PHYLOGEOGRAPHY OF *Mytilus galloprovincialis* IN THE ISTANBUL STRAIT SYSTEM

by

Aslı KURTULUŞ
B.S., Biology, Istanbul University, 2006

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

2009

PHYLOGEOGRAPHY OF *Mytilus galloprovincialis*
IN THE ISTANBUL STRAIT SYSTEM

APPROVED BY:

Assist. Prof. Raşit Bilgin . . . . . .
(Thesis Supervisor)

Prof. Dr. Nüzhet Dalfes . . . . . .

Prof. Dr. Orhan Yenigün . . . . . .

DATE OF APPROVAL (20/02/2009)

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor Assist. Prof. Raşit Bilgin for his continuing support, guidance and invaluable criticism throughout the study and for inspiring confidence and enthusiasm with his positive and kind approach in all situations, that gave rise to an enjoyable and productive research environment.

I would like to express my gratitude to the members of the thesis commitee, Prof. Dr. Nüzhet Dalfes and Prof. Dr. Orhan Yenigün for evaluating my study.

I am grateful to Evrim Kalkan for helping me in the laboratory and during fieldwork. It was a pleasure for me to work with her. I am also thankful to Öncü Maracı and Aydın Mammadov for their help and friendship. I would like to extend my gratitude to Liz Hemond for her kind help and support. I am indebted to Ünsal Karhan for his assistance in the fieldwork, and Arpat Özgül for helping with the preparation of the region map.

I would like to express my special thanks to my family for their encouragement, love and endless support.

This thesis project was supported by TÜBİTAK (Grant No: 107T626) and Boğaziçi University Research Fund (Grant No: 08M104) grants to Raşit Bilgin.

PHYLOGEOGRAPHY OF *Mytilus galloprovincialis* IN THE ISTANBUL STRAIT SYSTEM

Phylogeography of the commercially valuable species, the Mediterranean mussel, *Mytilus galloprovincialis* was investigated in the Black Sea, the Bosphorus and the Sea of Marmara using the CO3 region of mitochondrial DNA to understand whether the Istanbul Strait was a barrier or corridor to gene flow for this species. A total of 96 mussels representing eight localities extending from the Black Sea to the Sea of Marmara were analysed. In total, nine haplotypes were found, and the most common haplotype with almost 90% frequency, was shared among the Sea of Marmara, the Bosphorus and the Black Sea. The results showed that there is no genetic grouping among geographical regions and Istanbul Strait is a corridor to gene flow for the Mediterranean mussel, *Mytilus galloprovincialis*.

İSTANBUL BOĞAZIN'DAKİ *Mytilus galloprovincialis* TÜRÜ MİDYELERİN FİLOCOĞRAFYASI

Bu çalışmada, ticari değeri olan kara midyenin, *Mytilus galloprovincialis*, filocoğrafyası, İstanbul Boğazı'nın bu türde gen akımını engelleyip engellemediğini anlamak için mitokondrial DNA CO3 dizileri kullanılarak incelenmiştir. Karadeniz'den Marmara'ya uzanan sekiz örnekleme bölgesinden toplam 96 midyenin analizi yapılmıştır. Toplamda dokuz haplotip bulunmuş ve en yaygın olan, neredeyse bireylerin %90'ında bulunan haplotip, Karadeniz, Boğaz ve Marmara denizlerinde görülmüştür. Sonuçlar coğrafik bölgeler arasında genetik gruplaşma olmadığını ve İstanbul Boğazı'nın kara midyenin gen akımına koridor olduşturduğunu göstermiştir.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation
DNA	Deoxyribonucleic acid
mtDNA	Mitochondrial deoxyribonucleic acid
RNA	Ribonucleic acid
mRNA	Messenger ribonucleic acid
tRNA	Transfer ribonucleic acid
PCR	Polymerase Chain Reaction
dNTP	Deoxy-ribonucleotide triphosphate
bp	Base Pairs
ka	One thousand years ago
ϕ_{ST}	Coancestry coefficient

1. INTRODUCTION

The Istanbul Strait, connecting the Marmara Sea to the Black Sea, acts as a biological barrier, corridor and acclimization zone for various species (Öztürk & Öztürk, 1996). Geographical and hydrographical properties of the water bodies around the Istanbul Strait and the Sea of Marmara also create a special ecosystem for many unique organisms.

For some species, the Marmara Sea serves as a barrier. For example, *Posidonia oceanica* (Linnaeus, 1758) occurs only in the Mediterranean Sea and around the southern coasts of Australia. The Dardanelles Strait limits the northern distribution of this species (Augier, 1985). Likewise it limits the southern distribution of the cetacean species, harbor porpoise, *Phocoena phocoena* (Linnaeus, 1758), living usually in the Black Sea and partially in the Marmara Sea (Öztürk, 1995).

The Marmara Sea also plays an important role for migration of many species, as a corridor. Most of the Mediterranean zooplankton and phytoplankton species penetrate into the Black Sea with the Istanbul strait. Calycophoran siphonophore, *Muggiaea kochi* (Will, 1844), heteropod, *Solmundella mediterranea* (Quoy and Gaimard, 1833), calanoid copepods, *Eucalanus crassus* (Giesbrecht, 1888) and copepod, *Calanus tenuicornis* (Dana, 1849) are some examples of these migratory species (Nalbantoğlu, 1957). In the spring, the direction of this migration is from the Aegean Sea to the Black Sea through the straits and vice versa in the autumn.

In addition, the Istanbul Strait System and the Marmara Sea act as an acclimization zone. In the Black Sea, 150 zoobenthic species are of Mediterranean origin and are entirely discovered in the entrance of the Bosphorus Strait (Bacescu et al., 1971). The dispersion from the Mediterranean to the Black Sea implies that there is an adjustment to a new environment. This indicates that these 150 species were acclimatized gradually to the environmental conditions of the Black Sea.

In this study, genetics methods were used to make an inferences on the natural and evolutionary history of the the commercially valuable species, the Mediterranean mussel, *Mytilus galloprovincialis* (Lamarck, 1818), in the region. In this regard, whether the Istanbul Strait was a barrier or corridor for gene flow was investigated.

The relevant null hypothesis was framed as follows:

H0: Istanbul Strait is not a barrier to gene flow for the Mediterranean mussel, *Mytilus galloprovincialis*.

Under this null hypothesis, our expectation was that no genetic differences would be seen between the populations of the Black Sea and the Marmara.

Our alternative hypothesis was as follows:

H1: Istanbul Strait is a barrier to gene flow for the Mediterranean mussel, *Mytilus galloprovincialis*.

Under this alternative hypothesis, the expectation was that genetic differences would be seen between the populations of the Black Sea and the Marmara.

With this perspective in mind, the basic hydrological and geological charactersitics of the Istanbul Strait System, and the basic biology of the Mediterranean mussel will be discussed, a synopsis on the use of mtDNA on making inferences about evolution of species will be made and previous genetic research on and in the vicinity of the Istanbul Strait will be summarized.

1.1. Basic Characteristics of the Istanbul Strait System including the Bosphorus, the Black Sea and the Marmara Sea)

1.1.1. Hydrological Properties

The Marmara Sea is an inland sea that connects the Black Sea to the Aegean. It is linked to the Black Sea through the Bosphorus Strait and to the Aegean Sea through the Dardanelles Strait. Hydrography and the surface circulation of the Marmara Sea is influenced by the Black Sea and the Aegean. The Black Sea water causes an upper current while the Aegean water enters the Marmara Sea as a deeper layer flow (Beşiktepe et al., 1994). This pattern is also seen in the Istanbul Strait, forming a wedge of these two water bodies.

The distribution of these two layers of water effects temperature and salinity ranges of the Istanbul Strait. While the temperature of the surface water of the Bosphorus varies between 4 to 24 ^{0}C and the salinity ranges 10 and 18 PSU, the temparature and salinity of deeper water has significant changes. The salinity reaches 39 PSU at a depth of below 30 m. (Kocataş et al., 1993).

1.1.2. Geological History of the Marmara Sea and Istanbul Strait

Paleoceanography history of the Marmara Sea has been very complex from the last glacial maximum to the Holocene. Marmara Sea was isolated from both the Black Sea and the Aegean Sea. During the glacial periods, there were two different stages. These included the lake stage and gateway stage, depending on the levels of the global sea level. During the lake stage, global sea level decreased and the Marmara had no connection with the Straits of Bosphorus and Dardanelles. However, in the gateway stage the Marmara Sea reconnected through both straits.

These inferences were made by examination of sapropels, which are dark-coloured sediments that are rich in organic matter. Organic carbon concentrations in sapropels commonly exceed 2% in weight and they are thought to develop during episodes of

reduced oxygen availability in bottom waters. In the Marmara Sea, cores from two sapropel layers were identified as sapropel M2 and sapropel M1. M2 was deposited between ~29.5 and 23.5 ka while M1 was deposited between ~10.5 and 6 ka (Aksu et al., 2002a). According to micropaleontological data, sapropelic sediments at the basal section of the cores reflect the lake stage of the Marmara Sea, when it was isolated both the Black and Mediterranean Seas. Benthic foraminifera distribution data to deep sea core data have been utilized to identify paleoceanographic changes in intermediate and deep water circulation. During glacial periods, there was a reduction in deep-water circulation, which can cause a serious decrease in deep-water oxygen concentrations due to biological oxygen demand associated with the decay of organic matter that sinks into the deep sea. The absence of benthic foraminifera and the rare dysoxic species in sporel M1 suggest that deep water conditions during the deposition of sapropel M1 must have been close to anoxic. Micropaleontological results show that at ~10.5 ka the Black Sea rose to the breach depth of the Bosphorus and flowed into the Marmara Sea, developing a cooler, low-salinity surface layer,and promoting the deposition of sapropel M1 in the Marmara Sea (Aksu et al., 2002).

1.2. Basic Biology of the Mediterranean Mussel, *Mytilus galloprovincialis*

1.2.1. Taxonomy and Distribution

The common name “mussel” is used for members of several different families of clams or bivalve molluscs. Marine mussel species that belong to the family Mytilidae, in the order Bivalvia, live in intertidal and subtidal areas along coastlines worldwide.

The marine mussel genus *Mytilus* beloging to the family Mytilidae includes a complex of three sibling species, *Mytilus edulis, Mytilus trossulus* and *Mytilus galloprovincialis.* All three are now globally widespread and form hybrid zones where they overlap (McDonald et al., 1991, Sarver & Foltz, 1993, Hilbish et al., 2000). *M. edulis* (Linnaeus, 1758) and *M. galloprovincialis* are the most closely related according to genomic DNA (Martinez-Lage et al., 2002) and mtDNA analysis (Hilbish et al., 2000). *M. galloprovincialis* is the most widely distributed of the three sibling species (Daguin & Borsa, 2000). Taxanomy of this species is given in Table 1.1.

Table 1.1. Taxanomy of *Mytilus galloprovincialis.*

Kingdom	Animalia
Phylum	Mollusca
Class	Bivalvia
Order	Mytiloida
Family	Mytilidae
Genus	Mytilus
Species	galloprovincialis

The Mediterranean mussel, *Mytilus galloprovincialis*, is native to the Mediterranean coast, the Black Sea and the Adriatic, and was introduced to the South Africa, east and west North America, Hawaii, and north-eastern Asia (Branch and Steffani, 2004).

1.2.2. Morphology and Life History

The Mediterranean mussels (*Mytilus galloprovincialis)* are dark blue or brown to almost black, with two shells that are equal and nearly quadrangular. They are filter feeders, with gills that filter plankton and other microscopic sea creatures, which are free-floating in seawater. As the water flows through their bodies, sediments, organic matter, and pollutants are filtered out and ingested so they contribute to improve water quality. Like other mussels this species have gonochoristic reproduction where males and females spawn simultaneously. Fertilization occurs outside the body, with a larval stage that drifts for three weeks to six months, before settling on a hard surface as a young mussel.

1.2.3. Genetics

In species of the blue mussel family Mytilidae, transmission of mtDNA differs from the standart maternal inheritance system (Zouros et al., 1994). These species contain two types of mtDNA genomes, one that is transmitted maternally (the F type) and another that is transmitted paternally (Ladoukakis et al., 2002). In doubly uniparental inheritance (DUI) females are homoplasmic for an F type and males heteroplasmic for an F and M type mitochondrial genome.

Occasionally, F genome may invade the paternal route and create a heterogenous pool of paternally transmitted molecules (Steward et al., 1996). With the aid of the mtDNA variation in species *Mytilus galloprovincialis*, Ladoukakis et al. (2002) have shown that there is a strong differentiation between the Mediterranean and the Black Sea, but much smaller than between each of these seas from the Atlantic. The authors also pointed out "our results regarding the degree of population differentiation within the Mediterranean and the Black Sea, as well as among the tree major Seas, are based on small samples sizes. This is particularly true for the Black Sea, from where we were able to secure only one sample. Obviously, a more elaborate sampling scheme is needed before we could have a good idea of the gene flow of *M. galloprovincialis* between these large water bodies.'' This remark also comprised the starting point for this study.

1.3. Mitochondrial DNA Synopsis

The use of mtDNA data have facilitated progress in many areas in studies of organismal evolution. Following the discovery of the faster rates of evolution in mtDNA compared to other genomic regions (Brown et al., 1979), the discovery of polymerase chain reaction (PCR) (Mullis and Faloona, 1987) and development of universal primers (Kocher et al., 1989), and finally through advancements in sequencing technology, a greater degree of resolution has been obtained in terms of the extent of mtDNA data obtained, making it a useful and powerful tool in addressing a wide spectrum of evolutionary questions.

One of the most important uses of mtDNA is related to phylogenetic questions for resolving the taxonomic relationships of species and higher taxa. This is possible, especially because of the hierarchical branching structure of mtDNA gene trees due to maternal inheritance and clean branching structure due to lack of recombination (Avise, 1994). Therefore, mtDNA methods comprised additional and complementary tools to using morphological, ecological and behavioral traits, for assessing systematic relationships. The increase in computational power over the years helped the applications of relatively more sophisticated maximum parsimony and maximum likelihood approaches to the analyses of large data sets (Page and Holmes, 2000).

The mtDNA in animals is composed of 38 units (including 22 tRNAs and two rRNA genes), which evolve at different rates (Alexeyev et al., 2004; Avise, 1994). For investigating phylogenetic questions and hypotheses at species and higher levels, mentioned above, slowly evolving genes like cytochrome-b are useful, due to lower degrees of saturation (Irwin et al., 1991; Kocher et al., 1989; Yang and Yoder, 1999). On the other hand, analyses with faster evolving regions like the D-loop resulted in fine scale studies within species, culminating in founding of fields such as "intraspecific phylogeography" (Avise 2000), which investigates relationships between gene genealogies and geography. In animals, many of these studies showed extensive genetic structure, even within species, which were not distinguishable morphologically, as well as recognition of "cryptic species" (Johnson and Cicero, 2004; Narang et al., 1993; Rocha-Olivares et al., 1999; Wilcox et al., 1997).

In addition to the methods discussed above, which rely on building trees to a greater extent, population genetics methods have also been incorporated into analyzing mtDNA data, providing another venue for investigating mtDNA differentiation. Especially the development of mtDNA analogues of the conventional F-statistics (Wright, 1943), such as Φ_{st} (Excoffier et al., 1992) and N_{st} (Lynch and Crease, 1990), made it possible to use population genetics theory to draw conclusions on the causes of differentiation of populations, which could be interpreted in a similar way and along with nDNA data.

With the greater availability of sequence data, the theory behind analysis of mtDNA have become more creative and complex. The initial relatively simple symmetrical models of sequence evolution (Jukes and Cantor, 1969) were appended by more complicated models of evolution that take into account asymmetrical transition and transversion probabilities (Tamura and Nei, 1993), as well as tests to see how the data fit these models (Posada and Crandall, 1998). The coalescent theory provided a basis to draw inferences about population history based on the relationships of haplotypes (Rogers and Harpending, 1992). Again, the increase in computational power made it possible to use Bayesian statistics, through Markov Chain Monte Carlo procedures, in conjunction with the coalescent theory (Storz and Beaumont, 2002). Likelihood methods have also been developed to estimate historical population size and mutation rates (Kuhner et al., 1998), as well as to quantify directional levels of migration among populations (Beerli and Felsenstein, 1999)

1.4. The use of genetic methods in the studies about the Istanbul Strait system

Genetics methods have been useful in understanding the properties of strait systems as barriers or corridors to gene flow. For instance, the Gibraltar Strait's character in preventing or facilitating migration was examined by various genetic studies (Patarnello et al., 2007). Similarly various genetics studies helped to understand the properties of the Istanbul Strait System as barriers or corridors to gene flow. These were on *Zostera marina*, *Sagitta setosa*, *Engraulis encrasicolus*, and *Phocoena phocoena*. These studies are individually described below. A summary of the findings of these studies can be found in Table 1.2.

Table 1.2. A summary of the marine genetics studies encompassing the Black Sea and the Mediterranean.

Species	**Genetic marker**	**DNA Analyses**	**Black Sea vs. Mediterranean differentiation**
Zostera marina	Nuclear rDNA-internal transcribed spacer (ITS) and chloroplast matK-intron and nine microsatellite loci	Sequencing	No differentiation
Sagitta setosa	Mitochondrial DNA, four microsatellite markers	RFLP RAPD	Differentiation
Engraulis encrasicolus	Mitochondrial DNA	RFLP	No differentiation
Phocoena phocoena	Mitochondrial DNA	Sequencing	Strong differentiation
Mytilus galloprovincialis	Mitochondrial DNA COIII gene (20 restriction sites) and 16S RNA gene (4 restriction sites)	RFLP	Significant differentiation

Zostera marina (seagrass), the most basal and widely distributed species of the Zosteraceae family, originated in the Pacific between 8- 20 mya. *Zostera marina* had an opportunity to spread out into the Atlantic only within the past 3.5 mya, with the aid of the opening of the Bering Strait. In order to investigate North Atlantic phylogeography and large scale population differentiation of the seagrass *Zostera marina*, Olsen et al. (2004) collected samples from different locations across the Northern Hemisphere. Sampling was done from Baltic Sea-Finland, Wadden Sea-North Sea, Atlantic coast-Europe,

Mediterranean Sea–France, Black Sea-Azov Sea, and the Atlantic and the Pacific coasts of USA-Canada. Sequences were used from the nuclear rDNA-internal transcribed spacer (ITS) and chloroplast matK-intron and nine microsatellite loci. According to rDNA –ITS and matK intron sequences, no varition was detected among Atlantic isolates which ranged from the west Atlantic to the Black Sea. Based on microsatellite data, the neighbour-joining tree demonstrated that the east Pacific group consistently linked with the west Atlantic group and never with the east Atlantic. General allelic richness was higher in the Pacific than in the Atlantic. The highest genetic diversity was found in the Wadden Sea-North Sea –southwest Baltic region, although these areas could only have been recolonized within the past 7500 years.

In the central Baltic and the Black Sea, gene flow and genetic drift were not in equilibrium, because the respective populations had low clonal diversity and high population differentiation. The populations in the Aland Archipelago in the central Baltic were isolated and connected to the neighbouring subpopulations in the form of stepping stones. However, in the Black Sea or the Mediterranean Sea, no such stepping stones were observed. Also, in the Mediterranean Sea and the Black Sea, no significant isolation by distance was found.

Peijnenburg et al. (2006) studied the spatial and temporal structure of the planktonic *Sagitta setosa* (Chaetognatha) in the European seas. The researchers used both mitochondrial and DNA microsatellite markers (RFLP analysis for variation within the mitochondrial COII region and four nuclear-encoded microsatellite loci). They collected samples from the Northeast Atlantic, the Mediterranean Sea and the Black Sea. Combined analyses of mitochondrial and nuclear data for individuals indicated significant differentiation between populations from the NE Atlantic, the Mediterranean and Black Sea. This indicated that there probably was no gene flow between basins and populations were disjunct. The highest levels of differentiation were between the Mediterranean and Black Sea populations, as detected by microsatellite data.

When genetic diversity within basins was examined, no evidence of spatial and temporal genetic structuring was found within the NE Atlantic. However, significant structuring was found within the Mediterranean Sea that divided into several sub-basins

and having some fragmented neritic habitats. Within the Black Sea populations, data analyses were not conclusive.

Magoulas et al. (2006) studied the phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*) with the aid of the mitochondrial DNA. Samples of fish were collected from European Seas including the Black Sea , the Aegean Sea, the Mediterranean and the Northeastern Atlantic. Eighty-eight haplotypes determined two divergent clades (A and B). Clade B supported only a scenario of population growth while, Clade A showed a typical pattern of recent population expansion and had lower genetic varibility in the Black Sea. They concluded that clade A originated in the Black Sea after isolation from Mediterranean in the Pleistocene. The genetic differentiation among populations was high with the greatest differences being observed between basins. European anchovy occupies coastal seas, isolated from one another by peninsulas and narrow straits. This isolation is the most likely cause for the reduction of gene flow between regions. For example, the Adriatic and the Aegean, which are partially isolated from the Mediterranean, and the Bosporus Strait restrict gene flow between the Black Sea from the Mediterranean. The history, complex geography, and hydrography of the Mediterranean influence genetic variability among populations of this species.

As mentioned before, the Marmara Sea acts as a barrier for some species like the case of the Black Sea harbour porpoise (*Phocoena phocoena*). Viaud-Martinez et al. (2007) used cranial morphology and mitochondrial DNA variation in order to evaluate the degree of morphological and genetical differentiation of this species. Tissue samples were collected from the eastern North Atlantic, the Northern Aegean Sea, the Sea of Marmara and the Black Sea. The authors observed that male and female specimens were significantly smaller in the Black Sea than other populations for most of the measurements. They also found no shared haplotypes among the eastern Atlantic populations and the Black Sea population. Their results showed that harbor porpoises from the Black Sea, the Marmara Sea and the northern Aeagen Sea (eastern region) were reproductively isolated for thousands of years and diverged genetically. Consequently, they recommended a new subspecies classification for the eastern harbor porpoises.

Ladoukakis et al. (2002) examined the mediterranean mussel, *Mytilus galloprovincialis*, in a similiar study. They used mitochondrial DNA variation in samples that were collected from the Black Sea, the Mediterranean and the Spanish Atlantic coast. This species carried two types of mitochondrial genomes (F and M type). The F type was transmitted maternally to offspring of both sexes and the M type was transmitted paternally to male descendants only. It was an exception to the rule of uniparental maternal transmission of mtDNA, and is known as doubly uniparental inheritance (DUI; Zouros et al. 1994).The majority of males were heteroplasmic for an F and M genome. According to Nei haplotype distances and F_{st} values for the three major seas, populations were more differentiated for the M genome when compared to the F genome. The F genome showed higher degrees of diversity within populations, however, the M genome exhibited higher degrees of differentiation between populations. For some individuals, invasion of the paternal route by the F lineage was observed (masculinization of the F molecule). In the Black Sea sample, 20 of the 30 males did not contain the M mtDNA type. Because of the faster evolution and turn-over rate of the M lineage, combined with its liability to invasion from the F lineage, M mtDNA lineage was not reliable for phylogenetic and biogeographic studies of *Mytilus galloprovincialis.* The differentiation of the Mediterranean and the Black Sea was significant, but not higher than that between each of these seas from the Atlantic. Therefore, the authors concluded that gene flow among the three seas was restricted and not enough to erase the effects of mutation and random drift.

.

2. THESIS OBJECTIVE

This study aims to investigate phylogeography of the commercially valuable species, the Mediterranean mussel, *Mytilus galloprovincialis* in the Black Sea, the Bosphorus and the Sea of Marmara using the CO3 region of mitochondrial DNA to understand whether the Istanbul Strait acts as a barrier or corridor to gene flow for this species. In this vein various potential factors that could influence the genetic differentiation of the populations, such as salinity differences, isolation by distance, and historical bottlenecks will also be examined. Consequently, making inferences on the natural and evolutionary history of the Mediterranean mussel, *Mytilus galloprovincialis* in the vicinity of Bosphorus comprises the main objective of this thesis.

3. MATERIALS AND METHODS

3.1. Sample Collection

A total of 96 mussel samples were collected and analyzed from eight locations including two (Kilyos and Riva) in the Black Sea, three (A.Kavağı, R.Hisarı, Kuzguncuk) from the Bosphorus, and three (Kalamış, Burgazada, Mürefte) from the Sea of Marmara (Table 3.1. , Figure 3.1.).

Table 3.1. List of sampling location, sample size and geographic coordinates of the sites.

Code	Sampling site	N	Geographic coordinates (N, E)
1	Kilyos	5	41.25172°, 29.03846°
2	Riva	11	41.23384°, 29.22681°
3	Anadolu Kavağı	13	41.172986°, 29.088372°
4	Rumeli Hisarı	11	41.087944°, 29.056931°
5	Kuzguncuk	12	41.03848°, 29.03213°
6	Kalamış	17	40.977483°, 29.037969°
7	Burgazada	14	40.88585°, 29.060728°
8	Mürefte	13	40.67649°, 27.26489°

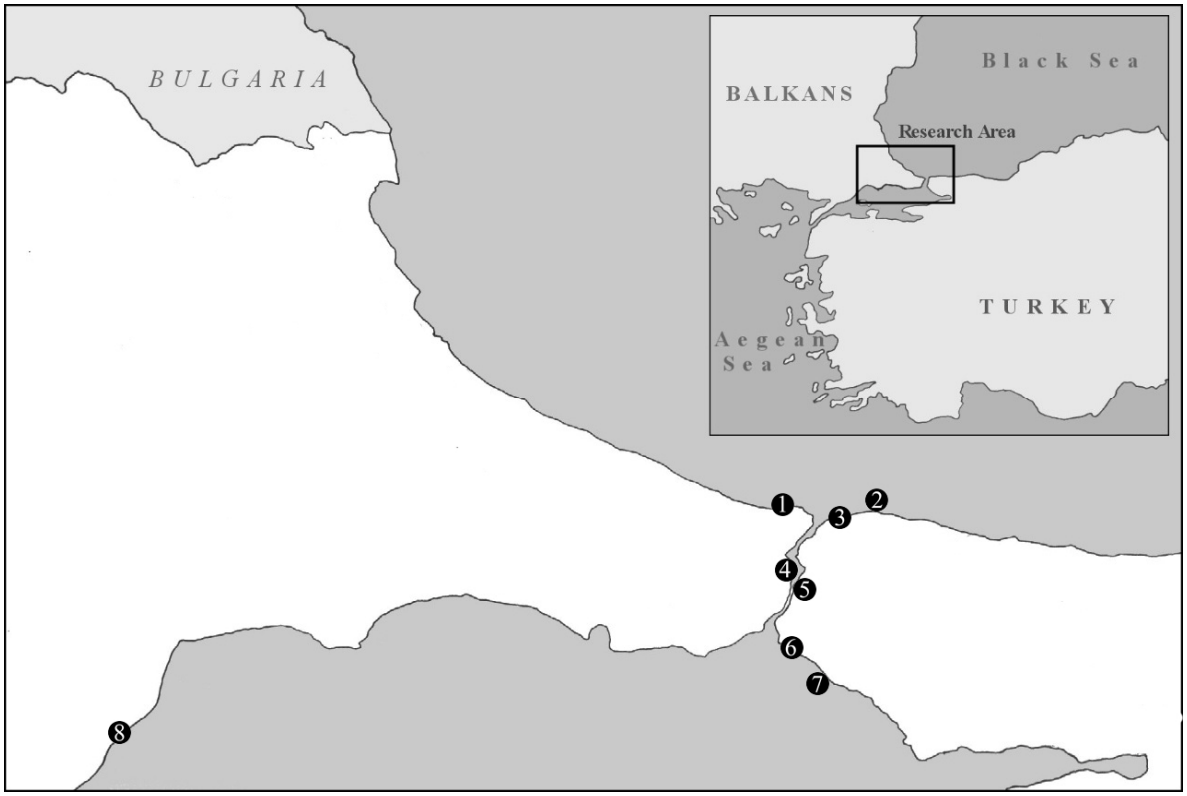

Figure 3.1. The map shows sampling locations. Numbers indicate the codes of the sampling sites as given in Table 3.1.

In order not to be influenced by the effects of isolation by distance, sampling points from the Black Sea and The Sea of Marmara were selected in the vicinity of the Bosphorus, and not from farther locations. The temperature and salinity of each sampling site were measured using the WTW Multi 340i set. The results for salinity and temperature are given in Table 3.2.

Table 3.2. Temperature and salinity values of sampling sites.

Sampling site	Temperature (ºC)	Salinity (PSU)
Kilyos	25.7	17.9
Riva	27.5	17.9
Anadolu Kavağı	21.5	17.6
Rumeli Hisarı	17	17.9
Kuzguncuk	24.7	18.5
Kalamış	24.5	23.1
Burgazada	22	23.2
Mürefte	26.5	23.4

3.2 Laboratory Protocols

Mussels were opened in the laboratory and tissue samples from gonad and gills were removed for genetic analysis. Roche DNA Extraction Kit (Mannheim, Germany) was used for the DNA extraction applying the manufacturer protocol. To a nuclease-free 1.5ml microcentrifuge tube, 25-50 mg tissue sample, 200 µl tissue lysis buffer and 40 µl proteinase K were added, mixed immediately and incubated for one hour at 55°C or until tissue is digested completely. After that 200 µl binding buffer was added to the tube and incubated for 10 minutes at 70°C. Before drawing part of the sample into a 1 ml disposable pipette tip, 100 µl isopropanol was added and mixed well. The pipette tip carrying the insoluble tissue segments was discarded. One high filter tube was inserted in one collection tube and the remainder of the liquid sample was transfered into the upper buffer reservoir into the filter tube. The entire high pure filter tube assembly was inserted into standart table top centrifuge and centrifuged 1 minute at 8000x g. After centrifugation, filter tube was removed and combined with a new collection tube. 500 µl inhibitor removal buffer was added to upper reservoir of the filter tube and centrifuged for minute at 8000× g. After centrifugation, filter tube was removed and combined with a new collection tube again. 500 µl wash buffer was added to upper reservoir of the filter tube and centrifuged for one minute at 8000x g. After this, the wash buffer procedure was applied once more. After discarding the flowthrough liguid, the entire high pure assembly was centrifuged for an additional 10 seconds at full speed. To elute the DNA, the filter tube was inserted into a clean, sterile 1.5 ml microcentrifuge tube and 200 µl prewarmed elution buffer was added and centrifuged for 1 min at 8000× g . The microcentrifuge DNA contained the eluted DNA. The eluted DNA were checked on a 1% agarose gel, prepared in 1XTBE (Tris base, boric acid, EDTA) buffer with ethidium bromide. The eluted DNA was mixed with 3 µl of 2X loading dye (Fermentas) and loaded on the gel. Sample were run at 90V for 40 minutes and the band images were taken under ultraviolet light by Biorad Gel Doc Imaging System (Figure 3.2).

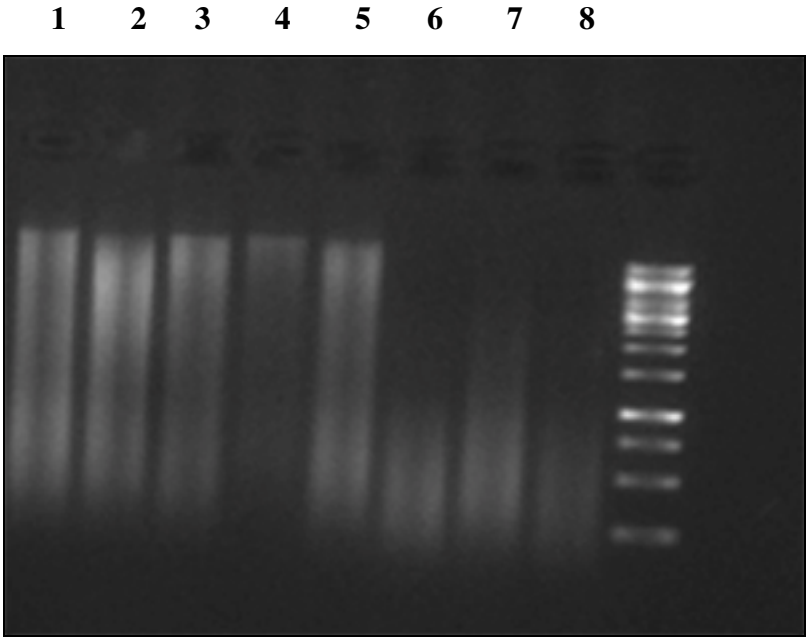

Figure 3.2. The gel electrophoresis of eluted DNA of eight of the samples

The CO3 region of mitochondrial DNA was amplified and sequenced in both directions using the primers 5'-TATGTACCAGGTCCAAGTCCGTG-3' and 5'-ATGCTCTTCTTGAATATAAGCGTAC –3' (Zouros et al., 1994). 2 µl of DNA was added to a 50 µl reaction mixture, containing 5 µl 10x high fidelity buffer, 6 µl $MgCl_2$ (25 mM), 1 µl of 10 mM deoxyribonucleotide triphosphate (dNTPs), 0.625 µl of each primer (20µM), 34,5 µl H_2O and 1.25 units of Taq DNA polymerase. The PCR thermal profile was 2 min at 94°C, 35 cycles of {1 min at 94°C, 45 sec at 54°C, 1 min at 72°C}, and 7 min at 72°C. The PCR products were purified by using invitrogen PureLink PCR Purification Kit. 200 µl of Purelink binding buffer with isopropanol was added to 50 µl of PCR product and mixed well. Sample with binding buffer was added to the Purelink spin column and the column was centrifuged at 10,00x g for 1 minute. After discarding the flowthrough liguid, the spin column was placed into the collection tube. 650 µl of wash buffer with ethanol was added to the column and the column was centrifuged at 10,000x g for 1 minute. 2.5 minutes additional centrifugation was applied at maximum speed, to remove any residual wash buffer. Subsequently, the spin column was placed in a clean 1.7-ml Purelink elution tube supplied with the kit. 50 µl elution buffer was added to the center of the column and incubated for one minute. Afterward, the column was centrifuged at maximum speed for two minutes. The purified PCR product was obtained in the elution tube. The purified PCR products were checked on 1% agarose gels, prepared in 1XTBE buffer with ethidium

bromide. The each purified PCR product was mixed with 3 µl of 2X loading dye and loaded on the gel. Samples were run at 90V for 40 minutes and the band images were taken under ultraviolet light by Biorad Gel Doc Imaging System.

Purified PCR products were sent to the Macrogen Inc. in South Korea for sequencing. The obtained sequences were cleaned manually with Sequencher v. 4.1 (Gene Codes Corp.), and aligned using Clustal X (Thompson et al., 1997).

3.3. Data Analysis

Only F type mtDNA genomes were used for data analysis. Comparing the sequences, when F like paternal genomes and M type genomes were detected, they were excluded from the analyses. In addition, F-type haplotypes that showed evidence for recombination were excluded from the analyses.

The mtDNA data were used to calculate basic descriptive statistics in DnaSP v. 4.0 (Rozas et al., 2003). Phylogenetic analysis were performed using distance (Neighbor-Joining) and maximum parsimony analysis as implemented in PAUP* v.4.0.6b (Swofford, 2001). The robustness of the topologies were tested by bootsrapping (1000 replicates). Two sequences (BS12, BS81) retrieved from GenBank, from the Azov Sea, Ukraine, were added to the mtDNA data to construct the trees that depict the relationships of individuals (GenBank accession numbers DQ445468 and DQ445474, respectively).

The evolutionary relationships among haplotypes were determined using network parsimony reconstruction in TCS v. 1.21 (Clement et al., 2000). Nested Clade Analysis (NCA) (Templeton, 1998) was used for examining haplotype differences within species. NCA uses a coalescent approach to analyze geographical structuring of haplotypes based on the hypothesis that there is a random relationship between them. If a non-random geographical and haplotype relationships are found, NCA can show which of the potential factors such as recent gene flow, past population fragmentation, habitat expansion or long distance colonization best explains these relationships. GeoDIS v. 2.0 (Posada et al., 2000) program and its inference key were used for this analysis. The latest inference key from December 11, 2008 was used for this analysis. The key can be found in Appendix B.

Furthermore, the extent of population genetic structure was evaluated by the **A**nalysis of **Mo**lecular **Va**riance (AMOVA). As an additional measure of population differentiation, pairwise ϕ_{ST} values between each sampling site were calculated in Arlequin v. 3.11 (Excoffier et al., 2005).

Within the same species, the relationship between genetic distances and geographical distances would be expected to be linear in the absence of any barriers (physical, climatic, social, etc.) to gene flow. The presence of such a correlation indicates that the populations differentiate under "isolation by distance" (Wright, 1943). The presence/absence of isolation by distance (IBD) was examined by plotting genetic distances against geographic distances, and examining their correlation. Genetic distance calculations were based on pairwise population ϕ_{ST} values. Four combinations of untransformed and log-transformed geographic and genetic distances were plotted, to account for different dimensional considerations (Rousset, 1997). The geographic distances were measured as straight lines along coastline between sampling sites, using Google Earth. Additionally, to test whether populations experienced selection or recent demographic expansions R_2 (Ramos-Onsins and Rozas, 2002), Fu's F_s (Fu and Li, 1993), and Tajima's D (Tajima, 1989) neutrality tests were performed using DnaSP v. 4.0 (Rozas et al., 2003).

4. RESULTS AND DISCUSSION

A total of 813 base pairs (bp) were sequenced as a maximum from all of the samples. The chromotogram of an example sequence (individual MG010) obtained from Macrogen Korea, is provided in Figure 4.1. 177 of 813 bp positions were considered as alignment gaps or missing data; hence 636 bp were sequenced for all individuals. The first PCR products had extra bands with low density (Figure 4.2.) however, the extra bands were not observed in the gel electrophoroses of purified PCR products (Figure 4.3.).

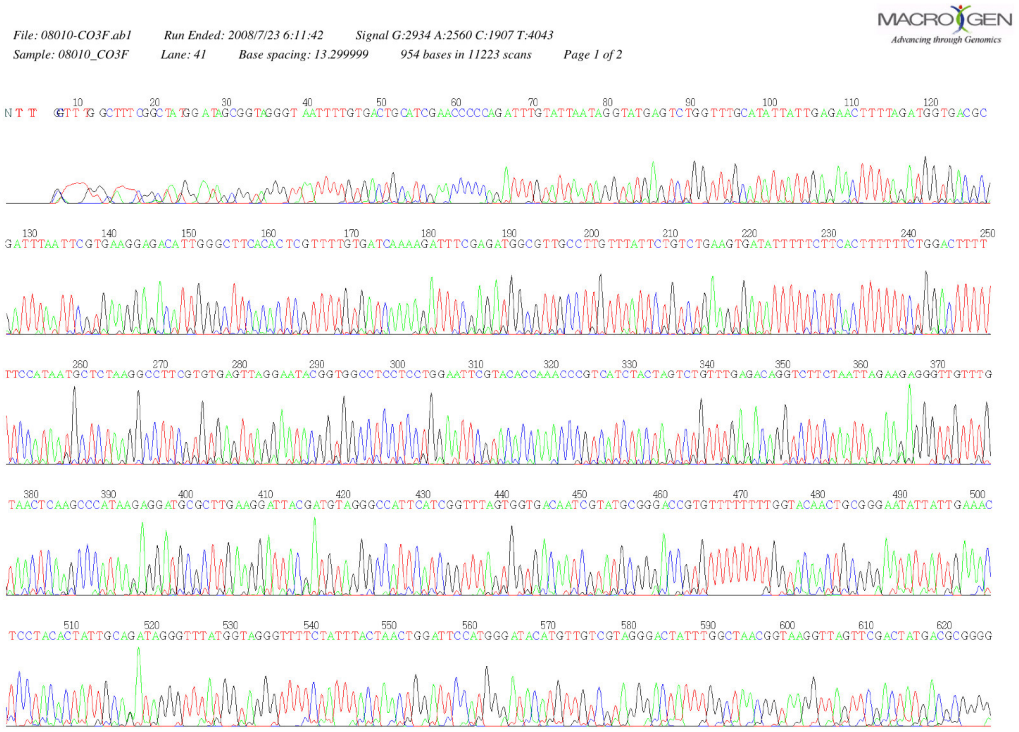

Figure 4.1. The chromotogram of an example sequence obtained from Macrogen Korea (individual MG010)

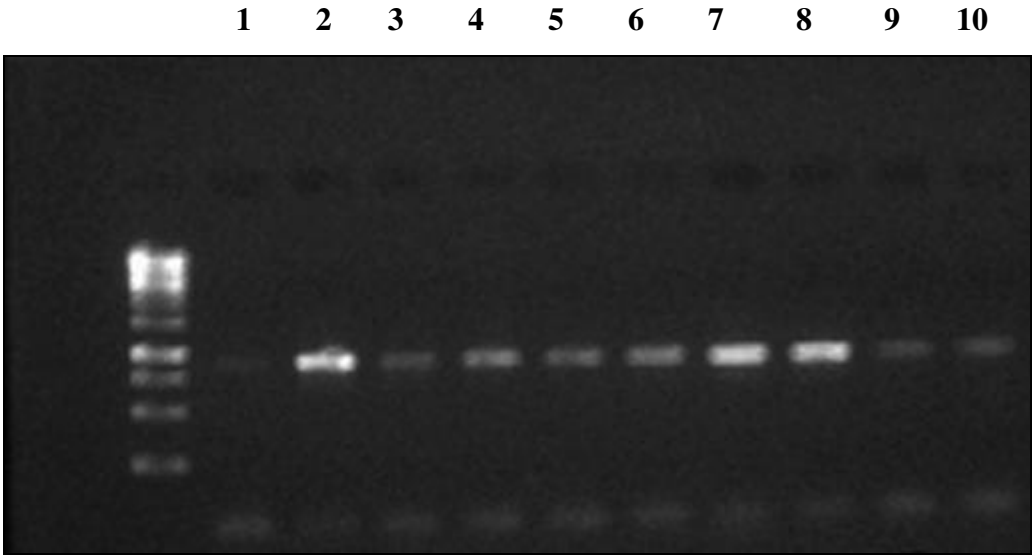

Figure 4.2. The gel electrophoresis of PCR products of ten of the samples

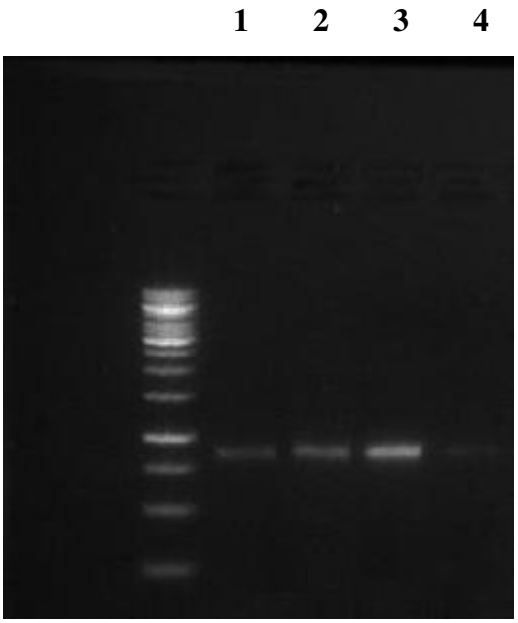

Figure 4.3. The gel electrophoresis of purified PCR products of four of the samples

In total, nine haplotypes were found in 96 mussels. Haplotypes base differences are shown in Table 4.1. The base sequences of each haplotype can be found in the Appendix C.

Table 4.1. Haplotypes base differences. Numbers refer to base positions where haplotypes differ from one another. Points indicate sharing same bases with BS012.

	12	31	53	86	90	102	128	356	389	483
BS012	T	G	T	G	C	T	C	T	C	C
BS081	.	C	G	.	.	.	.	.	.	.
MGH01	A	C	G	.	.	.	.	.	.	.
MGH02	-	C	.	.	.	.	.	.	.	.
MGH03	A	C	G	A	.	.	.	.	.	.
MGH04	A	C	G	.	T	.	.	.	.	.
MGH05	-	C	G	.	.	C	.	.	.	.
MGH06	A	C	G	.	.	.	G	.	.	.
MGH07	A	C	G	.	.	.	.	C	.	.
MGH08	-	-	G	.	.	.	.	.	T	.
MGH09	A	C	G	.	.	.	.	.	.	T

Both haplotype and nucleotide diversities were low. Haplotype diversity (h) was 0.230 (± 0.057 + 1 S.D.) and nucleotide diversity (π) was 0.00038, respectively. A total of eight polymorphic sites were detected within 636 base pairs, among which five sites were singleton variable sites with two variants, and three sites were parsimony informative sites with two variants. The Tajima's *D* (Tajima, 1989) test for the neutrality was not significant (D = -0.05286). R_2 (Ramos-Onsins and Rozas, 2002) and Fu's F_s (Fu and Li, 1993) values were also not significant, indicating no statistical support for population expansion.

The average numbers of nucleotide differences between regions, Black Sea and Bosphorus, Black Sea and the Sea of Marmara, Bosphorus and the Sea of Marmara were 0.250, 0.225 and 0.250 respectively. The average number (Dxy) and the net number (Da) of nucleotide substitutions per site between populations are provided in Table 4.2. Dxy values of Black Sea, Bosphorus and Bosphorus, Sea of Marmara were the same (Dxy = 0.00039 ± 0.00020). Dxy between Black Sea and Sea of Marmara was 0.00035 ± 0.00019. All Da values were equal to zero which means that no net nucleotide substitution per site exists between populations.

Table 4.2. The average number of (Dxy) and the number of net (Da) nucleotide substitution per site between populations.

	Black Sea	**Bosphorus**	**Sea of Marmara**
Black Sea		Da = 0.00000 ± 0.00020	Da = 0.00000 ± 0.00019
Bosphorus	Dxy = 0.00039 ± 0.00020		Da = 0.00000 ± 0.00016
Sea of Marmara	Dxy = 0.00035 ± 0.00019	Dxy = 0.00039 ± 0.00016	

Other statistical treatments of population differentaition also indicated lack of any genetic differentiation. Population pairwise ϕ_{ST} values were low and non significant (Table 4.3.).

Table 4.3. Population pairwise ϕ_{ST} values between each sampling site.

	Kilyos	**Riva**	**Kuzguncuk**	**R.hisarı**	**Burgazada**	**Murefte**	**Kalamış**	**A.Kavağı**
Kilyos	0.00000							
Riva	0.31250	0.00000						
Kuzguncuk	0.33386	0.00000	0.00000					
R.hisarı	0.01561	0.05000	0.05979	0.00000				
Burgazada	0.06789	0.01710	0.02484	0.00000	0.00000			
Murefte	0.35369	0.00000	0.00000	0.06888	0.03191	0.00000		
Kalamış	0.04213	0.02638	0.03363	0.00000	0.00000	0.04024	0.00000	
A.Kavağı	0.00470	0.09188	0.10184	0.00000	0.00000	0.11111	0.00000	0.00000

Analysis of Molecular Variance (AMOVA) showed that 97.65% of variation was contained within populations 5.59% among populations within groups, and no significant variation was seen among groups (Table 4.4). The corresponding fixation indices were calculated as $\phi_{ST} = 0.02352$, $\phi_{SC} = 0.05412$, $\phi_{CT} = -0.03236$.

Table 4.4. Analysis of molecular variance.

Source of variation	d.f	Percentage of variation
Among groups	2	-3.24
Among populations within groups	5	5.59
Within populations	88	97.65

Phylogenetic trees, used to analyze the relationships between the haplotypes indicated two different results. The Neighbor-Joining tree showed a polytomy, as none of the haplotypes grouped differently from the others. The maximum parsimony phylogram

(Figure 4.4.) showed greater genetic structure. Haplotype BS12 from the Azov Sea, Ukraine, was seen to be more closely related to MGH02 instead of BS81 from the Azov Sea, Ukraine. BS81 was grouped with MGH03, MGH04, MGH05, MGH06, MGH07, MGH08, MGH09 and MGH01 and this split occured in 100 % of trees. However, the haplotypes MGH03, MGH04, MGH05, MGH06, MGH07, MGH08, MGH09 and MGH01 were grouped together only in 55% of trees. This result supports the idea of no significant genetic grouping among geographical regions, even at a larger geographic scale, although there is evidence of the start of some accumulation of genetic differences, as observed in the variability of the haplotype BS12.

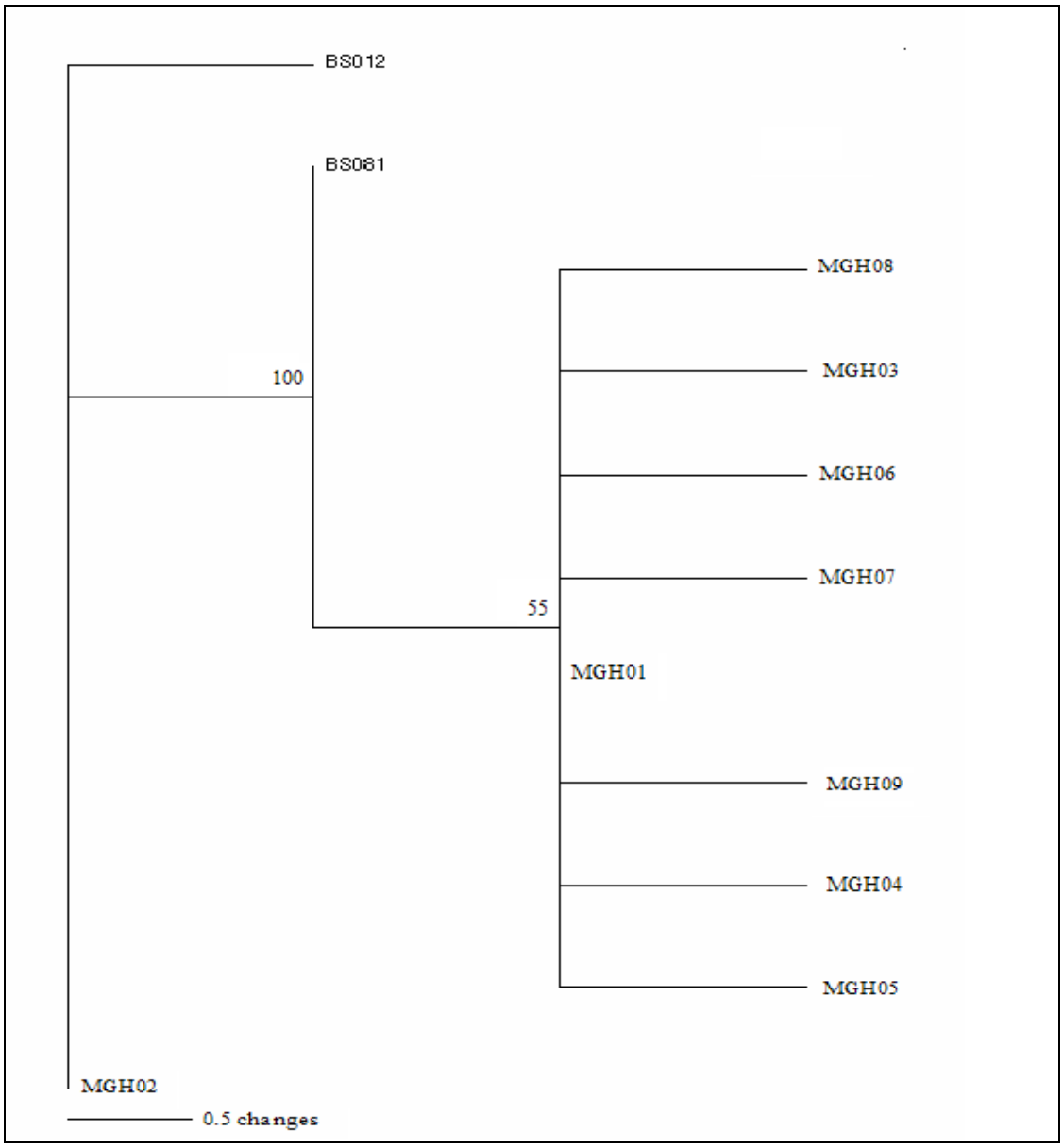

Figure 4.4. Maximum parsimony phylogram, showing respective branch lengths. Numbers refer to bootstrap values.

The haplotype network (Figure 4.5.) showed that the clade was characterized by a star genealogy centered on one geographically common haplotype, connected to less-frequent haplotypes with one mutation step. MGH01 was the most common haplotype and almost 90% of individuals had this haplotype. MGH01 was shared among the Sea of Marmara, the Bosphorus and the Black Sea. MGH03 and MGH08 haplotypes were each found in only one individual, and exclusively in the Black Sea. Haplotype MGH02 included two individuals from the Sea of Marmara and one from the Bosphorus. MGH07, MGH04, MGH09 were found only in the Bosphorus and the frequency of MGH07 was two times

higher than that of MGH04 and MGH09. The frequency of MGH05 and MGH07 were the same, they were found in two individuals only. MGH05 and MGH06 were also found only in the Sea of Marmara.

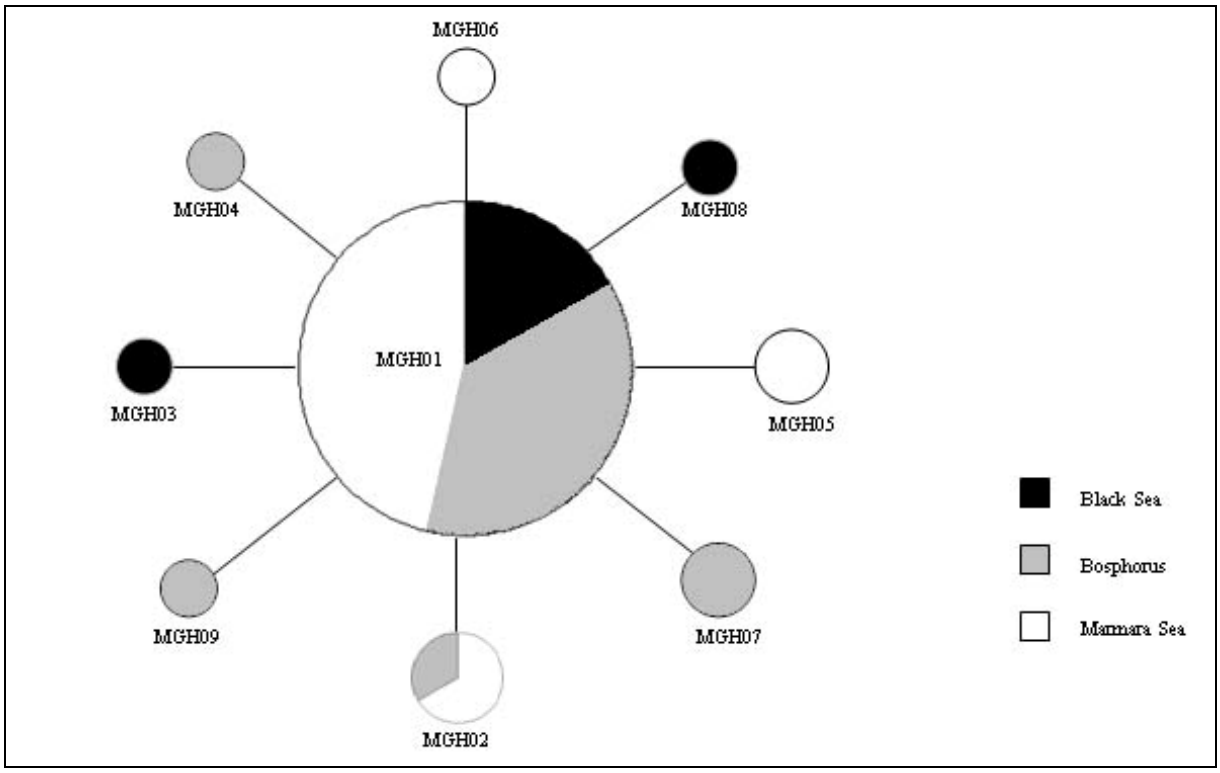

Figure 4.5. Haplotype network for the nine haplotypes in *Mytilus galloprovincialis.* Haplotypes were classified into four frequency sizes for representation. The area of the circles is proportional to the haplotype frequency. Each branch represents a single mutational event. Shading refers to the region in which haplotypes were found. In the cases where haplotypes are shared among regions, shading is proportional to the frequency of the haplotype in each region.

Although the haplotype network indicated a starlike network, indicative of a population expansion (Rogers and Harpending 1992), the population expansion statistics of R_2 and F_s did not suggest a demographic expansion in population size. This is probably due to the recency of the expansion. Considering the point that the Istanbul Strait was formed around 10.5 Kya most recently, it is likely that enough genetic differences to show signatures of statistically significant population expansion did not accumulate in this time frame. The genetic richness of populations in the region reinforce this idea, in

conformation with the theoretical (Ibrahim et al., 1996) and empirical (Hewitt, 1999) expectations from founder events where the recent migration front of species show low levels of genetic diversity.

Using the latest inference key of Templeton (2004), NCA inferred that there was restricted gene flow with isolation by distance (steps in inference key : 1, 3, 4 no, 2 yes c). However, a plot of genetic and geographical distances for all pairs of populations (Figure 4.6a-d) did not confirm this finding. The correlation between genetic and geographic distances, with various transformational combinations was low, with maximum R^2 value being equal to 0.031 (Figure 4.6a). This means that approximately three percent of the variation in the genetic distances can be explained by the geographic distances.

Within the same species, the relationship between genetic distances and geographical distances would be expected to be linear in the absence of any barriers to gene flow. As our data indicated the absence of any barriers to gene flow, we checked for the presence/absence of isolation by distance (IBD) between the sampling localities. According to NCA, there was evidence for IBD. However, plots of geographical and genetic distances indicated low levels of correlation, not supporting a scenario of IBD. A review study indicated that NCA does not always give accurate results (Knowles and Maddison, 2002), and we therefore adopted the result from the formal plot of geographic *vs.* genetic distances and rejected the presence of IBD in the studied region. Hence, genetic variation in the populations we studied cannot be explained by the IBD phenomenon. At a larger geographic scale, the haplotype BS12 is different from the rest of the samples in three base positions, and this in theory suggests the presence of IBD at a larger geographic scale. However, more sampling from the Black Sea coast, including Azov Sea, will be necessary to investigate this idea further.

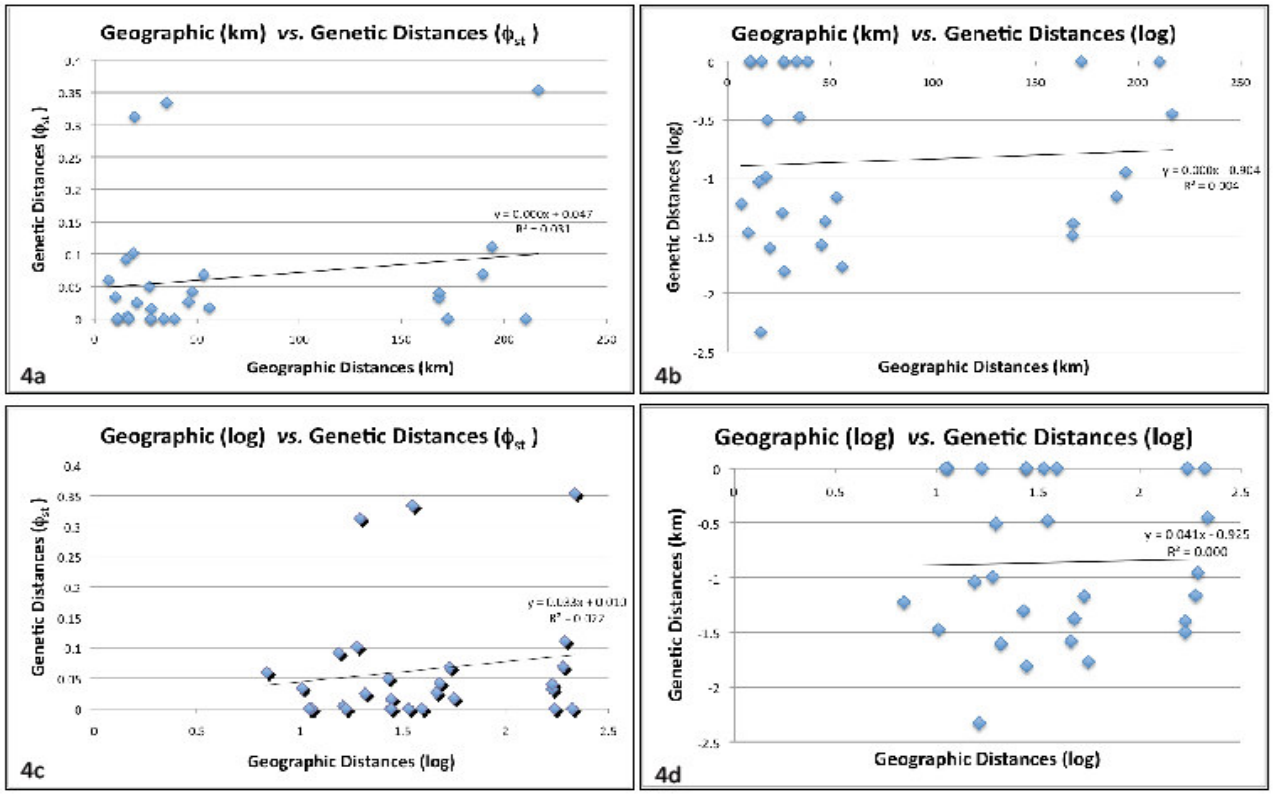

Figure 4.6. Plots of isolation by distance using a) Untransformed Genetic (ϕ_{st}) / Geographic (km) Distances b) Untransformed Geographic / Log Transformed Genetic Distances c) Log Transformed Geographic / Untransformed Genetic Distances d) Log Transformed Geographic / Log Transformed Genetic Distances.

5. CONCLUSIONS AND RECOMMENDATIONS

The results of the study suggest that the null hypothesis of the Istanbul Strait not being a barrier, but a corridor, to gene flow for the Mediterranean mussel, *Mytilus galloprovincialis*, cannot be rejected. Various lines of statistical evidence support this conclusion. The parsimony tree and the haplotype network showed that no genetic grouping exists among geographical regions. Other analytical methods, such as the average numbers of nucleotide differences between regions, the average number and the net number of nucleotide substitutions per site between populations all support this concept of no genetic grouping. The results of AMOVA and low population pairwise ϕ_{ST} values strengthened the idea that no statistically significant variation exists among the populations of the Black Sea and the Sea of Marmara. Hence it can be concluded that there is no restricted gene flow between the populations of *M. galloprovincialis* between the Black Sea, Bosphorus and the Sea of Marmara sampled in this study, and all of the mussels in the region of interest comprise a panmictic unit.

It is also a possibility that the mtDNA gene region that was investigated was not variable enough, and the genetic similarities of the regions that we investigated is due to the lack of variation, rather than extensive gene flow. However, two lines of evidence suggest that the region is informative. First, the study of Ladukakis et al. (2002) showed genetic differentiation in the CO3 region between the Black Sea and the Mediterranean populations of this species. It was not possible to make a direct comparison of these samples with those used in their study because the methods were not compatible; RFLPs were used in Ladukakis et al. (2002), whereas direct sequencing were used in this study. The only sequence information avaialable in the literature and submitted to Genbank was from the study of Venitis *et al.* (2006). These included the haplotypes BS12 and BS081 from the Azov Sea. BS12 and BS081 were different from our samples in that they had the base timine in the 12th site position, whereas the samples analyzed in this study had adenine in the same position. Also, in the samples that sequenced, although one haplotype (MG001) was dominant, there were still eight other variable positions. Hence, the mtDNA region that was looked at, CO3, captures the differences between the populations in the

Azov Sea and the vicinity of Bosphorus, is variable in the Bosphoric area, and the similarity due to lack of variability is not an issue.

In conclusion, the main finding of this study was that Istanbul Strait is a corridor to gene flow for the Mediterranean mussel, *Mytilus galloprovincialis*. Various statistical analyses showed that there were no genetic differences between the populations of the Black Sea and the Sea of Marmara. This result has important implications for further studies on gene expression related to salinity in the region. As there were no genetic differences in areas whose salinity differed between 17.9 - 23.9 PSU, any differences in gene expression found between the populations in these regions would not be due to evolutionary genetic differences. Hence the area comprises an excellent natural laboratory for studies of gene expression for *Mytilus galloprovincialis*. Future research should be carried out using microsatellites, and with a more extensive geographical coverage to better evaluate the phylogeography of *Mytilus galloprovincialis* and to understand the effect of the Dardanelles Strait and the Sea of Marmara to the gene flow between the Black Sea and the Aegean Sea.

REFERENCES

Aksu, A.E., Hiscott, R.N., Kaminski, M.A., Mudie, P.J., Gillespie, H., Abrajano, T., Yasar, D., 2002. Last glacial Holocene paleoceanography of the Black Sea and Marmara Sea: stable isotopic, foraminiferal and coccolith evidence. Marine Geology, 190, 119-149

Alexeyev, M. F., LeDoux, S. P., Wilson, G. L., 2004. Mitochondrial DNA and aging. Clinical Science London, 107, 355–364.

Avise, J. C., 1994. Molecular Markers, Natural History and Evolution, Chapman and Hall, New York.

Avise, J. C., 2000. Phylogeography: The History and Formation of Species, Harvard University Press, Cambridge, Massachusetts.

Bacescu, M.C., Müler, G.I., Gomoiu, M.T., 1971. *Ecologie Marina* vol. IV, Romanian Academy Press, Bucharest 357.

Beerli, P., Felsenstein, J., 1999. Maximum likelihood estimation of migration rates and population numbers of two populations using a coalescent approach. Genetics, 152, 763–773.

Besiktepe, S., Sur, H. I., Ozsoy, E., Latif, M. A., Oguz, T., and U. Unluata, 1994. The Circulation and Hydrography of the Marmara Sea. Progress In Oceanography, 34, 285-334.

Branch, G. M., Steffani, C. N., 2004. Can we predict the effects of alien species? A case-history of the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck), Journal of Experimental Marine Biology and Ecology, 200, 189–215.

Brown, W. M., George, M., Wilson, A. C., 1979. Rapid evolution of animal mtDNA. Proceedings of the National Academy of Science, USA, 76, 1967-1971.
Clement, M., Posada, D., Crandall, K. A., 2000. TCS**:** a computer program to estimate gene genealogies. Molecular Ecology, 9, 1657–1659.

Daguin, C., Borsa, P., 2000. Genetic relationships of *Mytilus galloprovincialis* Lmk. Populations worldwide: evidence from nuclear-DNA markers, Geological Society of London Special Publications, 177, 389–397.

Excoffier, L., Smouse P. E., Quattro, J. M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data, Genetics 131, 479–491.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin version 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1, 47–50.

Fu, Y. X., Li, W. H., 1993. Statistical tests of neutrality of mutations. Genetics, 133 (3), 693–709.

Hewitt G. M. 1999. Postglacial recolonization of European biota. Biological Journal of the Linnean Society. 68: 87-112.

Hilbish, T.J., Mullinax, A., Dolven, S.I., Meyer, A., Koehn, R.K., Rawson, P.D., 2000. Oriogin of the anti-tropical distribution pattern in marine mussels (*Mytilus* spp.): routes and timing of transequatorial migration, Marine Biology, 136., 69–77.

Ibrahim K. M., Nichols J. D. and Hewitt G. M. (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity. 77: 282-291.

Irwin, D. M., Kocher, T. D. and Wilson , A. C., 1991. Evolution of the cytochrome *b* gene of mammals. Journal of Molecular Evolution, 32, 128-144.

Johnson, N. K., Cicero, C., 2004. New mitochondrial DNA date affirm the importance of Pleistocene speciation in North American birds. Evolution, 58, 1122–1130.

Jukes, T. H., Cantor, C. R., 1969. Evolution of protein molecules. In: H.N. Munro, Editor, Mammalian Protein Metabolism 3**,** 21–132.

Knowles, L.L. and Maddison, W.P. 2002. Statistical phylogeography. Molecular Ecology. 11, 2623–2635.

Kocatas, A., Koray, T., Kaya M., and Kara, O.F, 1993. Review of the Fishery Resources and their Environment in the Sea of Marmara, Studies and Reviews, General Fisheries Council for the Mediterranean (GFCM), FAO, 64, 87–143.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X., Wilson, A. C.,1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceeding Of the National Academy of Sciences USA, 86, 6196–6200.

Kuhner, M. K., Yamato, J., Felsenstein, J., 1998. Maximum likelihood estimation of population growth rates based on the coalescent. Genetics, 149, 429–434.

Ladoukakis, M., Saavedra C., Magoulas A., and Zouros, E., 2002. Mitochondrial DNA variation in a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. Molecular Ecology, 11, 755-769.

Lynch, M., Crease, T. J.,1990. The analysis of population survey data on DNA sequence variation. Molecular Biology and Evolution, 7, 377–394.

Magoulas, A., Castilho, R., Caetano, S., Marcato, S., Patarnello, T., 2006. Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). Molecular Phylogenetics and Evolution, 39, 734-746.

Martínez-Lage, A., Rodríguez, F.,González-Tizón, Prats, E.,Cornudella, L., Méndez, J., 2002. Comparative analysis of different satellite DNAs in four *Mytilus* species. Genome, 45 ,922–929.

McDonald, J. H., Seed, R., Koehn, R. K., 1991. Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern Hemispheres, Marine Biology, 111, 323–333.

Mullis, K. B., Faloona, F. A., 1987. Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. Methods in Enzymology, 155, 335–350.

Nalbantoglu, U., 1957. Contents of the stomachs of mackerel (*Scomber scomber*) caught in the Marmara, the Bosphorus and in the areas of the Black Sea bordering the Bosphorus. General Fisheries Council for the Mediterranean (GFCM), FAO, Rome, 4, 181-192.

Narang, S. K., Klein, T. A., Perera, O.P., Lima, J. B., Tang, A. T., 1993. Genetic evidence for the existence of cryptic species in the Anopheles albitarsis complex in Brazil: allozymes and mitochondrial DNA restriction fragment length polymorphisms. Biochemical Genetetics, 31, 97-112.

Olsen, J. L., Stam, W.T., Coyer, J. A., 2004. Population differentiation and post-ice age recolonization of the North Atlantic by the seagrass *Zostera marina* L. Molecular Ecology, 13, 1923-1941.

Ozturk, B., 1995. Marine mammals inventory of Turkey. Abstract submitted to the Ninth annual conference of European Cetacean Society, Lugano.

Ozturk, B., Ozturk, A.A., 1996. On the biology of the Turkish Straits System. Journal Bulletin de I'Institut. Oceanographique, 17, 205-221.

Page, R., Holmes, E. , 2000. Molecular Evolution. Blackwell Science, Oxford.

Patarnello, T., Volckaert, Filip A. M. J., Castilho, R., 2007. Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographical break? Molecular Ecology, 16, 4426-4444.

Peijnenburg, K. T. C. A., Fauvelot, C., Breeuwer, J. A. J., and Menken, S. B. J., 2006. Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. Molecular Ecology, 15, 3319-3338.

Posada, D., Crandall, K. A., 1998. MODELTEST: testing the model of DNA substitution, Bioinformatics, 14, 817–818.

Posada, D., Crandall, K. A., Templeton, A. R., 2000. GeoDis: A program for the Cladistic Nested Analysis of the Geographical Distribution of Genetic Haplotypes. Molecular Ecology, 9(4),487-488.

Rogers, A. R., Harpending, H., 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution, 49, 552–569.

Ramos-Onsins, S. E., Rozas, J., 2002. Statistical properties of new neutrality tests against population growth. Molecular Biology and Evolution, 19, 2092–2100.

Rocha-Olivares, A., Rosenblatt, R.H., Vetter, R.D., 1999. Cryptic species of rockfishes (Sebastes: Scorpaenidae) in the southern hemisphere inferred from mitochondrial lineages. Journal of Heredity, 90, 404-411.

Rousset, F., 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance, Genetics, 145, 1219–1228.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods, Bioinformatics, 19, 2496–2497.

Sarver, S.K., Foltz, D.W., 1993. Genetic population structure of a species complex of blue mussels, Marine Biology, 117. 105–112.

Swofford, D. L., 2001. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Steward, D., Kenchington, E., Singh, R., Zouros, E., 1996. Degree of selective constraint as an explanation of the different rates of evolution of gender-specific mitochondrial DNA lineages in the mussel Mytilus. Genetics, 143**,** 1349–1357.

Storz, J. F., Beaumont, M. A., 2002. Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. Evolution, 56, 154–166.

Tajima, F., 1989. The effect of change in population size on DNA polymorphism. Genetics, 123, 597–601.

Tamura, K., Nei, M., 1993. Estimation of number of nucleotide substitutions in the control region of mtDNA in humans and chimpanzees. Molecular Biology and Evolution, 10, 512-526.

Templeton, A., 1998. Nested clade analysis of phylogeographic data : testing hypotheses about gene flow and population history. Molecular Ecology, 7, 381-397.

Templeton, A., 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. Molecular Ecology, 13, 789–809.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., Higgin, D. G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 25, 4876–4882.

Venetis, C., Theologidis, I., Zouros, E., Rodakis, G. C., 2006. No evidence for presence of maternal mtDNA in the sperm of *Mytilus galloprovincialis* males. Real Sociedad, 273, 2483-2489.

Viaud-Martinez, K. A., Vergara, M. M., Gol'din, P. E., Ridoux, V., Öztürk, A. A., Öztürk, B., Rosel, P. E., Frantzis, A., Komnenou, A., Bohonak, A. J., 2007. Morphological and genetic differentiation of the Black Sea harbour porpoise *Phocoena phocoena*. Marine Ecology Progress Series, 338, 281–294.

Wilcox, T.P., Hugg, L., Zeh, J.A., Zeh, D.W., 1997. Mitochondrial DNA sequencing reveals extreme genetic differentiation in a cryptic species complex of neotropical pseudoscorpions. Molecular Phylogenetics and Evolution, 7, 208-216.

Wright, S., 1943. Isolation by distance. Genetics, 28, 114–138.

Yang, Z., Yoder, A. D., 1999. Estimation of the transition/transversion rate bias and species sampling. Journal of Molecular Evolution, 48, 274–283.

Zouros, E., Ball, A. O., Saavedra, C., Freeman, K. R., 1994. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*, Proceedings of the National Academy of Science, U. S. A, 91, 7463–7467.

APPENDIX A: OUTPUT FILE OF NESTED CLADE ANALYSIS

Differentiating population structure from history - Geodis 2.5

Contact: David Posada, University of Vigo, Spain (dposada@uvigo.es)

Input file: /Users/euryale/Documents/rasit/students/Aslõ/data/NCA/geodis.txt
mytilus galloprovincialis mtDNA

Sun Jan 11 17:14:41 EET 2009

PERMUTATION ANALYSIS OF 1-1
BASED ON 1000 RESAMPLES

PART I. PERMUTATIONAL CONTINGENCY TEST:

OBSERVED CHI-SQUARE STATISTIC = 77.5958

THE PROBABILITY OF A RANDOM CHI-SQUARE BEING GREATER THAN OR EQUAL TO THE OBSERVED CHI-SQUARE = 0.0120

PART II. GEOGRAPHIC DISTANCE ANALYSIS:

GEOGRAPHICAL CENTERS	LATITUDE	LONGITUDE
1-1	41.0406	28.8508
1	41.0274	28.8172
2	40.9933	29.0531
3	41.2517	29.0385
4	41.0879	29.0569
5	40.9272	29.0505
6	40.9775	29.0380
7	41.1730	29.0884
8	41.2517	29.0385
9	41.1730	29.0884

CLADE 1 (Interior)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	42.3429	0.9750	0.0250
NESTED CLADE	40.5511	0.9740	0.0260

CLADE 2 (Tip)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	8.7611	0.2940	0.7190
NESTED CLADE	19.9392	0.1820	0.8310

CLADE 3 (Tip)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	0.0000	1.0000	1.0000
NESTED CLADE	28.2231	0.7580	0.2890

CLADE 4 (Tip)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	0.0000	1.0000	1.0000
NESTED CLADE	18.0496	0.4190	0.7010

CLADE 5 (Tip)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	5.1301	0.3010	0.7540
NESTED CLADE	21.2580	0.3670	0.6880

CLADE 6 (Tip)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	0.0000	1.0000	1.0000
NESTED CLADE	17.1867	0.2960	0.8710

CLADE 7 (Tip)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	0.0000	0.1140	1.0000
NESTED CLADE	24.7341	0.4700	0.5440

CLADE 8 (Tip)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	0.0000	1.0000	1.0000
NESTED CLADE	28.2231	0.7430	0.3170

CLADE 9 (Tip)

TYPE OF DISTANCE	DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	0.0000	1.0000	1.0000
NESTED CLADE	24.7341	0.7230	0.4140

PART III. TEST OF INTERIOR VS. TIP CLADES:

TYPE OF DISTANCE	I-T DISTANCE	PROB.<=	PROB.>=
WITHIN CLADE	39.2976	0.9880	0.0120
NESTED CLADE	18.1996	0.9550	0.0450

APPENDIX B : THE KEY FOR NESTED CLADE ANALYSIS

1

Inference Key for the Nested Haplotype Tree Analysis of Geographical Distances

Start with haplotypes nested within a 1-step clade and work up to clades nested within the total tree. If the tree is not rooted through an outgroup or if none of the clades nested at the total tree level have the sum of the outgroup probabilities of their haplotypes greater than or equal to 0.95, regard all clades nested at the total tree level as tips. When rooting is deemed reliable, interiors should also refer to the older clades in a nesting category, and tips to their evolutionary descendants.

This key is applied only if there are some significant values for Dc, Dn, or I-T within the nesting clade. If there are no statistically significant distances within the clade, the null hypothesis of no geographical association of haplotypes cannot be rejected (either panmixia in sexual populations, extensive dispersal in non-sexual populations, small sample size, or inadequate geographical sampling). In that case, move on to another clade at the same or higher level.

1. Are all clades within the nesting clade found in separate areas with no overlap?
• NO – Go to step 2.
• YES - Go to step 19.

2. Is at least one of the following conditions satisfied?
a. The Dc's for one or more tips are significantly small and the Dc's for one or more of the interiors are significantly large or non-significant.
b. The Dc's for one or more tips are significantly small or non-significant and the Dc's for some but *not* all of the interiors are significantly small.
c. The Dc's for one or more interiors are significantly large and the Dc's for the tips are either significantly small or non-significant
d. The I-T Dc is significantly large.
• NO - Go to step 11.
• YES - Go to step 3.
• Tip/Interior Status Cannot be Determined - **Inconclusive Outcome.**

3. Is at least one of the following conditions satisfied?
a. Some Dn and/or I-T Dn values are significantly reversed from the Dc values.
b. One or more tip clades show significantly large Dn's.
c. One or more interior clades show significantly small Dn's.
d. I-T has a significantly small Dn with the corresponding Dc value non-significant.
• NO - Go to step 4.
• YES - Go to step 5.

4. Are both of the following conditions satisfied?
a. The clades (or 2 or more subsets of them) with significantly small Dc or Dn values have ranges that are completely or mostly non-overlapping with the other clades in the nested group (particularly interiors).

2

b. The pattern of completely or mostly non-overlapping ranges in the above condition represents a break or reversal from lower level trends within the nested

clade series (applicable to higher-level clades only).
• NO - **Restricted Gene Flow with Isolation by Distance (Restricted Dispersal by Distance in Non-sexual species).** This inference is strengthened if the clades with restricted distributions are found in diverse locations, if the union of their ranges roughly corresponds to the range of one or more clades (usually interiors) within the same nested group (applicable only to nesting clades with many clade members or to the highest level clades regardless of number), and if the Dc values increase and become more geographically widespread with increasing clade level within a nested series (applicable to lower level clades only).
• YES - Go to step 9.

5. Are both of the following conditions satisfied?
a. The clades (or 2 or more subsets of them) with significantly small Dc values have ranges that are completely or mostly non-overlapping with the other clades in the nested group (particularly interiors).
b. The pattern of completely or mostly non-overlapping ranges in the above condition represents a break or reversal from lower level trends within the nested clade series (applicable to higher-level clades only).
• NO - Go to step 6.
• YES - Go to step 15.

6. Do clades (or haplotypes within them) with significant reversals or significant Dn values without significant Dc values define two or more geographically concordant subsets.
• No - Go to step 7.
• YES - Go to step 13.
• TOO FEW CLADES (< 2) TO DETERMINE CONCORDANCE - **Insufficient Genetic Resolution to Discriminate between Range Expansion/Colonization and Restricted Dispersal/Gene Flow** - Proceed to step 7 to determine if the geographical sampling is sufficient to discriminate between short versus long distance movement.

7. Are the clades with significantly large Dn's (or tip clades in general when Dn for I-T is significantly small) separated from the other clades by intermediate geographical areas that were sampled?
• NO - Go to step 8.
• YES - **Restricted Gene Flow/Dispersal but with some Long Distance Dispersal.**

8. Is the species absent in the non-sampled areas?
• NO - **Sampling Design Inadequate to Discriminate between Isolation by Distance (Short Distance Movements) versus Long Distance Dispersal**
3
• YES - **Restricted Gene Flow/Dispersal but with some Long Distance Dispersal over Intermediate Areas not Occupied by the Species; or Past Gene Flow Followed by Extinction of Intermediate Populations**.

9. Are the different geographical clade ranges identified in step 4 separated by areas that have not been sampled?
• NO - **Allopatric Fragmentation.** (If inferred at a high clade level, additional

confirmation occurs if the clades displaying restricted by at least partially non- overlapping distributions are mutationally connected to one another by a larger than average number of steps.)
• YES - Go to step 10.

10. Is the species absent in the non-sampled areas?
• NO - **Geographical Sampling Scheme Inadequate to Discriminate Between Fragmentation and Isolation By Distance.**
• YES - **Allopatric Fragmentation.** (If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted by at least partially non- overlapping distributions are mutationally connected to one another by a larger
than average number of steps.)

11. Is at least one of the following conditions satisfied?
   a. The Dc value(s) for some tip clade(s) is/are significantly large.
   b. The Dc value(s) for all interior(s) is/are significantly small.
   c. The I-T Dc is significantly small.
• NO - Go to step 17
• YES - **Range Expansion**, go to step 12.

12. Are the Dn and/or I-T Dn values significantly reversed from the Dc values?
• NO - **Contiguous Range Expansion.**
• YES - Go to step 13.

13. Are the clades with significantly large Dn's (or tip clades in general when Dn for I-T is significantly small) separated from the geographical center of the other clades by intermediate geographical areas that were sampled?
• NO - Go to step 14.
• YES – 1) **Long Distance Colonization, Past Larger Range Coupled with Subsequent Extinction in Some Intermediate Geographical Areas, or Past Range Expansion, All of Which Can Possibly Be Coupled with Subsequent Fragmentation** (subsequent fragmentation is indicated if the clades displaying restricted but at least partially non-overlapping distributions are mutationally connected to one another by a larger than average number of steps) or 2) **Past Fragmentation Followed by Range Expansion**. To see if secondary contact is involved in scenario 2), perform the supplementary tests given in Templeton, Molecular Ecology **10**: 779-791, 2001. To discriminate the type of movement leading to this pattern in scenario 1), go to step 21.
4

14. Is the species present in the intermediate geographical areas that were not sampled?
• YES - **Sampling Design Inadequate to Discriminate between Contiguous Range Expansion, Long Distance Colonization, and Past Fragmentation.**
• NO - **Long Distance Colonization and/or Past Fragmentation** (not necessarily mutually exclusive)**.** If inferred at a high clade level, fragmentation rather than colonization is inferred if the clades displaying restricted but at least partially non-overlapping distributions are mutationally connected to one another by a larger than average number of steps. If the branch lengths are short, a colonization event is inferred, perhaps associated with recent fragmentation. To discriminate the type of movement leading to this pattern, go to step 21.

15. Are the different geographical clade ranges identified in step 5 separated by areas that have not been sampled?
• NO - **Past Fragmentation and/or Long Distance Colonization** (not necessarily mutually exclusive)**.** If inferred at a high clade level, fragmentation rather than colonization is inferred if the clades displaying restricted but at least partially non-overlapping distributions are mutationally connected to one another by a larger than average number of steps. If the branch lengths are short, a colonization event is inferred, perhaps associated with recent fragmentation. To discriminate the type of movement leading to this pattern, go to step 21.
• YES - Go to step 16.

16. Is the species present in the intermediate geographical areas that were not sampled?
• YES - Go to step 18.
• NO - **Allopatric Fragmentation.** If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted by at least partially non- overlapping distributions are mutationally connected to one another by a larger than average number of steps.

17. Are either of the following conditions satisfied?
a. The Dn values for tip or some (but not all) interior clades are significantly small.
b. The Dn for one or more interior clades is/are significantly large.
c. The I-T Dn value is significantly large.
• NO - **Inconclusive Outcome.**
• YES - Go to step 4.

18. Are the clades found in the different geographical locations separated by a branch length with a larger than average number of mutational steps.
• NO - **Geographical Sampling Scheme Inadequate to Discriminate Between Fragmentation, Range Expansion, and Isolation By Distance.**
• YES - **Geographical Sampling Scheme Inadequate to Discriminate Between Fragmentation and Isolation By Distance.**

19. Is the species present in the areas between the separated clades?
5
• NO – **Allopatric Fragmentation.** If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted by at least partially non- overlapping distributions are mutationally connected to one another by a larger than average number of steps.
• YES - Go to step 20.

20. Was the species sampled in the areas between the separated clades?
• NO – **Inadequate Geographical Sampling.**
• YES – Go to step 2.

21. Are all of the following true?
a. Is it biologically realistic that the organism could have undergone long-distance movement?
b. Are the nested haplotypes that mark a potential long-distance colonization event within a

clade that shows evidence of population growth by other methods (such as mismatch distributions)?
c. At the level of the entire cladogram, does the clade *not* inferred to have produced long-distance colonization *not* show evidence of past population growth with other methods?
• YES – **Long-distance movement.**
• NO – Insufficient evidence to discriminate between long-distance movements of the organism and the combined effects of gradual movement during a past range expansion and fragmentation. If the case against long-distance movement is compelling, then the inference is **past gradual range expansion followed by fragmentation or a past larger range followed by extinction in intermediate areas.**

APPENDIX C : THE BASE SEQUENCES OF THE NINE DIFFERENT HAPLOTYPES (MGH01-MGH09)

MGH01

GGCTATTTCGGCTAATGGAATAGCGGTAGGGTTAATTTTGTGACTGCATCGAA
CCCCCAGATTTGTATTAATAGGTATGAGTCTGGTTTGCATATTATTGAGAACTT
TTAGATGGTGACGCGATTTAATTCGTGAAGGAGACATTGGGCTTCACACTCGTT
TTGTGATCAAAAGATTTCGAGATGGCGTTGCCTTGTTTATTCTGTCTGAAGTGA
TATTTTTCTTCACTTTTTTCTGGACTTTTTTCCATAATGCTCTAAGGCCTTCGTGT
GAGTTAGGAATACGGTGGCCTCCTCCTGGAATTCGTACACCAAACCCGTCATC
TACTAGTCTGTTTGAGACAGGTCTTCTAATTAGAAGAGGGTTGTTTGTAACTCA
AGCCCATAAGAGGATGCGCTTGAAGGATTACGATGTAGGGCCATTCATCGGTT
TAGTGGTGACAATCGTATGCGGGACCGTGTTTTTTTTGGTACAACTGCGGGAAT
ATTATTGAAACTCCTACACTATTGCAGATAGGGTTTATGGTAGGGTTTTCTATT
TACTAACTGGATTCCATGGGATACATGTTGTCGTAGGGACTATTTGGCTAACGG
TAAGGTTAGTTCGACTATGACGCGGGGAGTTTTCTAGTCAACGACACTTTGGGT
TTGAGGCTTGTATTTGGTACTGACATTTTGTAGATGTGGTATGAGTGGCATTGT
GGTGCTTAGTGTAC

MGH02

TAATGGAATAGCGGTAGGGTTAATTTTGTGACTTCATCGAACCCCCAGATTTGT
ATTAATAGGTATGAGTCTGGTTTGCATATTATTGAGAACTTTTAGATGGTGACG
CGATTTAATTCGTGAAGGAGACATTGGGCTTCACACTCGTTTTGTGATCAAAAG
ATTTCGAGATGGCGTTGCCTTGTTTATTCTGTCTGAAGTGATATTTTTCTTCACT
TTTTTCTGGACTTTTTTCCATAATGCTCTAAGGCCTTCGTGTGAGTTAGGAATAC
GGTGGCCTCCTCCTGGAATTCGTACACCAAACCCGTCATCTACTAGTCTGTTTG
AGACAGGTCTTCTAATTAGAAGAGGGTTGTTTGTAACTCAAGCCCATAAGAGG
ATGCGCTTGAAGGATTACGATGTAGGGCCATTCATCGGTTTAGTGGTGACAAT
CGTATGCGGGACCGTGTTTTTTTTGGTACAACTGCGGGAATATTATTGAAACTC
CTACACTATTGCAGATAGGGTTTATGGTAGGGTTTTCTATTTACTAACTGGATT
CCATGGGATACATGTTGTCGTAGGGACTATTTGGCTAACGGTAAGGTTAGTTC
GACTATGACGCGGGGAGTTTTCTAGTCAACGACACTTTGGGTTTGAGGCTTGTA

TTTGGTACTGACATTTTGTAGATGTGGTATGAGTGGCATTGTGGTGCTTAGTGT
AC

MGH03

TTTTTGTGGCTATTTCGGCTAATGGAATAGCGGTAGGGTTAATTTTGTGACTGC
ATCGAACCCCCAGATTTGTATTAATAGGTATAAGTCTGGTTTGCATATTATTGA
GAACTTTTAGATGGTGACGCGATTTAATTCGTGAAGGAGACATTGGGCTTCAC
ACTCGTTTTGTGATCAAAAGATTTCGAGATGGCGTTGCCTTGTTTATTCTGTCT
GAAGTGATATTTTTCTTCACTTTTTTCTGGACTTTTTTCCATAATGCTCTAAGGC
CTTCGTGTGAGTTAGGAATACGGTGGCCTCCTCCTGGAATTCGTACACCAAACC
CGTCATCTACTAGTCTGTTTGAGACAGGTCTTCTAATTAGAAGAGGGTTGTTTG
TAACTCAAGCCCATAAGAGGATGCGCTTGAAGGATTACGATGTAGGGCCATTC
ATCGGTTTAGTGGTGACAATCGTATGCGGGACCGTGTTTTTTTTGGTACAACTG
CGGGAATATTATTGAAACTCCTACACTATTGCAGATAGGGTTTATGGTAGGGTT
TTCTATTTACTAACTGGATTCCATGGGATACATGTTGTCGTAGGGACTATTTGG
CTAACGGTAAGGTTAGTTCGACTATGACGCGGGGAGTTTTCTAGTCAACGACA
CTTTGGGTTTGAGGCTTGTATTTGGTACTGACATTTTGTAGATGTGGTATGAGT
GGCATTGTGGTGCTTAGTGTAC

MGH04

TATTTCGGCTAATGGAATAGCGGTAGGGTTAATTTTGTGACTGCATCGAACCCC
CAGATTTGTATTAATAGGTATGAGTTTGGTTTGCATATTATTGAGAACTTTTAG
ATGGTGACGCGATTTAATTCGTGAAGGAGACATTGGGCTTCACACTCGTTTTGT
GATCAAAAGATTTCGAGATGGCGTTGCCTTGTTTATTCTGTCTGAAGTGATATT
TTTCTTCACTTTTTTCTGGACTTTTTTCCATAATGCTCTAAGGCCTTCGTGTGAG
TTAGGAATACGGTGGCCTCCTCCTGGAATTCGTACACCAAACCCGTCATCTACT
AGTCTGTTTGAGACAGGTCTTCTAATTAGAAGAGGGTTGTTTGTAACTCAAGCC
CATAAGAGGATGCGCTTGAAGGATTACGATGTAGGGCCATTCATCGGTTTAGT
GGTGACAATCGTATGCGGGACCGTGTTTTTTTTGGTACAACTGCGGGAATATTA
TTGAAACTCCTACACTATTGCAGATAGGGTTTATGGTAGGGTTTTCTATTTACT
AACTGGATTCCATGGGATACATGTTGTCGTAGGGACTATTTGGCTAACGGTAA
GGTTAGTTCGACTATGACGCGGGGAGTTTTCTAGTCAACGACACTTTGGGTTTG
AGGCTTGTATTTGGTACTGACATTTTGTAGATGTGGTATGAGTGGCATTGTGGT

GCTTAGTGTACGTGTGGTTTGGAGGATGATTATATATGTGATGGTTTAAGATAT
GGGACGGGGATGTTTATACATTTAAGTATCCGGA

MGH05

TTCGGCTAATGGAATAGCGGTAGGGTTAATTTTGTGACTGCATCGAACCCCCA
GATTTGTATTAATAGGTATGAGTCTGGTTTGCATACTATTGAGAACTTTTAGAT
GGTGACGCGATTTAATTCGTGAAGGAGACATTGGGCTTCACACTCGTTTTGTGA
TCAAAAGATTTCGAGATGGCGTTGCCTTGTTTATTCTGTCTGAAGTGATATTTT
TCTTCACTTTTTTCTGGACTTTTTTCCATAATGCTCTAAGGCCTTCGTGTGAGTT
AGGAATACGGTGGCCTCCTCCTGGAATTCGTACACCAAACCCGTCATCTACTA
GTCTGTTTGAGACAGGTCTTCTAATTAGAAGAGGGTTGTTTGTAACTCAAGCCC
ATAAGAGGATGCGCTTGAAGGATTACGATGTAGGGCCATTCATCGGTTTAGTG
GTGACAATCGTATGCGGGACCGTGTTTTTTTTGGTACAACTGCGGGAATATTAT
TGAAACTCCTACACTATTGCAGATAGGGTTTATGGTAGGGTTTTCTATTTACTA
ACTGGATTCCATGGGATACATGTTGTCGTAGGGACTATTTGGCTAACGGTAAG
GTTAGTTCGACTATGACGCGGGGAGTTTTCTAGTCAACGACACTTTGGGTTTGA
GGCTTGTATTTGGTACTGACATTTTGTAGATGTGGTATGAGTGGCATTGTGGTG
CTTAGTGTAC

MGH06

GGCTATTTCGGCTAATGGAATAGCGGTAGGGTTAATTTTGTGACTGCATCGAA
CCCCCAGATTTGTATTAATAGGTATGAGTCTGGTTTGCATATTATTGAGAACTT
TTAGATGGTGACGGGATTTAATTCGTGAAGGAGACATTGGGCTTCACACTCGT
TTTGTGATCAAAAGATTTCGAGATGGCGTTGCCTTGTTTATTCTGTCTGAAGTG
ATATTTTTCTTCACTTTTTTCTGGACTTTTTTCCATAATGCTCTAAGGCCTTCGT
GTGAGTTAGGAATACGGTGGCCTCCTCCTGGAATTCGTACACCAAACCCGTCA
TCTACTAGTCTGTTTGAGACAGGTCTTCTAATTAGAAGAGGGTTGTTTGTAACT
CAAGCCCATAAGAGGATGCGCTTGAAGGATTACGATGTAGGGCCATTCATCGG
TTTAGTGGTGACAATCGTATGCGGGACCGTGTTTTTTTTGGTACAACTGCGGGA
ATATTATTGAAACTCCTACACTATTGCAGATAGGGTTTATGGTAGGGTTTTCTA
TTTACTAACTGGATTCCATGGGATACATGTTGTCGTAGGGACTATTTGGCTAAC
GGTAAGGTTAGTTCGACTATGACGCGGGGAGTTTTCTAGTCAACGACACTTTG

GGTTTGAGGCTTGTATTTGGTACTGACATTTTGTAGATGTGGTATGAGTGGCAT
TGTGGTGCTTAGTGTAC

MGH07

TGGCTATTTCGGCTAATGGAATAGCGGTAGGGTTAATTTTGTGACTGCATCGAA
CCCCCAGATTTGTATTAATAGGTATGAGTCTGGTTTGCATATTATTGAGAACTT
TTAGATGGTGACGCGATTTAATTCGTGAAGGAGACATTGGGCTTCACACTCGTT
TTGTGATCAAAAGATTTCGAGATGGCGTTGCCTTGTTTATTCTGTCTGAAGTGA
TATTTTTCTTCACTTTTTTCTGGACTTTTTTCCATAATGCTCTAAGGCCTTCGTGT
GAGTTAGGAATACGGTGGCCTCCTCCTGGAATTCGTACACCAAACCCGTCATC
TACTAGTCTGTTTGAGACAGGTCTCCTAATTAGAAGAGGGTTGTTTGTAACTCA
AGCCCATAAGAGGATGCGCTTGAAGGATTACGATGTAGGGCCATTCATCGGTT
TAGTGGTGACAATCGTATGCGGGACCGTGTTTTTTTTGGTACAACTGCGGGAAT
ATTATTGAAACTCCTACACTATTGCAGATAGGGTTTATGGTAGGGTTTTCTATT
TACTAACTGGATTCCATGGGATACATGTTGTCGTAGGGACTATTTGGCTAACGG
TAAGGTTAGTTCGACTATGACGCGGGGAGTTTTCTAGTCAACGACACTTTGGGT
TTGAGGCTTGTATTTGGTACTGACATTTTGTAGATGTGGTATGAGTGGCATTGT
GGTGCTTAGTGTAC

MGH08

AGGGTTAATTTTGTGACTGCATCGAACCCCCAGATTTGTATTAATAGGTATGAG
TCTGGTTTGCATATTATTGAGAACTTTTAGATGGTGACGCGATTTAATTCGTGA
AGGAGACATTGGGCTTCACACTCGTTTTGTGATCAAAAGATTTCGAGATGGCG
TTGCCTTGTTTATTCTGTCTGAAGTGATATTTTTCTTCACTTTTTTCTGGACTTTT
TTCCATAATGCTCTAAGGCCTTCGTGTGAGTTAGGAATACGGTGGCCTCCTCCT
GGAATTCGTACACCAAACCCGTCATCTACTAGTCTGTTTGAGACAGGTCTTCTA
ATTAGAAGAGGGTTGTTTGTAACTCAAGCTCATAAGAGGATGCGCTTGAAGGA
TTACGATGTAGGGCCATTCATCGGTTTAGTGGTGACAATCGTATGCGGGACCG
TGTTTTTTTTGGTACAACTGCGGGAATATTATTGAAACTCCTACACTATTGCAG
ATAGGGTTTATGGTAGGGTTTTCTATTTACTAACTGGATTCCATGGGATACATG
TTGTCGTAGGGACTATTTGGCTAACGGTAAGGTTAGTTCGACTATGACGCGGG

GAGTTTTCTAGTCAACGACACTTTGGGTTTGAGGCTTGTATTTGGTACTGACAT
TTTGTAGATGTGG

MGH09

TGGCTATTTCGGCTAATGGAATAGCGGTAGGGTTAATTTTGTGACTGCATCGAA
CCCCCAGATTTGTATTAATAGGTATGAGTCTGGTTTGCATATTATTGAGAACTT
TTAGATGGTGACGCGATTTAATTCGTGAAGGAGACATTGGGCTTCACACTCGTT
TTGTGATCAAAAGATTTCGAGATGGCGTTGCCTTGTTTATTCTGTCTGAAGTGA
TATTTTTCTTCACTTTTTTCTGGACTTTTTTCCATAATGCTCTAAGGCCTTCGTGT
GAGTTAGGAATACGGTGGCCTCCTCCTGGAATTCGTACACCAAACCCGTCATC
TACTAGTCTGTTTGAGACAGGTCTTCTAATTAGAAGAGGGTTGTTTGTAACTCA
AGCCCATAAGAGGATGCGCTTGAAGGATTACGATGTAGGGCCATTCATCGGTT
TAGTGGTGACAATCGTATGCGGGACCGTGTTTTTTTTGGTACAATTGCGGGAAT
ATTATTGAAACTCCTACACTATTGCAGATAGGGTTTATGGTAGGGTTTTCTATT
TACTAACTGGATTCCATGGGATACATGTTGTCGTAGGGACTATTTGGCTAACGG
TAAGGTTAGTTCGACTATGACGCGGGGAGTTTTCTAGTCAACGACACTTTGGGT
TTGAGGCTTGTATTTGGTACTGACATTTTGTAGATGTGGTATGAGTGGCATTGT
GGTGCTTAGTGTAC