalavanan, L., Iehata, S., Fletcher, R., Stevens, E. H., & Zainathan, S. C. (2021). A review of marine viruses in coral ecosystem. In Journal of Marine Science and Engineering (Vol. 9, Issue 7). https://doi.org/10.3390/jmse9070711

Andrews, S. (2010). FastQC. Babraham Bioinformatics.

Bally, M., & Garrabou, J. (2007). Thermodependent bacterial pathogens and mass mortalities in temperate benthic communities: A new case of emerging disease linked to climate change. Global Change Biology, 13(10). https://doi.org/10.1111/j.1365-2486.2007.01423.x

Banin, E., Israely, T., Kushmaro, A., Loya, Y., Orr, E., & Rosenberg, E. (2000). Penetration of the coral-bleaching bacterium vibrio shiloi into Oculina patagonica. Applied and Environmental Microbiology, 66(7). https://doi.org/10.1128/AEM.66.7.3031-3036.2000

Ben-Haim, Y., & Rosenberg, E. (2002). A novel Vibrio sp. pathogen of the coral Pocillopora damicornis. Marine Biology, 141(1). https://doi.org/10.1007/s00227-002-0797-6

Ben-Haim, Y., Zicherman-Keren, M., & Rosenberg, E. (2003). Temperature-regulated bleaching and lysis of the coral Pocillopora damicomis by the novel pathogen Vibrio coralliilyticus. Applied and Environmental Microbiology, 69(7). https://doi.org/10.1128/AEM.69.7.4236-4242.2003

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution, 69(2). https://doi.org/10.1016/j.ympev.2012.08.023

Bourne, D., Iida, Y., Uthicke, S., & Smith-Keune, C. (2008). Changes in coral-associated microbial communities during a bleaching event. ISME Journal, 2(4). https://doi.org/10.1038/ismej.2007.112

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

Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., McGill, S. K., & Dougherty, M. K. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Research, 47(18). https://doi.org/10.1093/NAR/GKZ569

Cárdenas, A., Ye, J., Ziegler, M., Payet, J. P., McMinds, R., Vega Thurber, R., & Voolstra, C. R. (2020). Coral-Associated Viral Assemblages From the Central Red Sea Align With Host Species and Contribute to Holobiont Genetic Diversity. Frontiers in Microbiology, 11. https://doi.org/10.3389/fmicb.2020.572534

Chefaoui, R. M., Casado-Amezúa, P., & Templado, J. (2017). Environmental drivers of distribution and reef development of the Mediterranean coral Cladocora caespitosa. Coral Reefs, 36(4). https://doi.org/10.1007/s00338-017-1611-8

Çinar, M. E. (2003). Ecological features of Syllidae (Polychaeta) from shallow-water benthic environments of the Aegean Sea, eastern Mediterranean. Journal of the Marine Biological Association of the United Kingdom, 83(4). https://doi.org/10.1017/S0025315403007732h

Correa, A. M. S., Ainsworth, T. D., Rosales, S. M., Thurber, A. R., Butler, C. R., & Vega Thurber, R. L. (2016). Viral outbreak in corals associated with an in situ bleaching event: Atypical herpes-like viruses and a new megavirus infecting symbiodinium. Frontiers in Microbiology, 7(FEB). https://doi.org/10.3389/fmicb.2016.00127

Cuif, J. P., Lecointre, G., Perrin, C., Tillier, A., & Tillier, S. (2003). Patterns of septal biomineralization in Scleractinia compared with their 28S rRNA phylogeny: A dual approach for a new taxonomic framework. Zoologica Scripta, 32(5). https://doi.org/10.1046/j.1463-6409.2003.00133.x

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

Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Research, 45(4). https://doi.org/10.1093/nar/gkw955

Doney, S. (2009). The consequences of human-driven ocean acidification for marine life. F1000 Biology Reports, 1. https://doi.org/10.3410/b1-36

Dunigan, D. D., Fitzgerald, L. A., & Van Etten, J. L. (2006). Phycodnaviruses: A peek at genetic diversity. Virus Research, 117(1). https://doi.org/10.1016/j.virusres.2006.01.024

Elahwany, A. M. D., Ghozlan, H. A., Elsharif, H. A., & Sabry, S. A. (2015). Phylogenetic diversity and antimicrobial activity of marine bacteria associated with the soft coral Sarcophyton glaucum. Journal of Basic Microbiology, 55(1). https://doi.org/10.1002/jobm.201300195

Geo Pertea. (2018). gpertea/fqtrim: fqtrim release v0.9.7.

Glynn, P. W. (1993). Coral reef bleaching: ecological perspectives. Coral Reefs, 12(1). https://doi.org/10.1007/BF00303779

Goreau, T. J., & Hayes, R. L. (1994). Coral bleaching and ocean ″hot spots″. Ambio, 23(3). https://doi.org/10.2307/4314195

Gregg, W. W., Conkright, M. E., Ginoux, P., O'Reilly, J. E., & Casey, N. W. (2003). Ocean primary production and climate: Global decadal changes. Geophysical Research Letters, 30(15). https://doi.org/10.1029/2003GL016889

Greiner, S., Lehwark, P., & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research, 47(W1). https://doi.org/10.1093/nar/gkz238

Hahn, C., Bachmann, L., & Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach. Nucleic Acids Research, 41(13). https://doi.org/10.1093/nar/gkt371

Herniou, E. A., Olszewski, J. A., O'Reilly, D. R., & Cory, J. S. (2004). Ancient Coevolution of Baculoviruses and Their Insect Hosts. Journal of Virology, 78(7). https://doi.org/10.1128/jvi.78.7.3244-3251.2004

Hoegh-Guldberg, O. (1999). Climate change, coral bleaching and the future of the world's coral reefs. In Marine and Freshwater Research (Vol. 50, Issue 8). https://doi.org/10.1071/MF99078

Hoegh-Guldberg, O., & Bruno, J. F. (2010). The impact of climate change on the world's marine ecosystems. In Science (Vol. 328, Issue 5985). https://doi.org/10.1126/science.1189930

Hughes, A. L., Irausquin, S., & Friedman, R. (2010). The evolutionary biology of poxviruses. In Infection, Genetics and Evolution (Vol. 10, Issue 1). https://doi.org/10.1016/j.meegid.2009.10.001

Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., Babcock, R. C., Beger, M., Bellwood, D. R., Berkelmans, R., Bridge, T. C., Butler, I. R., Byrne, M., Cantin, N. E., Comeau, S., Connolly, S. R., Cumming, G. S., Dalton, S. J., Diaz-Pulido, G., … Wilson, S. K. (2017). Global warming and recurrent mass bleaching of corals. Nature, 543(7645). https://doi.org/10.1038/nature21707

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics, 11. https://doi.org/10.1186/1471-2105-11-119

Kersting, D. K., & Linares, C. (2012). Cladocora caespitosa bioconstructions in the Columbretes Islands Marine Reserve (Spain, NW Mediterranean): Distribution, size structure and growth. Marine Ecology, 33(4). https://doi.org/10.1111/j.1439-0485.2011.00508.x

Kimes, N. E., Grim, C. J., Johnson, W. R., Hasan, N. A., Tall, B. D., Kothary, M. H., Kiss, H., Munk, A. C., Tapia, R., Green, L., Detter, C., Bruce, D. C., Brettin, T. S., Colwell, R. R., & Morris, P. J. (2012). Temperature regulation of virulence factors in the pathogen Vibrio coralliilyticus. ISME Journal, 6(4). https://doi.org/10.1038/ismej.2011.154

Kružić, P., & Benković, L. (2008). Bioconstructional features of the coral Cladocora caespitosa (Anthozoa, Scleractinia) in the Adriatic Sea (Croatia). Marine Ecology, 29(1). https://doi.org/10.1111/j.1439-0485.2008.00220.x

Kružić, P., & Požar-Domac, A. (2003). Banks of the coral Cladocora caespitosa (Anthozoa, Scleractinia) in the Adriatic Sea. Coral Reefs, 22(4). https://doi.org/10.1007/s00338-003-0345-y

Kuhn, J. H., Radoshitzky, S. R., Bavari, S., & Jahrling, P. B. (2013). The International Code of Virus Classification and Nomenclature (ICVCN): Proposal for text changes for improved differentiation of viral taxa and viruses. Archives of Virology, 158(7). https://doi.org/10.1007/s00705-012-1582-6

Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Jutz, S., Pascelli, C., Botté, E. S., Bell, S. C., Peirce, T. E., Weynberg, K. D., van Oppen, M. J. H., Rattei, T., & Webster, N. S. (2018). Reef invertebrate viromics: diversity, host specificity and functional capacity. Environmental Microbiology, 20(6). https://doi.org/10.1111/1462-2920.14110

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4). https://doi.org/10.1038/nmeth.1923

Lawrence, S. A., Davy, J. E., Aeby, G. S., Wilson, W. H., & Davy, S. K. (2014). Quantification of virus-like particles suggests viral infection in corals affected by Porites tissue loss. Coral Reefs, 33(3). https://doi.org/10.1007/s00338-014-1168-8

Lefkowitz, E. J., Dempsey, D. M., Hendrickson, R. C., Orton, R. J., Siddell, S. G., & Smith, D. B. (2018). Virus taxonomy: The database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Research, 46(D1). https://doi.org/10.1093/nar/gkx932

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. Methods in Ecology and Evolution, 6(9). https://doi.org/10.1111/2041-210X.12410

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics, 31(10). https://doi.org/10.1093/bioinformatics/btv033

Macher, T. H., Beermann, A. J., & Leese, F. (2021). TaxonTableTools: A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data. Molecular Ecology Resources, 21(5). https://doi.org/10.1111/1755-0998.13358

Mahmoud, H., & Jose, L. (2017). Phage and nucleocytoplasmic large viral sequences dominate coral viromes from the Arabian Gulf. Frontiers in Microbiology, 8(OCT). https://doi.org/10.3389/fmicb.2017.02063

Messyasz, A., Rosales, S. M., Mueller, R. S., Sawyer, T., Correa, A. M. S., Thurber, A. R., & Vega Thurber, R. (2020). Coral Bleaching Phenotypes Associated With Differential Abundances of Nucleocytoplasmic Large DNA Viruses. Frontiers in Marine Science, 7. https://doi.org/10.3389/fmars.2020.555474

Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., Hingamp, P., Goto, S., & Ogata, H. (2016). Linking virus genomes with host taxonomy. Viruses, 8(3). https://doi.org/10.3390/v8030066

Montagna, P., McCulloch, M., Mazzoli, C., Silenzi, S., & Odorico, R. (2007). The non-tropical coral Cladocora caespitosa as the new climate archive for the Mediterranean: high-resolution (~weekly) trace element systematics. Quaternary Science Reviews, 26(3–4). https://doi.org/10.1016/j.quascirev.2006.09.008

Mönttinen, H. A. M., Bicep, C., Williams, T. A., & Hirt, R. P. (2021). The genomes of nucleocytoplasmic large dna viruses: Viral evolution writ large. In Microbial Genomics (Vol. 7, Issue 9). https://doi.org/10.1099/MGEN.0.000649

Morrow, K. M., Muller, E., & Lesser, M. P. (2018). How Does the Coral Microbiome Cause, Respond to, or Modulate the Bleaching Process? https://doi.org/10.1007/978-3-319-75393-5_7

Nasir, A., Kim, K. M., & Caetano-Anollés, G. (2012). Viral evolution: Primordial cellular origins and late adaptation to parasitism. Mobile Genetic Elements, 2(5). https://doi.org/10.4161/mge.22797

Nguyen-Kim, H., Bouvier, T., Bouvier, C., Bui, V. N., Le-Lan, H., & Bettarel, Y. (2015). Viral and bacterial epibionts in thermally-stressed corals. Journal of Marine Science and Engineering, 3(4). https://doi.org/10.3390/jmse3041272

Oksanen, J., Simpson Gavin, Blanchet F, Kindt Roeland, Legendre Pierre, Minchin Peter, O'Hara R.B., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour Matt, Bedward Michael, Bolker Ben, Borcard Daniel, Carvalho Gustavo, Chirico Michael, De Caceres Miquel, Durand Sebastien, … Weedon James. (2022). vegan: Community Ecology Package. R package version 2.6-4. Community Ecology Package, 2.6-4.

Özalp, H. B., & Alparslan, M. (2011). The first record of Cladocora caespitosa (Linnaeus, 1767) (Anthozoa, Scleractinia) from the Marmara Sea. Turkish Journal of Zoology, 35(5). https://doi.org/10.3906/zoo-0907-80

Pérez-Losada, M., Arenas, M., Galán, J. C., Bracho, M. A., Hillung, J., García-González, N., & González-Candelas, F. (2020). High-throughput sequencing (HTS) for the analysis of viral populations. In Infection, Genetics and Evolution (Vol. 80). https://doi.org/10.1016/j.meegid.2020.104208

Pollock, F. J., M. Wood-Charlson, E., Van Oppen, M. J. H., Bourne, D. G., Willis, B. L., & Weynberg, K. D. (2014). Abundance and morphology of virus-like particles associated with the coral Acropora hyacinthus differ between healthy and white syndrome-infected states. Marine Ecology Progress Series, 510. https://doi.org/10.3354/meps10927

Pratchett, M. S., McCowan, D., Maynard, J. A., & Heron, S. F. (2013). Changes in Bleaching Susceptibility among Corals Subject to Ocean Warming and Recurrent Bleaching in Moorea, French Polynesia. PLoS ONE, 8(7). https://doi.org/10.1371/journal.pone.0070443

Raina, J. B., Dinsdale, E. A., Willis, B. L., & Bourne, D. G. (2010). Do the organic sulfur compounds DMSP and DMS drive coral microbial associations? Trends in Microbiology, 18(3). https://doi.org/10.1016/j.tim.2009.12.002

Repullés, M., López-Márquez, V., Templado, J., Taviani, M., & Machordom, A. (2022). Genetic Structure of the Endangered Coral Cladocora caespitosa Matches the Main Bioregions of the Mediterranean Sea. Frontiers in Genetics, 13. https://doi.org/10.3389/fgene.2022.889672

Rodolfo-Metalpa, R., Richard, C., Allemand, D., Bianchi, C. N., Morri, C., & Ferrier-Pagès, C. (2006). Response of zooxanthellae in symbiosis with the Mediterranean corals Cladocora caespitosa

and Oculina patagonica to elevated temperatures. Marine Biology, 150(1). https://doi.org/10.1007/s00227-006-0329-x

Rong, C., Zhou, K., Li, S., Xiao, K., Xu, Y., Zhang, R., Yang, Y., & Zhang, Y. (2022). Isolation and Characterization of a Novel Cyanophage Encoding Multiple Auxiliary Metabolic Genes. Viruses, 14(5). https://doi.org/10.3390/v14050887

Rubio-Portillo, E., Gago, J. F., Martínez-García, M., Vezzulli, L., Rosselló-Móra, R., Antón, J., & Ramos-Esplá, A. A. (2018). Vibrio communities in scleractinian corals differ according to health status and geographic location in the Mediterranean Sea. Systematic and Applied Microbiology, 41(2). https://doi.org/10.1016/j.syapm.2017.11.007

Rubio-Portillo, E., Kersting, D. K., Linares, C., Ramos-Esplá, A. A., & Antón, J. (2018). Biogeographic differences in the microbiome and pathobiome of the coral Cladocora caespitosa in the Western Mediterranean Sea. Frontiers in Microbiology, 9(JAN). https://doi.org/10.3389/fmicb.2018.00022

Schuhmacher, H., & Zibrowius, H. (1985). What is hermatypic? Coral Reefs, 4(1). https://doi.org/10.1007/bf00302198

Simmonds, P. (2015). Methods for virus classification and the challenge of incorporating metagenomic sequence data. In Journal of General Virology (Vol. 96). https://doi.org/10.1099/jgv.0.000016

Simmonds, P., & Aiewsakun, P. (2018). Virus classification – where do you draw the line? Archives of Virology, 163(8). https://doi.org/10.1007/s00705-018-3938-z

Strand, M. R., & Burke, G. R. (2014). Polydnaviruses: Nature's genetic engineers. Annual Review of Virology, 1(1). https://doi.org/10.1146/annurev-virology-031413-085451

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution, 38(7). https://doi.org/10.1093/molbev/msab120

Thurber, R. V., Payet, J. P., Thurber, A. R., & Correa, A. M. S. (2017). Virus-host interactions and their roles in coral reef health and disease. In Nature Reviews Microbiology (Vol. 15, Issue 4). https://doi.org/10.1038/nrmicro.2016.176

Thurber, R. V., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R. A., Angly, F., Dinsdale, E., Kelly, L., & Rohwer, F. (2009). Metagenomic analysis of stressed coral holobionts. Environmental Microbiology, 11(8). https://doi.org/10.1111/j.1462-2920.2009.01935.x

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq - Versatile and accurate annotation of organelle genomes. Nucleic Acids Research, 45(W1). https://doi.org/10.1093/nar/gkx391

Vega Thurber, R. L., Barott, K. L., Hall, D., Liu, H., Rodriguez-Mueller, B., Desnues, C., Edwards, R. A., Haynes, M., Angly, F. E., Wegley, L., & Rohwer, F. L. (2008). Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral Porites compressa. Proceedings of the National Academy of Sciences of the United States of America, 105(47). https://doi.org/10.1073/pnas.0808985105

Vezzulli, L., Previati, M., Pruzzo, C., Marchese, A., Bourne, D. G., & Cerrano, C. (2010). Vibrio infections triggering mass mortality events in a warming Mediterranean Sea. Environmental Microbiology, 12(7). https://doi.org/10.1111/j.1462-2920.2010.02209.x

Weston, A. J., Dunlap, W. C., Shick, J. M., Klueter, A., Iglic, K., Vukelic, A., Starcevic, A., Ward, M., Wells, M. L., Trick, C. G., & Long, P. F. (2012). A profile of an endosymbiont-enriched fraction of the coral Stylophora pistillata reveals proteins relevant to microbial-host interactions. Molecular and Cellular Proteomics, 11(6). https://doi.org/10.1074/mcp.M111.015487

Weynberg, K. D., Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Rattei, T., Webster, N. S., & van Oppen, M. J. H. (2017). Coral-associated viral communities show high levels of diversity and host auxiliary functions. PeerJ, 2017(11). https://doi.org/10.7717/peerj.4054

Weynberg, K. D., Voolstra, C. R., Neave, M. J., Buerger, P., & Van Oppen, M. J. H. (2015). From cholera to corals: Viruses as drivers of virulence in a major coral bacterial pathogen. Scientific Reports, 5. https://doi.org/10.1038/srep17889

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. In Media (Vol. 35, Issue July).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43). https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2019). dplyr: A Grammar of Data Manipulation. R package version. In Media.

Woo, A. C., Gaia, M., Guglielmini, J., Da Cunha, V., & Forterre, P. (2021). Phylogeny of the Varidnaviria Morphogenesis Module: Congruence and Incongruence With the Tree of Life and Viral Taxonomy. Frontiers in Microbiology, 12. https://doi.org/10.3389/fmicb.2021.704052

Wright, R. M., Strader, M. E., Genuise, H. M., & Matz, M. (2019). Effects of thermal stress on amount, composition, and antibacterial properties of coral mucus. PeerJ, 2019(4). https://doi.org/10.7717/peerj.6849

Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., & Voolstra, C. R. (2017). Bacterial community dynamics are linked to patterns of coral heat tolerance. Nature Communications, 8. https://doi.org/10.1038/ncomms14213

APPENDIX A: THE DRAFT MITOCHONDRIAL GENOME OF *Cladocora caespitosa*

>*Cladocora_caespitosa*

AGTGGTTATGGGGTTATTTTAAATGGAGTTTCTTTCTTTTTTTTTTTTTTTTGGTACGATTG
GTTCCGGAATAATGGTTGTTTCTGCCTTAAATCCTGTTCTTTCTATTTTTTGGTTAGTTCTT
GTTTTTGTTAATTCTGCGGTTTTTTTTCTTTTATTAGGAGTTGATTTTCTTGCTTTAATGTTT
TTACTTGTTTATGTAGGAGCTATTGCTGTTTTGTTTTTATTTGTAGTAATGTTATTAAATTTA
ACAGACTACCCTCCTGTTTTTAGAGAAGAAATTGATATGACAAATTATATACCAATTGGGT
TTTTAATTGGAGTATTTTTTTTTTCAGAAATTGCTTCCAGTTGATTTATTATATCTCCTCGA
GTTGAAAGTTGAGATTTGTCTTTCCCTTGATTTCTTGTTTCTTATCATAATATTGAGGCACT
AGGGCAGGTGTTATATGTAGCTTGTTTTTGTTTATTTCTTTTAGCAAGTTTTGTTTTATTAG
TTGCTATGATTGGTGTTATTGTTTTAACTCAAGAAACCGAACCTTTAACTAAAAAACAAG
ATCTTTTTATTCAAATAAATAGATAATGAGTGGTTCTTCTTATTTTGAACAGTTTAATGTTG
TGTGATTATTTGGTTTGACAAATTCTGCTATAATGATGGTTCTTGTCATTTTTGTGGTTTTA
TTGTTTTTAAAAGGAATTGATTTGATTCCTAAAAGATGGCAATCTTTTTTTGAATTAGTAT
ATTTTCATTTTTATTATATGGCAAAAGATAACTTAGGTGTTGAGGGCTTAAAGTATTTTCCT
TTTATTTTTTCTCTTTTTTTTTTTTTTAGTCTTTTTAAATGTTTTTGGTTTATGTCCTTATGTCT
TTACCCCAACAACTCACATTATTGTAACTCTTGGGTTTTCTTTTTCCATTATTATTGGGGTT
ACTTTTGCTGGACTTTGGAGATTTAAAAAAGATTTTTTTAGTATTTTAATGCCCAGTGGCG
CCCCTTTAGTTTTAGCTCCTCTTTTGGTTCTAATAGAAACAGTAAGTTATTTTTCTCGGGC
TATTTCTTTAGGAGTTCGGTTGGCCGCAAATCTTTCTGCGGGTCATCTTTTATTTGCTATTT
TAGCCGGTTTTGGTGTAGTTTTTAAATCAGCAATTTTAATAATGATTTTTATTACACTATTA
GAAATTGCTGTTGCGATTATTCAAGCTTATGTTTTTTGTTTATTAACAACTATCTATTTGGC
GGACACACTTGTTTTACATTAATGAGTCTTTTTTGGTTAATTCTTTTAATTGGAGCAATTA
GTGTGATGAGGGTTTCGAGAGAAAAAAAAAAGTCTCTTAATAAAACGGGCCTTAGAGTG
GTCTTTGGCCATTTTTTTTAGTGCTTTGGTTTTATGGGGGGGATTTGATGGAGAAGGACA
TTTTCAATTTATAGAATTCGTTGAATGAGGAGGATTTTTAGATTGGGGCCCTCTCCTTTTT
GCTTTAGATGGTGTTTCTTTATTTTTTTTGCTTTTAACAACTTTTTTAATTCCGATTTGTATT
TTAATTAGTCAGAAATCGATAAAATTTTTATTTAAAGAGTTTTTATTATGTTTATTTTTTTTA
GAAGTTTTATTGGTTGGTGTTTTTTTAGTCTTTGATCTTCTTTTATTTTATATTTTTTTTGAG
GGTGTTTTAATACCAATGTTTTTTTTAATTGGCATTTGGGGCTCTCGAGAAGAAAAGGTC

CGTGCTTCTTTTTATTTTTTTTTTTTTACTTTTGCAGGCTCGGTTTTCTTTTTTTTTGTAATA
CTTTTTTTATATCAGACGACTGGGACAACAGATTATTTTCTTTTACTTAATATGAAGTTGTC
TTTTAATGTTCAAAAATGGGCACTTGTTGGTGTGTTTCTTAGTTTTGCTGTAAAATTACCA
CTTATTCCTTTTCATATTTGACTTCCGCAAGCACATGTAGAAGCCCCTGTGGCCGGTTCTG
TTATTTTAGCAGGTATTTTATTAAAACTAGGAGGTTATGGTCTTTTACGTTTTTCTTGGCCT
ATTTTCCCTGAGGCTTCTTTTTATTGATCCCCGGTTATTATTGTTTTTAGTGTTATTGCAGT
TGTTTATGGGGGGTTGATGACATGTCGTCAAATTGATTTTAAACGACTTGTTGCTTATTCT
TCTGTTGCACATATGGGTCTCGTTCCCCTGGGTATTTTTACACATATTATGGAGGGGTTAG
TTGGGGCTGTGTTTTTAATGTTAGCACATGGTTTTGTTAGTTCTGCTCTTTTTATTGGAGT
GACTTTTTTATATGATCGTCATCATACGCGTTTAATAAAATACTATCGGGGTTTGACATTAA
CTATGCCCCTTTTTGCTGTTTCCATGCTTGTTTTATCTTTATCAAATATGGGACTTCCTTTA
AGTTGTAATTTTGTTGGGGAGTTTTTTTCTTTACTTGCGGCATTTAAATATTATTATGGGGT
TGGGGTGCTTGTTGTTTTTGGAGCTCTTTTTTCTGCTATTTATTCTCTTAGTTTATTTAATC
GTATTTCTTTTGGAGGAGGGTCCAATTATCTTCTTTTTAATAGAGATTTAAATCGACAAGA
AGGCTTTGCAATGTTTCCTTTTCTTATAATAATTTTTTTAGGAGGGGTTGTACCTTTTTTTG
TTATTAATTTGGTTAAAAACAGTCTTGTTTTTAGCCCAGTCGGTTAGTCCTTTGGTTTTGG
GGATTAAAAAGCTTTTGGTCTGGGTAATACATGCTTGTAGATGCGAACAGGTGAGGGTG
AAAGAAAAAGTTTATTTTTTTTTATTGTATTAGGAGAAAAAGAAGTTTTTTTCTTTTTTCC
AAAGATGCATGGTAAAACCACATTTTCACTGAAACAAGGGAAAACATTGGCAGCAGTA
AAGAATTTTATGCAATATACGAAAGTAAGACATAGGTAGGACATTTTAACCTGTCAAATT
AAGTGCCAGCAGACGCGGTGAAACTTAAGGGTTTGTTTTTTAGAAAAAGCAGAAAGCG
TGTAAAGGTAAAACATTTAAAATAAATAGAATTTTTTTAGTAATAGTGAAATGTTAAAATG
AGAAAAAGAATTCTTTATGTGAAAACAATTTATTTTTTTCTTGAACACGAAGGTTCTGGG
AGCAAACAGGATTAGAGACCCTGGTAGTCGATACAGTAAAAGAAAGTCGCTTGAGTAGT
ACGGTCGCAAGATTAAAATTCAAAAAACTTGGCTGTTCATTGTTATTTAGAGGAGCGTGT
CGCTTAATTCGACAATCCGCGAATTACCTTACCAAGACTTGGTTGTCAGGTGTTGCATGG
CCGTCGTCAATTTATGTTTGATACAATAAATTAAACCCTAGAAAGGTTGAAGTCAGGTAT
TATTGCCCTTATGTCTTGGGGTTAGATGCGCTACATTTTTTTTTTAATAAAAAAAAAAGAGT
TCGGATTGTCTTTTAAAGAGGACGATTAAGTTGAATTTAATAGTAATTGAAAAGTAGAGC
GTCTCAATGAATTTTAAGCAATGTTAGTACAAATAGCCCGTCGCCTTTACTGAGGTAAAC
AAAGTAGAGTAAGTCGTAACATAGTGAGGGTGAGGGAACTGGCCCTTGATGCAATATCA
TCCATATCATTTAGTGGAGCCTTCTCCATGGCCTTTTGTTGGGGCAGCAGGAGCGTTTTT
CTTGACTATTGGTTCAGTGGTTTTTTTTCATTATGGGTTTCGTTTTTTTTTAGGGTTAGGGC
TTTTAGTTATGGTTGCAGTTATGTTCGTTTGATGACAAGATGTTATACGAGAGTCTACCTT
CCAAGGGCATCATTCTTTAATTGTTAAACAAGGAATTAAATATGGTATGCTTTTGTTTATT

CTTTCAGAAGTACTTTTTTTTTTTTCTTTTTTTTGGGCTTTTTTTCATAGTAGTTTGGCCCC
TGCGGTCGAATTGGGGGTTGTATGACCACCACAAGGTGTTTCTGCATTAAACCCTTTTTC
TGTTCCTTTATTAAATACTGCTGTGTTATTAAGTTCCGGGGCAACAGTGACATGGGCTCAT
CATGCTATTGTTTGTGGAGCTAAAAAAGAAGCGCAGTTGGGTTTGTTTGTAACACTTGTT
TTAGGAGTAATTTTTACTGGTTTACAGGCATTTGAGTATTATGAGGCGCCGTTTGCTTTGT
CAGATTCTGTTTATGGATCTACTTTTTTTGTTGCAACAGGGTTTCATGGCTTACATGTTATA
ATAGGAACAACTTTTTTATTTGTTTGTTTTCTTCGTTTGTTATCTAATCAATTTACTCGTCG
TCAACATGTTGGTTTTGAGGCCGCTAGTTGGTATTGACACTTTGTTGATGTCGTGTGGTT
ATTTTTATATCTCTGTATTTACTGATGAGGTTCTTAAAGATGTTTCAAGATGCGGCAGAGG
CTTGGTCATTGGGTTTTCAAGATGTGGCAGATCCAGTGGTTGAAGAAATTGTTTTTTTTC
ATGATCGGGTTATGTTTTTGTTGGTCATTATTGTCACTGTTGTTATGTGACTTATTGGCGAG
GCTTTAAAAAACAAATTTTATGATCGCTTTTTAGTTGATGGTACTTTTTTAGAAATTATTTG
AACTATAATACCGGCAGTTATTTTAGTTTTTATTGCATTACCTTCTCTTAAATTATTGTATTT
AATGGATGAAGTGGTTTCTCCTGCTTTAACCGTAAAGGTGATTGGACATCAGTGATATTG
GTCTTATGAATATTCTGATTATGAGGGGGAGACATTAGGGTTTGATTCTTATATGGTTCCA
ACATCAGATTTAACTTCAGGGGAGAATCGTTTGTTGGAAGTTGATAACAAACTTGTTCTT
CCAATTTTAACTCATACAAGATTTTTGGTTACAGGGGCAGATGTTTTGCATTCTTTTGCTG
TTCCTTCTTTAGGGTTAAAAATAGATGCGGTTCCAGGCCGTCTTAACCAAACGGGGGTTT
TTATTAAAAGGCCGGGGGTTTTTTTTGGACAATGTTCTGAAATCTGTGGTGCAAATCATT
CTTTTATGCCAATTGTTATAAAAGGAGTTTGGGTTGAAGAATATCTTCATTATTTGAAATG
TATTATAAATACCTAGTTCTGGTAGTTCTTTTATTTTTATTGGGGAGCTGGGGCATAATTTT
AAACCGAGGACATTTTATTATTATGATTGTTTCCATTGAACTGGTTTTATTGTCTACTTTTT
TTTTCTTTTTAATAAATTCTAAAGAAATTGATGTTTTAATAGAACAAGTGTTTATGGTAATG
GGTTTAACAATAGCGGCAGCCGAGTCTTCTATTGGTTTAGCTATTTTGGTTGCTTATTATC
GTATTCGAGGGACAATAGTTTTAAAATCTTTTAGTTCTTTACGAGGTTAATGATGAGTTTT
TTTGTTTTTTGTTTTTTGGCAATTGTTTTAGCGGGGTTGTTTTCTATTGTTTCTTTTTTTCT
AGGAGAGAAAGTGCCAGATCGAGAAAAGGTTTCTGCTTATGAGTGTGGTTTTGTGCCAT
TTAGTTTTTTAGGTCGTCCTTTTTCAATACGATTTTTTTTAATTGGCATTTTATTTTTAATTT
TTGATTTAGAAATATCTTTTTTTTTCCCTTGGTGTGTTTTATATAATTCAATTGCTCCTTTTG
GGTTTTGAACTATGGTTGGGTTTTTCTTTATATTAACTTTAGGTTTGGTTTATGAGTGGTTA
AAGGGGGGTTTAGAGTGAGAGTAAAAAAAAAAGAGTAGAAGAAAAAGTCCTGTCTATTA
TGGGAGTAGTAGTTATTAGTATACCAACTCCGGTTTCTGCTTTAATCCATGCGGCAACAAT
GGTTACGGCGGGTGTGTTTTTATTAATTCGAGCATCTCCTTTGTTTGATGTTGTTCCTCTT
ATGTTAATTATTGTAACAATAATTGGGGTTTTAACGGTGTTTCTTGCTGGTACAATTGGTT
TAGTTCAAAATGATGTGAAAAAAATAATCGCTTATTCTACTTGTAGTCAACTTGGATATAT

GGTTGTTGCTTGTGGGCTTTCTCAATTTTCTATTGGTTTGTTTCATTTGATGAACCATGCT
TTTTTTAAAGCTTTGTTGTTTTTGAGTGCGGGTTCTTTAATCCATGCGGTGGTTGATGAAC
AAGATGTAAGGAAAATGGGGGGGTTATCCTCTTTTCTTCCTTTAACTTATATTTTTTTTATT
ATAGGGTCTTTTTCTTTAATGGGACTTCCTTTTTTAACCGGGTTTTATTCAAAAGATTTGA
TTTTAGAGTTTGCTTTTGGGCAATATTATTTAATTTTTGTTTATTGACTTGGGTGTTTTTCT
GTTTTATTAACAGCAATATATTCTATGCGTCTAATTTATTTATCTTTTTTATCCAATACAAAC
TTAAAACGAGGAGTATTGTCTTCTCTTAAGGAAGGAGAGTGACTTCTTTTAGCTCCTTTG
GGTATATTGGCTTTGGGAAGTCTTTTTTGGGGTTATTTTTGTAAAGAAATGATCTGAAGTT
TTCAAATAGGCACTTTTCCTGTGCTTTTTTTAAAAATTAAACTTCTTCCTGTTTTTTTTAGT
TTAGTAGGGGTTTTAGGGGTACTTTTTTTTTTTTATTTTTTTTTGTCTAAAACTTTGTTTTT
CTTTTTTTCTATTTATAGTTTTTTAGGCTCTGCTTGACAAATTAATTATTTTTATAATTTTTTT
TTTATAAAAAAAAATATATAAAATGGGACATATAATTACAAATTTAACTCTTGATAAAGGGG
TGTTAGAGCTTATTGGACCAAAGGGAATTGTTAATTTTTTGATTTTACAAGTGCAAAGAT
TAAGTAGTTTTCAGTCTGGGCTTGTTTTTAATTATGCTTTGGTTTTCTTTATTGGAGTTATT
GTTTTTATTGTTTTATTATAGAGAATTCGGTTAAAGTAAGCCATTGGCCTTCAAAGCTAAA
AATGTAGGTGCAAGTCCTACTTTCTCTGAAAATGCCACAGTTAAACACAACAATGTTTTT
AAATCAATATGGGTGTACTTTTTTTTTATTTTTTGTTCTTTTATTTTTATTTTTGTTTATTCTT
TTACCTCAAGTAAAAAAAAATTTTTTTTTTAGAAAAAAAAATAAGTACAAAGGGGGTTTC
TAAAGAATCTCTTTTAATAAAAGAAGTTGTTTTTATTTGATTATAATGAAAAATAATTATAT
TATTCGTTGGGTTTTTTCTACAAATCATAAAGATATTGGTACTTTATATTTAGTCTTTGGTG
TTGGGGCAGGGCTAATTGGGACTGCTTTTAGTATGCTTATACGACTGGAGCTTTCTGCGC
CAGGCGCTATGTTAGGTGATGATCATCTTTATAATGTAATTGTAACAGCACATGCTTTTATT
ATGATTTTCTTTTTAGTAATGCCGGTTATGATTGGGGGATTTGGAAACTGGTTGGTGCCAT
TATATATTGGGGCACCGGATATGGCGTTCCCCCGATTAAATAATATTAGTTTTTGGTTGTTA
CCGCCTGCTTTGCTTTTATTGTTAGGCTCTGCTTTTGTTGAACAAGGTGCAGGAACGGGA
TGGACGGTTTATCCTCCTCTTTCTGATATTTATGCTCATTCTGGGGGTTCTGTTGATATGGT
TATCTTTAGTCTTCATTTAGCTGGGGTTTCTTCTATCTTAGGAGCAATAAATTTTATTACAA
CGATTTTTAATATGCGAGCTCCTGGTGTTTCTTTTAATAGAATGCCTTTGTTTGTTTGGTCT
ATTTTAATAACTGCGTTTTTATTGCTTTTATCTTTGCCTGTATTAGCGGGTGCAATCACTAT
GTTGTTAACAGATCGAAATTTTAATACAACTTTTTTTGATCCTTCTGGGGGTGGAGATCCT
ATTTTATTCCAACATTTATTTTGGTTTTTTGGTCATCCAGAAGTTTATATTTTAATTTTGCCT
GGTTTTGGTATGATTTCTCAAATAATTCCGACTTTTGTTGCTAAAAAACAGATTTTTGGGT
ACTTAGGAATGGTTTATGCTATGCTTTCTATTGGGCTTCTTGGATTTATTGTTTGGGCTCAT
CATATGTTTACTGTTGGGATGGATGTAGATACAAGAGCATATTTTACTGCTGCCACTATGA
TTATTGCTGTGCCAACTGGGATTAAAGTTTTTAGTTGGTTGGCAACTATTTATGGAGGTGT

TCTTAGGTTAGATACTCCAATGCTTTGGGCTATGGGGTTTGTTTTTTTATTTACAGTAGGT
GGTTTGACCGGAGTTATATTAGCAAATAGTTCTCTTGATATTGTTTTACATGACACATATTA
TGTAGTTGCGCATTTTCATTATGTTCTTTCTATGGGAGCTGTTTTTGCTATTTTTGGGGGAT
TTTACTATTGAATTGGGAAAATCAGTGGTTATTGTTATAATGAATTTTTTGGGAAAGTGCA
TTTTTGATTAATGTTTATCGGGGTTAATCTAACCTTTTTCCCTCAACATTTCTTAGGTTTAG
CAGGGTTTCCCAGACGATACTCTGATTTTGCAGACGCTTTTGCCGGTTGAAATTTAATAA
GCTCTTTAGGGTCTGTTATTTCAATAATAGGTGTTATTTGGTTTTTATATATTGTATTTGATC
TTTTTGCTAAAGAAGAAAAATTCTTAGGTTGAAAAGATGGGAGCTCTTTAGAATGGATC
CATTCTTCTCCTCCCTTATTCCATACTTATGAGGAGTTACCTTTTGTGTGAAAAAACTAAA
AAAAGATATTGTAGTTTTTAGTAAGAAAGACTAAGGGTCAGTCTTCGGGCTCATGCCCCG
AGAAAAGTGAGTTCAAGTCTCACTCTTACAAAATTTTGTCTTTGAAAAAATAAAAAAAA
AGAAGAATGCTAAGATAAACTAAGACCAGACGATAGACTCTTCGAAACGGCGGAAAGA
AGCATTACTCGTTGCCTCTGCGGAAAACCGAGCTTGAAATTTGAATTGAAACATCTTAAT
ACCAAGTAAAAATCAAACGAGATTCCGAGAGTAGTGGCGAGCGAAATTGGAGTTGTATA
AAAGGTGATCATAGATCTAGACTAAACGTTAGTTTATACTGATAGTGAGAGTACTGTAAA
GGAAAGTTGAAAGAGAGTTGAAAAGAGCAGGAATCTTTTTTTTTTAAGCAGCTTTAAAA
CAGCGTACCTTTTGTATAATGGGTGAAAGAGATTAATTTGGCATATTTAAAAAAAGGAAA
ATTAAAAGAGATTTTTTTTAGTCAAAAATACCCGAAACCAAGTGATTTAATCATGGATAG
TAAATTAGAACGAACTGGTGGTTGTGGCAAAACTCTCAGATAATCTGTGATTAGGGGTG
AAAAACTAATCGAACTTGGACAGAGCTGGTTTTTTGCGAAAACTATTTAAGTAGTGCGC
TTTTTATTCTATTTTTTTTGTAATAAAATAAATAATAAAATAAAAAAAAAGTAGACAAACA
AGAAATGAAAAGATTTTTTGTTAAAAGAGAAAGCTCCAATCAGAAGTTAAGGTCATTTA
TTTCTTTTTAGTGTAAAAAAAAAATAGTCGGTTTAGACAATAAAAAAGTAGGCTTGGAG
CCAGCCATCTTTTAAAAAGTGCGTAGTTGTTTATTTAATAAATTAAATTTTTTTTTAATTTT
GGTGGTTAAAATGAAACTAAAACTCTGAAAATATAATTATTTAGTAGCAAAATACACTGT
TTTTCAGTAGAAACAATTTTTACATGAGTAATCTAAATATTATTTTTTTTTCTTTATATTACC
ATTTTTGGGTTATACAGACCCTAAGAACTTTTTTCTATGGGAATTTTTAAAAAATATATTAT
ATCCGGAAAAAATAAAAAAAGGAGCTGTTTTTCTAACTAAAGAAAGAAAAAAAAAAAGA
GACACATTTTTTTAAGGAACTCGGCAAAAGAGAACTTCGACTGTTTACCAAAAACATAG
CTTTTTGATTTTTATAAAAGGTGAGACCTGCCCTATGGTTGTATCTAAAAAAGTTTGTTTT
GCTTACTAATAAAGACAATTAAATGGCCGCGGTAACACTAACTGTGAAAATGTAGCGTAA
TCAATTGTCAATTAATTGTTGACCGGTATGAATGGTGTCACGAAAGTTCCTCTGTCTTAA
AAAAATACTTAATGAAATTGAATTTGTAGTGAAGATGCTACATTAAAATTGTAAGACGAG
AAGTCCCCATGGAGCTTTACTGAAAGCTTAAGATTTTCTATAAGCGGGACAGTTTTGTTG
GGGCGACAGTTTTTTAAAAAGTAACGAAGACGAACTATGGCGGAGCTTCACCCTGAAA

AATTTTTTAAGGGACATCTTTGGTGTGTTTTACGATCCGTTGTTTGGAATGAAAAAAATA
ACGAAAACAAATAAAAGTTACCCTGGGGATAACAGCGCAATAACGTTTGAGAGTTAATT
AACGACGGTGTTTGCGACCTCGATGTTGAATTGCAGCGTCCTAGGGAGTAGTCACTCCT
AAGGGTTGGTTTGTTCATCCATTAAAGCTGTACATGATTTGAGTTAAAAGCGTGGTAACA
CAGCTTGGTTTCTATCTACAATTATAAAACAAAGATATTTTGTTTTTTAGTACGAAAGGAT
CAAAAATCAATAGTTCTCTAAAGATAACTATTCTTTAAATTAGGATTGCTTCTAAAAAAAA
AAAGAGATATTTTTTGTTTAGTTTTATTAATTTTGTTTAGTTTTATTTATGTATCTTTTAATTT
TATTTTTTCCTTTAATTGGAGCTATTTTAACTGGTTGTTTTGGAAGACATATAGGAGAAAA
GGGAGCAGGGATTTTAACTTCAGGTTGTTTAATTATTGGTTTATCTTATTCTTTTTTAGTAG
GAATAGAAATTTTATTTAATTCAACGGCAACATATCTTAGATTATGAAAGTGATTTGATTC
TGGGTTTTTTCTTGTTTTTTTTGATTTACAGTTTGATGGTTTAGTTGCTATTATGTTGTTTG
TGGTTTTTCTTGTTTCTACTGTGGTTCATGTTTTTTCTATTGCTTATATGCGAGGGGATCCT
CATGTGCCTCGGTTTATGGCATATCTTTCTTTGTTTACTTTTTTAATGGTTTTGTTAGTTAC
TAGTGCCAATTTTCTTCAGTTGTTTATTGGATGAGAAGGTGTTGGTCTTTGTTCTTATTTAT
TAATTAATTTTTGATTAACAAGACTGGAGGCAAATCGAGCAGCAATTAAAGCAATGTTGG
TTAATAGAGTTGGTGACATTGGATTGCTTTTAGCAATGTTTCTTCTTTGAAAAACTTTTGG
GACTTTAGATTTTTCTTCTGTTTTTAATTTAGTTACTCCTTCTAATGAAGTTTTTTTTATTTG
TTTATTTTTATTTTTTGGAGTAATGGGTAAATCAGCTCAATTAGGGTTACATACTTGGTTGC
CGGATGCTATGGAGGGTTGTTGGGCCTCTTAATTAAAAAATTGATTAAAAAAACCACTAT
ACACAGGAAAGACAATTGTTTACCAGTGGGCAATAAAATGTTTGTTTTAATGCCTAAGA
GACTTTATGTGGTCTACTTTTTTTTATTTGATTAAAGCTGTTTTTGTTATTGTTCCTTTACTT
ATTGCTGTGGCATTTTTAACTTTAGCAGAACGAAAGGTTTTAGGATATATGCAAATGAGA
AAAGGGCCAAATGTTGTGGGGGGCGGGTTACTTCAGCCTTTTGCAGATGGGATAAAATT
ATTTCTTAAAGAAATGGTCCTTCCTCATCAAGCAAGTGCTTTTGTTTACTTTTTTGCTCCA
ATTGCCTCTTTTACATTAGCATTTATTGTTTGGGGGATTCTTCCTTATGGAAGAGGAATCG
GTATAAGTGACTTTAATATTGGTCTTTTGTGGGTTTTAGCTATTTCTTCTATTGGTGTTTAT
GCTGTTTTAATGTCTGGGTGGGGAAGTCGTTCCAAATATGCTTTTTTAGGTGCTGTTCGT
GCGGCGGCGCAAATGATTAGTTATGAGGTTTCGATTGGGTTAATTTTGATCTCTGTTCTTT
TATGTGTTGGCTCTTTGGCTTTTAATAAAATTGTTTTAGCACAAAATTTTATTTGATTCTTT
ATTCCTTTTTTTCCTGTTGTTGTGATGTTTTTAGTTTCTATTTTAGCAGAAACAAATCGTGC
TCCATTTGATTTAACGGAAGGAGAGTCGGAGCTTGTTTCGGGTTATAATGTTGAATATGC
CTCTATGTCTTTTGCTTTGTTTTTTCTAGCAGAATATGCACATATTATTTTTATGAGTTGTTT
GACTATTCTTTTATTTTTTGGGGGATGGTTGATTTTTCCTTTTGGTTTAAAAGTTGTTTTTA
TTGTTTGGTTTTTTTTATGGGCGCGAGCTTCTTTTCCAAGGATTCGATACGACCAGTTAAT
GGCCTTATTATGGAAAGGGTATTTGCCTTTTAGTTTGGGTGTTGTTCTTTTTGTTGCTAGT

GTTTTGTTTGGATTTAATGGTCCTTCTCCAATATAGAAATGCCATTACGAAAAGATAATCC
GCTTTTATCTCCGGTTAACGGGTTTCTCGTAGATTTGCCTTCTCCTTCAAATATAAGTTATT
TTTGAAACTTTGGATCTTTATTAGGATTATGTTTAGTTTTACAAATTATAACAGGGTGTTTT
TTATCCATGCACTATTGTTCTGATGCATATCTTGCGTTTGCTTCTGTTGGGCATATAATGCG
AGATGTTAATTATGGATTTTTATTAAGATATTTACATGCAAATGGTGCTTCTTTGTTTTTTTT
TTGTCTTTATGTTCATATTGGACGGGGGTTATATTACGGGGGGTATTTAAAATTTCATGTTT
GAAGCGTTGGAGTTTTGCTTTTTTTATTGACTATGGCAACCGCTTTTATGGGTTATGTTTT
GCCTTGGGGCCAAATGTCTTTTTGGGGCGCGACTGTTATTACAAATTTATTATCTGCAATA
CCTTATTTTGGGGTGGATATTGTTCAATGGGTTTGAGGTGGTTTTAGTGTTTCTGGGGCTA
CATTAAATCGGTTTTTTAGTTTACACTTCTTGCTTCCTTTTATTTTAGCTTTTCTTGCTATTA
TTCATTTAGTTTATTTACATGTTGGTGGGTCAAATAATCCTATCGGCTTAAACTCTTCGATG
GACAGTGTTTCTTTTCATACTTTTTATACTTCAAAAGATCTTTTTGGTTTATTTTTTTTATTT
TTTTTATTTTTTTTTTTCGTTTTTTTTTTACCTAATTATTTAGGAGACGCAGAAAATTTTATT
CAGGCGAATTCTTTAGTTACTCCTGTACATATTCAACCAGAGTGGTACTTTTTATTTGCTT
ATGCCATTTTACGTTCGATACCAAATAAATTGGGGGGGGTTGTTGCTATGTTTTGTAGTAT
TTTAGTTTTATTTTTATTGCCCATTTTACATACAAGTGTTTTAAAAGGGTTATCTTTTCGTC
CTTTGGGCCGAATGGCGTTTTGATTTTTGATAGTTAACTTTGGTCTTTTGACTTGAATTGG
GGCGCAAGTAGTTGAAGAACCCTTTATTTTAATTGGGCAGGTTCTTTCTTTTTTCTATTTT
TTTTATTTTTTAGTTCTTTTGCCTGTCTTGGGGGTTTTAGAAAATCAATTATTAAAAAGAA
ATTAGCGAAATTTTTCTTTTAGGCGGTTTTATATTGAGTTTGGTGGTTTTAGGCTTTCGTG
GTTTTCTTTTAAAATTTTCTTTTTTTTTTTTATTGGTTTTCTTTTATTTTTTTTTTTTGAGTT
AGAGTGAGCAAAAATTTTTGTTTTAGTTGGGGGCTTGGCGGTTTTACTTTTATTAGGACC
AAAAAAAGACGAACAAACAGATATTCCTATTTTAAACTTAATTGTTGTTTTAGGTATTTTC
TGTTTAATTTCTTCTAAAAATTGACTTTCTGTTTATTTAGCAATAGAGCTCTCCACTCTTTG
TTTCTTTGTTTTAATTGCCCGAGGGTCTGGTTATAGTGCAGAGGCGGGGTTAAAATATTTT
GTTTTAGGGGCTCTTTCTTCTGGTTTATTTTTATTTGGTTGTGCTTTGTTGTGTGGAACGG
GGGGCAGTGTATATTTTTCTTATATAGAATTGCTTTTTAATTCAAAACAAAGTTTTTCTGAT
GTTTGTGTGCCAATTGGGTATCTTTTGATTATTGCGGCTCTTTTTTTTAAATTGTCTGTTGC
TCCTTTTCATATGTGGGTACCTGATGTTTATGAAGGGGCACCAACAAAAACGGTTGTTTT
ATTAGCTACTGTTCCTAAAATAGGTATTTTTTCTCTTTTATTTTCAATTGGGTTGCCTATTA
GTTTTTTTTTAATTGGAGTTGTTTTTTCTCTTTTCGTTGGGGCTTTAGGGGCTTTAAATCA
AACAAGAATAAAACGGCTTTTAGCTTATAGTGGGATTGGTCATATGGGTTTTATTTTATGG
GGCTTAGTAAATGGCTCTTTTGAAAGCCTACAAGCCAGTTTGGTCTATCTTTTTATATATAT
TATTATGACAATTTGTGTTTTTTCTATAATTTTGAGTTTTAATGTGTATAAAAATTTACTTAT
TGAATTTAGTGGGTTGTCTCGATTTTTACCTCTTCTCTCAATAACTTTTGGGATTGTTTTTT

TTTCTATTGCTGGGATTCCTCCTTTTGCAGGGTTTTTAGGAAAGTGGGTGGTTTTATTATC
GGGCGTTCTTTCTAAATCTTTTTTTATTTTTTTTTTTGCTGTTTTTTGTTCTGTAATTGCTG
GGGTTTATTATGTTAGAATTGTAAAAATACTTTTTTTTCAAAAAAAATTCTTTTTTTTTAATT
GCAACAAAAGCTTTAAAAAAAGAGCCCAAACTAAATTTTAAAAGGGTCTTTTTGATTGG
TCTTTGTGTTTATTTTATTCTTTTTCTTTTTTTTTCTCCCCATTTTGGTTTTTTTTTCACTTC
GCA