ABSTRACT

Title of Dissertation: THE MITOCHONDRIAL GENOME OF THE BLUE CRAB (*CALLINECTES SAPIDUS)*, AN INFORMATIVE GENETIC MARKER FOR THE EVOLUTIONARY BIOLOGY AND POPULATION GENETICS OF THE SPECIES

Xiaojun FENG Doctor of Philosophy 2009

Directed By: Professor Allen Place
Marine, Estuarine and Environmental Sciences Program

The blue crab (*Callinectes sapidus*) is a widely distributed decapod which ranges from Nova Scotia to the northern Argentina coasts. It is one of the most abundant estuarine invertebrates, supporting both commercial and recreational fisheries along the Atlantic and Gulf coasts.

This thesis presents data clearly establishing the unprecedented hyper-variability in the mitochondrial genome of *C. sapidus*. This variation extended to multiple regions, including the *cox1*, *nad2*, and *nad4* protein coding loci as well as ribosomal 12S RNA molecule. The haplotype diversity of the *nad2* gene approached 1, with a nucleotide diversity approaching 1%. This hyper-variability in the mtDNA allows using a single mtDNA gene (*nad2*) to distinguish hatchery-produced crabs from wild crabs after release to the wild.

I found no dominant mtDNA haplotypes in wild populations but instead a distribution of a few low-frequency recurrent haplotypes with a large number of singletons. Because of this high diversity and extensive population mixing, the geographic structure in wild populations exhibits panmixia from the Atlantic to Gulf of Mexico.

Some of the high genetic diversity found seems to stem from the heteroplasmic nature of the blue crab mtDNA. By cloning high fidelity PCR products, I confirmed single individual crab and megalopa harbored dozens of copies of mitochondrial haplotypes. A copy number analysis indicates discovery of unique haplotypes was probably not saturated with the possibility of inadequate sampling. The heteroplasmy in the blue crab appears to be under maternal inheritance without paternal contribution. While minor haplotypes are represented in wild populations, other minor haplotypes contained stop codons and/or non-synonymous substitutions which may influence the viability of the mitochondria.

Given the blue crab inhabits a broad variety of environments and that the mtDNA genome appears to be under selective pressure, the potential for mtDNA functional correlates with this genetic diversity maybe at the basis for the robust physiological capability of the species.

The mitochondrial genome of the blue crab (*Callinectes sapidus)*, an informative genetic marker for the evolutionary biology and population genetics of the species

By

Xiaojun FENG

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2009

Advisory Committee:
Professor Allen Place, Chair
Professor Feng Chen
Professor Patrick Gaffney
Professor Anson Hines
Professor Brian Masters
Professor Dan Terlizzi, Dean's Rep

Acknowledgements

I would like to thank all people who have helped and inspired me during my doctoral study.

I especially want to thank my advisor, Prof. Allen Place, for his guidance during my research and study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Feng Chen, Prof. Patrick Gaffney, Prof. Anson Hines, and Prof. Brian Masters, for their encouragement and insightful comments.

I would like to express my gratitude to all those who help me on sampling, without them, this thesis would not be possible. They are Eric Johnson from Maryland (SERC), Romuald Lipcius from Virginia (VIMS), Dave Eggleston from North Carolina, Jessica Reichmuth from New Jersey, Richard Darden from Mississippi, and Bruce Estrella from Massachusetts. In particular, I am indebted to Oded Zmora and his aquaculture crew at COMB/UMBI.

All my colleagues from Place Lab made it a convivial place to work. Especially I am obliged to Collin Stevens, Tsetsto Bachvaroff, and Ernest Williams for their assistance, guidance and valuable hints on my study. All other folks, Jessica Hill, Jason Adolf, Hiro Nonogaki, Andrea Allman, Kristy Lidie, Holly Bowers, Nona Kruparki, and Qiong Tang,

had inspired me in research and life through our interactions during the long hours in the lab. Thanks.

My deepest gratitude goes to my family for their unflagging love and support throughout my life.

Table of Contents

List of Tables

List of Figures

1 General introduction

1.1 Animal mitochondrial genomes

Less than 1% of the DNA found in eukaryotic cells is located in the organelle called the mitochondrion. The mitochondrial and nuclear DNA have different evolutionary origins, with the endosymbiotic origin hypothesis stating that the mitochondrial genome is a remnant of the circular genomes of Proteobacteria engulfed by the eukaryotic cells some 1.5 X 10^9 years ago with a subsequent reduction in coded genes (Gray 1992).

1.1.1 The size and content of animal mitochondrial genomes

The animal mitochondrial genome (mtDNA) is a double-stranded, circular DNA molecule with a relatively small and stable size, which is a sharp contrast to the mitochondrial DNA found in plants and fungi that can exhibit vast variability in size (i.e. up to several hundred kbps). According to a recent review paper (Gissi, Iannelli et al. 2008), 1206 metazoan species have their mitochondrial genomes fully sequenced ranging from 32 115 bp of *Placopecten magellanicus* (Mollusca Bivalvia) to 11 423 bp of *Paraspadella gotoi* (Chaetognatha), and with an average genome size of about 17 kbp. Among all the taxa, Chordata, Echinodermata, Arthropoda and Platyhelminthes are characterized by stable mtDNA genome size while Mollusca, enoplean Nematoda and Porifera show a much larger range in genome size (Figure 1-1).

The animal mitochondrial genome generally consists of 37 genes composed of 13 protein-coding genes (seven NADH dehydrogenase subunits, three cytochrome oxidase subunits, ATPase 6, ATPase 8, and cytochrome b), 2 ribosomal RNAs (rRNA; 12S and

16s), and 22 transfer RNAs (tRNAs). The total number of genes encoded in eukaryotic mitochondrial genomes ranges from 14 to 53 (Gissi, Iannelli et al. 2008) where most of the variability in the gene content is due to different numbers of tRNA genes. Although it is rare, it does happen that the animal mitochondrial genome misses some protein-coding genes, instances being ATPase 8 left out in a nematode (Okimoto, Macfarlane et al. 1992) and *nad6* in a Antarctic fish (Papetti, Lio et al. 2007). In addition to all these genes, a control region is present in the genome that contains the origin of heavy-stand replication (known as the D-loop region in vertebrates) (Billington 2003). In some mtDNAs all genes are transcribed from one strand (H or heavy strand - guanine-rich), whereas in others the gene transcription is distributed between the two strands (H and L, light strand - cytosine-rich). There are no introns present within the genes and the non-coding inter-gene fragments are short compared to coding regions of nuclear DNA.
In addition to the heterogeneity of genome content, the gene order in the genome is also variable. The positions of tRNAs change fairly frequently (Taylor, McKechnie et al. 1993; Nardi, Carapelli et al. 2001; Place, Feng et al. 2005) while other genes are translocated as large DNA blocks containing multiple genes (*nad1*, 12S, 16S and some tRNAs) as has occured in arthropods (Black and Roehrdanz 1998).

The mechanism and the biological and evolutionary significance of gene rearrangement (translocation, loss, and acquisition) is of considerable interest. The rearrangement of tRNAs are the most commonly observed alteration with many cases involving tRNAs adjacent to a coding region or more often surrounding a particular tRNA gene cluster (trnA-trnR-trnS(AGN)-trnE-trnF) that is considered to be the

replication origin of the mtDNA light strand (Boore 1999). This hotspot of rearrangement implicates genome replication as a major factor driving gene movement.

1.1.2 Homoplasmy and maternal inheritance of mtDNA

The mtDNA in most animals consists of multiple copies (2 to 10 per mitochondrion) with several hundred to thousands of mitochondria per cell. In this way the mitochondrial DNA in one animal can be considered a hierarchical population (Figure 1-2). These millions of copies of the mitochondrial genome result from the replications of a limited number of genome copies in the maternal germline.

It is generally accepted that this immense number of mitochondrial genome copies in an animal is homoplasmic. The confidence comes mostly from two facts: the uniparental inheritance and the bottleneck effect in maternal germ cells.

The uniparental inheritance pattern of the mitochondrial genome in animals is primarily based on analogies to the transmission of chloroplasts in plants and mitochondria in plants and fungi (Birky, Demko et al. 1978; Birky 1995; White, Wolff et al. 2008). The evidence for the cytoplasmic and uniparental mode of mitochondrial DNA first came from observations by yeast geneticists in the 1950s (Williamson 2002) and the direct proof for the maternal inheritance of mitochondrial DNA came about in 1972 (Dawid 1972; Dawid and Blackler 1972). Thus, the maternal inheritance of mtDNA in animals has been extensively proven and the concept is generally accepted (Egger and Wilson 1983; Rosing, Hopkins et al. 1985; Palmer 1988; Lin, Cheng et al. 1990; Rotig, Bessis et al. 1992; Lee and Taylor 1993; Lightowlers, Chinnery et al. 1997; Sligh, Levy et al. 2000; Marchington, Scott Brown et al. 2002; Taylor, McDonnell et al. 2003; Safrany,

Csongei et al. 2007; Stewart, Freyer et al. 2008). The inheritance pattern is mostly due to the prevalence of maternal mtDNA in the zygote (Sodmergen, Zhang et al. 2002). For example, the mtDNA in mammalian female gametocyte outnumbers its counterpart in male gametocyte by the order of 10^3 -10^4 (White, Wolff et al. 2008). Therefore, given that there are only a few million copies of the mitochondrial genome per cell, it would be very difficult to detect the male contribution to the mtDNA of the offspring if it even exists. The bottleneck effect of mitochondrial DNA also helps to keep the homogeneity of mtDNA in animals. During mammalian oogenesis, the immense number of germ cells at maturity originates from a limited number of progenitor germ cells, each of which contains only up to one hundred or so mitochondria. The numbers of mitochondria and germ cells increase drastically upon expansion, ending with about 200 000 mtDNA copies in mature oocytes. Furthermore, during embryogenesis, the large number of mtDNA copies is distributed among the daughter cells. Only a small fraction of all cells (the inner cell mass) will contribute to embryo's mtDNA (Shoubridge and Wai 2007; Wai, Teoli et al. 2008; White, Wolff et al. 2008) (Figure 1-3). These two aspects of mitochondrial genome transmission reduce the actual number of mtDNA copies that are passed on from generation to generation. Most low frequency mutated haplotypes accumulated in the F1 generation are lost in the bottleneck, which, as result, enhances the homoplasmy of the mtDNA copies in the next generation.

1.1.3 Doubly Uniparental Inheritance

Although maternal inheritance is the predominant form of animal mtDNA transmission, there are frequent exceptions. Doubly Uniparental Inheritance (DUI) occurs in marine mussels of the order Mytiloida, freshwater mussels of the superfamily Unionoidea, and

marine clams of the order Veneroida, which has gained a lot of attention in the past few years (Zouros, Oberhauser Ball et al. 1994; Ladoukakis, Saavedra et al. 2002; Quesada, Stuckas et al. 2003; Breton, Burger et al. 2006; Obata, Kamiya et al. 2006; Breton, Beaupre et al. 2007; Safrany, Csongei et al. 2007; Theologidis, Fodelianakis et al. 2008; Cao, Ort et al. 2009).

There are two types of distinct gender-associated mtDNAs in DUI-featured species and those two types of mtDNAs are inherited either maternally or paternally. These female-transmitted and male-transmitted mitochondrial genomes (referred to for convenience as 'F genomes' and 'M genomes', respectively) are different from each other at the nucleotide level by greater than 20%. Female mussels usually inherit their F genome only from their mother, but they transmit this F genome to either sex of offspring. Male offspring inherit their mtDNA from both parents, but they sort the mixture of mitochondrial genomes present such that they only pass on the M genome inherited from their father to the next generation. So the result is the paternal M genome will only be passed on by way of the sperm and only male organisms can keep passing it on. This is possibly caused by the limited replications of the M genome in females. Females are essentially homoplasmic for the F genome whereas males are heteroplasmic for both F and M genomes (Breton, Beaupre et al. 2007).

It seems both female and male mussels are heteroplasmic in their initial stage of development. But females shift from being heteroplasmic zygotes to essentially homoplasmic adults containing only F genome while mature males contain varying ratios of the F and M genomes in all tissues: predominantly M genome in testes while predominantly F genome in somatic tissues. Although the sperm mtDNA is very limited

compared to the oocyte mtDNA, males manage to aggregate the paternal mtDNA in a single blastomere that is to develop into the male germ line. The DUI phenomenon is only observed in bivalves to this date.

1.1.4 Heteroplasmy of mtDNA

Besides the heteroplasmy caused by Doubly Uniparental Inheritance in bivalves, other animals show mtDNA heteroplasmy as well (Kondo, Satta et al. 1990; Vilkki, Savontaus et al. 1990; Zeviani, Gellera et al. 1990; Lightowlers, Chinnery et al. 1997; Nesbo, Arab et al. 1998; Shitara, Hayashi et al. 1998; St John and Schatten 2004), although in a less common and consistent way. It is usually thought to result from paternal leakage, shown to be about 0.1% in *Drosophila* (Kondo, Satta et al. 1990), and mutation during replication. The mtDNA is known to have a higher rate of evolution than that of the single-copy genes of the nuclear genome by a factor of about 10, estimated at about 2% sequence divergence per million years in mammals (Brown, George et al. 1979; Ingman, Kaessmann et al. 2000; Mishmar, Ruiz-Pesini et al. 2003; Kivisild, Shen et al. 2006). This high mutation rate is attributable to the reactive oxygen environment produced by products of oxidative phosphorylation as well as an error-prone polymerase and limited DNA repair (Bogenhagen 1999). The high mutation rate plus the vast number of mitochondrial genome copies in an animal will certainly result in a considerable number of mtDNA variants. Whether this heteroplasmic pattern only occurs within the somatic tissues of an animal or if it can be also passed on to its offspring is not fully clarified. Given the mtDNA bottleneck, it is likely that most of the variants if not all would disappear in the offspring.

1.1.5 Use of a genetic marker in biological researches

In the classic paper *INTRASPECIFIC PHYLOGEOGRAPHY: The Mitochondrial DNA Bridge Between Population Genetics and Systematics* by the Avise group in 1987 (Avise, Arnold et al. 1987), the authors argued that the desired property for the ideal molecular marker for phylogenetic analysis should contain the following characteristics:

The molecule should:

(a) be distinctive, yet ubiquitously distributed, so that secure homologous comparsions could be made among a wide variety of organisms;

(b) be easy to isolate and assay;

(c) have a simple genetic structure lacking complicating features such as repetitive DNA, transposable elements. pseudogenes, and introns;

(d) exhibit a straightforward mode of genetic transmission, without recombination or other genetic rearrangements;

(e) provide suites of qualitative character states whose phylogenetic interrelationships could be inferred by reasonable parsimony criteria;

and, for purposes of microevolutionary analysis, (f) evolve at a rapid pace such that new character states commonly arise within the lifespan of a species.

Based on the thousands of articles published using mtDNA , this molecular marker meets all the descriptions above: mtDNA is generally maternally inherited, not subject to recombination, usually homoplasmic, of a rapid pace of evolution, and has extensive intraspecific polymorphisms (Avise, Arnold et al. 1987; Moritz, Dowling et al. 1987).

Thus, mtDNA has been extensively used as a genetic marker ever since and has distinct qualities as a genetic marker.

Compared with nuclear markers, a mtDNA marker has smaller effective population size – only a quarter of that of the bisexually inherited diploid nuclear genome if the sex ratio in populations is 1:1. That makes mtDNA more responsive to reductions in genetic variability due to bottleneck (Birky, Maruyama et al. 1983).

Based on the maternal inheritance nature of mtDNA, it is only natural that mtDNA should be a sensitive indicator of female-mediated gene flow, founder events, and other population level matrilineal processes. MtDNA markers are expected to show greater differences than nuclear markers between demes, especially where females are less mobile than males (Moritz, Dowling et al. 1987). Comparisons of the spatial distribution of nuclear and mtDNA markers can therefore provide information on sexual bias in dispersal events.

Two assumptions to justify the use of mtDNA as a good marker are that mtDNA markers are neutral and that homoplasy is negligible. The effect of slight deviations from neutrality and low degree of homoplasy is not clear. But mtDNA is a single locus marker because of the lack of recombination, thus the whole large genome as one marker should be able to supply considerably adequate neutral sites such as silent mutated sites. And as a marker consisting of fully linked sites, it is hard to believe all sites are saturated with

homoplasmy. However, it is certainly a challenge to exclude the sites affected by selection and homoplasy.

1.2 Introduction to the blue crab

The blue crab *Callinectes sapidus (Crustacea, Decapoda, Portunidae)* inhabits estuarine and nearshore environments from Nova Scotia to northern Argentina (Williams 1974). It is one of the most abundant estuarine invertebrates, supporting valuable commercial and recreational fisheries along the Atlantic and Gulf coasts. The total value of United States blue crab landings in 2002 was more than $167 million (McMillen-Jackson and Bert 2004).

A lot of work has been done upon the biology and ecology of blue crabs. Their life span is about 2-3 years. Their mating takes place in brackish areas from late spring to early fall, and then females migrate to high salinity areas such as the mouth of the Chesapeake Bay to extrude eggs in the fall, or they over-winter there and then spawn in the following spring/summer (Branco and Masunari 2000; Lipcius and Stockhausen 2002). Zoea go through 7-8 pelagic larval stages spanning a few weeks before turning into the post-larval (megalopae) stage. Megalopae are subsequently transported back to low salinity areas where they settle and grow into juveniles and then adults.

1.2.1 Population genetics of blue crabs

There are only a few published articles on the population genetics of the blue crab: (Cole and Morgan 1978), (Nelson and Hedgecock 1980), (Kordos and Burton 1993), (McMillen-Jackson, Bert et al. 1994), (McMillen-Jackson, Bert et al. 1994).

In the earliest research (Cole and Morgan 1978) protein polymorphisms were measured to detect the difference in gene frequencies in the Chesapeake Bay and Chincoteague Bay blue crab populations. Twenty five and fifty crabs were collected respectively from Patuxent and Choptank River, representing Chesapeake Bay, and fifty adult males were obtained from commercial sources captured in Chincoteague Bay (Figure 1-4). In general there was no significant gene frequency difference observed between these two Bays. There were only three enzymes that were able to produce reliably scorable zymograms out of 11 polymorphic enzymes examined, which might be a limit of their method. They argued that the observed genetic similarity among Chincoteague and Chesapeake Bay blue crabs probably resulted from larval intermixing by the long pelagic development period of *C. sapidus* and currents at the mouths of Chesapeake and Delaware bays.

Nelson and Hedgecock (Nelson and Hedgecock 1980) conducted a very broad study including 44 species of decapod crustaceans over 26 allozyme loci with an average of 24 individuals per species. The overall heterozygosity of the blue crab was 8.8%, ranked 8th among 44 species. The goal of their work was to study the correlation among those loci and between the enzyme loci and environmental variables. Based on results of 10 most repeatable enzymes they concluded the heterozygosities of central metabolic enzymes

were positively correlated with “environmental descriptors” such as euryhalinity and somewhat negatively correlated with the “organismic descriptors” such as size.

Kordos and Burton (Kordos and Burton 1993) surveyed the population structure of blue crabs along the Texas coast of the Gulf of Mexico. Their work used three polymorphic, enzyme-encoding gene loci. Significant temporal and spatial variation was observed at all three loci. They observed megalopal allelic frequencies differing significantly from those of neighboring adult populations; larval allelic frequencies appeared to vary seasonally, stronger in the summer months but less variable in winter; allelic frequencies among adult populations were significantly heterogeneous, but only one locus showed significant temporal variation; juvenile and adult crabs sampled within one bay showed no size-specific differences in allelic frequencies. The spatial heterogeneity in allelic frequencies suggests that interpopulation gene flow is not sufficient to overcome population differentiation resulting from genetic drift and/or natural selection. Temporal variation in larval allelic frequencies suggests seasonal changes in larval source populations, which may result from population differences in spawning seasons or developmental times, or from seasonal changes in coastal current patterns.

McMillen-Jackson team (McMillen-Jackson, Bert et al. 1994) performed electrophoretic allozyme analysis on 750 individuals collected from 16 nearshore locations ranging from New York to Texas, USA. Twenty enzymes and non-enzymatic proteins coded by 31 presumptive loci were examined. Twenty-two loci were either monomorphic or polymorphic at less than the $P<0.05$ level; alleles for these polymorphic loci were

geographically dispersed. Three of the remaining polymorphic loci were homogeneous over all populations. Phenograms generated by the UPGMA (unweighted pair-group method using arithmetic averages) and distance Wagner methods exhibited no geographic pattern in the clustering of populations. Estimates of the effective number of migrants per generation between populations indicated substantial gene flow with values sufficiently high to infer panmixia between all blue crab populations from New York to Texas.

However, despite this high level of gene flow, they did observe genetic patchiness in 5 loci and a clinal variation. Genetic patchiness in blue crabs is likely to be the result of the pre-settlement formation and subsequent settlement of genetically heterogeneous patches of larvae. In the Atlantic Ocean, a regional latitudinal cline of EST-2 allele frequencies was superimposed on the range-wide genetic patchiness exhibited by that locus. This pattern against a background of high gene flow is highly likely to be maintained by selection. The Gulf of Mexico showed no apparent cline, perhaps due to long-distance migration of females in some regions of the Gulf, or to masking by genetic patchiness. These results emphasize the importance of both ecological and evolutionary time scales and structuring mechanisms in determining genetic population structure. In general, despite low allozyme variability, moderate population genetic structuring was observed: a cline in allele frequencies at an esterase locus was evident along the Atlantic coast, range-wide genetic patchiness of allele frequencies occurred at several loci, and high regional gene flow was superimposed over lower long-distance gene flow.

Following the allozyme-based works, McMillen-Jackson and Bert (McMillen-Jackson and Bert 2004) conducted a study on the DNA level. They genotyped 176 blue crabs which were collected from New York to Mexico employing RFLP analysis on the mtDNA genome. Five restriction enzymes (*Sac*I, *Hind*III, *Hae*III, *Msp*I, and *Eco*RV) were used. Surprisingly, this big spatial scale survey on blue crab genetics did not achieve much resolution of populations. Although high genetic diversity and nuclear diversity were observed, the assumed Atlantic/Mexico split was not supported by either nested analysis or haplotype cluster analysis. However, those 176 individuals were from 14 sites, so the sample sizes from a few sites were 6 or lower. The majority of sample sizes however range from 8 to 13. This might have weakened the power of that analysis. A cut map of these five restriction enzymes on the whole mtDNA genome is shown as Figure 1-5, indicating that the variable nucleic acid sites were distributed throughout the whole genome.

1.2.2 The blue crab ‘s mitochondrial genome

The mitochondrial genome of the blue crab has been completely sequenced and annotated (Place, Feng et al. 2005) (Figure 1-6). It is a typical animal mitochondrial genome existing as a circular, double-stranded molecule. It is 16,263-bp in length, A+T-rich (69.1%), encoding 37 genes (for 13 proteins, 22 tRNAs, and two rRNAs) plus a large (1434 bp), hypervariable putative control region that is 78.2% A+T. Gene order and arrangement is similar to other arthropods (e.g. Artemia) but dramatically different from the hermit crab, which has a unique gene order among arthropods. As in the mtDNA of the swimming crab, *Portunus trituberculatus*, *trnH* is located between the *trnE* and *trnF*

genes, rather than at its ancestral position upstream of *nad5* (Place, Feng et al. 2005). Figure 1-7 shows the phylogenetic tree generated by Maximum Parsimony (MP) analysis of the amino acid sequences from the concatenated 13 protein coding genes. This tree was the same as what NJ analysis produced and with similar bootstrap values.

1.2.3 The Blue Crab Advanced Research Consortium

This work is a part of the Blue Crab Advanced Research Consortium (BCARC). BCARC is led by the Center of Marine Biotechnology (COMB), cooperating with Smithsonian Environmental Research Center (SERC) in Edgewater, Maryland, the University of Southern Mississippi, the Virginia Institute of Marine Sciences (VIMS) and North Carolina State University, with initial funding from the State of Maryland and from industry (Phillips Food Inc.), and with the additional federal funding provided through the Chesapeake Bay Office of NOAA. One goal of BCARC is to develop reliable hatchery technologies to produce large numbers of blue crab juveniles and then, to assess the potential to enhance the abundance and harvests of blue crabs in the Chesapeake Bay through releases of hatchery-produced juveniles.

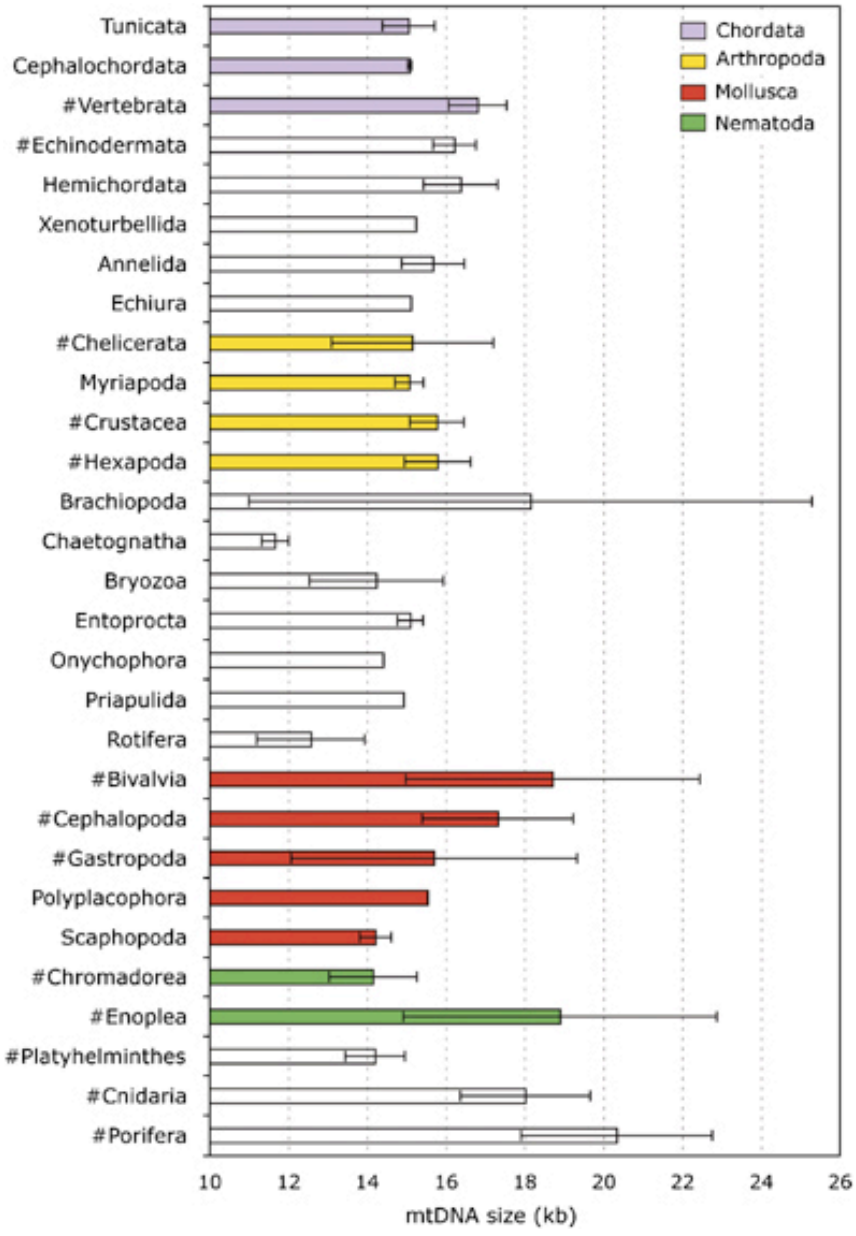

Figure 1-1. The mtDNA genome sizes of metazoan (Gissi, Iannelli et al. 2008).

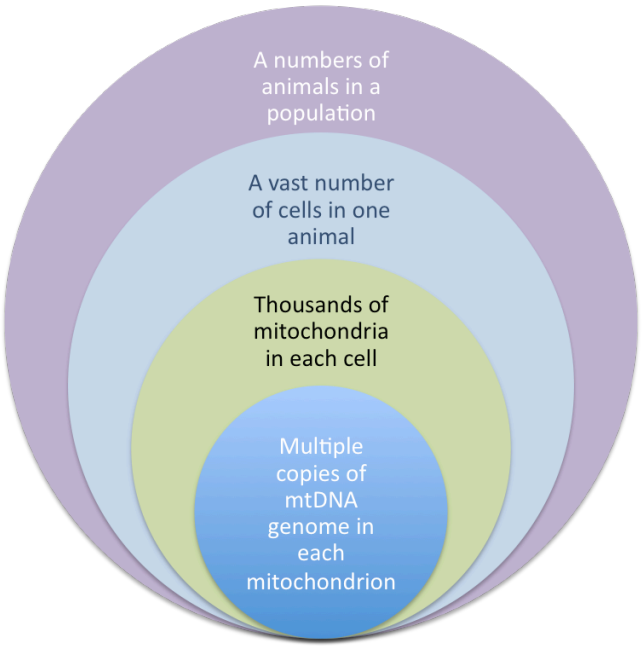

Figure 1-2. The mtDNA hierarchical population in animals.

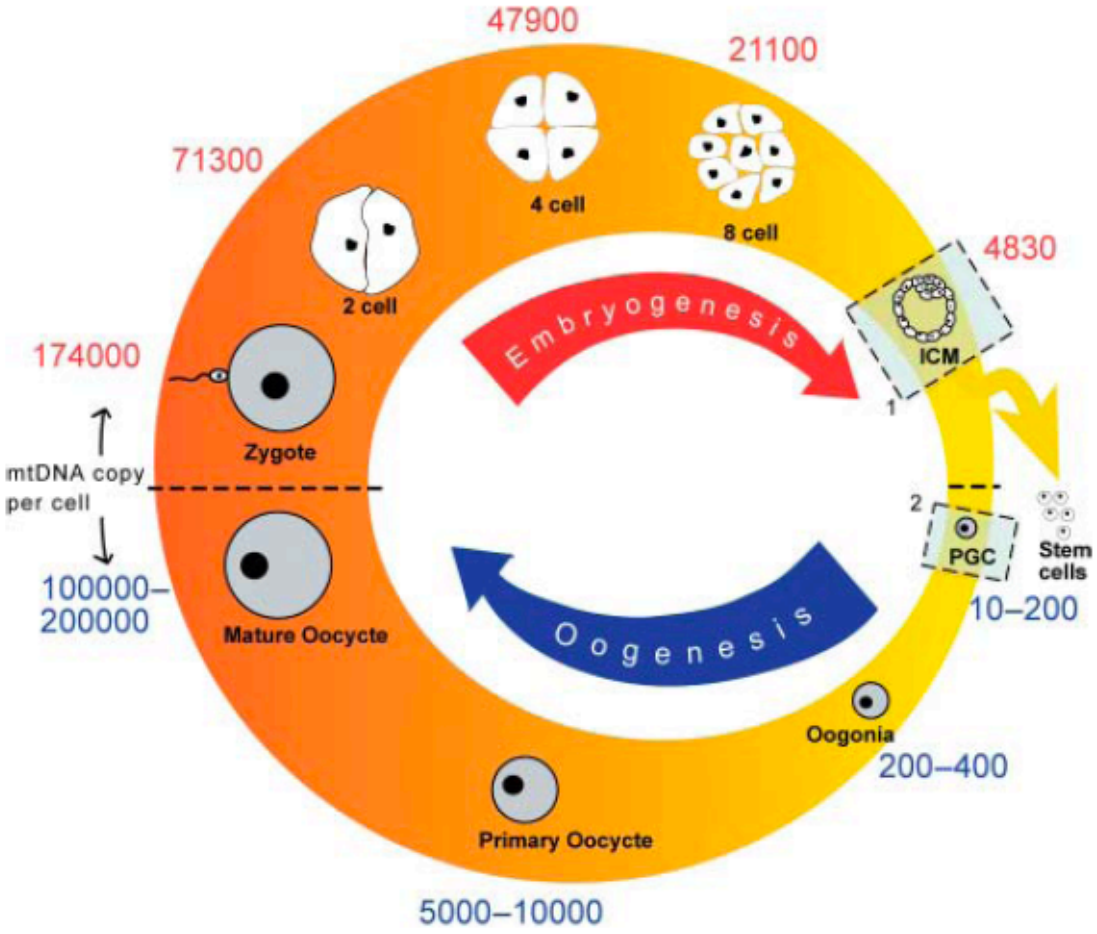

Figure 1-3. The mitochondrial bottleneck. 1 and 2 represent the stages that have the strongest effect on the haploytpe segregation (White, Wolff et al. 2008). The numbers are the estimates of the mtDNA copy numbers per cell during different developmental stages. ICM: inner cell mass; PGC: progenitor germ cells.

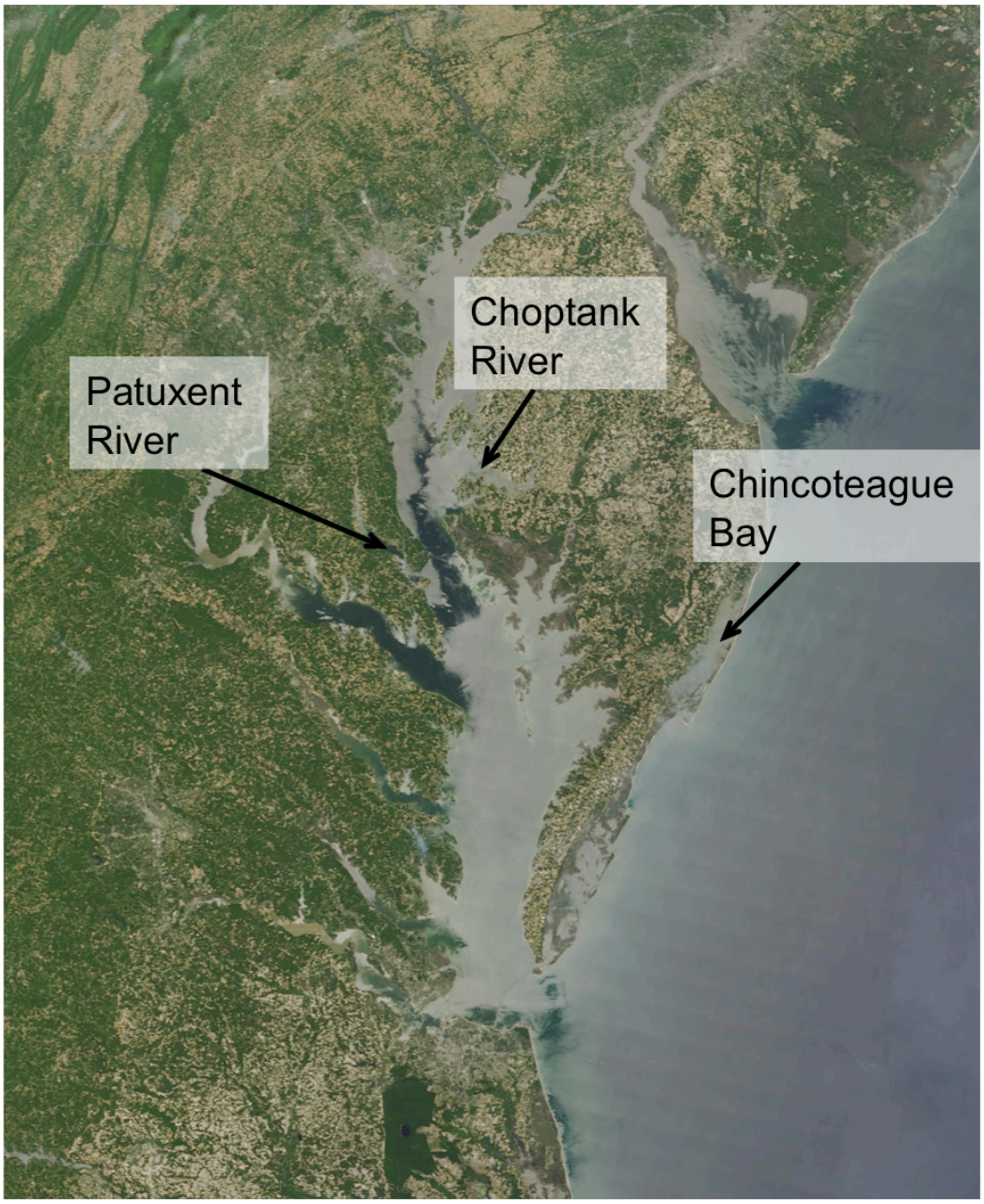

Figure 1-4. The sampling sites in Cole group's work (Cole and Morgan 1978).

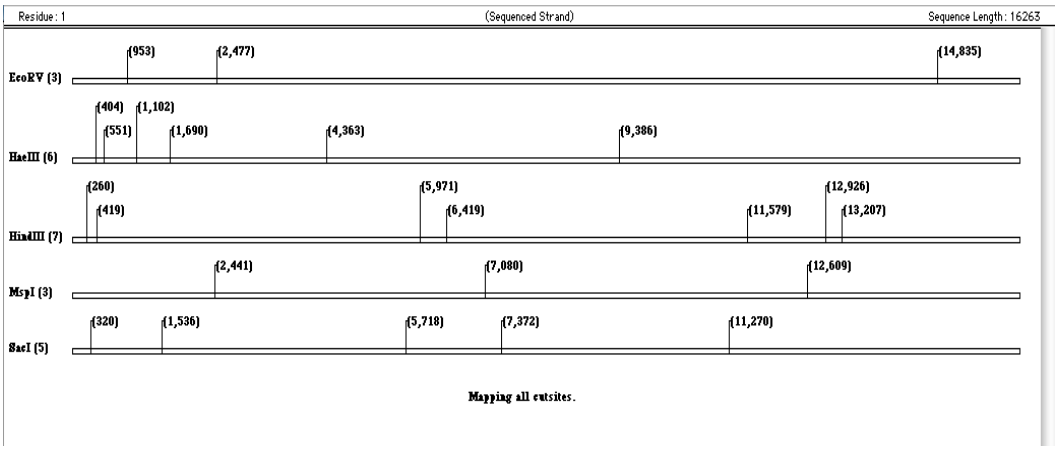

Figure 1-5. The restriction enzyme cut map of the 5 enzymes (*Sac*I, *Hind*III, *Hae*III, *Msp*I, and *Eco*RV) on the whole mtDNA genome of the blue crab.

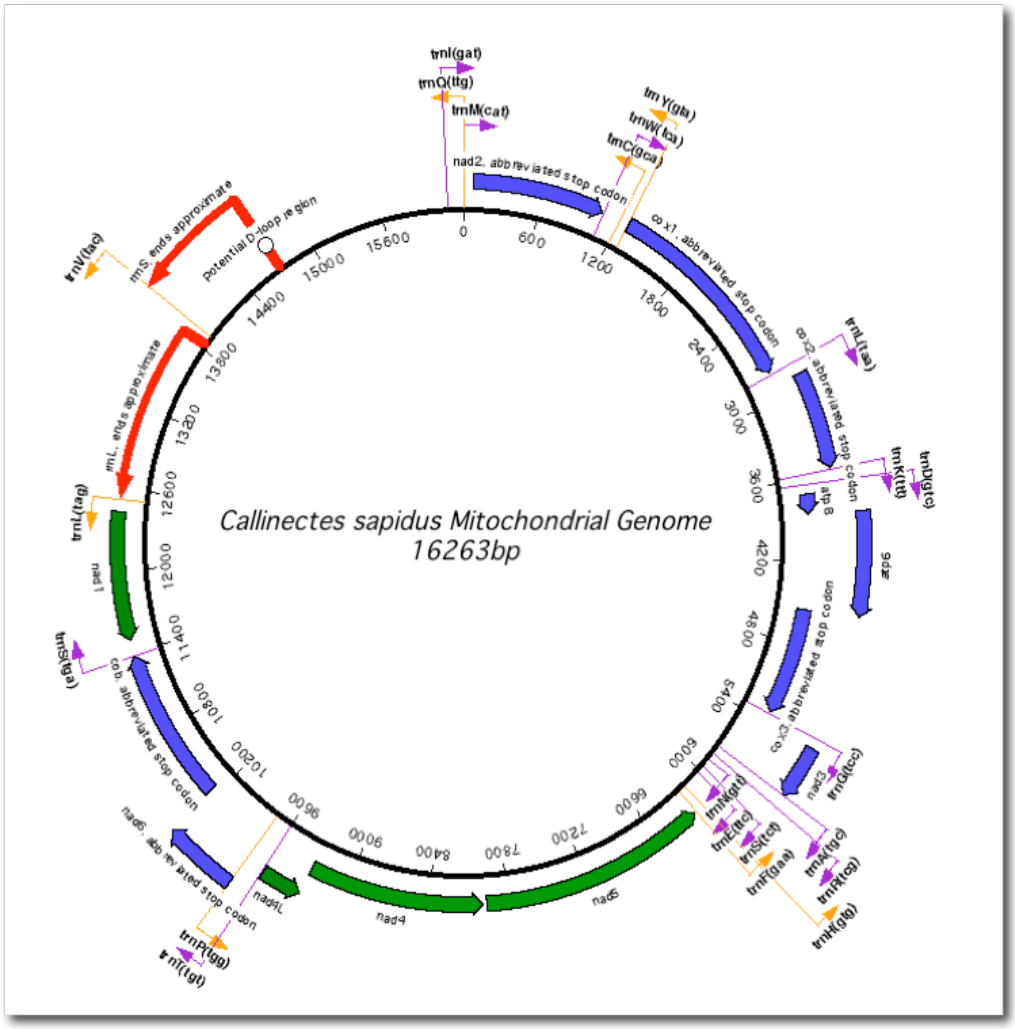

Figure 1-6. The annotated mitochondrial genome of the blue crab (Place, Feng et al. 2005).

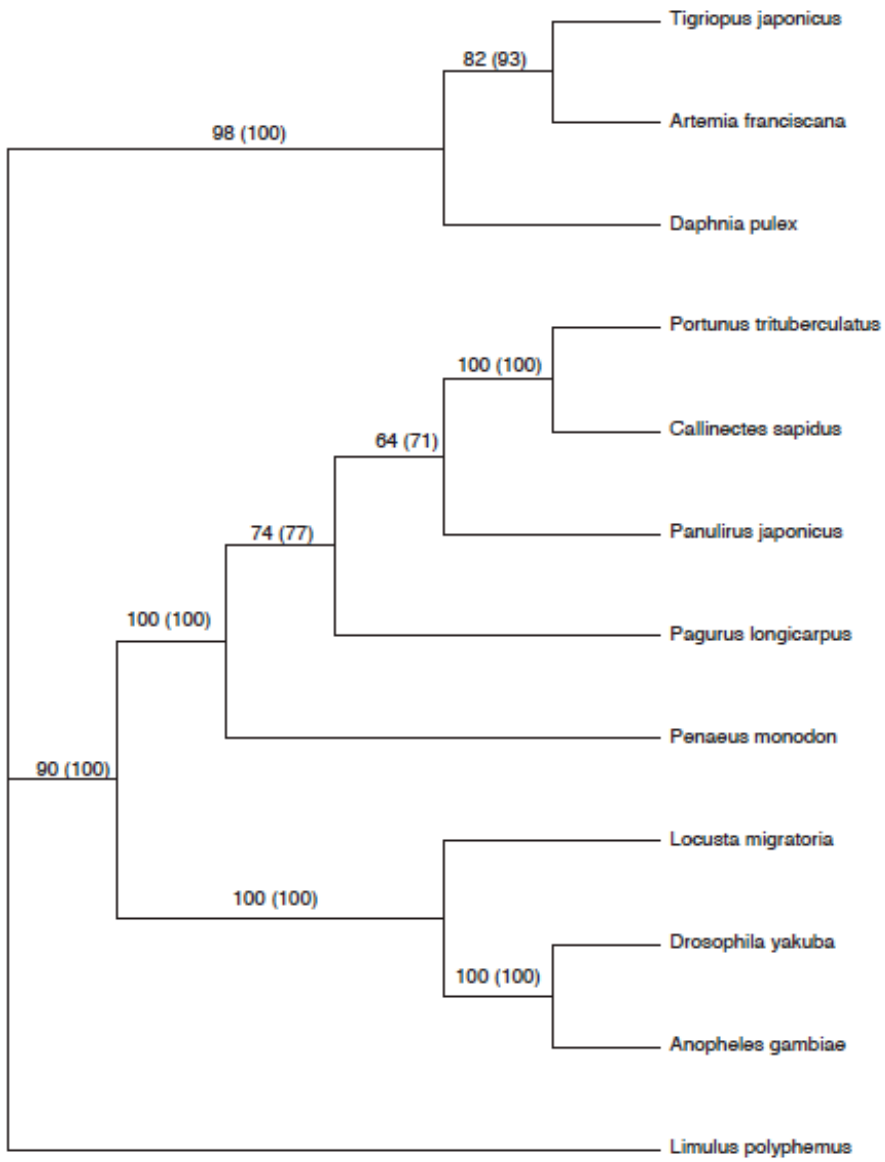

Figure 1-7. A MP tree for the Crustacea taxa inferred from concatenated amino acid sequences of the 13 protein-coding mitochondrial genes. The NJ tree was identical to that of the MP tree. Numbers beside internal branches indicate percentage of bootstrap values for each analysis obtained for 1000(MP), and 2000(NJ) replicates (Place, Feng et al. 2005).

2 Using molecular markers to track hatchery-produced crabs upon release

2.1 Introduction

The blue crab (*Callinectes sapidus*) inhabits estuarine and nearshore environments from Nova Scotia to northern Argentina (Williams 1974). It is one of the most abundant estuarine invertebrates that support valuable commercial and recreational fisheries along the Atlantic and Gulf coasts. The total value of United States blue crab landings in 2002 was more than $167 million (McMillen-Jackson and Bert 2004).

The Chesapeake Bay, as the largest western Atlantic estuary is supporting major production of blue crabs especially for the States of Maryland and Virginia. However, the landing of blue crabs from the Chesapeake Bay through 1990's has decreased with catch per unit effort increasing (Miller and Houde 1998). A long-term study in Chesapeake Bay from 1945-1995 showed a similar trend (Rugolo, Knotts et al. 1998). A study focusing only on females showed there was a dramatic reduction in Chesapeake Bay blue crab spawning stocks: 81% reduction in adult female abundance between 1988-1991 and 1992-2000 and 84% reduction in spawning stock biomass between 1988-1991 and 1994-2000 (Lipcius and Stockhausen 2002). In order to manage the blue crab fishery better, more and more attention and effort have been drawn to the studies of biology, ecology, population genetics and fisheries of the blue crab. The Center of Marine Biotechnology (COMB), University of Maryland, Biotechnology Institute (UMBI) in Baltimore has led the establishment of the Blue Crab Advanced Research Consortium (BCARC), cooperating with Smithsonian Environmental Research Center (SERC) in Edgewater, Maryland, the University of Southern Mississippi, the Virginia Institute of Marine

Sciences (VIMS) and North Carolina State University, with initial funding from the State of Maryland and from industry (Phillips Food Inc.), and with the additional federal funding provided through the Chesapeake Bay Office of NOAA. As a major component of BCARC, a blue crab hatchery facility has been set up in COMB. COMB and SERC managed to capture female crabs and produce off-season larval crabs in captivity via re-circulating, closed-loop tank systems. When the larvae grow up into juvenile crabs (C5-C7), they are released into suitable environments. The general goal of BCARC is to strengthen understanding of the blue crab's biology, genetics and ecology and so as to try to enhance the abundance and harvest of blue crabs in the Chesapeake Bay through releases of hatchery-produced juveniles.

In order to assess the success of the stock enhancement, a convenient and economic tagging technique is necessary. Coded microwire tag has been applied in tagging juvenile blue crabs rather than external tags due to some merits: it is retained during the molts of crabs and doesn't affect the survival and growth of the crabs (Fitz and Wiegert 1991). But use of this kind of tag is expensive and more importantly extremely labor intensive. It becomes infeasible when hundreds of thousands of crabs or juveniles less than C5 are to be tagged. In contrast inherent genetic tags are universal regardless of size or number. A variety of genetic tags have been used in mark-recapture studies such as microsatellite markers, AFLP markers, and mitochondrial DNA haplotypes.

A significant advantage in using mtDNA markers is the maternal heritage as well as the lack of recombination. Although there have been reports of mtDNA recombination (instances include: in the flatfish, (Hoarau, Holla et al. 2002); in the mussel, (Ladoukakis and Zouros 2001); this phenomenon is still considered as exceptionally rare, especially

while it becomes strong enough to be detected with current DNA sequencing techniques. Thus in principal, without considering mutations, all the offspring from a female crab are predicted to share the same mtDNA haplotype. Therefore, mtDNA seems like a perfect tool for tracking hatchery-produced crabs after they are released into the field.

2.2 Materials and Methods

2.2.1 Collecting blue crab samples

From 2003 to 2005, our colleagues at SERC conducted sampling for us. In the summers of 2003, 2004 and 2005, 41, 47 and 26 wild blue crabs were collected from Rhode River, Maryland **(**Figure 2-1) respectively. For 2003, 16 Batch H hatchery produced blue crabs were caught for our experiment before that batch was released. In 2004, 42 Batch J and 8 Batch A crabs were collected before the release; 47 Batch M and 49 Batch L crabs were sampled in 2005. All the hatchery crabs were produced from the hatchery facility at COMB. Hatchery crabs had been implanted with microwire tags in their walking legs for the identification. Crabs labeled with the same batch letter were supposed to be from the same brood.

Year of sample	2003	2004	2005
N (Wild)	41	47	26
N (Hatchery)	18	8 (Batch A) 40 (Batch J)	49 (Batch L) 47 (Batch M)
Total	59	95	122

Table 2-1. The sample sizes: 59, 95 and 122 crabs were collected in total from 2003 to 2005.

2.2.2 DNA extraction

Crabs were preserved in 95% ethanol until DNA extraction. For each wild crab, 100 mg muscle tissue from a walking leg was used for DNA extraction with Qiagen DNeasy Blood and Tissue Kit. For the hatchery crabs, the full crab body (if CW<10mm) or a half of the body (if CW>10mm) was used for the extraction.

2.2.3 Mitochondrial gene Polymerase Chain Reaction (PCR)s and DNA sequencing

Mitochondrial PCRs were performed to amplify an 750-bp segment of *nad2* gene (NADH dehydrogenase subunit 2) of mtDNA, using primers F296 (5'-TGCTTTATTATTCAACCCCG-3') and R1046 (5'-CCGAATAGATTGATTGAAGT-3') designed according to the complete *C. sapidus* mitochondrial genome sequence (Place, Feng et al. 2005). PCRs were done in a final volume of 30µL, containing 15µL PCR 2X Master Mix (Promega), 10pmol of each primer, and 20ng of template DNA under the following cycling conditions: 94ºC for 30s; followed by 32 cycles at 94ºC for 15s, 48ºC for 30s, 68ºC for 1min; followed by 72ºC for 5min. The size of PCR products was estimated by electrophoresis in 1% (w/v) agarose gels.

PCR products were precipitated by adding equal volumes of 20% polyethylene glycol (PEG) 8000 and 2.5 M NaCl, incubating for 15min at room temperature, and centrifuging at 2,500g for 30min. The supernatant was removed and the DNA pellet was washed with 70% ethanol, followed by another centrifuge at 2,500g for 15min. The supernatant was decanted and the purified PCR products were air dried and dissolved with 15- 20µL water. The sequencing reaction was done in 10µL volume, consisting of 40-150ng PCR product, 3pmol of primer, 0.5µL Big Dye v3.1 sequencing mix and 1.5µL 5X sequencing

buffer (Applied Biosystems). The cycling parameters were 95ºC for 5min; followed by 50 cycles at 95ºC for 15s, 50ºC for 15s, 60ºC for 4min. The sequencing product was cleaned by adding 60μL 100% isopropanol and 30μL H_2O, mixing thoroughly, incubating at room temperature for 30min, and centrifuging at 2000g for 30min. The supernatant was discarded and 100μL 70% isopropanol was added to wash the DNA, followed by another centrifugation at 2000g for 14min and the disposal of the supernatant. Cleaned DNA was air dried for 20-30min before 10μL formamide was added. The mixture was heated at 95ºC for 2min and then put on ice immediately. After it was chilled down, the denatured sequencing product was loaded into an ABI 3130xl genetic analyzer. For each PCR product, we conducted both forward and reverse sequencing.

2.2.4 Microsatellite genotyping

6 loci were used for microsatellite analyses using primers in Table 2-2.

Primer	Sequence (5'-3')	Tm (ºC)
1-121F	GAATAAGAGAACAAACACACGGGG	56
1-121R	AACTGCTTGCCTTCCTTCCATC	
1-035F	GACTGGAGAAACGATAGGTG	46
1-035R	GAACAAGGAGATTACACGGATTC	
1-073F	GCCTATTTGCCTCGCTACCCC	55
1-073R	GTCACCAAAGTTGAGCAAGACTCTCT	
2-001F	ATTGGGTGGTTGCTTCATTC	60
2-001R	ACGAACGAGGAGACGAGGA	
2-004F	AAACAACGGTAATTGTACGAGAAA	60
2-004R	AGGCTAATGCCACCATCATC	
2-007F	GGGACAAACAACATGAAAGTGG	60
2-007R	GAAAACCTATTCCGGGAAGC	

Table 2-2. The PCR primers of 6 microsatellite loci.

PCR for amplification of the microsatellite region was carried out in 10μL reactions in a 96 well polypropylene plate using Amplitaq gold (Applied Biosystems 4311806) with 4.6μL water, 1μL 10x buffer, 0.8μL supplied 25mM $MgCl_2$ solution, 1μL 10mM DNTPs, 1μL of primer mix at 5μM each for the forward and reverse primer, 0.1μL amplitaq gold, and 1.5μL template DNA. Thermal cycling began with a 10min denaturation at 95°C to activate the enzyme followed by 36 cycles of 95°C for 30s, the primer specific annealing temperature for 30s, and 72°C for 1min; and ending with a 10min polishing step at 72°C. The PCR products were precipitated with 1 volume of isopropanol/0.1 volume 3M sodium acetate pH 5.2 and spun at 3000xg for 20min at RT. The supernatant was decanted and the pellet was washed with 150μL of 70% ethanol and spun at 3000xg for 5min at RT. The supernatant was decanted and the plate was spun inverted at 50xg for 2min to remove residual ethanol. The dried pellet was resuspended in 15μL HiDi formamide and 0.5ul Genescan 500 LIZ size standard (Applied Biosystems 4311320 and 4322682 respectively). The products were analyzed using Genemapper software (Applied Biosystems 4370784) on an Applied Biosystems 3130XL genetic analyzer. Peak and allele calling was first performed by the Genemapper software as a rough estimate followed by manual curation and binning of the results.

2.2.5 Data analysis

We checked the quality of all DNA sequences and aligned them using the program SEQUENCHER®4.8 (Gene Codes Corporation) before exported the consensus sequences of forward and reverse sequences. The published *C. sapidus* mtDNA complete genome sequence was used as the reference to align these 627-bp fragments.

The number of variable sites, haplotype frequency distribution, and haplotype diversity were calculated using the program ARLEQUIN version 3.11 (Excoffier, Laval et al. 2005). The haplotype diversity was based on a formula: $h=(1-\sum x_i^2)n/(n-1)$, where x_i is the frequency of a haplotype and n is the sample size (Nei and Tajima 1981). Mean number of differences (π) between all pairs of haplotypes in the sample was given by

$$\hat{\pi} = \frac{n}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_j \hat{d}_{ij}$$

where d_{ij} is an estimate of the number of mutations having occurred since the divergence of haplotypes i and j, k is the number of haplotypes, is the frequency of haplotype I, and n is the sample size (Tajima 1993).The open source software Network 4.5.1.0.(fluxus-engineering.com) was used to produce minimum spanning network of all haplotypes with Median Joining calculation (Bandelt, Forster et al. 1995).

2.3 Results

For the samples from 2003, 2004, and 2005, respectively, 59, 95, and 122 wild and hatchery crabs were successfully sequenced in two directions on a 627 base-pair fragment of *nad2* gene.

Table 2-3 shows the size and sex ratio of all crab subsamples. For all 3 years, wild crabs were significantly larger than hatchery crabs with the only exception of Batch A hatchery crabs in 2004. The CW range of 2004 Batch A was 50-101mm, close to the range of 2004 wild crabs, but the average CW of Batch A (77.2mm) was still much lower than the wild (94.9mm). A strong sex ratio skew towards male crabs was present in all data. The male/female ratio in wild samples varied from 2.3 to 3.7 while the range for all hatchery

samples were 1.2-3.5. The size distributions in two sexes were totally overlapped, indicating there was no size difference between sexes.

Sample	N	CW range (mm)	Average CW (mm)	Sex ratio (M:F)
2003 Wild	41	35-68	53.4	3.5
2003 Hatchery	18	11-26	15.7	3.5
2004 Wild	47	59-153	94.9	3.7
2004 Hatchery	8 (Batch A)	50-101	77.2	1.7
	40 (Batch J)	12-40	30.2	2.1
2005 Wild	26	51-98	71.3	2.7
2005 Hatchery	49 (Batch L)	9-25	15.5	2.3
	47 (Batch M)	11-25	17.3	1.2

Table 2-3. The size and sex ratio of all wild and hatchery crab samples in 3 years.

2.3.1 mtDNA sequence variability in wild crabs.

As shown in Table 2-4 the *nad2* region of the blue crab was highly variable, with haplotype diversity very close to 1.00 for all three samples. 2003 sample had much larger mean number of pairwise differences (9.57±4.47) than 2004 (7.74±3.66) and 2005 (6.38±3.11). Although 2003 sample had fewer crabs (41) than 2004 sample (47), it had more variable sites (67) compared to 2004 (59).

	2003	2004	2005
Sample size	41	47	26
Number of polymorphic sites	67	59	43
Mean number of pairwise differences	9.57±4.47	7.74±3.66	6.38±3.11
Number of haplotypes	37	39	25
Haplotype diversity	1.0000±0.0054	1.0000±0.0044	1.0000±0.0107

Table 2-4. Variability of mtDNA *nad2* sequences in wild crabs. The number of polymorphic sites, mean number of pairwise differences, number of haplotypes, and haplotype diversity were calculated for 3 wild samples.

Among 41, 47, and 26 wild crabs from the year of 2003, 2004 and 2005, we observed 37, 39 and 25 distinct haploytpes. Among those haplotypes, there were respectively 35, 35 and 24 singleton haplotypes. Only 2-4 haplotypes in each year of sample occurred more than once.

Besides the richness of singleton haplotypes, lack of prevalent haplotypes within populations, there was also a lack of shared haplotypes among populations. Among the 80 haplotypes in total, only 3 haplotypes were observed in all these 3 samples and 6 others were shared by 2 of 3 samples (Table 2-5).

Haploytpes (Occurrence)	2003	2004	2005
0357(6)	2	3	1
0321(5)	4	1	0
0322(5)	1	2	2
044(5)	0	5	0
0372(3)	1	1	1
0335(3)	1	2	0
0455(2)	0	1	1
0355(2)	1	0	1
0425(2)	0	1	1

Table 2-5. There were 9 mtDNA haplotypes that occurred more than once. Three of them were shared in 3 years of samples. The most common haplotype within sample changed from year to year.

Furthermore, only 53.6%, 74.4%, and 61.5% of haplotypes were coding for the same protein sequence respectively in the sample of 2003, 2004, and 2005. Four protein sequences were observed across years (Table 2-6) while there were 4, 5, and 3 singleton protein sequences in 2003, 2004, and 2005 samples respectively.

Protein sequences (Occurrence)	2003	2004	2005
1(74)	23	35	16
2(22)	13	5	4
3(4)	1	1	2
4(2)	0	1	1

Table 2-6. Four protein sequences occurred more than once and three of them were presented in all 3 years of samples.

F-statistics showed the *F*st P value was significantly different from zero (Table 2-7) only between the 2003 and 2004 sample.

Sample combinations	***F***st	*P*
2003 vs. 2004	**0.0429**	**0.0030±0.0006***
2004 vs. 2005	0.0043	0.2871±0.0045
2003 vs. 2005	0.0172	0.1014± 0.0029

Table 2-7. *F*st and *P* value among wild caught samples. * denotes significant P value at 5% level for 10000 permutations.

2.3.2 Identifying hatchery-produced crabs.

We use the program Network 4.5.0.0 to produce a minimum spanning network (MSN) for each sample, using Median Joining calculation (Bandelt, Forster et al. 1995). For the 2003 sample, all 18 hatchery-produced Batch H crabs but one individual shared the same *nad2* sequences in the MSN (Figure 2-2). Not a single wild crab carried caught at the same time exhibited this haplotype. The single hatchery crab (RR0373) outlier had 6 nucleotides substitutions different from this brood's maternal haplotype while it had the

same protein sequence as other hatchery crabs (Figure 2-3). Either this represented a hatchery crab from a different brood or implied extensive mutation during hatchery grow out.

For the 2004 sample, 39 out of 40 Batch J hatchery crabs were carrying the same haploytpe as observed with one wild crab (RR0472) as well as being identical with two Batch A crabs (RR0412 and RR0420). Only one crab of Batch J (RR0471) was left outside of this major type, different by 4 bases, which made it have a different protein sequence (Figure 2-4).

However, Batch A crabs showed a more complicated pattern. For only 8 of Batch A crabs, 2 of them had same haplotype as Batch J major type, 3 of them grouped with 5 wild crabs and 3 others randomly dispersed in the network (Figure 2-4). Among those 8 crabs, only one had a nonsynonymous substitution in the DNA sequence (Figure 2-5).

For the 2005 sample, 47 Batch M crabs all shared the same haplotype that was distinct from all wild crabs; 48 of 49 Batch L crabs were of the same *nad2* sequence that was also different from any wild sample. The only one outlier of Batch L (RR05264) had two different bases from its supposed siblings (Figure 2-6) and a different protein sequence (Figure 2-7).

Male and female wild crabs were presented in different colors in these networks (Figure 2-2, Figure 2-4, and Figure 2-6). No sorting pattern according to sex was observed.

Using variation in this 627-bp of the *nad2* locus as a genetic marker, greater than 90% of hatchery-produced crabs could be identified from the wild individuals. We defined "False negatives" as those crabs that were tagged as hatchery but genetically different from the hatchery haplotypes, "false positives" as those wild ones sharing haplotypes with hatchery samples. As we can see from the Table 2-8, there were no false positives in 2003 or 2005, and only one false negative for each year. The 2004 sample was not as clean, but mostly due to the Batch A data. If we only consider Batch J data, there was also only 1 false negative and 1 false positive found. In total, the success ratio was 92.6% for three years of samples, 5 batches included (Table 2-8).

	2003	2004	2005
Number of hatchery-produced crabs	18	48	96
False negative/ False positive	1/0	4/6	1/0
Percentage of success	92.6%		

Table 2-8. Success ratio of hatchery crabs identification by using *nad2* gene as the single genetic marker. "False negatives" were those crabs that were tagged as hatchery but genetically different from the hatchery haplotypes, "false positives" as those wild ones sharing haplotypes with hatchery samples.

2.4 Discussion

In the mitochondrial genome of the blue crab there are 13 protein-coding genes, 22 tRNA genes, 2 ribosomal RNA genes and one hypervariable D-loop region (Place, Feng et al. 2005). To select an appropriate segment to be used as a genetic marker, we were looking for a relatively variable region so it would have high resolution when to be used to track down individual crabs; but at the same time we needed it to be relatively conserved so

mutation won't happen too often from generation to generation because we are developing the markers based on the sponging female crabs to track down their offspring. We performed our preliminary experiments on different regions of mitochondrial genome. The D-loop region, as a hypervariable region, was really difficult to design reliably amplifiable PCR primers on. There was insert/deletion between individuals (Place, Feng et al. 2005). So we moved on to coding regions. By comparing three regions, *cox1*, *cytb* and *nad2*, we finally selected *nad2* as our marker since among these three coding regions it had the highest nucleotide diversity value that was about 0.01 while *cox1* was 0.003-0.006 and *cytb* was 0.006-0.008.

Our first observation is that the genetic composition of the wild blue crab population is highly diverse. The haplotype diversity is nearly 1.0. Very few haplotypes occurred more than once which means most of sampled wild crabs have distinct haploid genotypes from each other. This is the basis of that we are able to use this single-locus genetic marker - *nad2* as our genetic tag to distinguish the hatchery-produced crabs from the wild. And apparently this is the reason for the very low false positive ratio for the identification.

As we indicated above, there are some discrepancies between our grouping pattern and the tagged groups. Theoretically, it is possible that PCR and sequencing artifacts might have played a role here but it is not of high likelihood because those troubling samples usually carried more than one differentiating bases. It is more likely we are picking up signals underestimated in the manual tagging process.

Our microsatellite data (Table 2-9) of the 2004 samples confirmed the reliability of the mtDNA marker. For the batch A sample, the crab RR0412 that was tagged as batch A hatchery crab but grouped as wild based on mtDNA haplotype, showed different alleles from the other 7 Batch A crabs. At four loci 1-035, 1-073, 2-004, and 2-007, it had the extra fifth/sixth alleles to the 4 parental alleles inferred from other 7 crabs. This proved RR0412 was not a batch A crab. Similarly, the individual RR0471 in the 2004 sample, the crab being tagged as a hatchery batch J crab but carrying a different haplotype from other batch J crabs, also had very different alleles at 5 loci. It had the fifth/sixth alleles at 3 loci. Thus we have enough confidence that RR0471 was not a hatchery Batch J crab. These two crabs were mis-tagged manually but we were able to point out this mis-tagging by using the genetic tag.

Another wild crab (RR0472) was shown as 'wild' since it didn't contain any tag, but it was grouped with the hatchery batch J crabs based on their identical haplotypes. This crab was very likely a sibling to those batch J crabs based on their microsatellite genotypes. It didn't have any outlier alleles at any of the 6 loci. Given the low frequencies of its alleles in wild population, although we still can not fully rule out the possibility of it was a wild crab happening to have the same genetic background as the Batch J crabs, it sounds that it could be just a hatchery crab having lost its physical tag. Then, by using microsatellite data, we proved the robustness of the mtDNA marker to some degree.

Locus	1-121	1-035	1-073	2-001	2-004	2-007
Batch A parental alleles	202	176	203	335	176	180
	204	198	252	348	180	186
		205	274	360	193	196
		228	282			198
Crab RR0412 alleles	202	**182**	**254**	344	**184**	**208**
		227			**203**	
Batch J parental alleles	200	144	254	329	182	196
	202	174	256	340	195	204
	204	180	262	347	203	
		196	276		223	
Crab RR0471 alleles	202	**186**	**235**	350	**180**	190
		202	**266**			
Crab RR0472	202 (82.8%)	174 (7.9%)	254 (1.3%)	329 (5.4%)	N/A	196 (1.5%)
alleles		196 (1.1%)	262 (1.3%)			

Table 2-9. The microsatellite genotype summary of Batch A and Batch H crabs at 6 loci. The Batch A parental alleles were inferred from all Batch A crabs excluding the crab RR0412 and the Batch J parental alleles were inferred from all Batch J crabs excluding RR0471. The crab RR0472's alleles were shown with their respective frequencies in wild crabs in parentheses. Bold font numbers showed the fifth/sixth alleles at one locus that indicated the alleles were from two pairs of parents.

There are a few possible causes for the physical mistagging. When crabs are kept in the hatchery facilities, to be specific, in a couple of huge tanks, they are sometimes

transferred accidently between different tanks, thus causing mixed batches. If the latter batches became our research batches, then these crabs could have been the 'false negatives'. And for the 'false positives', they can either be truly wild crabs but with the same haplotype/genotype as the hatchery ones, or, they might be from the hatchery, but somehow have lost their physical microwire tags then be regarded as wild samples. Microwire tag retention is not 100% and if the crab lost that swimming leg that contains the tag, then the tag would be lost. The confusing grouping pattern of Batch A crab of 2004 might be due to these explanation.

As we know, we do not have the ability to distinguish the wild crabs from our hatchery crabs while they have the same haplotypes if we only use mtDNA tags. So it is a big concern for our identifying methodology. If our 'mother' crabs have a haplotype that is very common in wild populations, then the power of our genetic marker will be really weakened. However, we think this may not be a big problem. Looking at the population genetic compositions of those three years of samples we see the major components are singletons. The frequency of the most common haplotype is lower than 8%. So even if the sponging females for the hatchery unfortunately happens to be of that most common haploytpe in wild, the false positives caused by wild individuals with the same haplotype will not be too much of a concern.

This *C. sapidus nad2* haplotype frequency distribution pattern is very interesting and has never gained enough attention. The possible mechanism could be a sudden expansion of a small population with a reasonably high mutation rate; or it could be result from subset of

a huge and highly diverged population that has reached equilibrium; or some other unclear reasons. This will be the subject of future chapters in my thesis.

In conclusion, despite the experimental error inherent in fingerprinting, mtDNA of the blue crab is an excellent marker for distinguishing hatchery rear crabs from wild crabs.

Figure 2-1. The location of Rhode River, a tidal tributary in the Chesapeake Bay western shore.

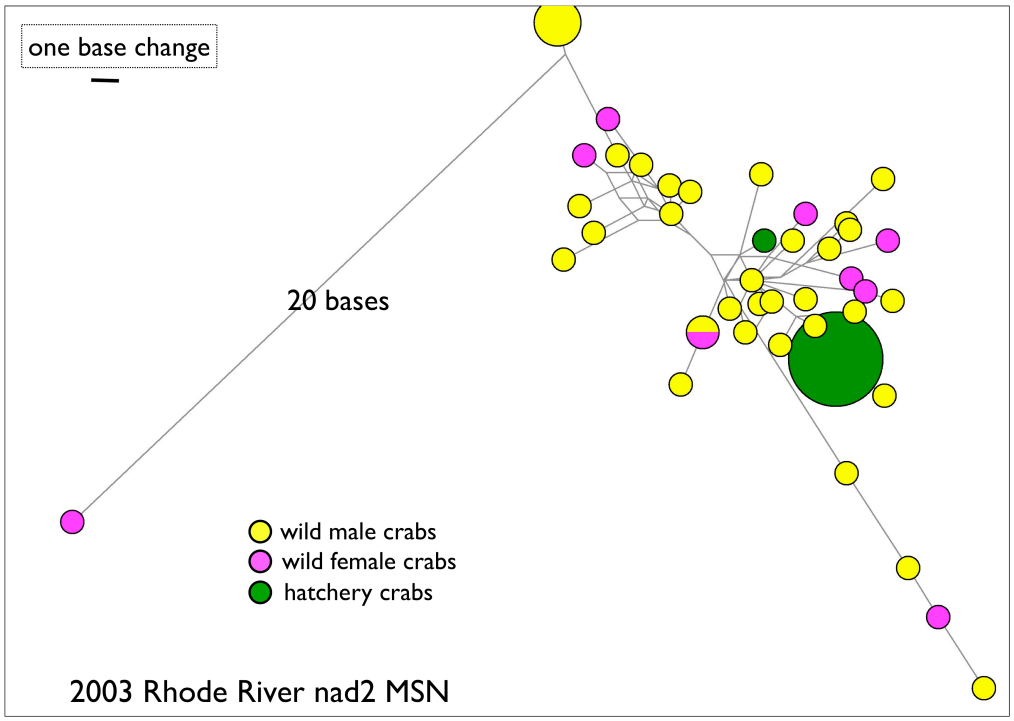

Figure 2-2. Minimum Spanning Network of 2003 Rhode River crabs *nad2* DNA sequences. Different colors denote the origin of sequences. Each circle stands for one distinct haplotype; the size of circles is proportional to the frequency of halotypes; and the branch length is proportional to the genetic distance.

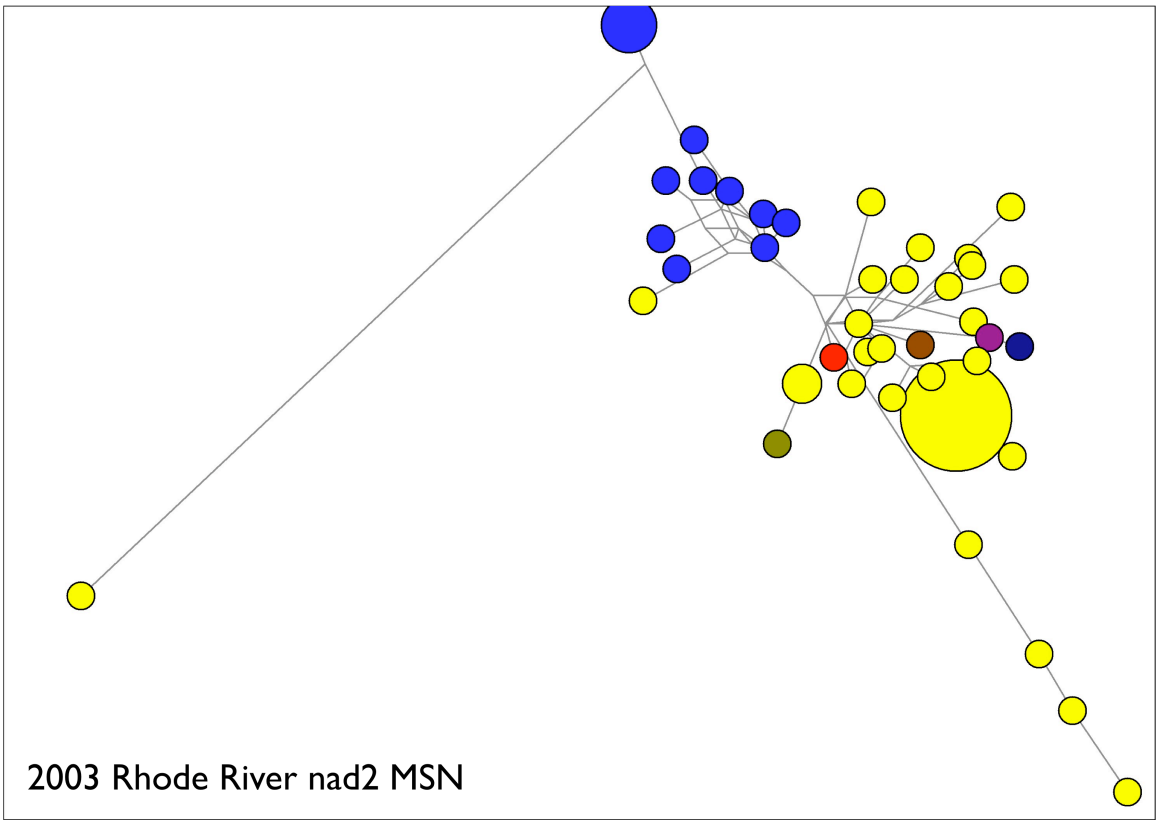

Figure 2-3. Modified Minimum Spanning Network of 2003 Rhode River samples based on Figure 2-2. Different colors stand for different protein sequences. The color scheme is consistent in Figure 2-3, 2-5, and 2-7. All haplotypes with same color in these three figures are coding for identical proteins.

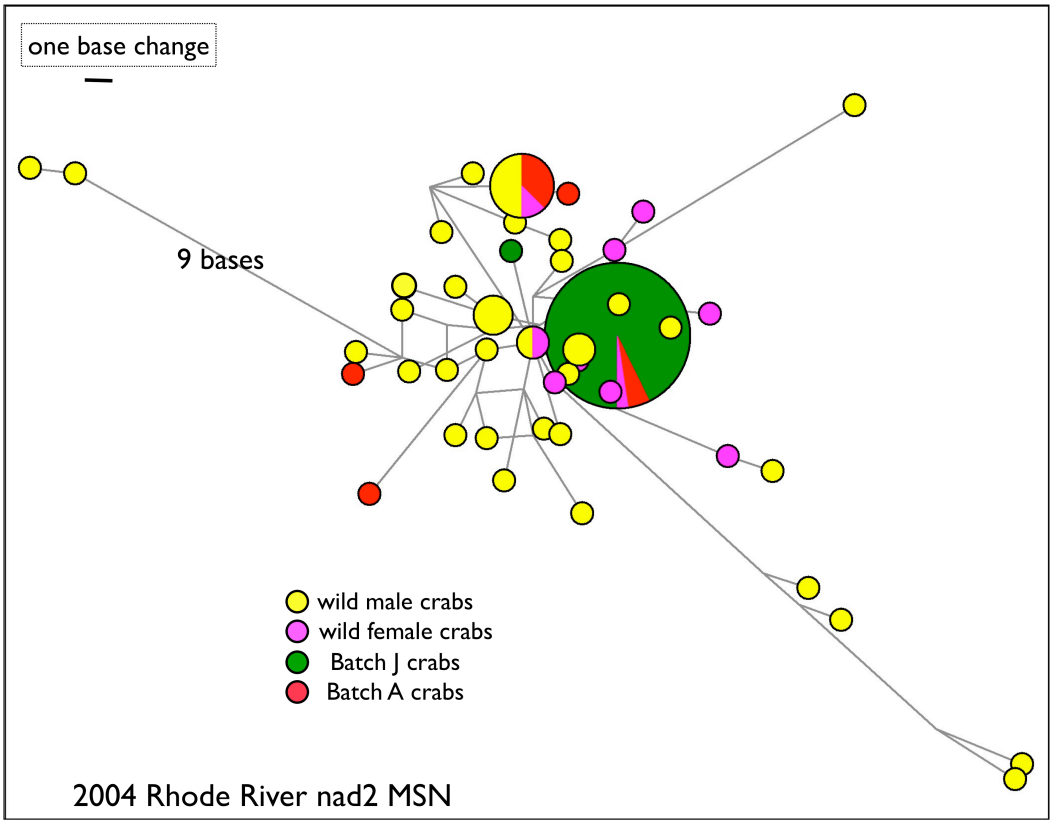

Figure 2-4. Minimum Spanning Network of 2004 Rhode River crabs *nad2* DNA sequences. Different colors denote the origin of sequences. Each circle stands for one distinct haplotype; the size of circles is proportional to the frequency of halotypes; and the branch length is proportional to the genetic distance.

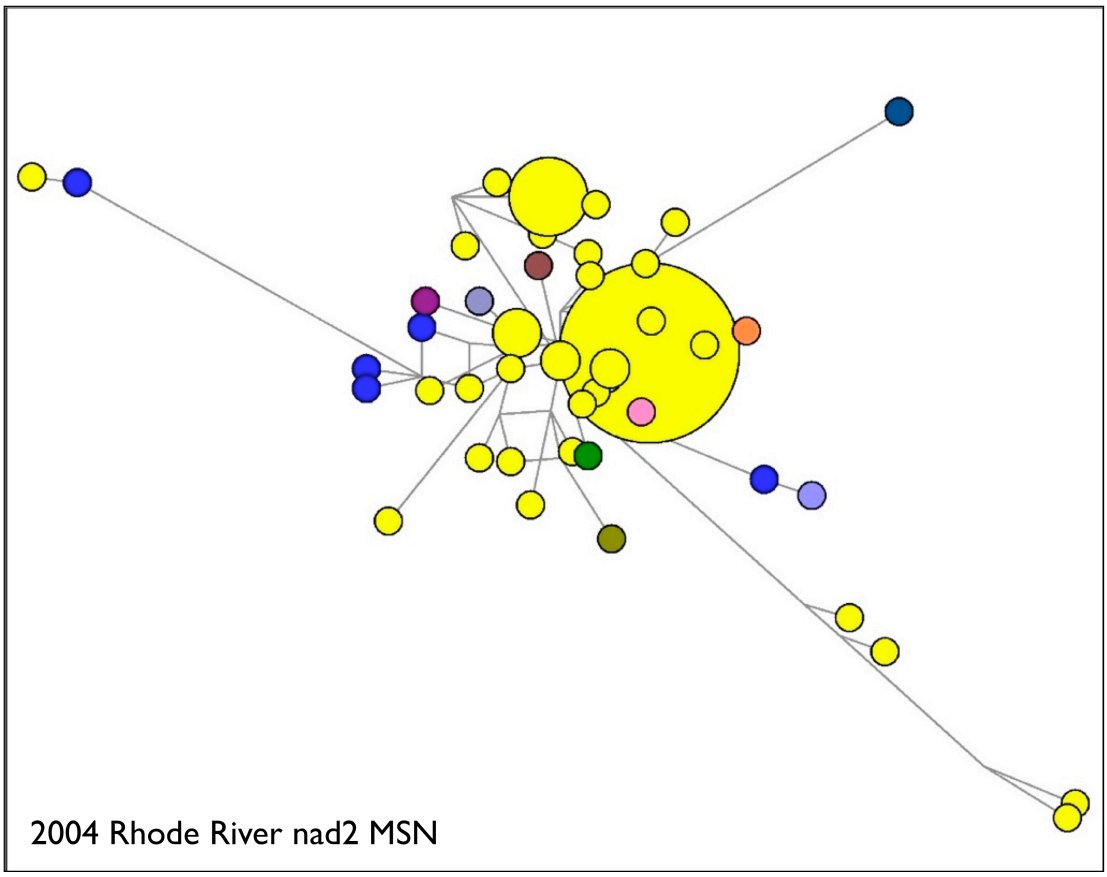

Figure 2-5. Modified Minimum Spanning Network of 2004 Rhode River samples based on Figure 2-4. Different colors stand for different protein sequences. The color scheme is consistent in Figure 2-3, 2-5, and 2-7. All haplotypes with same color in these three figures are coding for identical proteins.

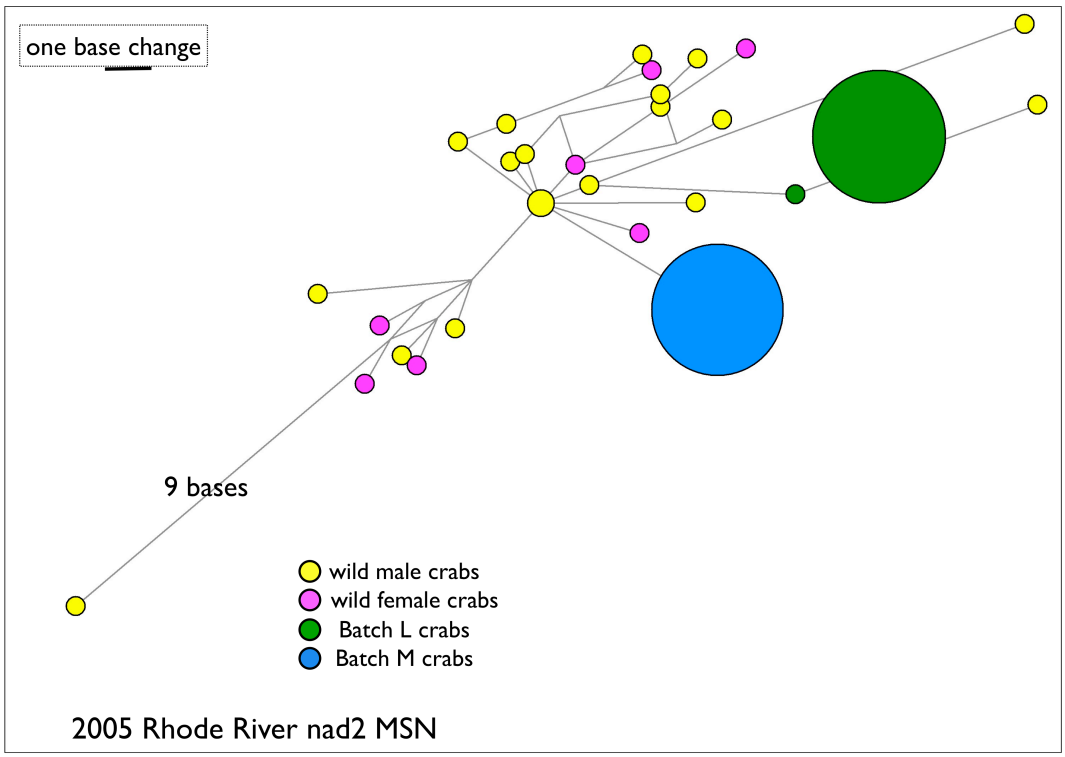

Figure 2-6. Minimum Spanning Network of 2005 Rhode River crabs *nad2* DNA sequences. Different colors denote the origin of sequences. Each circle stands for one distinct haplotype; the size of circles is proportional to the frequency of halotypes; and the branch length is proportional to the genetic distance.

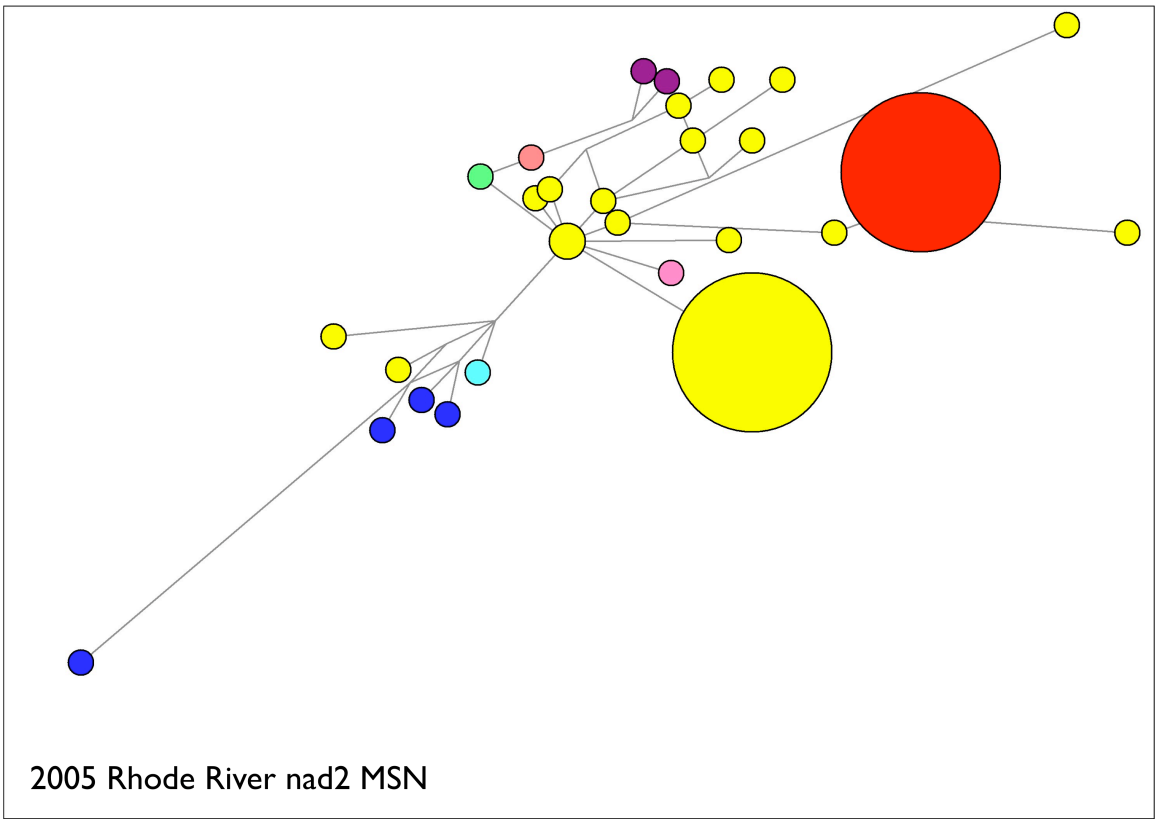

Figure 2-7. Modified Minimum Spanning Network of 2005 Rhode River samples based on Figure 2-6. Different colors stand for different protein sequences. The color scheme is consistent in Figure 2-3, 2-5, and 2-7. All haplotypes with same color in these three figures are coding for identical proteins.

3 Variation in mutation rate and polymorphism among four mitochondrial genes of the blue crab *Callinectes sapidus*

3.1 Introduction

Animal mitochondrial genome has been considered as a rapidly evolving molecule compared to the nuclear DNA (Brown, George et al. 1979; Ingman, Kaessmann et al. 2000; Tatarenkov and Avise 2007). Hence it is very useful for dating recent evolutionary events and population history (Avise, Arnold et al. 1987). It is possibly because of the reactive oxygen environment due to those by-products of oxidative phosphorylation and an error-prone polymerase and limited DNA repair (Bogenhagen 1999).

When mtDNA variations are used to estimate the divergence times among animal species, a molecular clock was assumed for the whole mitochondrial genome (Billington 2003). But now it looks like the mutation rate of mitochondrial genome varies distinctively among lineages (Nabholz, Glemin et al. 2009). Furthermore, different mitochondrial regions also feature different mutation rate. In fact the D-loop (control region) is well known to be much more variable than all genes so it is used to measure recent divergence among animals where other genetic markers may fail to detect (Nesbo, Arab et al. 1998; Cao, Ort et al. 2009).

Among the 13 protein-coding and 2 rRNA genes, *cox1* (cytochrome c oxidase subunit 1) is most commonly used for population genetics for their relatively high variability in contrast to the more conserved 12S and16S regions that are more used for inter-species phylogenetics research.

In the present work, we study the species-wide polymorphism and substitution rates in 4 of the mitochondrial genes in the blue crab *Callinectes sapidus* (Brachyura; Portunidae): *cox1* (cytochrome c oxidase subunit 1), *nad2* (NADH dehydrogenase subunit 2), *nad4* (NADH dehydrogenase subunit 4), and 12S rRNA.

3.2 Materials and methods

3.2.1 Collecting blue crab samples

From 2003 to 2005, our colleagues in Smithsonian Environmental Research Center conducted sampling for us. In the summers of 2003, 2004 and 2005, 30, 33 and 26 wild blue crabs were collected from Rhode River, Maryland (Figure 3-1).

3.2.2 DNA extraction and PCRs

Crabs were preserved in 95% ethanol until DNA extraction. About 100 mg muscle tissue from a walking leg in each crab was used for DNA extraction with Qiagen DNeasy Blood and Tissue Kit.

PCR primers were designed based on the published complete *C. sapidus* mitochondrial genome sequence (Place, Feng et al. 2005), targeting four genes—*nad2, cox1, nad4,* and 12S. Primer sequences are shown in Table 3-1.

Gene	Amplicon Length(bp)	Primers (5'- 3')	Forward Primer
nad2	750	F	TGCTTTATTATTCAACCCCG
		R	TGCTTTATTATTCAACCCCG
cox1	782	F	ACTGTTTACCCTCCCCTTGC
		R	AATACAGCGCCCATGGATAG
nad4	816	F	CATTTACTTTTCCTGAACAACATGA
		R	TGCCTTTGTTGGTGTCTTTG
12S	790	F	CCAGGTTCACTTTCCAGTAA
		R	ATGAGAGTTGTAGCGGGTAA

Table 3-1. The PCR primers used for 4 genes (*nad2, cox1, nad4,* and 12S).

PCRs were done in a final volume of 30µL, containing 15µL PCR 2X Master Mix (Promega), 10pmol of each primer, and 20ng of template DNA under the following cycling conditions: 94ºC for 30s; followed by 32 cycles at 94ºC for 15s, 48ºC for 30s, 68ºC for 1min; followed by 72ºC for 5min. The size of PCR products was estimated by electrophoresis in 1% (w/v) agarose gels.

3.2.3 DNA sequencing

PCR products were precipitated by adding equal volumes of 20% polyethylene glycol (PEG) 8000 and 2.5 M NaCl, incubating for 15min at room temperature, and centrifuging at 2,500g for 30min. The supernatant was removed and the DNA pellet was washed with 70% ethanol, followed by another centrifuge at 2,500g for 15min. The supernatant was decanted and the purified PCR products were air dried and dissolved with 15- 20µL water. The sequencing reaction was done in 10µL volume, consisting of 40-150ng PCR product, 3pmol of primer, 0.5µL Big Dye v3.1 sequencing mix and 1.5µL 5X sequencing buffer (Applied Biosystems). The cycling parameters were 95ºC for 5min; followed by

50 cycles at 95ºC for 15s, 50ºC for 15s, 60ºC for 4min. The sequencing product was cleaned by adding 60μL 100% isopropanol and 30μL H_2O, mixing thoroughly, incubating at room temperature for 30min, and centrifuging at 2000g for 30min. The supernatant was discarded and 100μL 70% isopropanol was added to wash the DNA, followed by another centrifugation at 2000g for 14min and the disposal of the supernatant. Cleaned DNA was air dried for 20-30min before 10μL formamide was added. The mixture was heated at 95ºC for 2min and then put on ice immediately. After it was chilled down, the denatured sequencing product was loaded into an ABI 3130xl genetic analyzer. For each PCR product, we conducted both forward and reverse sequencing.

3.2.4 Data analysis

We checked the quality of all DNA sequences and aligned them using the program SEQUENCHER®4.8 (Gene Codes Corporation) before exported the consensus sequences of forward and reverse sequences. The actual exported sequences were shorter than the PCR amplicons in length because only the DNA sequences unambiguously covered by two directions were used in alignments and then exported. The published *C. sapidus* mtDNA genome sequence was used as the reference sequence.

The polymorphism index calculation and neutrality test were performed by using program DnaSP v5 (Librado and Rozas 2009).

The following measures were calculated for haplotype/nucleotide diversity,:

S, the number of segregating (polymorphic) sites;

Haplotype (gene) diversity and its sampling variance (Nei 1987); the standard deviation (or standard error) is the square root of the variance;

Nucleotide diversity, Pi (π), the average number of nucleotide differences per site between two sequences and its sampling variance (Nei 1987); the standard deviation (or standard error) is the square root of the variance;

The mutation parameter Theta (θ , per site or per gene) calculated from S or Eta (η), the total number of mutations (Watterson 1975; Nei 1987);

k, the average number of nucleotide differences between two sequences (Tajima 1983); and G+C Content in coding/noncoding regions.

Some measures or parameters like Pi, Theta were also calculated with the sliding window method across the concatenated alignment. In this method a window (segment of DNA) is moved along the sequences in steps and the parameter is calculated in each window, and the value is assigned to the nucleotide at the midpoint of the window. When this method was applied in this study, the window size was 100 bp and the step size was 25 bp.

As for neutrality tests, following tests were conducted:

Tajima's D that tests the relationship between the number of segregating sites S and the average number of nucleotide differences k estimated from pairwise comparison (Tajima 1989) and its statistical significance;

Fu and Li's D* test statistic is based on the differences between η_s, the number of singletons (mutations appearing only once among the sequences), and η , the total number of mutations (Fu and Li 1993) and its statistical significance;

Fu and Li's F* test statistic is based on the differences between η_s, the number of singletons (mutations appearing only once among the sequences), and k, the average number of nucleotide differences between pairs of sequences (Fu and Li 1993) and its statistical significance;

Fu's Fs test statistic is based on the haplotype (gene) frequency distribution conditional the value of Theta, (Fu 1997).

The free software Network 4.5.1.0.(fluxus-engineering.com) was used to produce minimum spanning network of all haplotypes with Median Joining calculation (Bandelt, Forster et al. 1995).

Neighbor Joining trees were produced using program MEGA4 (Tamura, Dudley et al. 2007). All sites were equally weighted. For NJ tree construction, only 9 blue crab sequences were randomly chosen from 89 Rhode River wild crabs, in addition to the reference blue crab that was sampled from Mississippi area and one outlier Rhode River crab RR0382. In the NJ trees, 6 ourgroups were used, including 4 species with complete mitochondrial genome sequences available on NCBI website: *Portunus trituberculatus* (Brachyura; Portunidae; accession #: NC_005037); *Scylla olivacea* (Brachyura; Portunidae; accession #: NC_012569); *Scylla serrata* (Brachyura; Portunidae; accession #: NC_012565); *Eriocheir sinensis* (Brachyura; Varunidae; accession #: NC_006992); and two *Callinectes* species that had partial *cox1* sequences deposited:

Callinectes arcuatus (Brachyura; Portunidae; accession #: AY465911) and *Callinectes bellicosus* (Brachyura; Portunidae; accession #: AY465909).

E. sinensis was used as the root.

Pairwise Fst statistics based on distance were conducted using program Arlequin3.11 (Excoffier, Laval et al. 2005).

3.3 Results

3.3.1 The mtDNA sequence polymorphism in 4 genes.

For 3 protein-coding genes, the G+C content varied from 0.3073 (*nad4*) to 0.3601 (*cox1*) and it was 0.2888 in 12S, generally consistent with the high AT content of the blue crab mitochondrial genome (Place, Feng et al. 2005).

The DNA sequence variability of 4 genes is shown in Table 3-2. Among 4 genes, *nad2* had the most segregating sites, 14.8% of total 627 bases being variable; *nad4* had 13.3% of 744 bases variable; *cox1* had 9.4% for 678 bases, which made it the least variable gene in these three coding genes. But *cox1* was still carrying two-fold higher variability of 12S that had only 4.6% of 712 bases showing polymorphism. The mutation numbers were generally very close to the segregating numbers, meaning all the mutations but very few took place at different sites. All values of S, Eta, Pi, k and Theta were consistently showing *nad2* was the most polymorphic coding gene while *cox1* being the least polymorphic one but still much more variable than 12S region.

The haplotype diversity was very close to 1 (0.929-0.992) for all three protein-coding genes although much lower for 12S (0.8), indicating a generally highly diversified genetic composition of the blue crab populations.

The Pi and Theta values calculated through sliding window method showed the trend of variability across these 4 genes (Figure 3-2). In addition to the general pattern of higher variability in *nad2* and nad4, lower in *cox1* and 12S, it was shown that there were some relatively hot spots in *cox1* and 12S, especially in the latter half DNA of 12S.

Gene	*nad2*	*cox1*	*nad4*	12S	Total
N	90	90	90	90	90
Sites	627	678	744	712	2761
S	93	64	99	33	289
Eta	96	68	100	34	298
NHap	73	50	64	27	84
Hd	0.992	0.929	0.988	0.8	0.998
VarHd	0.00001	0.00048	0.00002	0.00155	0.00001
Pi	0.01305	0.00498	0.00915	0.00229	0.00724
K	8.1808	3.3775	6.8105	1.6295	19.9983
ThetaNuc	0.03019	0.01978	0.0265	0.00942	0.02128
ThetaG	18.9295	13.4084	19.7182	6.7042	58.7602

Table 3-2. The DNA diversity of 90 blue crabs (including the reference crab that we sequenced the whole genome from). N, the sample size; S, the number of Segregating Sites; Eta, the total number of mutations; NHap, number of haplotypes; Hd, Haplotype (gene) diversity and VarHd, its sampling variance; Pi (π), Nucleotide diversity; k, the average number of nucleotide differences of two sequences; Theta (per nucleotide site or per gene), the mutation factor.

3.3.2 Haplotype minimum spanning networks

The DNA minimum spanning networks are shown in Figure 3-3 through Figure 3-6 for 4 genes separately. One crab whose ID was RR0382 was always placed in the corners of the networks with very long branches, showing its distinctness from all other crabs. The *cox1* and 12S both had pronounced major haplotypes as well as less numbers of minor haplotypes compared with the *nad2* and *nad4* networks. This pattern was in congruence with their lower variability shown in Figure 3-2.

3.3.3 Phylogenetics analysis

The *C. sapidus* clade

Neighbor Joining trees were produced by MEGA4 on the concatenated sequences of 4 genes (Figure 3-7) and on 4 single genes (Figure 3-8); all sites were included and equally weighted. Four outgroup sequences were used in the phylogenetics analysis along with 10 *C. sapidus* sequences. The genealogies agreed with each other fairly well when all the *C. sapidus* crabs were concerned. All of them were grouped together as a monophyletic clade for either concatenated sequences or separate genes, indicating that although the crab RR0382 showed a long distance to its same species fellows in the minimum spanning networks, it was even far more distantly related to the 4 outgroups. The distance between RR0382 and other blue crabs was obviously shorter than the distance between two congeners *S. olivacea* and *S. serrata,* which implied the difference between RR0382 and other blue crabs could be attributed to within-species variation.

Only two other *Callinectes* species had their partial *cox1* sequences deposited on NCBI website. A Neighbor Joining tree (Figure 3-9) was made on this partial *cox1* sequence alignment that was 268bp in length, of the three *Callinectes* species plus *P. trituberculatus* while the latter one was used as the root. Similarly, all the presumed *C. sapidus* individuals including RR0382 formed a monophyletic group and the distance among all members in this group were much shorter than them to the two congener species. This confirmed the crab RR0382 was *C. sapidus* despite its distinctness from its fellows.

The phylogenetics based on different genes and different position codons

Comparing Figure 3-7 to Figure 3-8, there were three types of topologies in the trees. The tree based on 4-gene concatenated sequences agreed with *nad2* and *cox1* single gene trees, supporting that *P. trituberculatus* was more closely related to *Scylla* crabs than to *C. sapidus*. While 12S single gene tree supported the opposite that *P. trituberculatus* was more close to *C. sapidus*. The *nad4* tree was surprisingly supporting another different tree topology in that two *Scylla* species did not group into a monophyletic clade. This incongruence despite the fact that the mitochondrial DNA was a single locus marker without recombination suggested the strong background noise due to the high variability in the genome.

We broke down the sequences by genes and by codon positions. Figure 3-10 and Figure 3-11 showed the Neighbor Joining trees based on the 1^{st} + 2^{nd} codons and the 3^{rd} codons in the 4-gene concatenated sequences. Figure 3-12 and Figure 3-13 were the Neighbor Joining trees on two sets of codons for each coding genes.

The NJ trees on the 3^{rd} positions in each of three coding genes (Figure 3-13) disagreed with each other, supporting three types of tree topologies. Among them, only the *nad2* 3^{rd} codon tree was in congruence with some of the trees supported by other sequences. The *cox1* 3^{rd} codon tree basically lost all information to resolve any structure within all *Portunidae* crabs. And the *nad4* 3^{rd} codon tree showed another distinct topology that was not supported by any other sequences. This strongly suggested that *cox1* and *nad4* 3^{rd} codon was saturated with substitutions that they were not informative in resolving phylogenetic relationship among those crab species.

The NJ trees based on concatenated 4-gene sequences of two sets of codons (Figure 3-10,Figure 3-11) were again showing two incongruent topologies. Each of them was agreed on with some other trees in Figure 3-7 and Figure 3-8.

In a summary, the topology A in which *P. trituberculatus* was more closely related to *C. sapidus* was supported by 12S single gene full sequences, 1st and 2nd codons of 4-gene concatenated sequences, 1st and 2nd codons of *nad2* single gene sequences, and 1st and 2nd codons of *nad4* single gene sequences; the topology B in which *P. trituberculatus* was more closely related to two *Scylla* species was supported by 4-gene all sites concatenated sequences, *nad2* single gene full sequences, *cox1* single gene full sequences, , 1st and 2nd codons of *cox1* single gene sequences, 3rd codons of 4-gene concatenated sequences, and 3rd codons of *nad2* single gene sequences.

3.3.4 Neutrality tests

4 neutrality statistics, Tajima's D, Fu and Li's D*, Fu and Li's F*, and Fu's Fs tests were performed using DnaSP on 90 blue crab sequences. All tests show these 4 genes are off selective neutrality (Table 3-3).

	nad2	*cox1*	*nad4*	12S	Total
TajimaD	-1.8861	-2.4478	-2.1775	-2.3599	-2.2483
SigD	*	**	**	**	**
FuLiD*	-3.4325	-4.5669	-3.8094	-4.7531	-4.4654
SigD	**	**	**	**	**
FuLiF*	-3.3489	-4.4414	-3.7573	-4.5762	-4.2177
SigF	**	**	**	**	**
FuFs	-83.113	-59.344	-66.479	-26.34	-33.92

Table 3-3. Neutrality tests results. *, P<0.05; **, P<0.02.

Nonsynonymous and synonymous mutations were counted for three coding genes within 90 blue crab sequences as well as among blue crabs and other species.

The values of mutations and dN/ds are shown in Table 3-4. Although three coding genes all had dN/ds<<1, *cox1*'s dN/ds was significantly smaller than the values of *nad2* and *nad4* ($p<0.01$).

gene	sequences included (N)	sites included	segregating sites percentage	Nsyn. mut.	Syn. mut.	dN/ds
nad2	*C. sapidus* (90)	624	14.9%	19	77	0.247
	+ *P. trituberculatus* (91)	609	33.5%	73	155	0.471
	+ *S. olivacea* (92)	516	37.6%	58	170	0.341
	+ *S. serrata*(93)	483	38.5%	49	172	0.285
	+ *E. sinensis* (94)	378	42.3%	54	146	0.370
cox1	*C. sapidus* (90)	675	9.5%	4	64	0.063
	+ *P. trituberculatus* (91)	672	20.2%	8	150	0.053
	+ *S. olivacea* (92)	666	26.0%	8	203	0.039
	+ *S. serrata*(93)	660	27.9%	8	225	0.036
	+ *E. sinensis* (94)	651	32.3%	18	255	0.071
nad4	*C. sapidus* (90)	744	13.3%	17	83	0.205
	+ *P. trituberculatus* (91)	735	24.9%	41	155	0.265
	+ *S. olivacea* (92)	705	28.9%	52	181	0.287
	+ *S. serrata*(93)	687	32.0%	50	212	0.236
	+ *E. sinensis* (94)	627	36.7%	60	226	0.265

Table 3-4. The ratio of nonsynonymous and synonymous mutations within *C. sapidus* and among species.

3.3.5 Population differentiation

Pairwise Fst statistics were conducted using the program Arlequin, results being shown in Table 3-5. Only RR03 and RR04 showed significant population differentiation, at the

genes of *nad2* ($p<0.02$) and *nad* 4($p<0.1$). Both *cox1* and 12S failed to detect any population difference among the 3 Rhode River samples.

The *nad2* marker also showed a possible differentiation between RR04 and RR05 samples, although p value was not statistically significant (0.10254) at 0.05 level.

		nad2	*cox1*	*nad4*	12S
RR03: RR04	Fst	0.05255	0.01227	0.01657	0.00223
	P	0.00293±0.0016	0.13281±0.0135	0.08594±0.0080	0.35645±0.0144
	SigP	**		#	
RR03: RR05	Fst	0.01121	0.00540	-0.01259	-0.01139
	P	0.16992±0.0109	0.24609±0.0142	0.85742±0.0094	0.84473±0.0129
	SigP				
RR04: RR05	Fst	0.01523	0.01167	0.01015	-0.00034
	P	0.10254±0.0082	0.14648±0.0121	0.18555±0.0138	0.41797±0.0155
	SigP				

Table 3-5. Pairwise Fst and P values for three groups of samples. #, P<0.1; *, P<0.05; **, P<0.02. The sample sizes for RR03, RR04, and RR05 were respectively 30, 33, and 26.

3.4 Discussion

All 4 genes showed very high variability for the blue crabs. Even 12S sequences among blue crab individuals varied a lot; among 90 crabs, there were 28 distinct haploptes for 12S region. For more variable regions as *nad2* and *nad4*, the haplotype diversity was almost 1. The *nad2* gene is located close to the presumed replication origin of the blue crab's mitochondrial genome, which could contribute to the hypervariability of this gene. Similarly, the *nad4* gene, being on the light strand of the genome, is located close to the

origin of the light strand. But this apparently cannot fully explain the extreme diversity of all the 4 genes. And the hot spot in the latter half of 12S gene is obviously far from both of the replication origins.

The phylogenetic analysis reveals two types of topologies: the topology A in which *P. trituberculatus* was more closely related to *C. sapidus* and the topology B in which *P. trituberculatus* was more closely related to two *Scylla* species.

It is strongly suggested the 3rd codons in the three coding genes were saturated with substitutions and so it may not be of help to include them to resolve the phylogenies among these crab species.

By excluding all 3rd codons, most trees agreed on the topology A with each other. Those trees included the NJ trees based on the 1st and 2nd codons in 4-gene concatenated sequences and in *nad2*, *nad4* single gene sequences, and 12S whole sequences. The only exception was the 1st and 2nd codons in *cox1* gene, which supported topology B. But in the neutrality test based on dN/dS rations, it was showed that *cox1* was under stronger selection pressure compared with *nad2* and *nad4*. This may explain why the 3rd codons in cox1 lost all information on the phylogeny resolution among three *Portunidae* species. Furthermore, it may also explain the incongruence between the 1st and 2nd codon tree on *cox1* and other genes.

Figure 3-1. Sampling location.

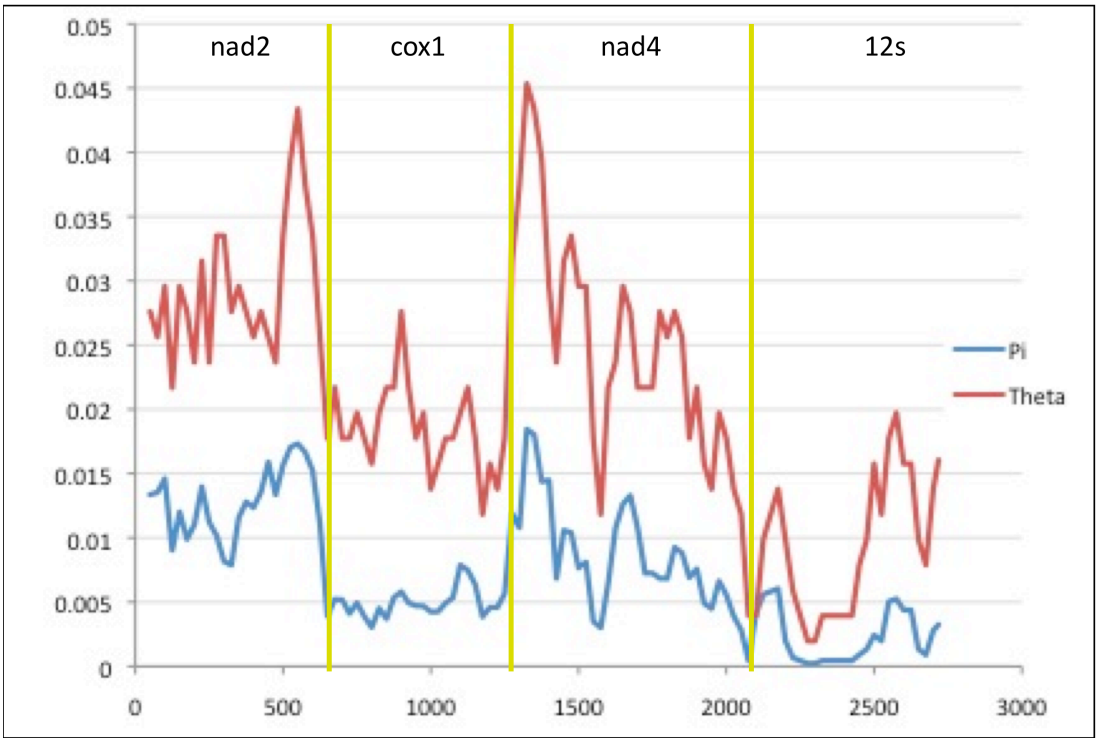

Figure 3-2. The Pi and Theta (per site) values across the concatenated DNA sequence. The X-axis is the concatenated DNA sequence in unit of base pair(bp). The region from position 1-627 was *nad2*, 628-1305 being *cox1*, 1306-2049 being *nad4*, and 2050-2761 being 12S. The size of sliding window was 100 bp and the step size was 25 bp.

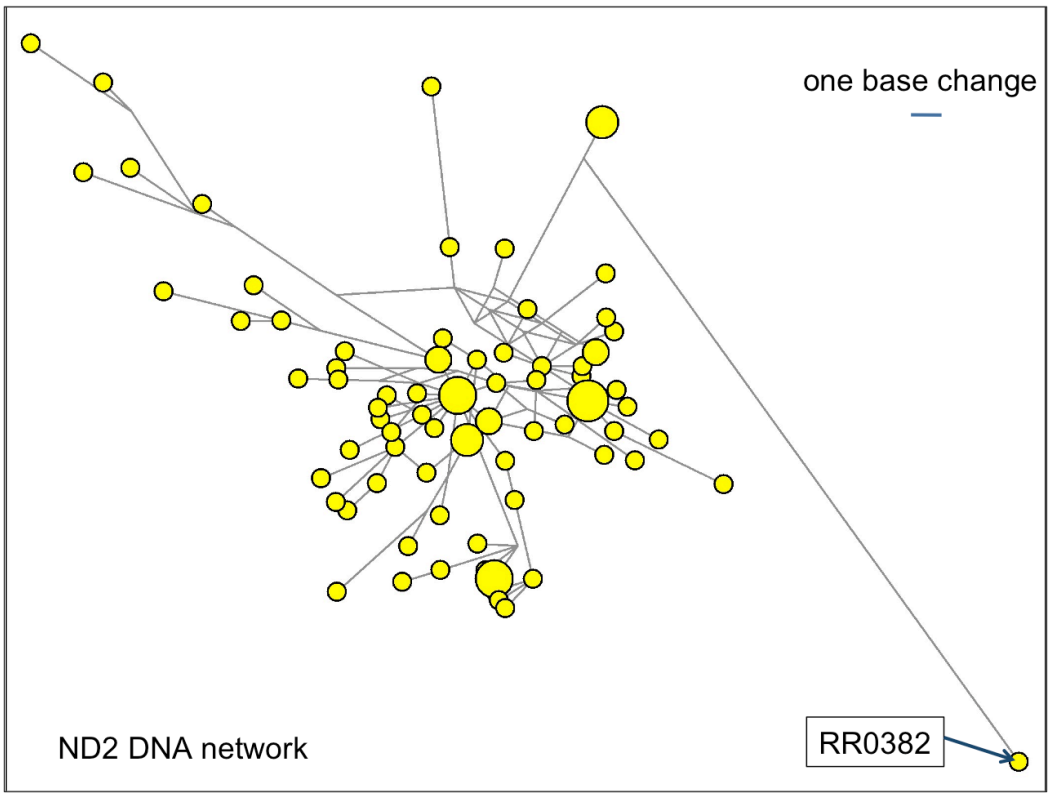

Figure 3-3. The *nad2* minimum spanning network. Each circle stands for one distinct haplotype and the size of circles is proportional to the frequency of the haplotype. The branch length indicates the genetic distance between pair of haplotypes with the scale bar denoting the branch length of one base change.

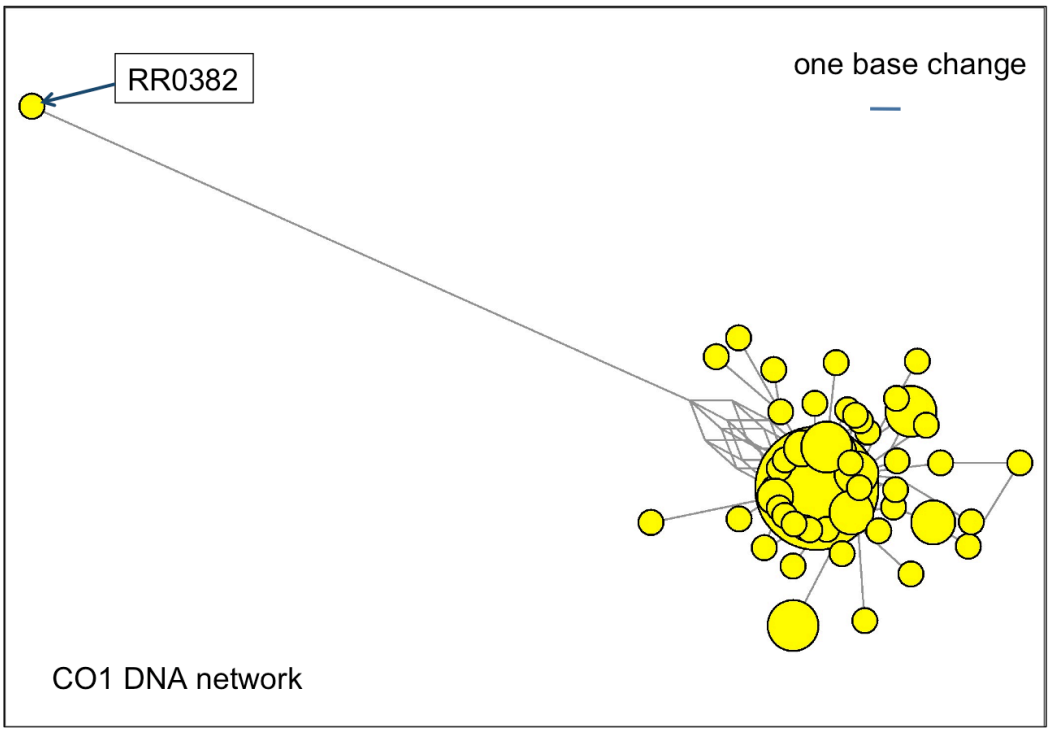

Figure 3-4. The *cox1* minimum spanning network. Each circle stands for one distinct haplotype and the size of circles is proportional to the frequency of the haplotype. The branch length indicates the genetic distance between pair of haplotypes with the scale bar denoting the branch length of one base change.

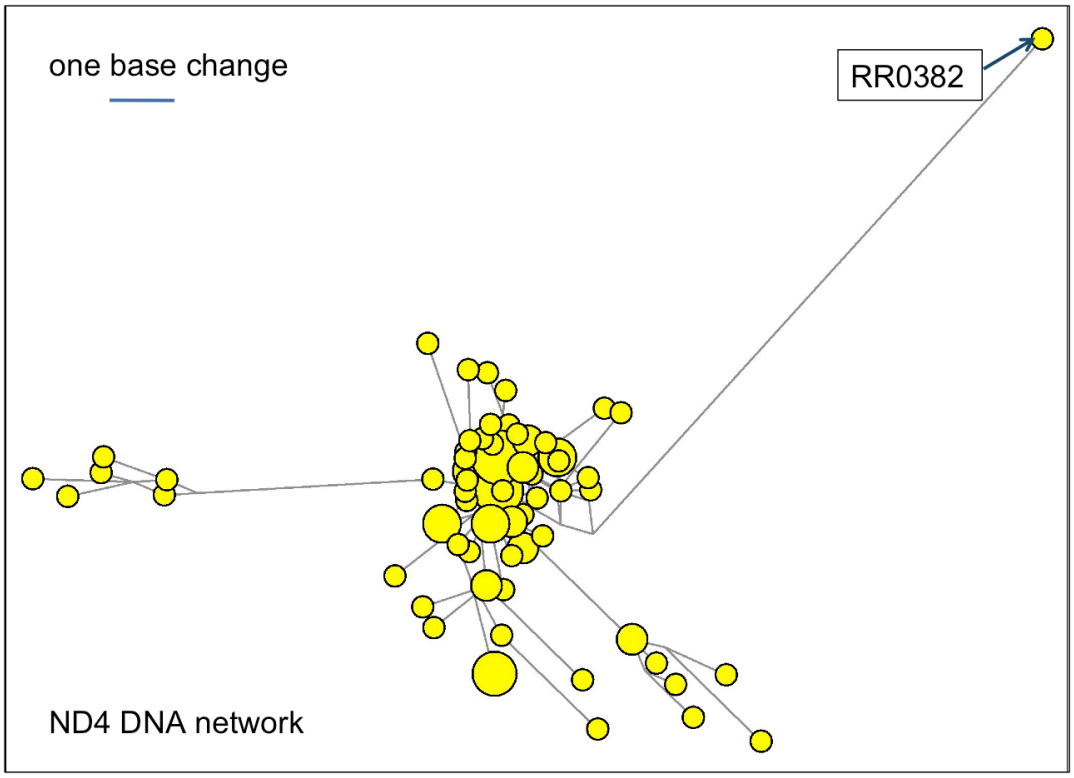

Figure 3-5. The *nad4* minimum spanning network. Each circle stands for one distinct haplotype and the size of circles is proportional to the frequency of the haplotype. The branch length indicates the genetic distance between pair of haplotypes with the scale bar denoting the branch length of one base change.

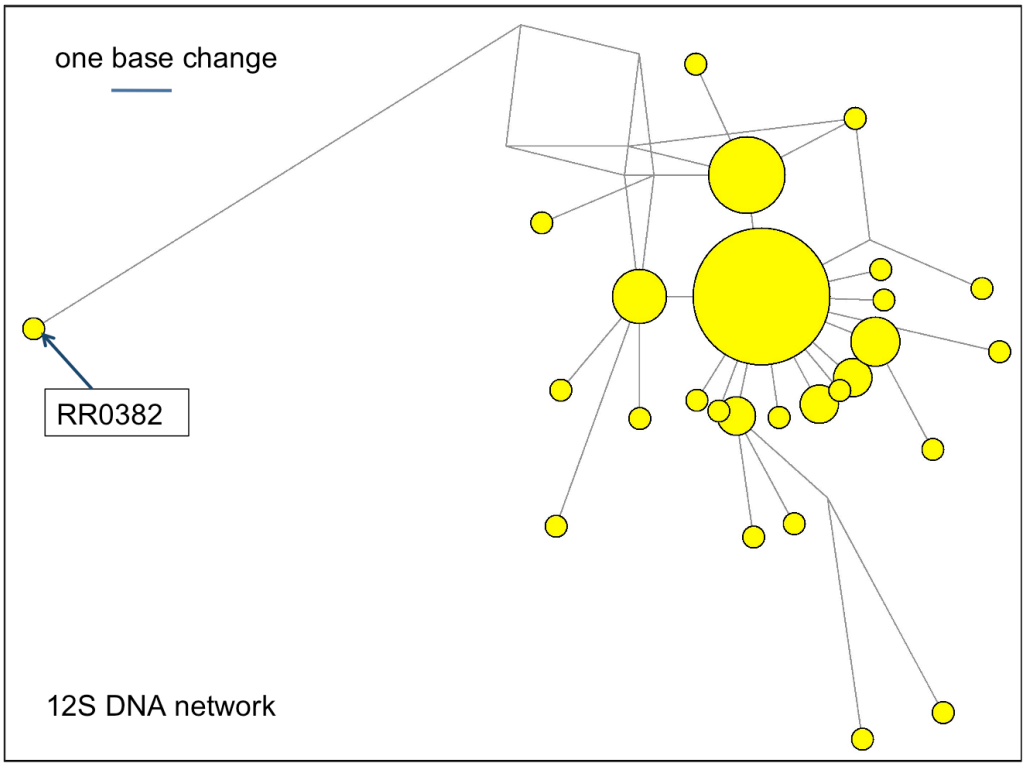

Figure 3-6. The 12S minimum spanning network. Each circle stands for one distinct haplotype and the size of circles is proportional to the frequency of the haplotype. The branch length indicates the genetic distance between pair of haplotypes with the scale bar denoting the branch length of one base change.

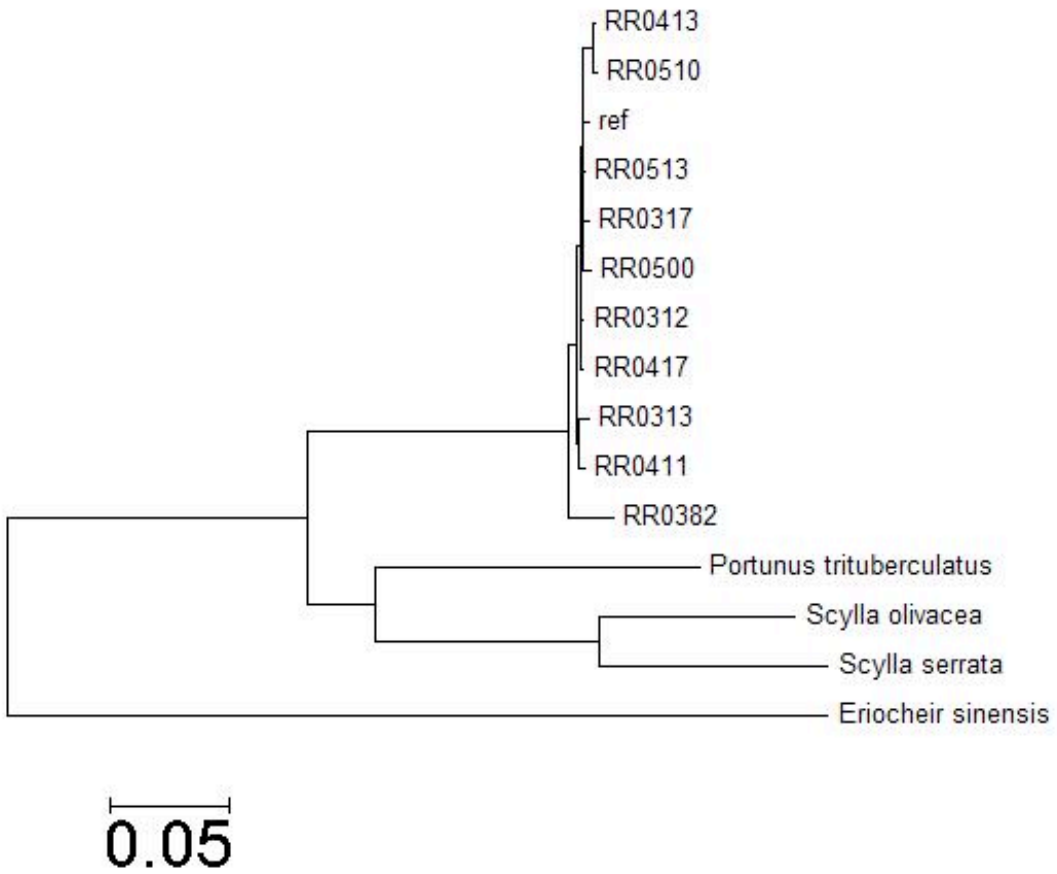

Figure 3-7. NJ tree on concatenated sequences of 4 genes including all sites. RR samples were crabs collected from Rhode River area; ref was the respective sequence of the completed genome.

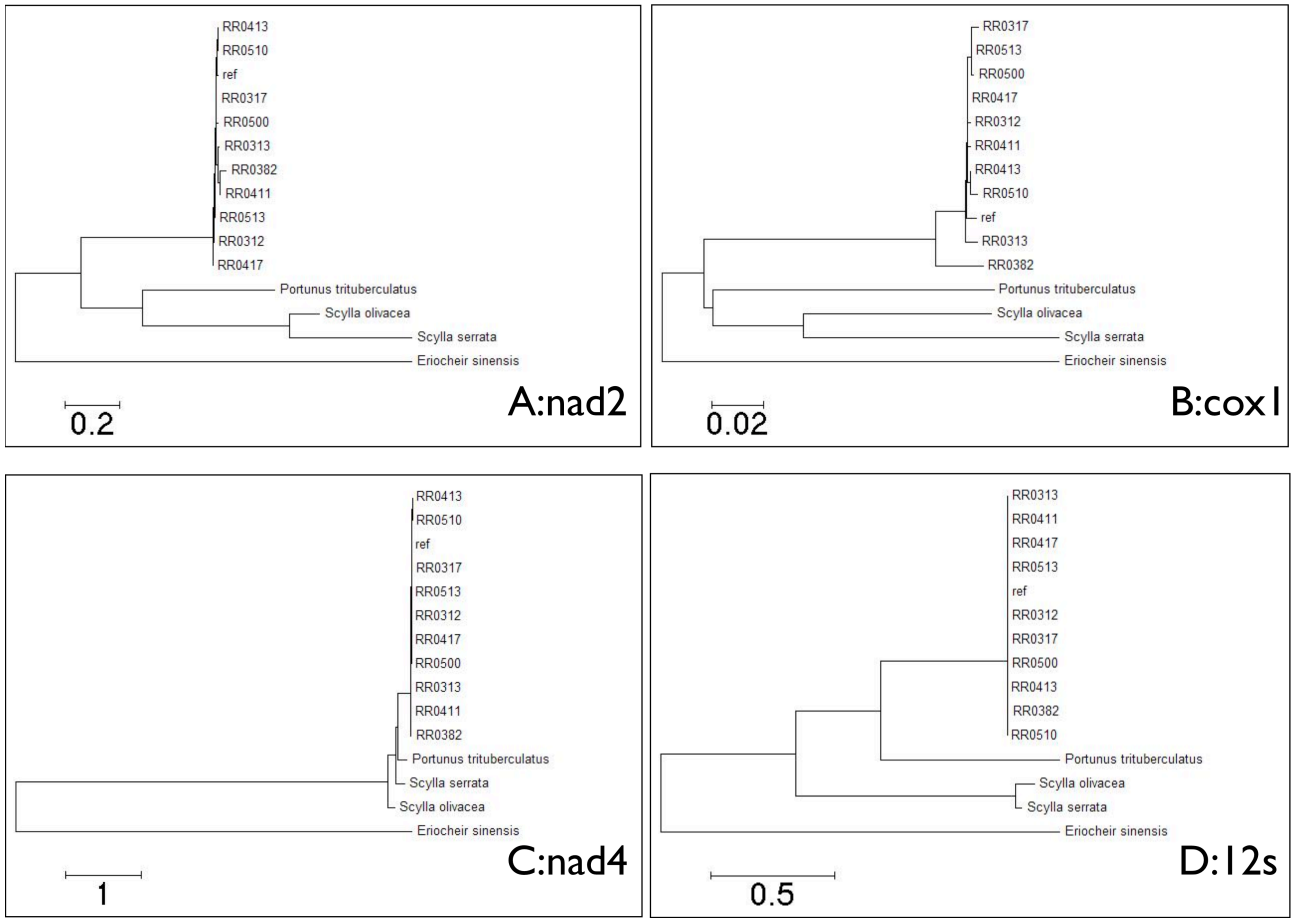

Figure 3-8. NJ trees of 4 genes including all sites. RR samples were crabs collected from Rhode River area; ref was the respective sequence of the completed genome.

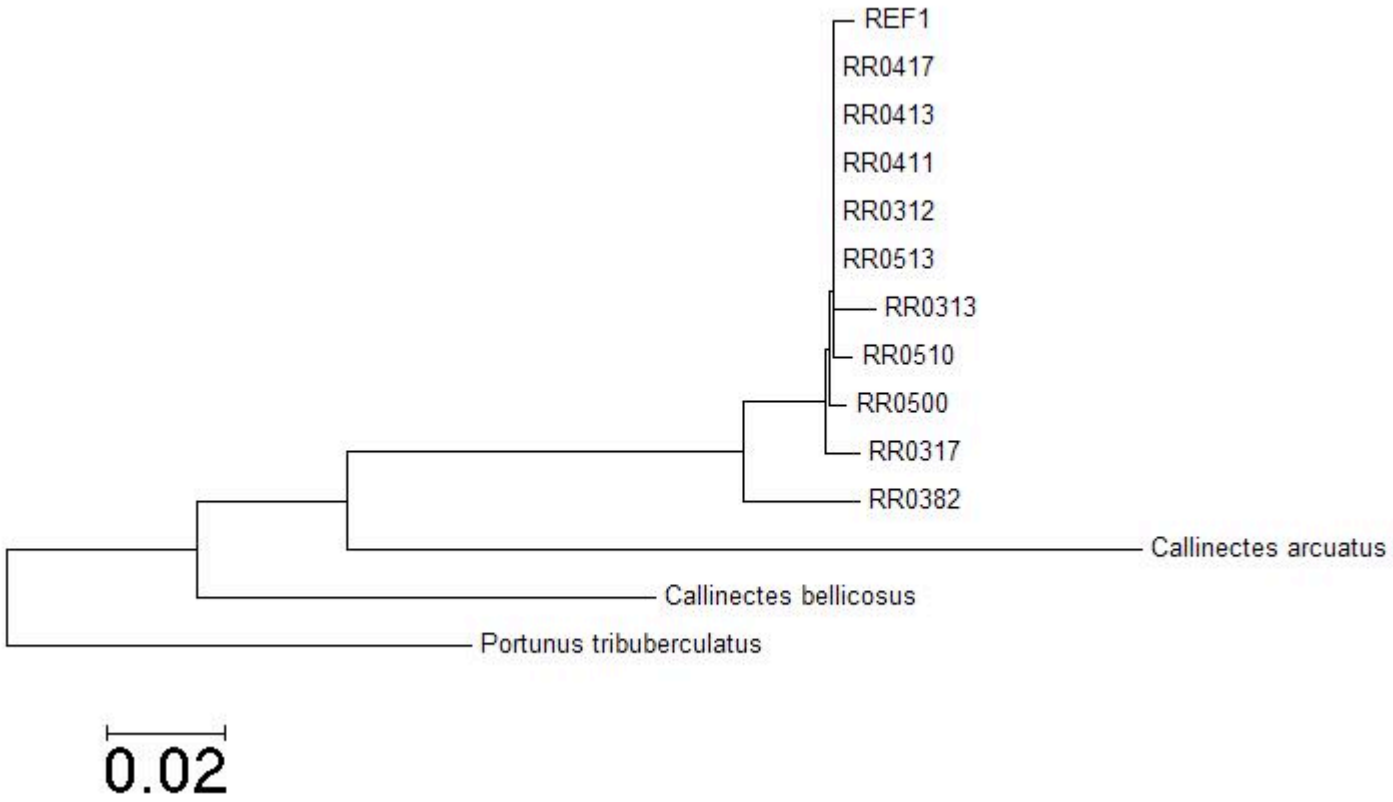

Figure 3-9. NJ tree on partial *cox1* gene. All sites (268bp) were included. RR samples were crabs collected from Rhode River area; ref was the respective sequence of the completed genome.

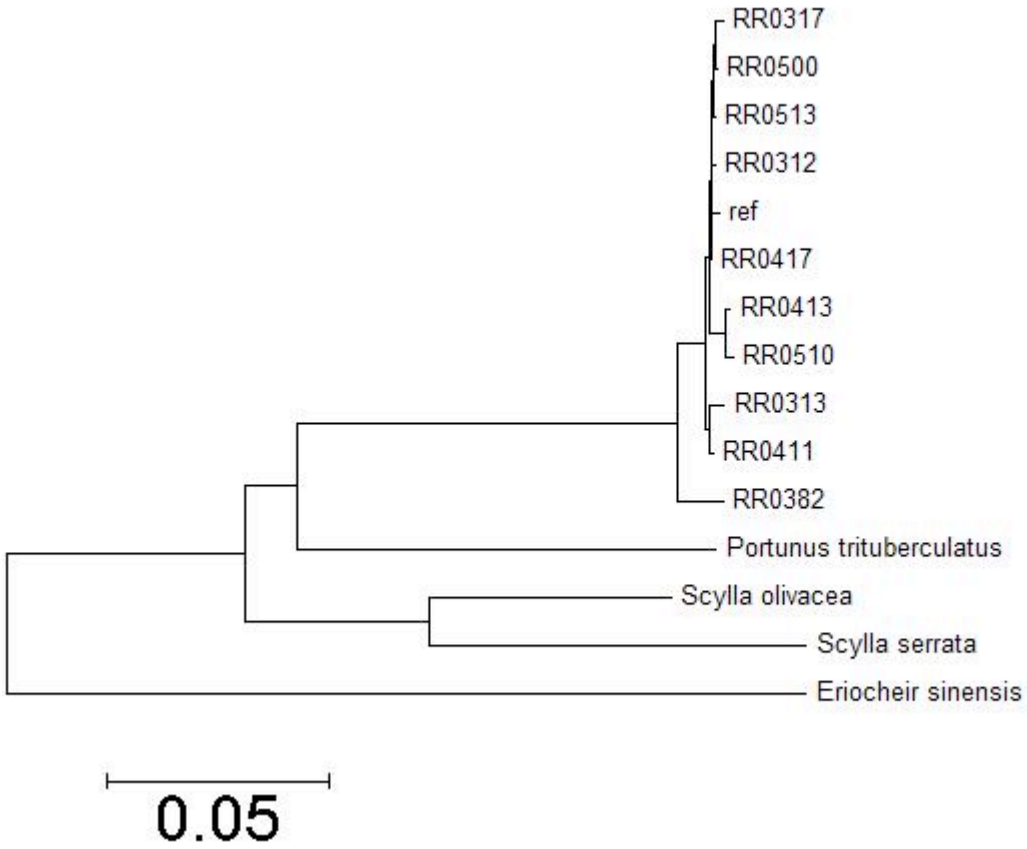

Figure 3-10. NJ tree 1st and 2nd codons at 4 gene concatanated sequences. RR samples were crabs collected from Rhode River area; ref was the respective sequence of the completed genome.

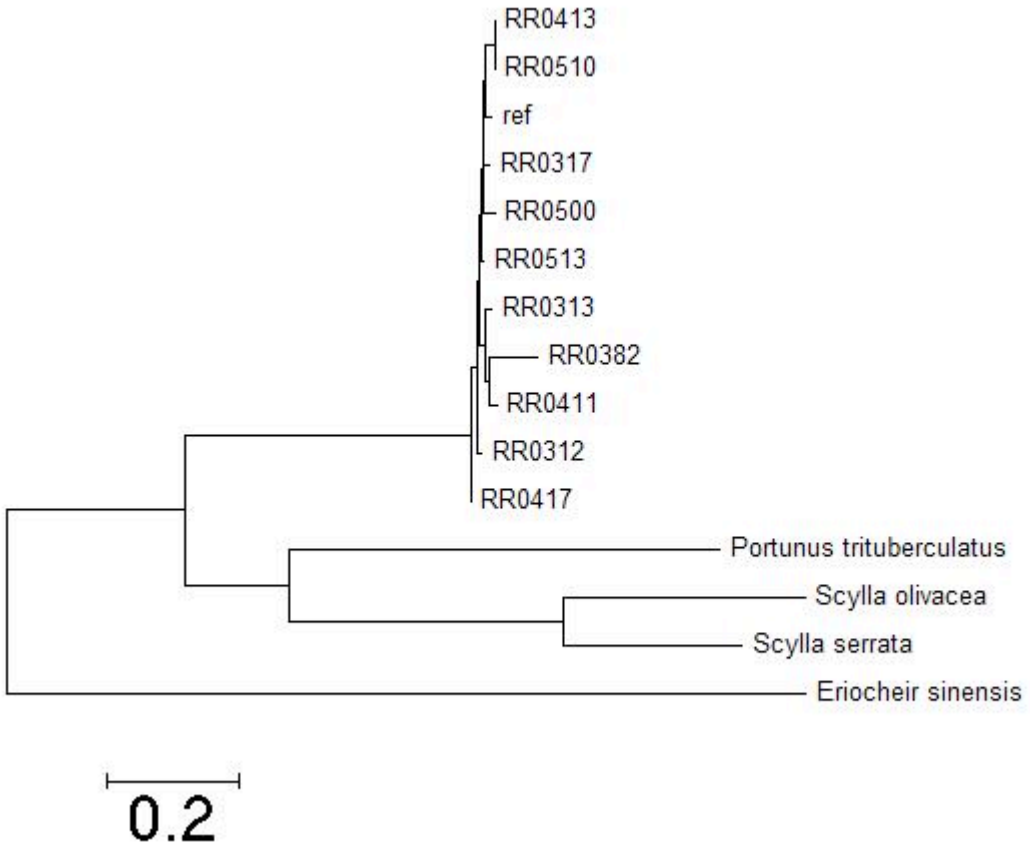

Figure 3-11. NJ tree 3rd codons at 4-gene concatenated sequences. RR samples were crabs collected from Rhode River area; ref was the respective sequence of the completed genome.

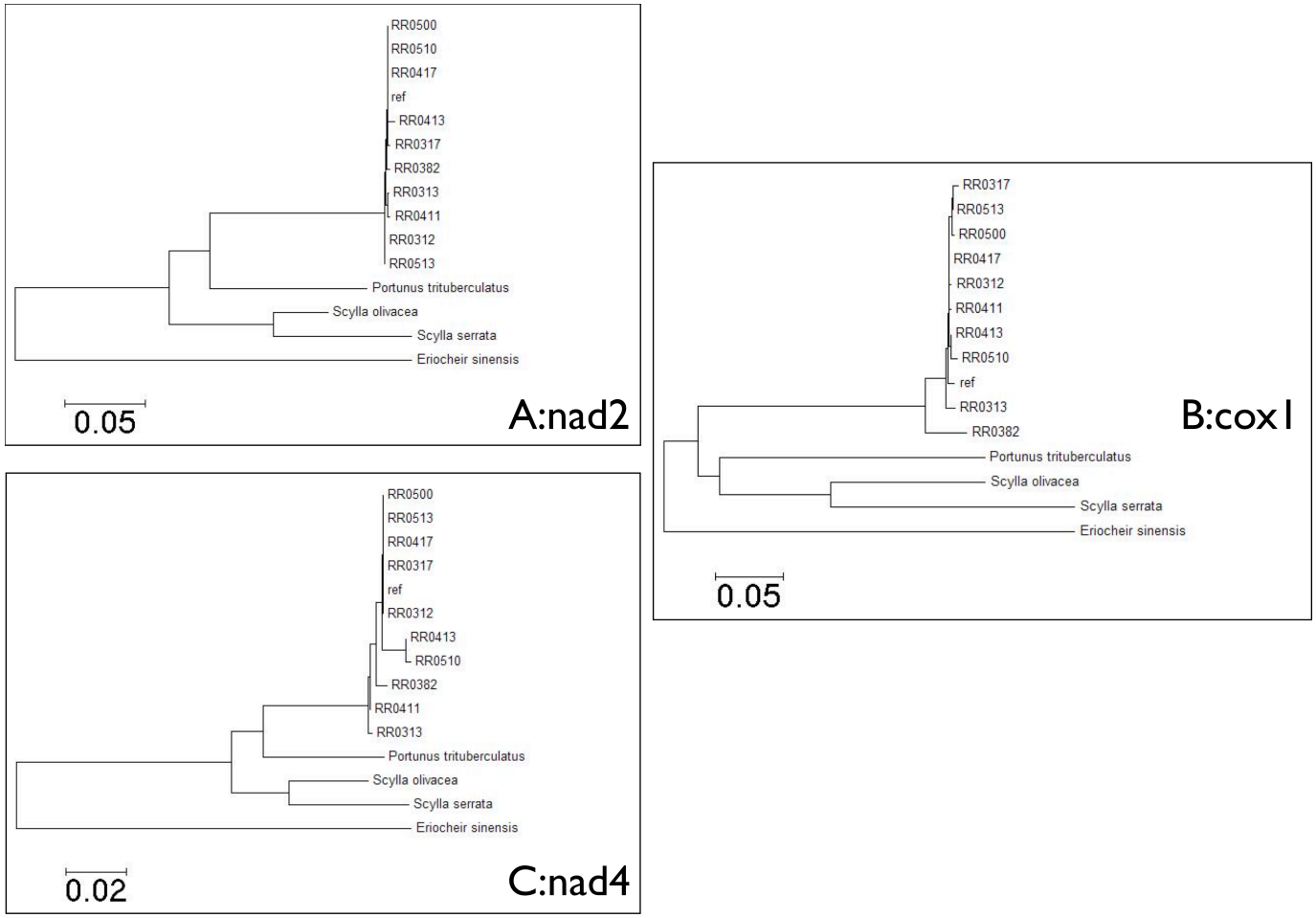

Figure 3-12. NJ trees on 1st+2nd codons of 3 coding genes. RR samples were crabs collected from Rhode River area; ref was the respective sequence of the completed genome.

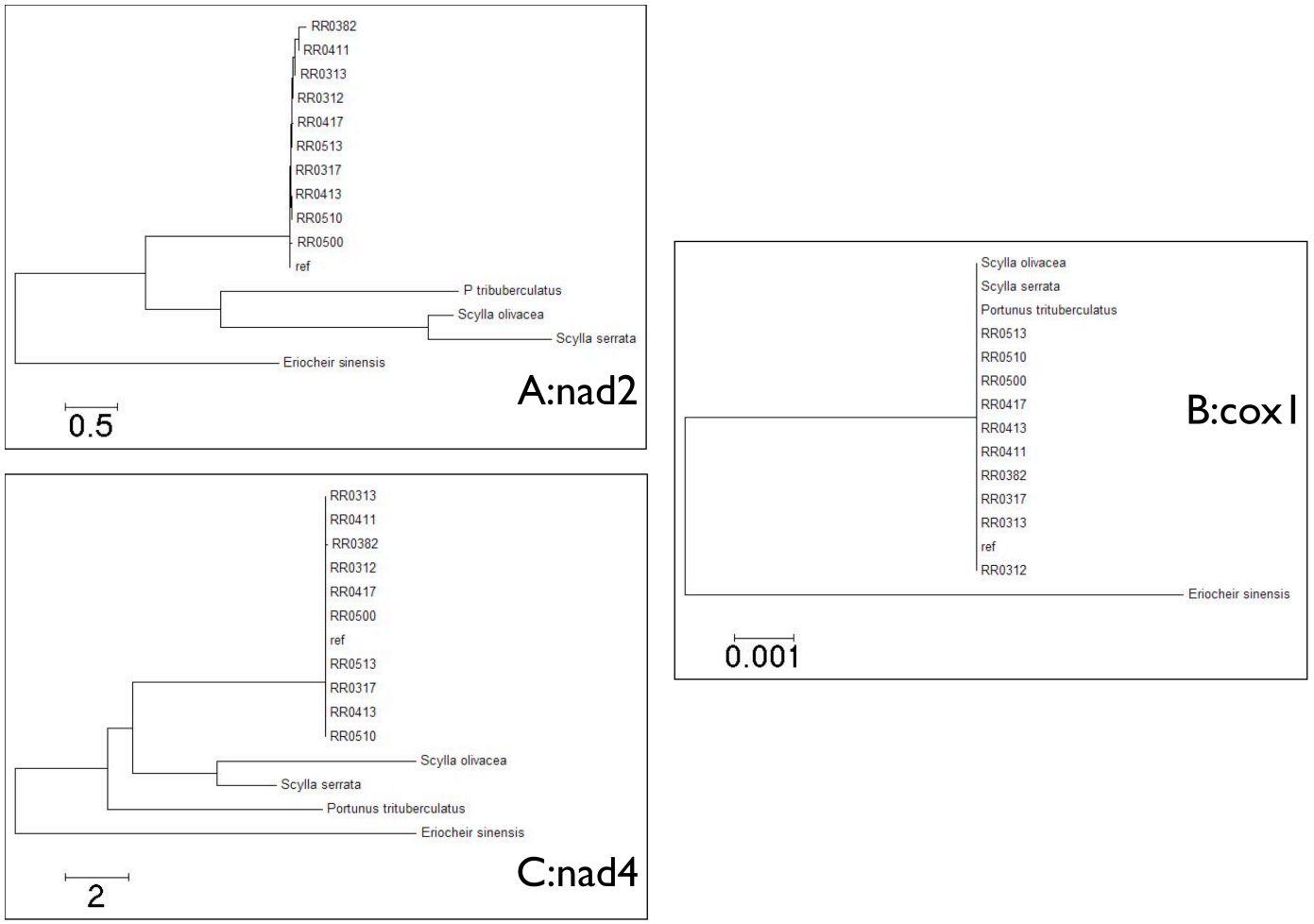

Figure 3-13. NJ trees on 3rd codons of 3 coding genes. RR samples were crabs collected from Rhode River area; ref was the respective sequence of the completed genome.

4 The population genetics of the blue crab: from the Atlantic coast to the Gulf of Mexico

4.1 Introduction

4.1.1 The biology

The blue crab (*Callinectes sapidus*) is a widely distributed estuarine and coastal marine crustecean which ranges from the western Atlantic in Nova Scotia to the northern Argentina coast (Williams 1974; Williams 1984). It lives in tropical, subtropical, and temperate zones but its presence in the far northern and southern borders is seasonal (Williams 1984; McMillen-Jackson and Bert 2004). The blue crab is an important link in the food chain, feeding on fish, aquatic vegetation, mollusks, crustaceans, and annelids while they serve as prey to mammals, birds, and fishes. The generation time of the blue crab is approximately one year as the life span is generally 2-3 years with exceptions of up to 6 years (Pellegrin, Guillory et al. 1999). It grows to adult size with 9.9–23.8 cm carapace width, after 18 to 20 molts (Ryer, vanMontfrans et al. 1997).

Adult blue crabs usually exhibit restricted movement within an estuary or inshore waters as they generally inhabit the low salinity area of the estuaries (Williams 1984). Mature female crabs after mating migrate to high salinity regions to spawn (e.g. mouth of the estuary). The number of eggs produced by each female is estimated to be between 700,000 and five million (Williams 1984) . Development through seven to eight zoea stages and the postlarval megalopa stage takes 30 to 50 days, taking place in coastal waters outside the estuary. The fact that females release their eggs in the interface

between estuary and open ocean and the larvae spend several weeks as plankton indicates the possibility of high gene flow via currents and population mixing among estuaries. When the megalopa stage is reached, the crab can employ a vertical migration and selective tidal stream transport to migrate back into estuary (Forward and Rittschof 1994). Megalopae stay in the lower water column during the ebbing tidal cycles to avoid to be carried away from the shore and move to the surface in the onshore currents so as to be transported back to the shore.

4.1.2 The fishery

Year	Pounds	Year	Pounds
1970	26,515,200	1990	46,838,347
1971	27,605,100	1991	48,627,671
1972	25,056,500	1992	31,353,002
1973	21,052,100	1993	57,625,281
1974	26,481,900	1994	45,546,516
1975	25,917,900	1995	42,162,413
1976	20,903,200	1996	37,701,414
1977	21,322,900	1997	41,306,363
1978	17,459,000	1998	26,219,204
1979	25,782,000	1999	32,131,249
1980	26,451,775	2000	21,661,198
1981	59,694,803	2001	24,549,197
1982	43,662,425	2002	25,011,003
1983	52,470,315	2003	26,430,273
1984	48,770,764	2004	33,495,212
1985	56,416,858	2005	32,060,201
1986	47,339,456	2006	29,017,729
1987	43,874,202		
1988	42,994,686		
1989	43,268,875		

Table 4-1. Maryland's commercial harvest (Chesapeake Bay and Ocean combined) by calendar year, data for year(s) omitted are not currently available (From Maryland DNR website).

The blue crab is one of the most valuable fishery species. The total United States blue crab landings in 2002 were valued at more than $167 million; in the State of Maryland alone, the annual commercial landing of blue crabs is about 20 to 50 million pounds (Maryland DNR data). From the middle of 1990s, the blue crab landing has decreased in Maryland with a slight recovery since the year 2004 (Table 4-1).

The crab population size in Chesapeake Bay has been estimated via annual winter dredge survey conducted by the states of Maryland and Virginia (Table 4-2). A trend of population size reduction is clearly shown since the late 90s except for 2009.

Year	Millions of Crabs	Year	Millions of Crabs
1990	791	2000	281
1991	828	2001	254
1992	367	2002	315
1993	852	2003	334
1994	487	2004	280
1995	487	2005	415
1996	661	2006	324
1997	680	2007	260
1998	353	2008	283
1999	308	2009	418

Table 4-2. The estimated number of crabs living in the bay. Managers estimate the total number of crabs living in the Chesapeake Bay each year by multiplying the estimated density of all crabs by the area of the bay. Data are from Maryland DNR.

4.1.3 The genetic structure of the blue crab populations

Knowledge of the population genetics is necessary for the successful management of ecologically and economically important species. Surprisingly few studies of the blue crab *Callinectes sapidus* genetics have been reported despite this species' significance:

(Cole and Morgan 1978; Nelson and Hedgecock 1980; Kordos and Burton 1993; McMillen-Jackson, Bert et al. 1994; McMillen-Jackson and Bert 2004). Only McMillen-Jackson group's work in 2004 was done at the DNA sequence level, and their samples covered a large geographic range from Atlantic coast to the Gulf of Mexico. Their results showed a high diversity within samples and a lack of population differentiation between the east coast and the Gulf coast samples. However, their sample sizes were limited, with no more than 15 individuals being genotyped, typically around 10 crabs for each sampling.

4.2 Material and methods

4.2.1 Collecting blue crab samples

We collected crabs samples from three major regions: the mid-Atlantic states (including New Jersey, Maryland, Virginia and North Carolina), the Gulf of Mexico and Massachusetts (Figure 4-1). In the mid-Atlantic states, 9 samplings were done in Chesapeake Bay area from the year 2003 to 2007 (Figure 4-2); three other samples were collected in year 2005 from North Carolina and New Jersey (location, size, year, and abbreviation in Table 4-3). All 13 samples from the Gulf of Mexico were collected in year 2002 (Figure 4-4; Table 4-3). Samples in Massachusetts were collected in year 2008, from 5 locations (Figure 4-4; Table 4-5).

Region	Sampling location	Year	Sample size	Abbreviation
Chesapeake Bay states	Rhode River, MD	2003	41	RR03
		2004	47	RR04
		2005	26	RR05
	Potomac River, MD	2005	83	PTM05
	Patuxent River, MD	2005	73	PTX05
	South River, MD	2005	10	SR05
	York River, VA	2005	47	VIMS05
	Hooper's Island, MD	2006	49	HI06
		2007	90	HI07
Other mid-Atlantic states	Raleigh, NC	2005	35	NC05
	Hackensack Meadowlands, NJ	2005	16	NJHM05
	Tuckerton, NJ	2005	73	NJTK05

Table 4-3. Samples collected in mid-Atlantic coast states. The location, sample size, year, and abbreviation were indicated.

Region	Sampling location	Year	Sample size	Abbreviation
Gulf of Mexico	Port Charlotte, FL	2002	35	GOMPC
	Tampa, FL		35	GOMTPA
	Crystal River, FL		35	GOMCR
	Steinhatchee, FL		35	GOMSH
	Panacea, FL		35	GOMPN
	Apalachicola, FL		35	GOMAP
	Destin, FL		30	GOMDS
	Pensacola, FL		34	GOMCOLA
	Mobile, AL		29	GOMMDB
	Gautier, MS		35	GOMGAU
	Gulfport, MS		35	GOMGP
	Brownsville, TX		15	GOMBSL

Table 4-4. Samples collected from Gulf of Mexico. The location, sample size, year, and abbreviation were indicated.

Region	Sampling location	Year	Sample size	Abbreviation
Massachusetts	Westport River	2008	43	MAWTR
	Cotuit Bay		40	MACB
	Agawam River		37	MAAR
	Bass River		36	MABR
	Oyster Pond (Martha's Vineyard)		40	MAMV

Table 4-5. Samples collected from Massachusetts. The location, sample size, year, and abbreviation were indicated.

4.2.2 DNA extraction and PCRs

Crabs were preserved in 95% ethanol until the DNA extraction. About 100 mg muscle tissue for each crab was used for DNA extraction with Qiagen DNeasy Blood and Tissue Kit.

Mitochondrial PCRs were performed to amplify a 750-bp segment of *nad2* gene (NADH dehydrogenase subunit 2) of mtDNA, using primers F296 (5'-TGCTTTATTATTCAACCCCG-3') and R1046 (5'-CCGAATAGATTGATTGAAGT-3') designed according to the complete *C. sapidus* mitochondrial genome sequence (Place, Feng et al. 2005). PCRs were done in a final volume of 30μL, containing 15μL PCR 2X Master Mix (Promega), 10pmol of each primer, and 20ng of template DNA under the following cycling conditions: 94ºC for 30s; followed by 32 cycles at 94ºC for 15s, 48ºC for 30s, 68ºC for 1min; followed by 72ºC for 5min. The size of PCR products was estimated by electrophoresis in 1% (w/v) agarose gels.

4.2.3 DNA sequencing

PCR products were precipitated by adding equal volumes of 20% polyethylene glycol (PEG) 8000 and 2.5 M NaCl, incubating for 15min at room temperature, and centrifuging at 2,500g for 30min. The supernatant was removed and the DNA pellet was washed with 70% ethanol, followed by another centrifuge at 2,500g for 15min. The supernatant was decanted and the purified PCR products were air dried and dissolved with 15- 20µL water. The sequencing reaction was done in 10µL volume, consisting of 40-150ng PCR product, 3pmol of primer, 0.5µL Big Dye v3.1 sequencing mix and 1.5µL 5X sequencing buffer (Applied Biosystems). The cycling parameters were 95ºC for 5min; followed by 50 cycles at 95ºC for 15s, 50ºC for 15s, 60ºC for 4min. The sequencing product was cleaned by adding 60µL 100% isopropanol and 30µL H_2O, mixing thoroughly, incubating at room temperature for 30min, and centrifuging at 2000g for 30min. The supernatant was discarded and 100µL 70% isopropanol was added to wash the DNA, followed by another centrifugation at 2000g for 14min and the disposal of the supernatant. Cleaned DNA was air dried for 20-30min before 10µL formamide was added. The mixture was heated at 95ºC for 2min and then put on ice immediately. After it was chilled down, the denatured sequencing product was loaded into an ABI 3130xl genetic analyzer. For each PCR product, we conducted both forward and reverse sequencing.

4.2.4 Data analysis

We checked the quality of all DNA sequences and aligned them using the program SEQUENCHER®4.8 (Gene Codes Corporation) before exporting the consensus

sequences of forward and reverse sequences. The actual exported sequences were shorter than the PCR amplicon in length because only the DNA sequences unambiguously covered by two directions were used in alignments and then exported. The published *C. sapidus* mtDNA genome sequence was used as the reference sequence (Place, Feng et al. 2005).

The polymorphism diversity calculation and neutrality test were performed by using program DnaSP v5 (Librado and Rozas 2009).

The following measures were calculated for the haplotype/nucleotide diversity,:

S, the number of segregating (polymorphic) sites;

Haplotype (gene) diversity and its sampling variance (Nei 1987);

Nucleotide diversity, Pi (π), the average number of nucleotide differences per site between two sequences and its sampling variance (Nei 1987);

k, the average number of nucleotide differences between two sequences (Tajima 1983).

The free software Network 4.5.1.0. (fluxus-engineering.com) was used to produce minimum spanning network of all haplotypes with Median Joining calculation (Bandelt, Forster et al. 1995).

Population genetics analysis was done with software ARLEQUIN V3.11 (Excoffier, Laval et al. 2005). Haplotype number, haplotype diversity, and nucleotide diversity were calculated for all sequences combined and separately for each collected sample.

Population pairwise Fst and P value of Fst based on 1000 permutation were calculated to detect population differentiation. AMOVA was performed with ARLEQUIN as well to detect possible population subdivisions.

Mismatch distributions were computed in which the numbers of pairwise differences between haplotypes were plotted according to their relative frequencies to determine if either species had experienced a sudden population expansion (Slatkin and Hudson 1991; Rogers and Harpending 1992). The sum of squared deviations (SSD) between the observed and the expected distributions was used as a test statistic, and the significance was assessed after 1000 bootstraps.

To look into the phylogenetic relationship of haplotypes from different regions, a Bayesian tree was constructed using the program MRBAYES (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). Thirty randomly picked sequences from each pool of 2005 mid-Atlantic, 2002 Gulf of Mexico, and 2008 Massachusetts samples were used for Bayesian inference. 500000 generations and 4 chains were employed; all characters weighed equally.

4.3 Results

In total, 1153 wild blue crabs from 30 samplings were sequenced in a 627bp fragment of *nad2* gene in mitochondrial genome on both forward and reverse directions. No stop codon or insert/deletion was observed in the DNA sequences. The average G+C content was 32.3%.

4.3.1 DNA sequence polymorphism

Among 1153 sequences for the 627 bp of the *nad2* locus examined, there were 259 variable sites, 296 mutations, 102 singleton variable sites, and 157 parsimony informative sites. There were 62.7%, 24.2%, and 10.6% of nucleotide sites variable respectively at 3^{rd}, 2^{nd}, and 1^{st} position of the codon. There were 623 distinct haplotypes. The Hd (haplotype diversity) was 0.9876, Pi (nucleotide diversity) 0.01223, k (average number of nucleotide difference between pairs of sequences) 7.642.

Polymorphism indices within each sample including numbers of substitutions, segregating sites, transversion, transition, and average pairwise distance were calculated for each sample (Table 4-6).

The number of substitutions was very close to number of segregating sites, indicating most segregating sites represented only one type of mutation. The ratios of substitution vs. sample size and segregating site vs. samples size were relatively stable for most samples (Figure 4-5), except for three samples that all had significantly smaller sample size (around 10, in contrast to usual sample size of 30 or more). When plotted against the sample size, the number of segregating sites seemed to reach a plateau (Figure 4-6). Plotting the mutation number against sample size, it seemed new mutations kept emerging along with newly included sequences (Figure 4-7). There was not any tendency for site saturation at least for this dataset.

The transversion/transition values varied from around 0.03(NC05 and HI06) to nearly 0.15 (GOMMDB and HI07). It didn't show any geographic pattern as two years of HI

samples showed both extreme high and low ratios (Figure 4-8). The number of private segregating sites showed a weak correlation with sample size (Figure 4-9).

Haplotype diversity (Hd), nucleotide diversity, and average pairwise difference (k) were calculated for each sample (Table 4-7). In general, all samples showed the gene diversity (haplotype diversity) close to 1.0, indicating the diversified genetic composition of each sample. The plot of number of haplotypes against sample size (Figure 4-10) exhibited a significant linear correlation between them with R^2>0.98.

The nucleotide diversity varied from 0.008799 (08MACB) to 0.015654 (NC05). And 08MACB was the only one sample having Pi value lower than 0.01 (Figure 4-11) and k value lower than 6 (Figure 4-12).

sample	N	sub	S	ts	tv	tv/ts	pri. S	k
RR03	41	70	68	64	6	0.094	2	9.262
RR04	47	61	59	54	7	0.130	1	7.736
RR05	26	48	48	42	6	0.143	4	6.705
PTM05	83	104	98	95	9	0.095	7	8.684
PTX05	73	71	67	63	8	0.127	2	6.782
SR05	10	31	31	28	3	0.107	1	7.444
VIMS05	47	73	69	66	7	0.106	5	8.042
HI06	49	60	59	58	2	0.034	3	6.969
HI07	90	92	87	80	12	0.150	11	8.154
NC05	35	65	63	63	2	0.032	2	9.815
NJHM05	16	30	30	28	2	0.071	0	6.742
NJTK05	73	68	65	63	5	0.079	3	6.638
GOMBSL	15	37	36	34	3	0.088	1	7.257
GOMCOLA	34	46	45	43	3	0.070	1	8.793
GOMCR	35	58	55	51	7	0.137	5	7.612
GOMDS	30	56	56	53	3	0.057	5	7.738
GOMGAU	35	50	48	46	4	0.087	0	7.434
GOMGP	35	47	46	44	3	0.068	3	6.718
GOMMDB	29	48	47	42	6	0.143	2	7.251
GOMPC	35	52	51	45	7	0.156	2	7.311
GOMPN	36	55	54	49	6	0.122	5	6.517
GOMSH	35	61	60	56	5	0.089	7	8.739
GOMTPA	10	33	31	29	4	0.138	0	7.867
GOMAP	35	47	45	43	4	0.093	1	6.528
08MAAR	37	55	54	50	5	0.100	5	6.914
08MABR	36	50	49	47	3	0.064	5	6.637
08MACB	40	50	49	47	3	0.064	0	5.517
08MAMV	40	60	59	55	5	0.091	5	8.208
08MAWTR	43	65	63	59	6	0.102	6	8.599

Table 4-6. Polymorphism indices of all samples. N, sample size; sub, number of substitutions; S, number of variable sites; ts, number of transistions and tv, number of transversions; pri. S, number of S only occurring within one sample.

sample	Hd	Pi	k
RR03	1.0000 +/- 0.0054	0.014772 +/- 0.007698	9.262
RR04	1.0000 +/- 0.0044	0.012339 +/- 0.006496	7.736
RR05	1.0000 +/- 0.0107	0.010693 +/- 0.005805	6.705
PTM05	1.0000 +/- 0.0019	0.013850 +/- 0.007158	8.684
PTX05	1.0000 +/- 0.0023	0.010816 +/- 0.005713	6.782
SR05	1.0000 +/- 0.0447	0.011873 +/- 0.006859	7.444
VIMS05	1.0000 +/- 0.0044	0.012826 +/- 0.006731	8.042
HI06	1.0000 +/- 0.0041	0.011115 +/- 0.005898	6.969
HI07	1.0000 +/- 0.0017	0.013004 +/- 0.006746	8.154
NC05	1.0000 +/- 0.0068	0.015654 +/- 0.008161	9.815
NJHM05	1.0000 +/- 0.0221	0.010752 +/- 0.005993	6.742
NJTK05	1.0000 +/- 0.0068	0.011455 +/- 0.006100	6.638
GOMBSL	1.0000 +/- 0.0243	0.011574 +/- 0.006441	7.257
GOMCOLA	1.0000 +/- 0.0071	0.014024 +/- 0.007374	8.793
GOMCR	1.0000 +/- 0.0068	0.012140 +/- 0.006449	7.612
GOMDS	1.0000 +/- 0.0086	0.012341 +/- 0.006581	7.738
GOMGAU	1.0000 +/- 0.0068	0.011856 +/- 0.006311	7.434
GOMGP	1.0000 +/- 0.0068	0.010714 +/- 0.005754	6.718
GOMMDB	1.0000 +/- 0.0091	0.011565 +/- 0.006208	7.251
GOMPC	1.0000 +/- 0.0068	0.011660 +/- 0.006215	7.311
GOMPN	1.0000 +/- 0.0065	0.010395 +/- 0.005593	6.517
GOMSH	1.0000 +/- 0.0068	0.013939 +/- 0.007326	8.739
GOMTPA	1.0000 +/- 0.0447	0.012547 +/- 0.007216	7.867
GOMAP	1.0000 +/- 0.0068	0.010411 +/- 0.005606	6.528
08MAAR	1.0000 +/- 0.0063	0.011028 +/- 0.005897	6.914
08MABR	1.0000 +/- 0.0065	0.010585 +/- 0.005686	6.637
08MACB	1.0000 +/- 0.0056	0.008799 +/- 0.004800	5.517
08MAMV	1.0000 +/- 0.0056	0.013090 +/- 0.006886	8.208
08MAWTR	1.0000 +/- 0.0050	0.013715 +/- 0.007176	8.599

Table 4-7. Haplotype diversity (Hd), nucleotide diversity (Pi), and average pairwise distance (k) of all samples.

4.3.2 Population genetic structure

Mid-Atlantic samples

The frequency of non-singleton haplotypes for each sample occurring in the upper Chesapeake Bay in 2005 is presented in Table 4-8. The single most prevalent haplotype (haplotype 1) was shared by approximately 6.7% (13/192) of all individuals examined, represented by 3 samples with the exception of PTM05. The second and third most common haplotypes were shared by 4.2% and 3.6% individuals, respectively, and present in all 4 samplings. There were only 18 haplotypes that occurred more than once with 70 crabs, and 122 haplotypes that were singletons.

Haplotype (occurrence)	RR05	PTM05	PTX05	SR05
1(13)	3	0	9	1
2(8)	1	6	1	0
3(7)	1	4	1	1
4(5)	1	2	2	0
5(4)	0	2	2	0
6(4)	0	1	3	0
7(3)	0	2	0	1
8(3)	0	3	0	0
9(3)	0	2	0	1
10(3)	0	1	2	0
11(3)	1	0	2	0
12(3)	0	0	3	0
13(2)	0	2	0	0
14(2)	0	2	0	0
15(2)	0	0	2	0
16(2)	0	1	1	0
17(2)	0	1	1	0
18(2)	0	1	1	0

Table 4-8. Frequency of shared haplotypes within 2005 upper bay samples

	RR05	PTM05	PTX05	SR05
RR05		0.14648±0.0118	0.62988±0.0159	0.59570±0.0167
PTM05	0.00886		0.13184±0.0114	0.35352±0.0162
PTX05	0.00354	0.00493		0.79102±0.0132
SR05	0.00803	0.00209	-0.01715	

Table 4-9. Population pairwise Fsts (below diagonal) and P values (above diagonal) among 4 2005 upper bay samples. No significant Fst values was detected at 0.05 level for 1000 permutations.

The population pairwise Fsts (Table 4-9) showed there was no any significant differentiation between any pair of samples.

The DNA minimum spanning network of all haplotypes, showed most of haplotypes were closely related except that there were 6 haplotypes carried by10 individuals which were more distant (Figure 4-13). These 6 haplotypes were represented in all 4 samplings. Among all other haplotypes there was no implied lineage sorting based on geography. This MSN (Figure 4-13) further showed that all population samples collected from 2005 upper bay area had similar genetic composition without any detectable differentiation.

There was no detectable difference between upper bay and lower bay population samples based on population pairwise Fst test (Table 4-10).

VIMS05	RR05	CB05PTM	CB05PTX	CB05SR
Pairwise Fst	-0.00721	-0.00247	-0.00413	-0.00877
P-value	0.73242±0.0167	0.58496±0.0145	0.73047±0.0130	0.58496±0.0164

Table 4-10. Population pairwise Fsts and P values between VIMS05 and 2005 upper bay samples. No significant Fst values was detected at 0.05 level for 1000 permutations.

Table 4-11 showed population pairwise Fsts among all mid-Atlantic coast samples and Table 4-12 showed the respective P values based on 1000 permutations. The population sample collected in Rhode River area in year 2003 (RR03) showed significant difference from all other samples except RR05 and SR05. Among all other samples, only NC05 and NJTK05 showed a significant Fst value against each other.

Gulf of Mexico samples (Table 4-13 and Table 4-14) showed population pairwise Fst values and respective P values between all the Gulf of Mexico samples. Only 4 significant Fst values were observed, all involving the sample GOMCOLA (collected in Pensacola, FL).

Massachusetts samples

Table 4-15 and Table 4-16 showed population pairwise Fst values and respective P values among all Massachusetts samples. All 5 samples were undistinguishable except for the pair of 08MACB (collected in Cotuit Bay) and 08MAWTR (collected in Westport River).

Among three areas

Mismatch analysis produced a distribution of pairwise haplotype differences for each of pooled 2005 Chesapeake Bay, 2002 Gulf of Mexico, and 2008 Massachusetts samples (Figure 4-14, Figure 4-15 and Figure 4-16). The average difference varied from 7 to 8 with a variance of 15. The three mismatch plots generally showed a unimodal curve but

02GOM and 08MA both had a much lower secondary peak between 10-15.

AMOVA did not detect any significant population subdivision. 99.5% of the genetic variation occurred within samples. There was no assumed Atlantic/Gulf split. In the Bayesian tree (Figure 4-17), haplotypes from mid-Atlantic, Massachusetts, and Gulf collections were dispersed on all branches of the tree. There was no clustering according to the geographic proximity. My dataset supports the contention that the blue crab populations lacked strong geographical structuring.

	RR03	RR04	RR05	SR05	PTM05	PTX05
RR03	0					
RR04	0.0294	0				
RR05	0.011	0.0038	0			
SR05	0.0154	-0.0137	-0.008	0		
PTM05	0.0446	0.0051	0.0089	0.0021	0	
PTX05	0.0289	-0.0007	-0.0035	-0.0172	0.0049	0
VIMS05	0.0289	-0.0023	-0.0072	-0.0088	-0.0025	-0.0041
HI06	0.048	0.0005	0.007	-0.0017	-0.0028	0.0017
HI07	0.0278	0.0045	-0.0001	0.0144	-0.0014	0.0065
NC05	0.0371	-0.001	0.0196	-0.0108	0.0048	0.0081
NJHM05	0.0084	-0.0042	-0.0096	0.0221	0.0129	0.0091
NJTK05	0.0369	0.0067	-0.008	-0.0007	0.005	0.0009

	VIMS05	HI06	HI07	NC05	NJHM05	NJTK05
VIMS05	0					
HI06	-0.0021	0				
HI07	-0.0057	0.0039	0			
NC05	0.0006	0.0115	0.012	0		
NJHM05	-0.0092	0.011	-0.0053	0.0086	0	
NJTK05	-0.0006	-0.0011	0.0026	0.0238	0.007	0

Table 4-11. Population pairwise Fst values of all mid-Atlantic coast samples.

P-value	RR03	RR04	RR05	SR05	PTM05	PTX05
RR03	*					
RR04	**0.00977±0.0026**	*				
RR05	0.14258±0.0109	0.30371±0.0112	*			
SR05	0.20898±0.0110	0.64844±0.0131	0.59570±0.0167	*		
PTM05	**0.00000±0.0000**	0.18359±0.0135	0.14648±0.0118	0.35352±0.0162	*	
PTX05	**0.00293±0.0016**	0.45215±0.0168	0.62988±0.0159	0.79102±0.0132	0.13184±0.0114	*
VIMS05	**0.00781±0.0024**	0.52051±0.0133	0.73242±0.0167	0.58496±0.0164	0.58496±0.0145	0.73047±0.0130
HI06	**0.00000±0.0000**	0.40039±0.0135	0.20410±0.0117	0.47852±0.0150	0.64355±0.0143	0.30859±0.0135
HI07	**0.00195±0.0014**	0.18652±0.0111	0.43359±0.0149	0.18652±0.0101	0.54004±0.0183	0.10156±0.0104
NC05	**0.00586±0.0022**	0.47559±0.0133	0.06445±0.0068	0.57324±0.0170	0.25586±0.0133	0.13672±0.0086
NJHM05	0.23926±0.0146	0.51465±0.0207	0.66406±0.0130	0.21875±0.0116	0.13574±0.0097	0.20312±0.0130
NJTK05	**0.00195±0.0014**	0.14746±0.0116	0.79395±0.0120	0.40332±0.0168	0.14258±0.0116	0.34082±0.0185

P-value	VIMS05	HI06	HI07	NC05	NJHM05	NJTK05
VIMS05	*					
HI06	0.54297±0.0120	*				
HI07	0.82910±0.0111	0.18262±0.0115	*			
NC05	0.40137±0.0147	0.09082±0.0082	0.07812±0.0085	*		
NJHM05	0.64844±0.0148	0.19238±0.0129	0.54785±0.0181	0.25488±0.0126	*	
NJTK05	0.45117±0.0160	0.49414±0.0168	0.22070±0.0164	**0.00879±0.0029**	0.26465±0.0133	*

Table 4-12. P values of population pairwise Fst values of all mid-Atlantic coast samples. Significant P values <0.05 were indicated with bold font.

	GOMBSL	GOMCOLA	GOMCR	GOMDS	GOMGAU	GOMGP
GOMBSL	0					
GOMCOLA	0.01853	0				
GOMCR	-0.02409	0.03586	0			
GOMDS	-0.0199	0.01991	0.01053	0		
GOMGAU	-0.01977	0.00482	0.00676	-0.00943	0	
GOMGP	-0.01344	0.0344	0.00816	-0.00048	-0.00401	0
GOMMDB	-0.01269	0.00761	-0.00346	0.00343	-0.00392	0.00359
GOMPC	-0.02343	0.01975	0.0038	-0.00479	-0.01233	-0.01255
GOMPN	-0.02056	0.05349	-0.01251	0.01021	0.01185	0.00287
GOMSH	-0.01685	0.01632	0.00297	-0.00002	-0.00314	-0.00111
GOMTPA	-0.04624	-0.02331	-0.0389	-0.03296	-0.03417	-0.01949
GOMAP	-0.00514	0.02988	0.00064	0.00821	0.01471	0.01889

	GOMMDB	GOMPC	GOMPN	GOMSH	GOMTPA	GOMAP
GOMMDB	0					
GOMPC	-0.00441	0				
GOMPN	-0.00079	0.00497	0			
GOMSH	-0.01372	-0.00788	0.00221	0		
GOMTPA	-0.05252	-0.03585	-0.02395	-0.03966	0	
GOMAP	-0.0169	0.01187	-0.00152	-0.00313	-0.03966	0

Table 4-13. Population pairwise Fst values of all Gulf of Mexico samples.

	GOMBSL	GOMCOLA	GOMCR	GOMDS	GOMGAU	GOMGP
GOMBSL	*					
GOMCOLA	0.20721±0.0408	*				
GOMCR	0.99099±0.0030	**0.02703±0.0139**	*			
GOMDS	0.89189±0.0214	0.10811±0.0227	0.12613±0.0545	*		
GOMGAU	0.89189±0.0345	0.30631±0.0411	0.27928±0.0497	0.78378±0.0361	*	
GOMGP	0.68468±0.0474	**0.01802±0.0121**	0.18018±0.0407	0.45946±0.0344	0.45045±0.0650	*
GOMMDB	0.66667±0.0493	0.19820±0.0227	0.61261±0.0297	0.34234±0.0327	0.41441±0.0667	0.34234±0.0402
GOMPC	0.91892±0.0328	0.10811±0.0326	0.33333±0.0665	0.56757±0.0543	0.90090±0.0304	0.89189±0.0287
GOMPN	0.91892±0.0228	**0.00901±0.0091**	0.87387±0.0238	0.24324±0.0430	0.08108±0.0286	0.36937±0.0497
GOMSH	0.76577±0.0366	0.12613±0.0309	0.33333±0.0385	0.36036±0.0407	0.53153±0.0394	0.46847±0.0402
GOMTPA	0.99099±0.0030	0.69369±0.0508	0.94595±0.0154	0.96396±0.0142	0.98198±0.0096	0.77477±0.0454
GOMAP	0.46847±0.0354	**0.03604±0.0201**	0.42342±0.0592	0.13514±0.0434	0.11712±0.0333	0.11712±0.0273

	GOMMDB	GOMPC	GOMPN	GOMSH	GOMTPA	GOMAP
GOMMDB	*					
GOMPC	0.48649±0.0388	*				
GOMPN	0.42342±0.0508	0.26126±0.0550	*			
GOMSH	0.89189±0.0165	0.78378±0.0408	0.34234±0.0424	*		
GOMTPA	0.99099±0.0030	0.95495±0.0151	0.89189±0.0345	0.97297±0.0184	*	
GOMAP	0.90991±0.0253	0.12613±0.0242	0.45045±0.0559	0.59459±0.0474	0.97297±0.0125	*

Table 4-14. P values of population pairwise Fst values of all Gulf of Mexico samples. Significant P values <0.05 were indicated with bold font.

	08MAAR	08MABR	08MACB	08MAMV	08MAWTR
08MAAR	0				
08MABR	-0.00771	0			
08MACB	-0.0025	0.01326	0		
08MAMV	0.00643	-0.00186	0.01617	0	
08MAWTR	-0.00122	-0.00305	0.0236	0.00772	0

Table 4-15. Population pairwise Fst values of all Massachusetts samples.

	08MAAR	08MABR	08MACB	08MAMV	08MAWTR
08MAAR	*				
08MABR	0.73874±0.0403	*			
08MACB	0.50450±0.0578	0.10811±0.0227	*		
08MAMV	0.18919±0.0438	0.40541±0.0365	0.12613±0.0201	*	
08MAWTR	0.41441±0.0338	0.57658±0.0511	**0.04505±0.0152**	0.15315±0.0273	*

Table 4-16. P values of population pairwise Fst values of all Massachusetts samples. Significant P values <0.05 was indicated with bold font.

4.4 Conclusion and discussion

High diversity in blue crab mtDNA

The mtDNA gene *nad2* of the blue crab showed extremely high polymorphism. The genetic diversity was nearly one and a new haplotype occurred for about every two crabs sampled. The average pairwise difference was about 7 nulceotides, but it could be up to 26 nulceotides, which was 4% of the sequenced gene fragment. There were very few prevailing haplotypes whose frequency was no higher than 15%. About 70% sequences were singletons. It seemed numbers of mutations and haplotypes were not reaching saturation yet with these data consisting of 1153 crab sequences.

We compared the mtDNA nucleotide diversity (Pi) data of 11 crustaceans (Table 4-17). The whole mitochondrial genome RFLP data of the blue crab (McMillen-Jackson and Bert 2004) and coconut crab (Lavery, Moritz et al. 1996) showed a Pi value close to 0.01, which was comparable with the *nad2* Pi (0.008-0.015) in our study. All other reports for coding region *cox1* had Pi values around 0.001 and two results on control region of shrimps had Pi value around 0.02. The ten-fold difference between *cox1* gene in other species and *nad2* gene in the blue crab seems intriguing. In fact, in the study presented in Chapter 3 that we performed using the *cox1* gene showed the Pi value was around 0.005, lower than *nad2* Pi but still much higher than the Pi values in other crabs. What drives and keeps the blue crab's high variability of the mitochondrial genome of needs to be further studied.

Species	Nucleotide diversity	Common name	Region used	Paper cited
Portunus trituberculatus	0.001-0.002	swimming crab	*cox1*	(Liu, Liu et al. 2009)
Erimacrus isenbeckii	0.001-0.002	hair crab	*cox1*	(Azuma, Kunihiro et al. 2008)
Birgus latro	0.0075	coconut crab	mtDNA-RFLP	(Lavery, Moritz et al. 1996)
Callinectes sapidus	0.01-0.026	blue crab	mtDNA-RFLP	(McMillen-Jackson and Bert 2004)
Maja brachydactyla	0.001-0.005	spiny spider crab	*cox1*	(Sotelo, Moran et al. 2008)
Paratya australiensis	0.00043	freshwater shrimp	*cox1*	(Cook, Bunn et al. 2007)
Crangon crangon (L.)	0.001-0.02	common shrimp,	*cox1*	(Luttikhuizen, Campos et al. 2008)
Artemia salina	0.001-0.01	Mediterranean brine shrimp	*cox1*	(Munoz, Gomez et al. 2008)
Litopenaeus setiferus	0.021	white shrimp	CR	(McMillen-Jackson and Bert 2003)
Farfantepenaeus aztecus	0.028	brown shrimp	CR	(McMillen-Jackson and Bert 2003)

Table 4-17. MtDNA variability of some other crustaceans.

Inter-annual variation within Chesapeake Bay

The F statistics showed that within the Chesapeake Bay, crabs collected in year 2003 were significantly different from samples of the other four years (2004-2007). No significant difference among the other years was detected. Whether the population from 2003 was different or not is unclear at this point.

Lack of geographic genetic structure

In general there was no population subdivision among samples collected from different regions: New England, mid-Atlantic coast and the Gulf of Mexico. Almost all (99.5%) of the genetic variation was attributed to within-sample variations. Phylogenetics inference also showed no clustering pattern according to regions. The lack of geographic population structure has been observed in earlier studies with the blue crab (McMillen-Jackson, Bert et al. 1994; McMillen-Jackson and Bert 2004).

Even when we examined individuals at the northern extreme of its distribution, Massachusetts samples didn't show statistically significant reduction in genetic variability. The haplotype diversity (Hd), nucleotide diversity (Pi), pairwise differences (k) were all fairly close to the indices for other samples with one exception (08MACB), which showed much lower Pi/k values than all other samples (Figure 4-11 and Figure 4-12). In general, Massachusetts, mid-Atlantic and Gulf of Mexico individuals show no significant differences in genetic structure.

This observation is not congruent with previous reports (McMillen-Jackson, Bert et al. 1994; McMillen-Jackson and Bert 2004). In both of their works, they observed a trend of latitudinal cline along the Atlantic coast in haplotype diversity. Since their works were on allozymes and RFLP data of mtDNA, this discrepancy may be due to the sensitivity of the markers employed and low sample numbers.

Low genetic differentiation across large geographic area is not uncommon in marine species. For example, brown shrimp (*Farfantepenaeus aztecus*) and pink shrimp (*Farfantepenaeus duorarumand*) (McMillen-Jackson and Bert 2003; McMillen-Jackson

and Bert 2004) both lack of genetic differentiation between samples collected from Atlantic coast to Gulf of Mexico. In both reports, the genetic diversity was extremely high (haplotype diversity was around 1) and there was no spatial population subdivision detected, which was very similar to our finding. The difference was the genetic markers employed: in these two shrimp studies the control region of the mtDNA genome was used while in our study the protein-coding region *nad2* gene was used. The CR is thought to be of much higher mutation rate than coding regions but the Pi value of blue crab in our study (around 1%) was in fact comparable to the Pi value of pink shrimp (1.8%), which suggested the *C. sapidus* may have even higher polymorphism in its control region.

Along with the brown shrimp, its sympatric species white shrimp was studied with the same sampling strategy and methodology (McMillen-Jackson and Bert 2003). Although its genetic diversity was so very high (haplotype diversity 0.96; nucleotide diversity Pi 2.1%), the white shrimp displayed a clear geographic structure. The two lineages of its haplotypes had distinguished frequencies between Atlantic/eastern Gulf and western Gulf. The authors attributed this to the difference in biology of these two shrimps. White shrimps preferentially inhabit lower salinity waters and are less cold tolerant. Thus this species may have gone through severe fluctuations in population size and distribution in Pleistocene era and this history has left a signature in its genetic background.
The white shrimp showing a geographic structure despite the high genetic diversity suggests that the extreme diversity of genetic markers doesn't necessarily impede the detection of population subdivisions if there were. *C. sapidus*, as a widely distributed

species inhabiting a large variety of environments of different salinity and temperature with a planktonic life stage, the lack of genetic spatial differentiation may be due to long-distance gene flows over long time periods.

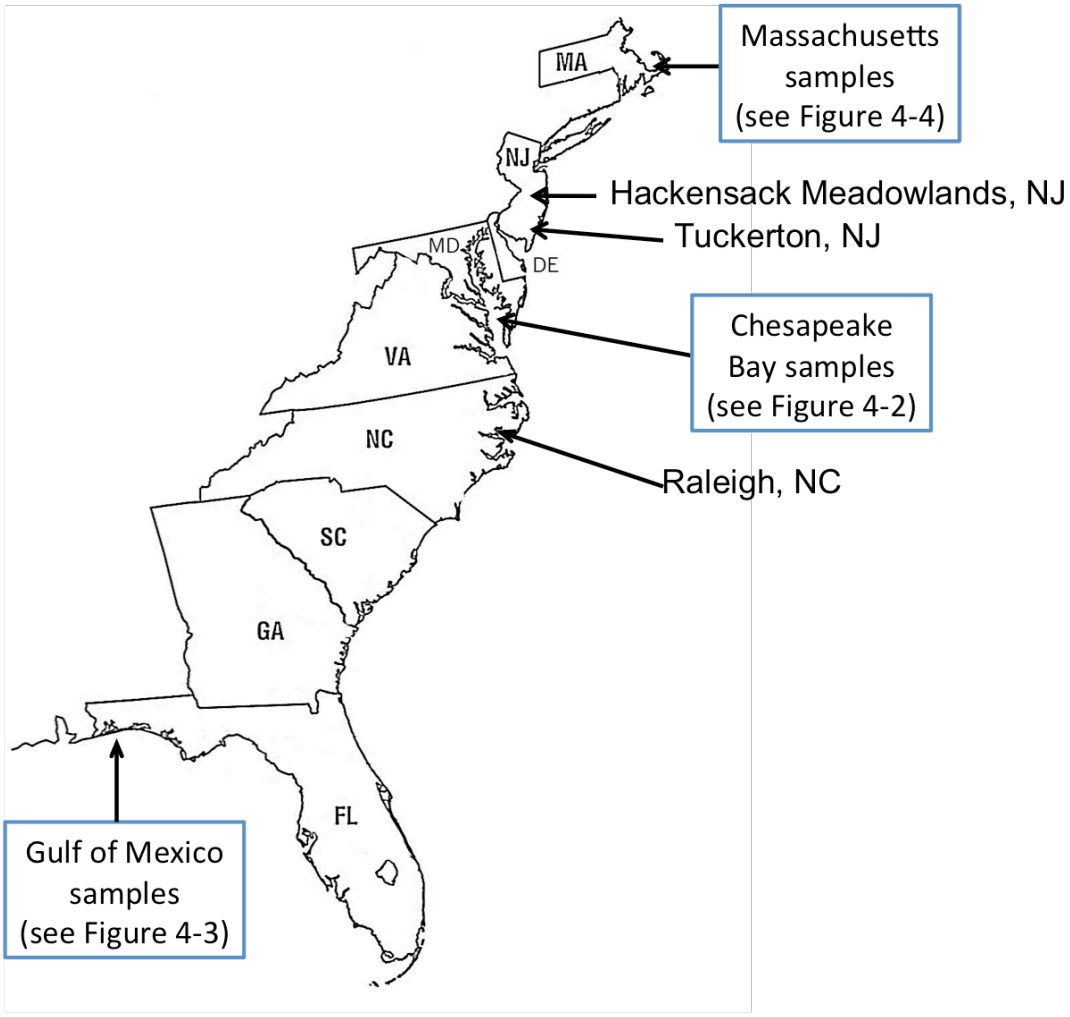

Figure 4-1. Three major sampling areas: mid-Atlantic states, including New Jersey, Chesapeake Bay states-Maryland and Virginia, and North Carolina; Gulf of Mexico; and Massachusetts. The detailed sampling location map for each area is shown as the following figures.

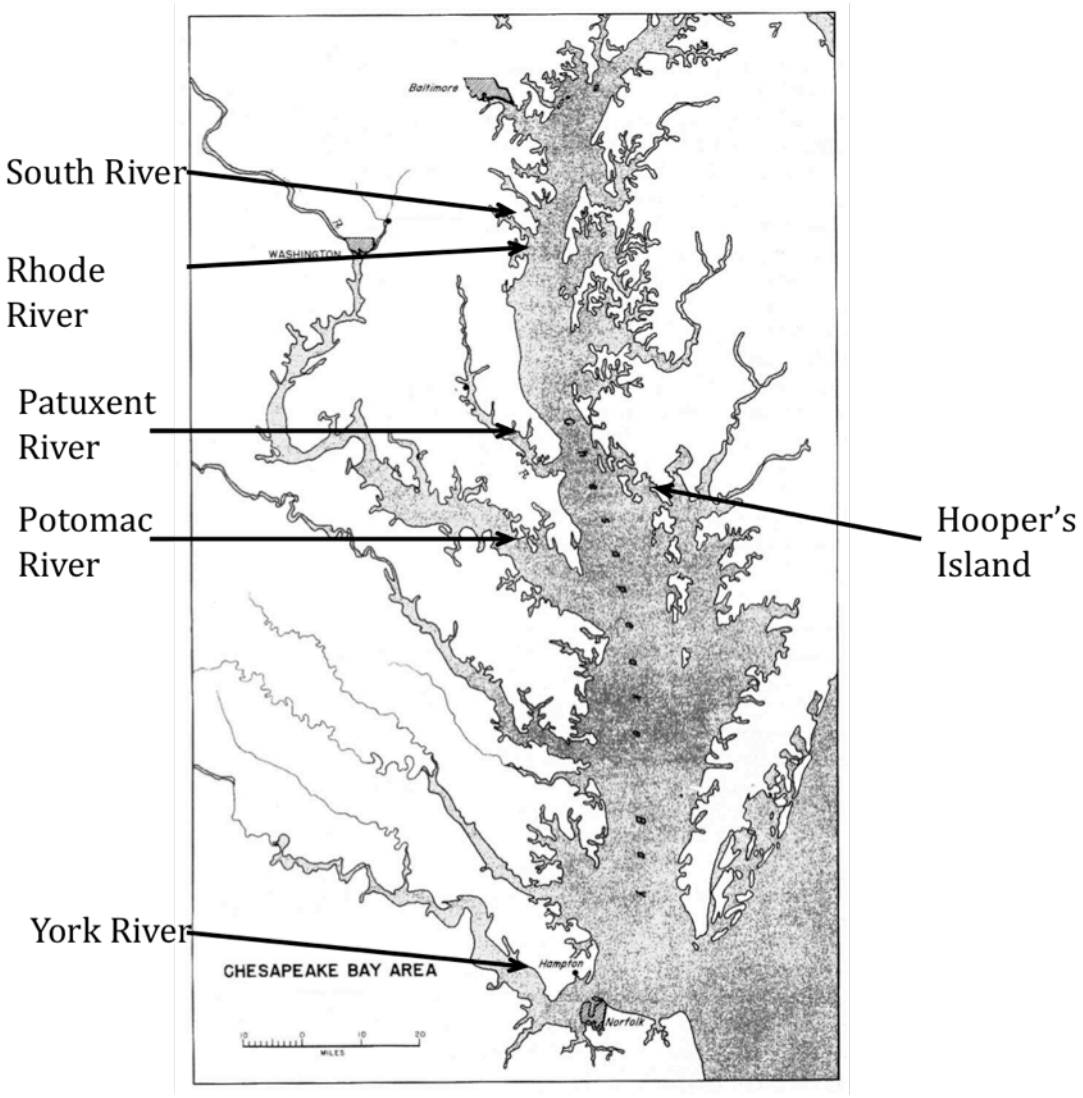

Figure 4-2. Sampling locations in Chesapeake Bay area.

Figure 4-3. Sampling locations in Gulf of Mexico area.

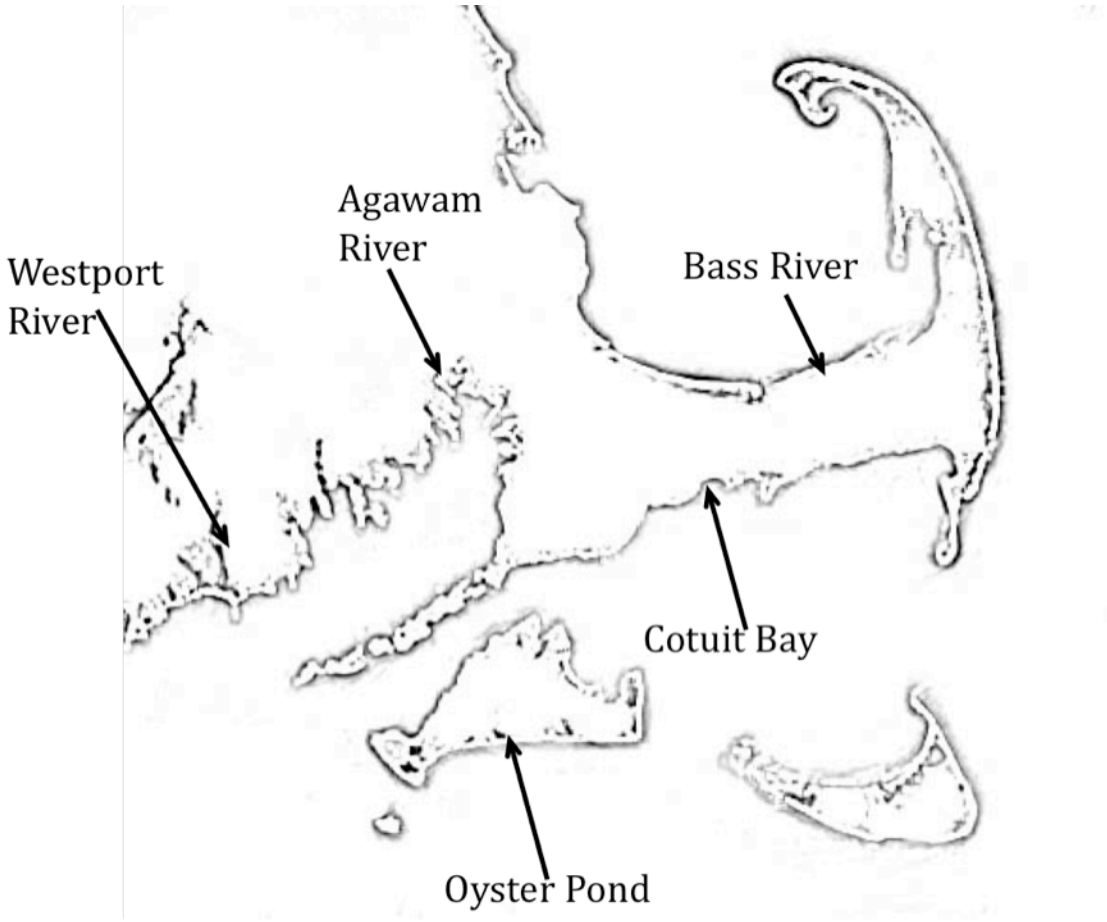

Figure 4-4. Sampling locations in Massachusetts.

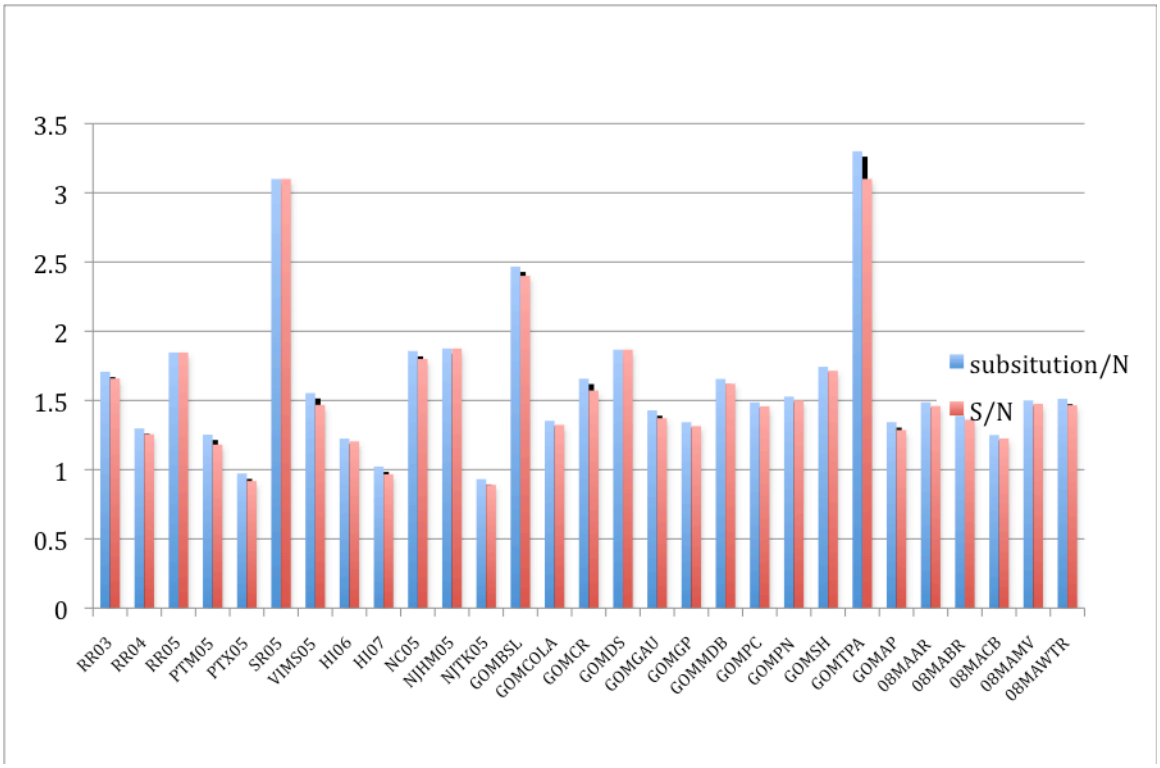

Figure 4-5. The ratio of substitution and segregating site versus sample size within each sample. The blue columns are the ratios of the numbers of total substitutions versus sample sizes while the red are the ratios of the numbers of segregating sites versus sample sizes. N denotes sample size and S denotes number of segregating site.

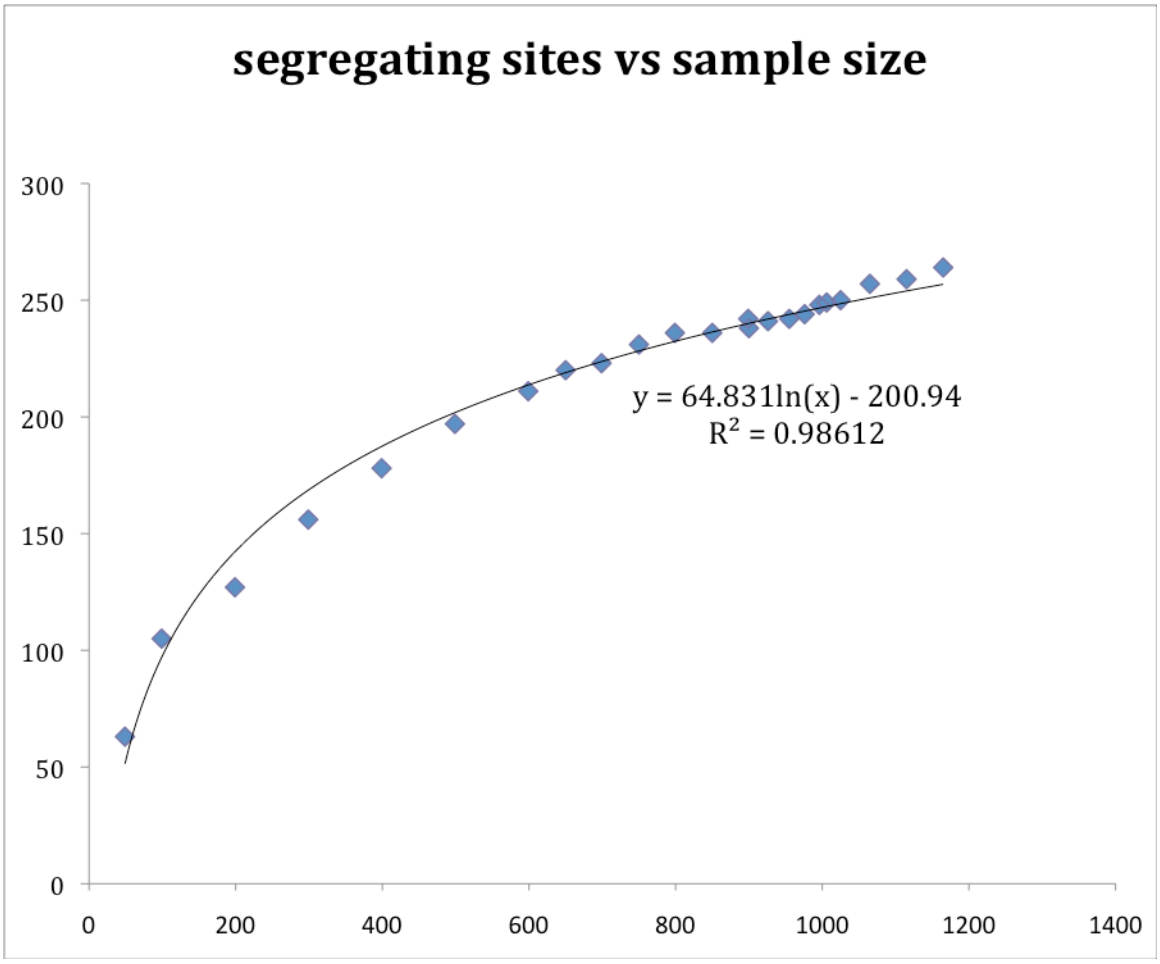

Figure 4-6. Plot of segregating sites against sample size. The X-axis is the sample size and the Y-axis is the number of segregating sites.

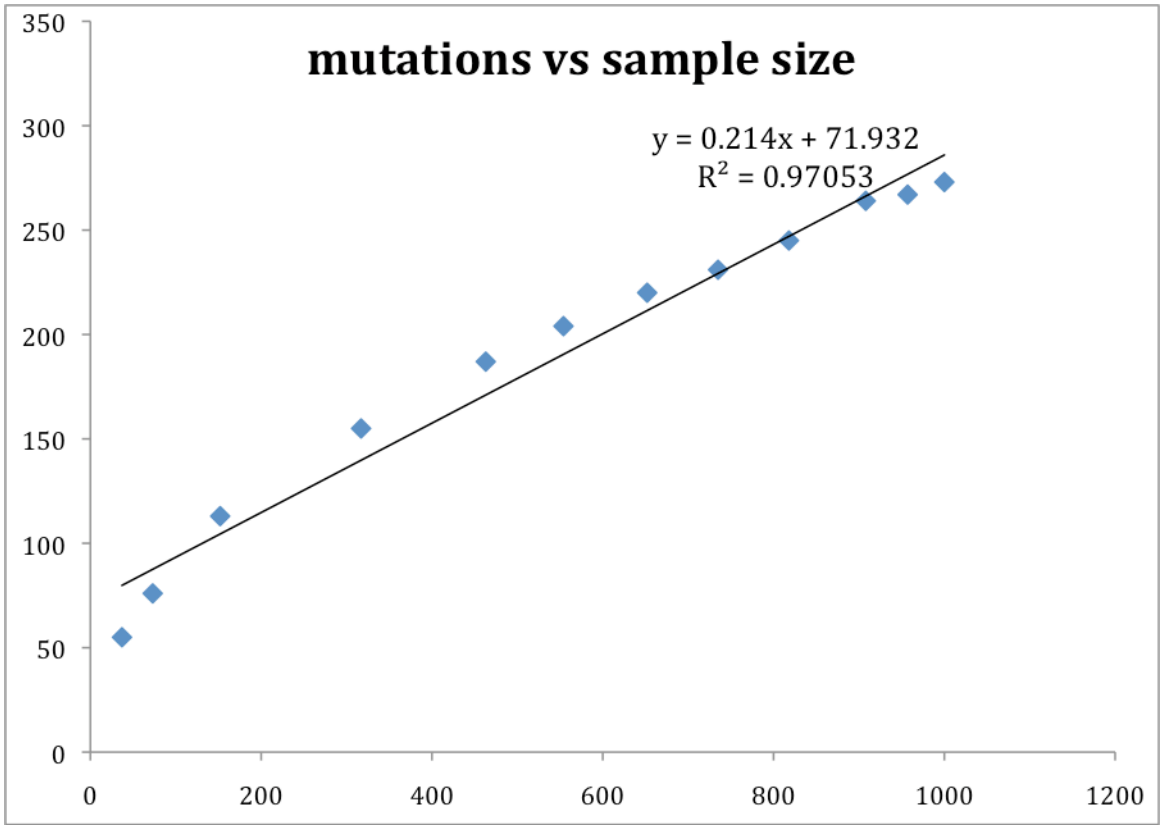

Figure 4-7. Plot of mutations against sample size. The X-axis is the sample size and the Y-axis is the number of total mutations.

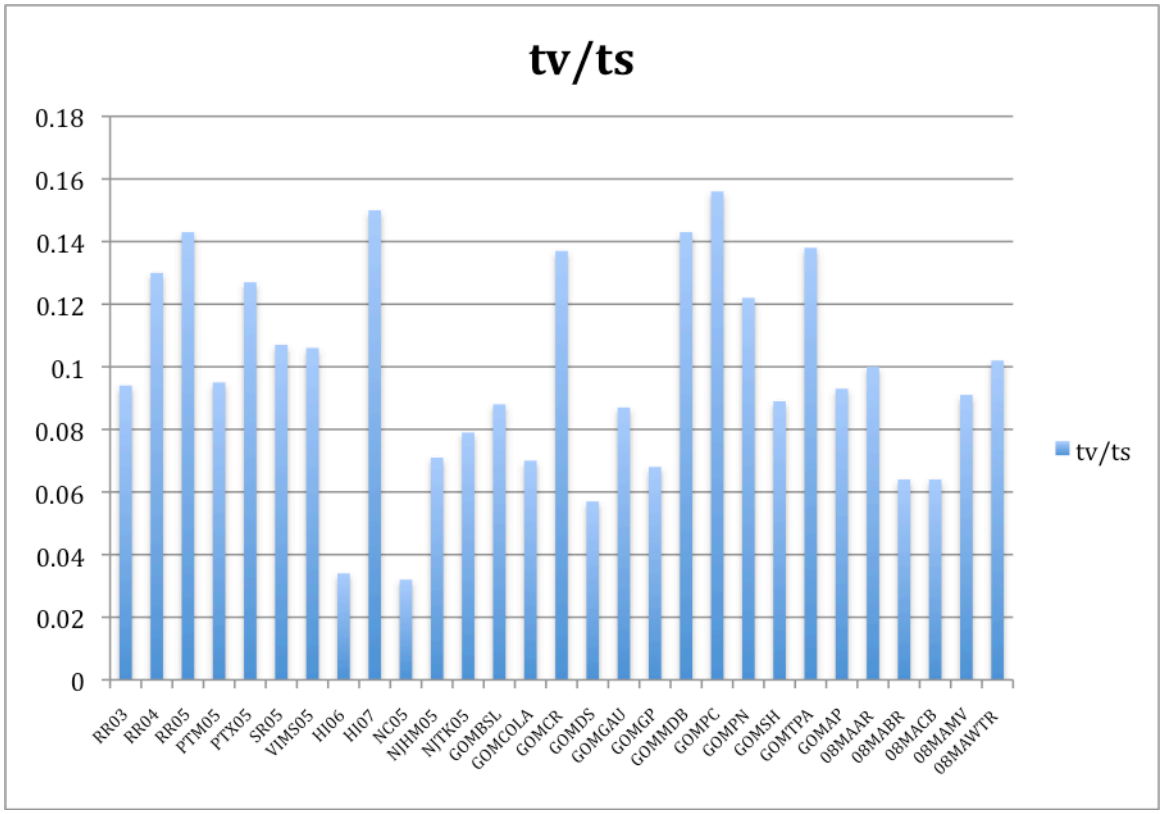

Figure 4-8. The transversion/transition ratio for each sample.

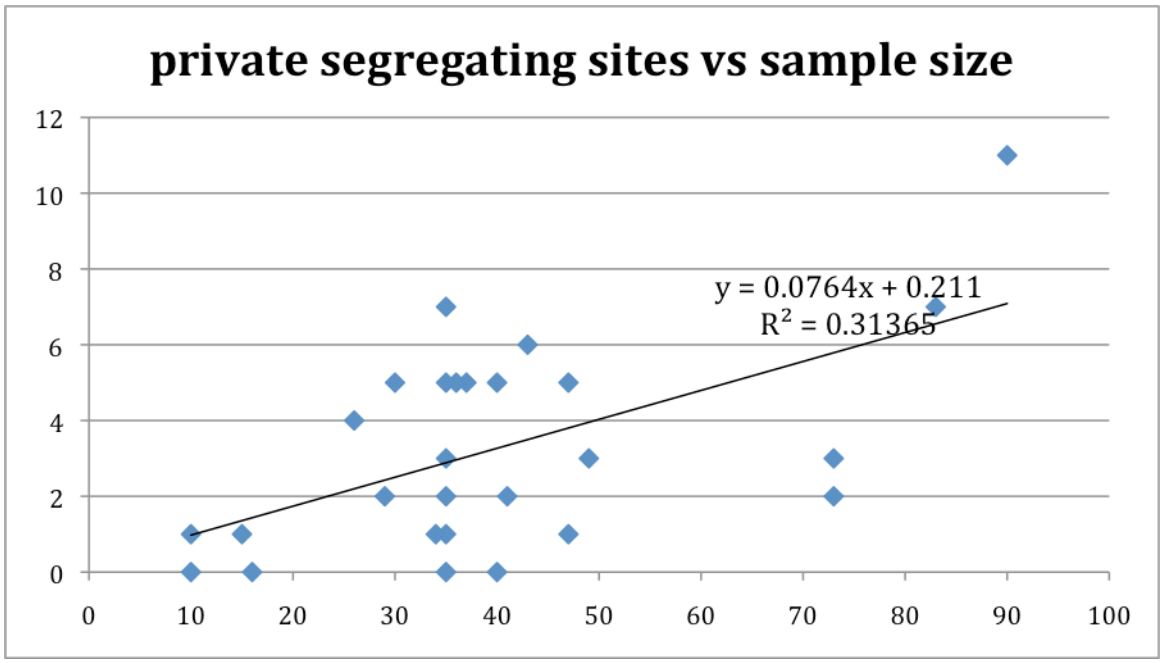

Figure 4-9. Plot of private variable sites vs sample size. The X-axis is the sample size and the Y-axis is the number of private segregating sites. Private mutations occurred within only one sample.

-

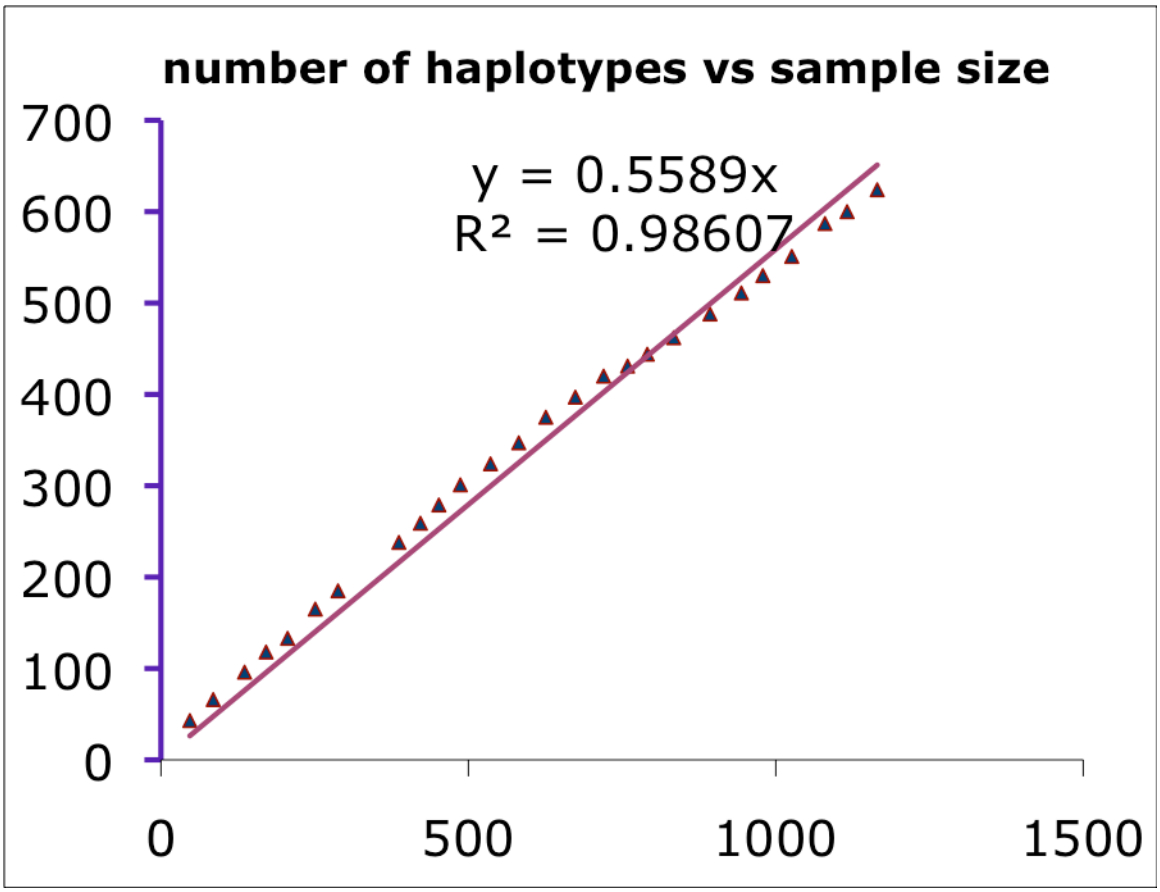

Figure 4-10. Plot of number of haplotypes against sample size. The X-axis is the sample size and the Y-axis is the number of haplotypes.

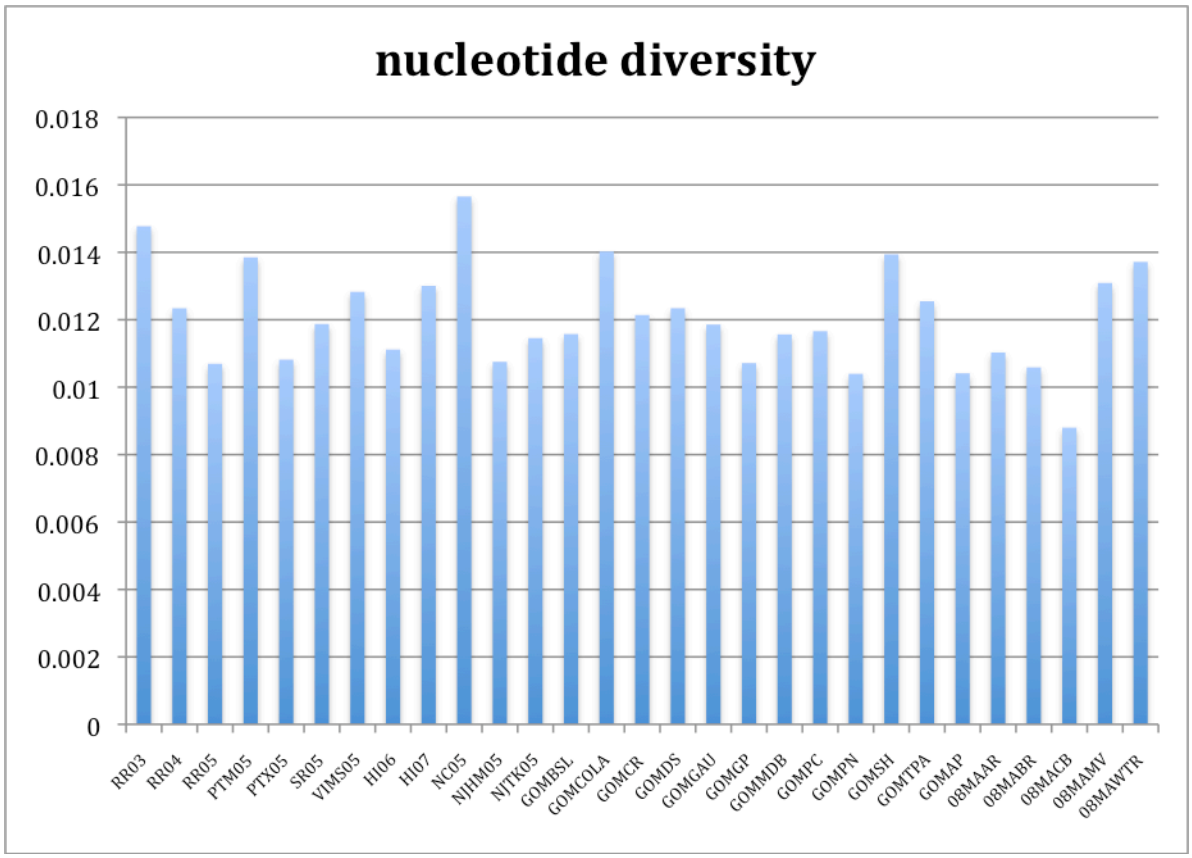

Figure 4-11. The nucleotide diversity of each sample.

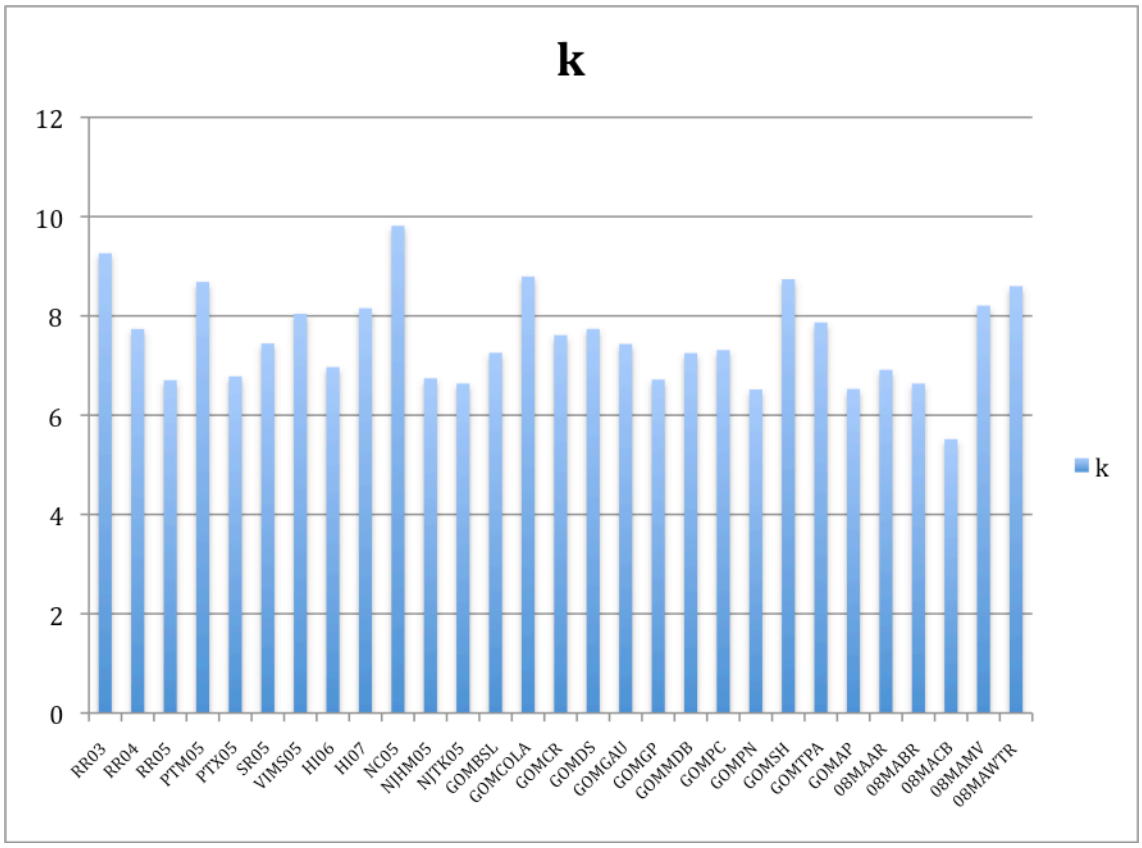

Figure 4-12. The average pairwise distance (k) of each sample.

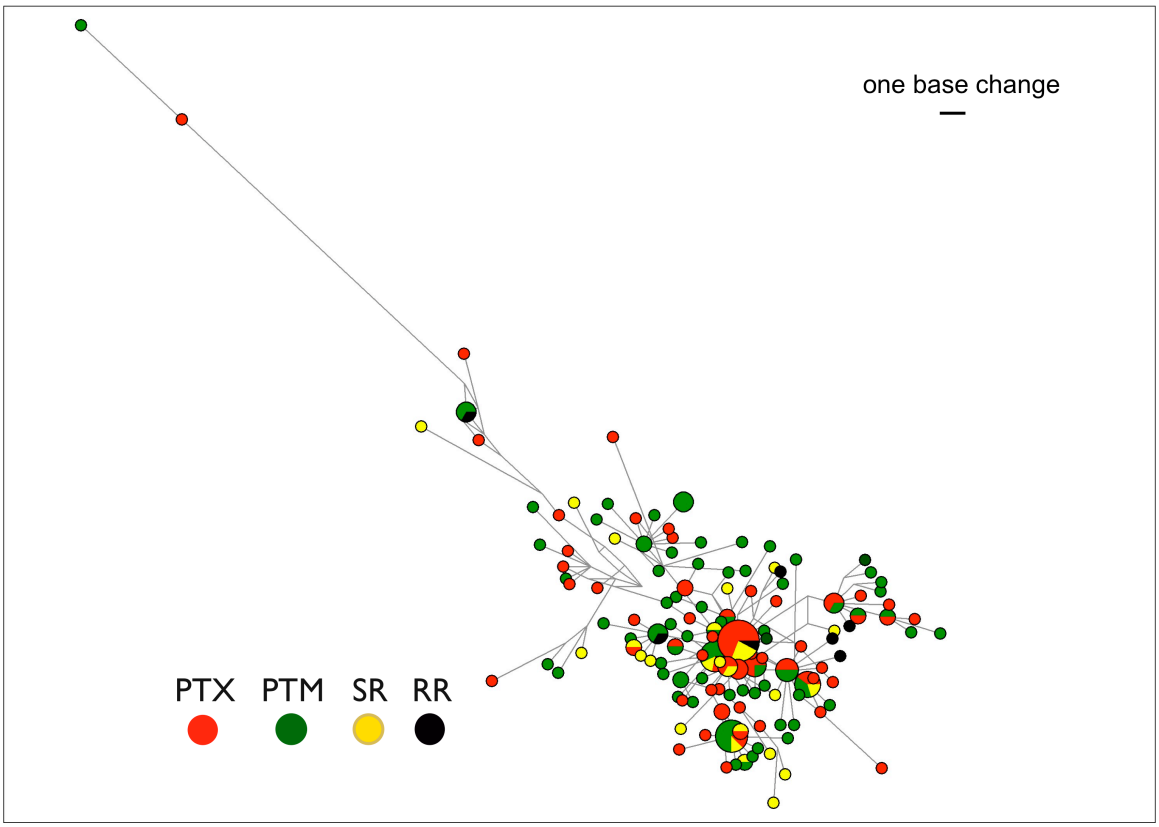

Figure 4-13. Minimum Spanning Network of 2005 upper Chesapeake Bay samples. Each circle stands for one distinct haplotype and the size of circles is proportional to the frequency of the haplotype. The branch length indicates the genetic distance between a pair of haplotypes, the scale bar denoting distance of one base change. PTX, the 2005 Patuxent River sample; PTM, the 2005 Potomac River sample; SR, the 2005 South River sample; RR, the 2005 Rhode River sample.

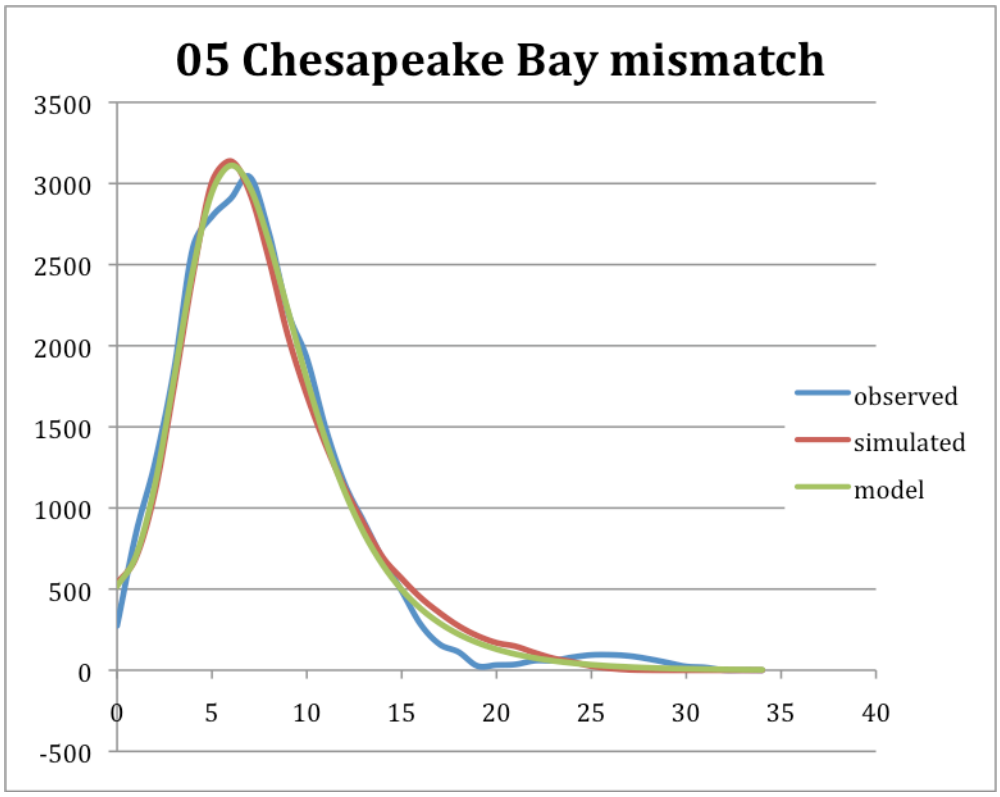

Figure 4-14. Mismatch plot of 05 Chesapeake Bay samples. The X-axis is the number of mismatches and the Y-axis is the frequency.

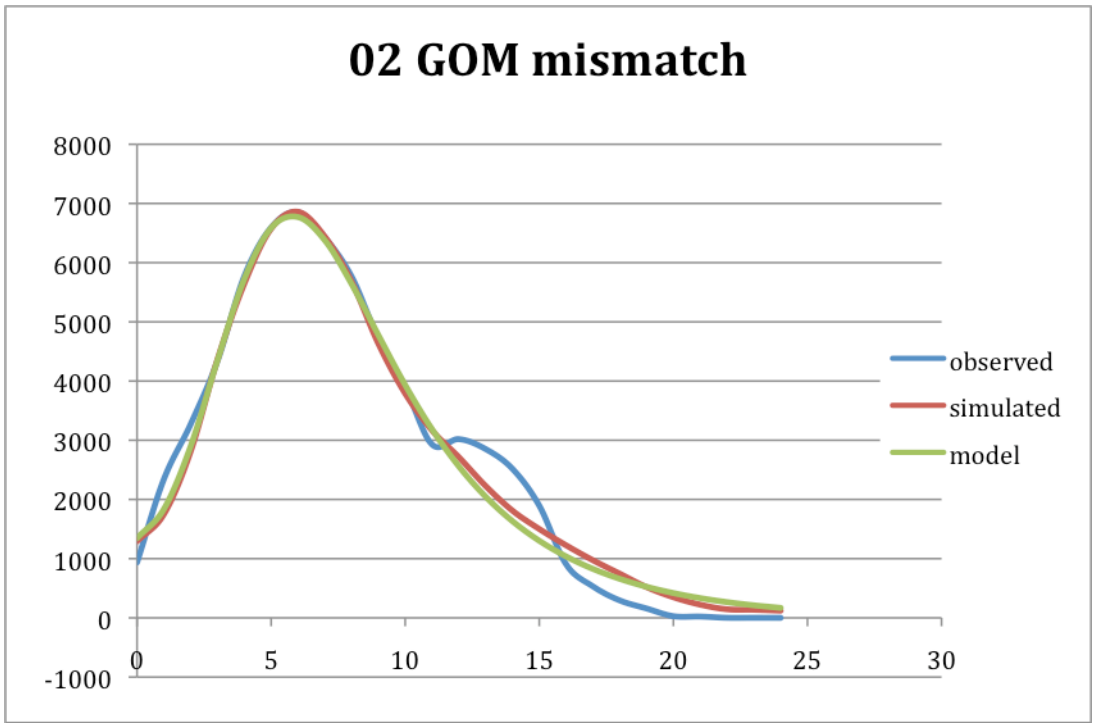

Figure 4-15. Mismatch plot of 02 Gulf of Mexico samples. The X-axis is the number of mismatches and the Y-axis is the frequency.

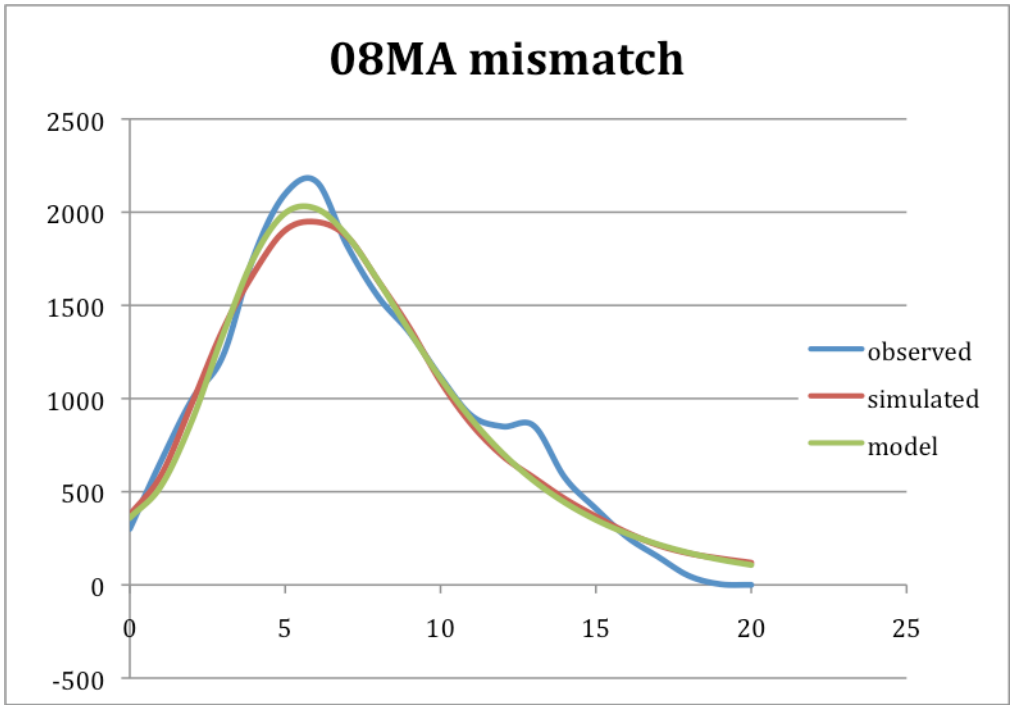

Figure 4-16. Mismatch plot of 08 Massachusetts samples. The X-axis is the number of mismatches and the Y-axis is the frequency.

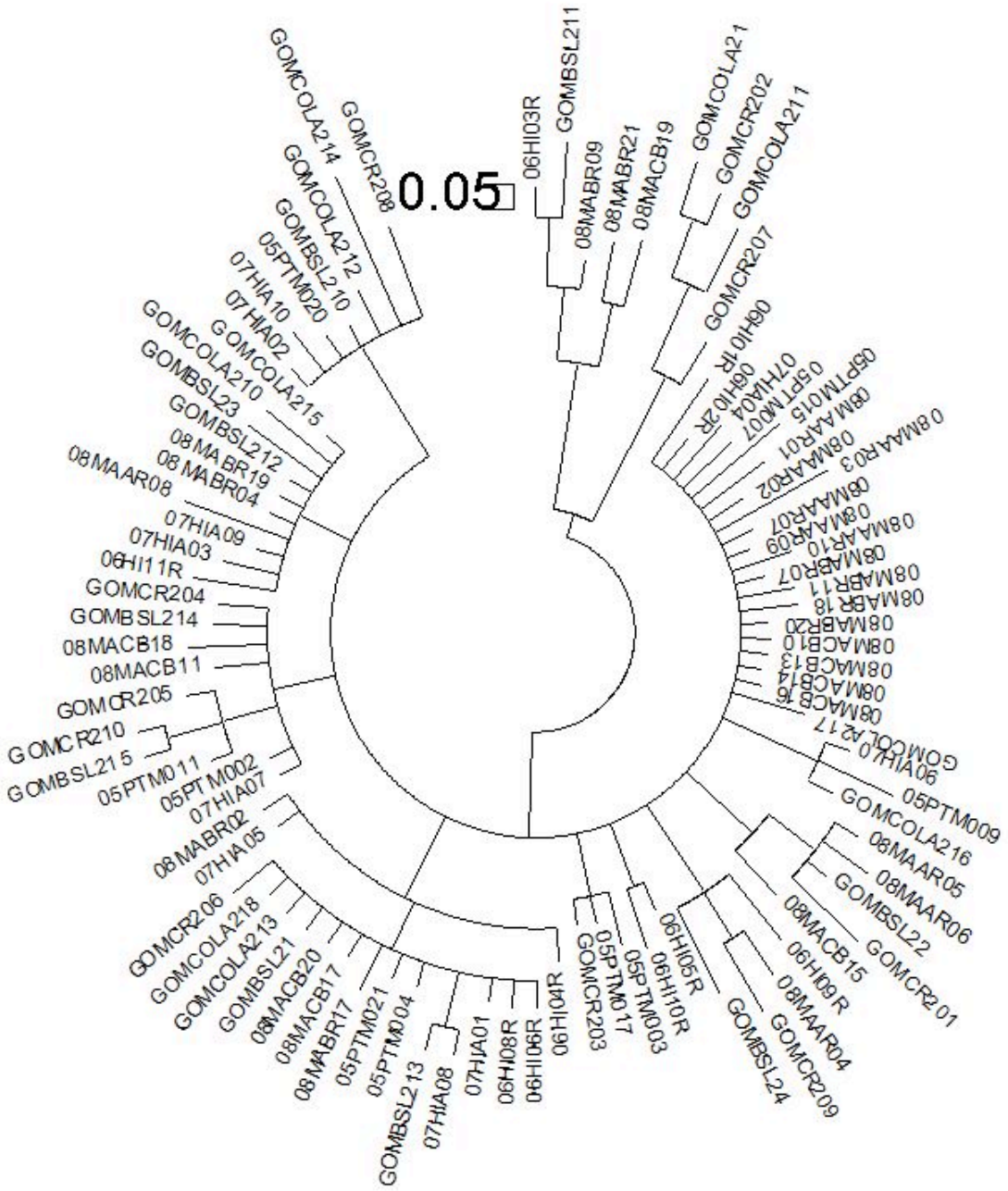

Figure 4-17. Baysian tree on 90 individuals consisting of 30 crabs from Chesapeake Bay, 30 from GOM and 30 from MA.

5 Evidence for extensive heteroplasmy for mtDNA in the blue crab

5.1 Introduction

The mitochondrial DNA has been extensively used in animal research of inter-specific phylogenetics and intra-specific population genetics. It is generally agreed that the mtDNA has a rapid evolution rate, lacks recombination and is homoplasmic within individual animals.

However, mitochondrial heteroplasmy has been reported in a wide range of species including insects (fruit flies: (Solignac, Monnerot et al. 1983) beetles: (Boyce, Zwick et al. 1989); birds (gulls: (Crochet and Desmarais 2000; Kvist, Martens et al. 2003); fish (red drums: (Gold and Richardson 1990); mammals (Hauswirth and Laipis 1982; Boursot, Yonekawa et al. 1987; Moraes, Schon et al. 1989; Holt, Harding et al. 1990); and consistently mussels (Fisher and Skibinski 1990; Quesada, Stuckas et al. 2003; Breton, Burger et al. 2006; Obata, Kamiya et al. 2006).

The mitochondrial DNA heteroplasmy in mussels in the genus *Mytilus* is due to an unusual type of inheritance called Doubly Uniparental Inheritance by which both paternal and maternal mitochondrial DNA may be passed on to the offspring. In other animals exhibiting mtDNA heteroplasmy, the reason is thought to be either paternal leakage (Kvist, Martens et al. 2003), germ-line mitochondrial molecules segregation (Hauswirth and Laipis 1982; Wu, Smith et al. 2000), mitochondrial gene duplication (Abbott, Double et al. 2005), somatic mutation (Chatterjee, Mambo et al. 2006), or sometimes, a false mtDNA heteroplasmy caused by nuclear copies of mitochondrial genes. Among them

NUMTs (nuclear copies of mitochondrial genes) has been increasingly reported (Lopez, Cevario et al. 1996; Williams and Knowlton 2001; Parr, Maki et al. 2006).

The most common form of mtDNA heteroplasmy involves length variation caused by the variable number of tandem repeats in the non-coding control region of the mitochondrial genome (Nesbo, Arab et al. 1998). This length variation is thought to occur through slipped strand mispairing during replication (Densmore, Wright et al. 1985). On the other hand large deletions or single nucleotide substitutions in both coding genes and non-coding regions are frequently documented in pathogenic mitochondrial diseases in humans (Van Hove, Freehauf et al. 2008; Wang, Gu et al. 2008; Ye, Gao et al. 2008). The nature of these forms of heteroplasmy is hard to determine, especially when it may require distinguishing somatic mutations within an individual from the homoplasmic form found in the oocytes.

The blue crab (*Callinectes sapidus*) is a species inhabiting estuarine and near shore environments from Nova Scotia to northern Argentina, being one of the most ecologically and economically important animals in the State of Maryland. This species' complete mitochondrial genome sequence has been published (Place, Feng et al. 2005). I will present data with the blue crab mtDNA using sequence variation at multiple loci showing the DNA sequences of this genome are heteroplasmic and that juvenile crabs from the same batch of one mature female crab had very similar heteroplasmic character in their DNA sequencing chromatograms (Figure 5-1).

5.2 Material and method

5.2.1 The blue crab samples

A known three-member crab family was used for the study. The virgin female adult Y16 and the male adult R were mated in the laboratory in 2003 to produce the offspring megalope M9.

5.2.2 DNA extraction and PCRs

Adult crabs were preserved in 95% ethanol until the DNA extraction. About 100 mg muscle tissue for each crab was used for DNA extraction with Qiagen DNeasy Blood and Tissue Kit. Individual megalopa was rinsed in distilled water and placed into 100 μL of extraction buffer (3M urea, 4M guanidine thiocyanate, 20mM EDTA, 50mM Tris pH 8.0, 1% Tween 20 (Sigma P9416, St. Louis MO.)) where it was crudely homogonized by repeated passage through a pipette tip with a bore approximately the same diameter as that of the megalopa. Homogenization was performed to expose the soft tissue to the extraction buffer and not for mechanical cell lysis. The samples were incubated for 3 hours at 60ºC, and proteins and lipids were removed by purification through a Zymo Clean and Concentrator -5 column (D4013, Orange Ca.).

PCR primers were designed based on the published complete *C. sapidus* mitochondrial genome sequence (Place, Feng et al. 2005), targeting three genes—*nad2, cox1*, and *nad4*. Primer sequences are shown in Table 5-1.

Gene	Amplicon Length(bp)	Forward Primer (5'- 3')	Reverse Primer (5'- 3')
nad2	685	TGCTTTATTATTCAACCCCG	CCGAATAGATTGATTGAAGT
cox1	678	TTGCTGCTGCTATTGCTCAC	AATACAGCGCCCATGGATAG
nad4	713	CATTTACTTTTCCTGAACAACATGA	TGCCTTTGTTGGTGTCTTTG

Table 5-1. PCR primers for 3 genes.

PCRs were done in a final volume of 30 µL using high fidelity cloned Pfu polymerase (Stratagene), under conditions as the instruction indicated. The size of PCR products was estimated by electrophoresis in 1% agarose gels.

5.2.3 Cloning and DNA sequencing

PCR products of three genes were cloned by using StrataClone Blunt PCR Cloning Kit (Stratagene). Colonies were screened with PCR primers and positive colonies were selected and cultured for mini-prep and DNA sequencing.

The sequencing reaction was done in 10 µL volume, consisting of 40-150 ng PCR product, 3 pmol of primer, 0.5 µL Big Dye v3.1 sequencing mix and 1.5 µL 5X sequencing buffer (Applied Biosystems). The cycling parameters were 95ºC for 5 min; followed by 50 cycles at 95ºC for 15 s, 50ºC for 15 s, 60ºC for 4 min.

The sequencing reaction product was cleaned by adding 60µL 100% isopropanol and 30µL H_2O, mixing thoroughly, incubating at room temperature for 30 min, and centrifuging at 2,000g for 30 min. The supernatant was discarded and 100 µL 70% isopropanol was added to wash the DNA, followed by another centrifuge at 2,000g for 14 min and the disposal of the supernatant. Cleaned DNA was air dried for 20-30 min before

10μL formamide was added. The mixture was heated at 95ºC for 2 min and then put on ice immediately. After it was chilled down, the denatured sequencing product was loaded into an ABI 3130xl genetic analyzer.

For each PCR product, we conducted both forward and reverse sequencing.

5.2.4 Data analysis

We checked the quality of all DNA sequences and aligned them using the program SEQUENCHER®4.8 (Gene Codes Corporation) before exported the consensus sequences of forward and reverse sequences. The actual exported sequences were shorter than the PCR amplicons in length because only the DNA sequences unambiguously covered by two directions were used in alignments and then exported. The published *C. sapidus* mtDNA genome sequence was used as the reference sequence.

The polymorphism diversity calculation and neutrality test were performed by using program DnaSP v5 (Librado and Rozas 2009).

The following measures were calculated for haplotype/nucleotide diversity,:

S, the number of segregating (polymorphic) sites;

Haplotype (gene) diversity and its sampling variance (Nei 1987); the standard deviation (or standard error) is the square root of the variance;

Nucleotide diversity, Pi, the average number of nucleotide differences per site between two sequences and its sampling variance (Nei 1987); the standard deviation (or standard error) is the square root of the variance;

The mutation parameter Theta (per site or per gene) from S or Eta, the total number of mutations (Watterson 1975; Nei 1987);

k, the average number of nucleotide differences between two sequences (Tajima 1983).

The free software Network 4.5.1.0.(fluxus-engineering.com) was used to produce minimum spanning network of all haplotypes with Median Joining calculation (Bandelt, Forster et al. 1995).

5.3 Results

The total number of clone sequences used in this study for each gene and individual is shown in Table 5-2. In total, 217, 105, and 32 clones were sequenced in both directions for *nad2, cox1* and *nad4* genes for three crabs/megalopa. Among them multiple haplotypes were observed for each gene in each organism. The maximum number of different genotypes occurred for single gene in one individual was 24 for nad2 in M9 out of 66 clones; the minimum number was 2 for nad4 in Y16 from 8 clones. In general, the number of distinctive haplotypes was over 20% of sequenced clones except for nad2 in R (8.7%) and nad4 in R (15.8%).

Gene	# of haplotype/(# of colonies)			
	Total	Y16	R	M9
nad2(685bp)	44(217)	16(82)	6(69)	24(66)
cox1(678bp)	28(105)	3(15)	10(43)	17(47)
nad4(713bp)	7(32)	2(8)	3(19)	3(5)

Table 5-2. Number of sequenced colonies for each gene.

The haplotype diversity (Hd) varied from 16.7%-78.7% (Table 5-3). For all three genes (*nad2, cox1* and *nad4*), M9 always showed the highest haplotype diversity (0.787, 0.622 and 0.700) and R had the lowest (0.167, 0.568 and 0.205) except for *cox1* where Y16 had the lowest diversity (0.257). Comparing the average nucleotide diversity (Pi) and pairwise difference (k) in three genes across all clones, *nad2* and *nad4* had much higher values of Pi (0.00838 and 0.00736) and k (5.732 and 5.246) than *cox1* (0.00363 and 2.453), showing *cox1* had lower diversity than 2 *nad* subunit genes.

		Total	Y16	R	M9
nad2	Hd	0.747	0.481	0.167	0.787
	Pi	0.00838	0.00528	0.00142	0.00810
	k	5.732	3.620	0.973	5.539
cox1	Hd	0.763	0.257	0.568	0.622
	Pi	0.00363	0.00059	0.00258	0.00161
	k	2.453	0.400	1.748	1.090
nad4	Hd	0.635	0.250	0.205	0.700
	Pi	0.00736	0.00541	0.00236	0.00112
	k	5.246	2.500	1.684	0.800

Table 5-3. Polymorphism parameters for three genes in three organisms.

There was one dominant haplotype in each animal for each gene, shown as those largest circles or the big slices in major circles in Figure 5-2, Figure 5-3, and Figure 5-4. For *nad2*, the frequencies of dominant haplotype in Y16, R, and M9 were respectively 72.0%, 91.3%, and 43.9%; for cox1, the dominant genotype frequencies were 86.7%, 62.8%, and 61.7%; and for nad4 they were 87.5%, 89.5%, and 60.0%. M9 had the lowest relative abundance in its dominant haplotype for all three genes when compared to the two other individuals. This result was in consistence with the generally higher haplotype diversity of M9.

The major haplotypes for all three genes of Y16 were always identical to M9's whereas the major haplotypes of R were always distinct from them. Y16 and M9 also shared a few secondary haplotypes in the nad2 gene. R didn't share any minor haplotypes with either Y16 or M9 but only one exception clone that was of same sequence to Y16's and M9's dominant haplotype in *cox1* gene.

None of the dominant haploytpes included any stop codons but there were 3 minor haplotypes (5 clones) of *nad2*, 2 haploptypes (2 clones) of *cox1* and none of *nad*4 having stop codons. There were a few haplotypes had non-synonymous substitutions as Figure 5-5 shown as one example. This figure showed the minimum spanning network of *nad2* DNA haplotypes, color-coded for distinct amino acid sequences. There were 5 minor sequences (3 DNA haplotypes) shown in dark circles including stop codons. All other 212 sequences (41 DNA haplotypes) were without stop codons, encoding for 18 different protein sequences. The major haplotypes of Y16 and M9 (in white) were the same, having one amino acid difference from the major type of R (in brown).

5.4 Conclusion and discussion

5.4.1 The true heteroplasmy, paternal leakage, and NUMTs

Heteroplasmy of mitochondrial DNA has been reported in a large range of species. One of the first studies was in Holstein cows when people noticed the mitochondrial DNA polymorphism within a maternal lineage (Hauswirth and Laipis 1982). Later studies confirmed (Wu, Smith et al. 2000) that different mitochondrial DNA existed within individual cows by sequencing cloned large fragments including 12S, 16s, tRNAs, and D-loop. In these studies, the possible cause of the heteroplasmy due to paternal leakage

was generally ruled out and point mutation was considered as the most likely reason. At the same time, a lot of research discovering the heteroplasmy in other animals confirmed that paternal leakage could contribute to this phenomenon especially when it happened in animals living in hybridization zones (Fontaine, Cooley et al. 2007). It was possibly because of the reduction of recognition and the following destroy to the alien sperm mitochondria from different species. If the paternal leakage happened within one species, it might be taking place at a very low level. A study on the paternal leakage of mitochondrial DNA in Chinook salmon (Wolff, Gandre et al. 2008) suggested the maximum frequency of leakage was 0.03% (power of 0.95) and 0.05% (power of 0.99) by screening more than 10,000 fish. This was apparently below the detection capability of most works.

The nuclear copies of mitochondrial genes (NUMTs) have been an issue when mitochondrial genes were employed for phylogenetic research since when NUMTs sequences were mistakenly thought as orthologue mitochondrial gene, the inference of evolutionary scenarios could be severely misled. A conventional thought is the high copy number of mitochondrial genome over the corresponding nuclear loci can mute the co-amplification of NUMTs. However, it turns out there may be also big numbers up to dozens or even hundreds of pseudo-gene copies in nuclear genomes (Parr, Maki et al. 2006; Antunes, Pontius et al. 2007). Combining with the more feasible and more sensitive DNA sequencing technology, caution on the co-amplification of NUMTs is necessary in the mitochondrial gene studies.

5.4.2 Maternal inheritance pattern of the clones

We confirmed the phenomenon of mitochondrial DNA heteroplasmy in the blue crab by cloning the high fidelity PCR products of *nad2*, *cox1* and *nad4* genes within a three-member crab family. The dominant haplotypes of the mother always matched up with the offspring's; the mother and the offspring also shared a few intermediate frequency haplotypes; and the father's major haplotypes were always distinct from the mother/offspring's. This indicates the offspring's mitochondrial genotypes were passed on from the mother. Most of the secondary haplotypes of the mother and the offspring didn't match up, which might be due to the limited number of clones being sequenced. It was obvious that secondary halpotypes were of very low frequency and they were easy to be missed out when only a certain number of clones were picked out for sequencing. And it seemed the number of new haplotypes had not been saturated yet.

The fact that the father and the offspring never shared any secondary haplotypes strongly suggested that the heteroplasmy is not a result of genetic material transferred from the father (paternal leakage). It also strongly suggests the mitochondrial heteroplasmy represents only mitochondrial DNA rather than nuclear pseudo-genes of mitochondrial genes because nuclear pseudo-genes should have been showing bi-parental inheritance.

5.4.3 M9's higher haplotype diversity

The offspring M9 always had higher diversity than its parents for all three genes. Given that we used only muscle tissue of the adults for DNA extraction but the whole body of the megalopa, this may indicate the tissue distribution of the mitochondrial DNA heteroplasmy. However, it is also possible that the megalope who just went through a

rapid expansion of its mitochondrion population had more new mutations than the adults that lost some of them along their development.

5.4.4 Comparison of the clones to wild crabs

We produced a minimum spanning network by combining the *nad2* clones' sequences of Y16 with sequences of wild crabs collected from Massachusetts, Gulf of Mexico and Chesapeake Bay (Figure 5-6). All wild crabs' *nad2* sequences were results of direct sequencing on PCR products without cloning process.

The major haplotype of Y16 was the same as one Chesapeake Bay wild crab. Furthermore, one of the Y16's secondary haplotype (presented by 3 clones) was shared with some wild crabs. And interestingly, this haplotype was presented in all three sampling areas (Massachusetts, Gulf of Mexico and Chesapeake Bay), one or two individuals from each location. This indicated the secondary haplotype of Y16 was existent in wild crabs as their major type, fully functioning. Similarly, there was one secondary haploype of Y16 *cox1* was represented in one wild Chesapeake Bay crab (data not shown). None of minor haplotypes of R or M9 was observed in wild crabs though.

Neighbor Joining trees on full sequences of *nad2, cox1*, and *nad4* were shown in Figure 5-7, Figure 5-8, and Figure 5-9. Each of them included 30 clone sequences of Y16, R and M9 and 20 PCR product sequences of wild crabs collected from Rhode River in Chesapeake Bay area. There was no deep sorting pattern among these 50 sequences in any tree, which implied the lack of ancient separation. If NUMTs sequences were represented in the 30 clones, it is reasonable to expect to see them grouped together and/or simply showing longer distance to the real mitochondrial gene sequences, in this

case, represented by the PCR product sequences, because they should have been undertaking different evolutionary paths.

The fact that minor haplotypes have stop codons may inhibit the function of downstream genes and may result in semi or non-functional mitochondrial genomes. Also, these stop codons taken together with non-synonymous substitutions implies a reduced selection pressure for fully functional mitochondrial genomes.

Mitochondrial DNA in the blue crab still can be used as molecular markers despite heteroplasmy because the major haplotype is overwhelmingly dominant in frequency and when minor haplotypes are detected by sequencing they occur consistently within a brood. Thus, these minor haplotypes can be used to define a new distinct haplotype which helps to differentiate individual crabs and gives mitochondrial DNA markers higher resolution.

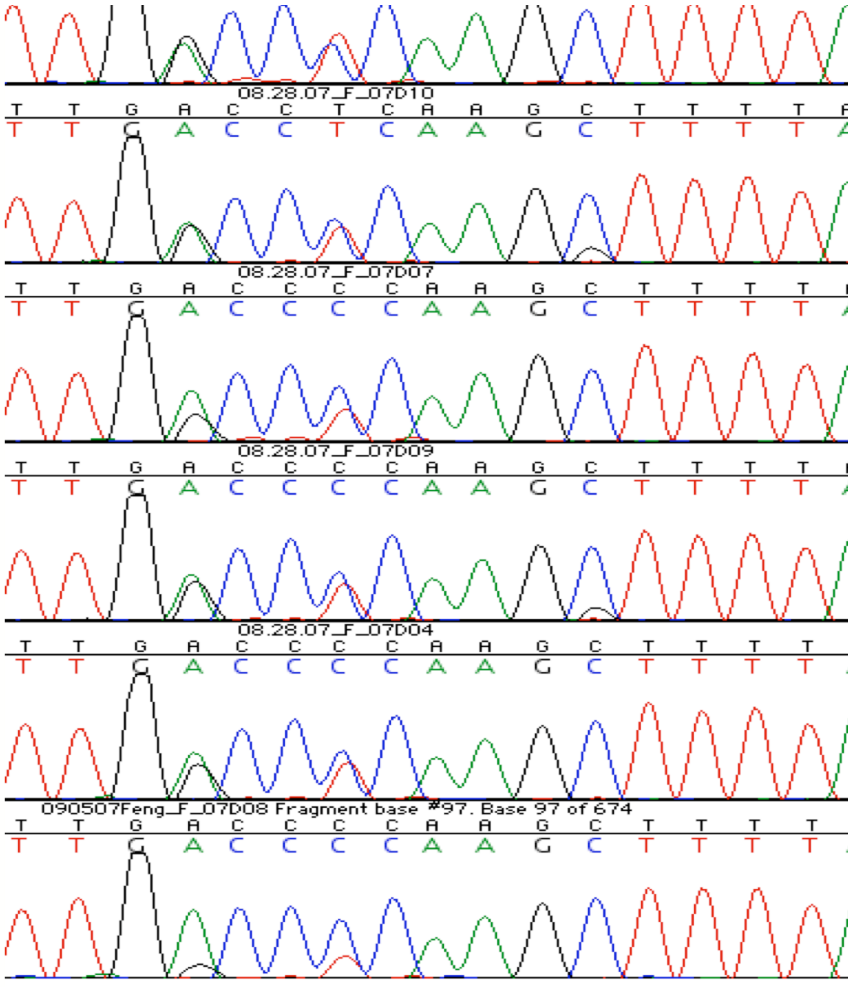

Figure 5-1. The *nad2* gene DNA sequencing chromatograms of six juveniles crabs from a single brood. Two heteroplasmic sites were shown in this figure.

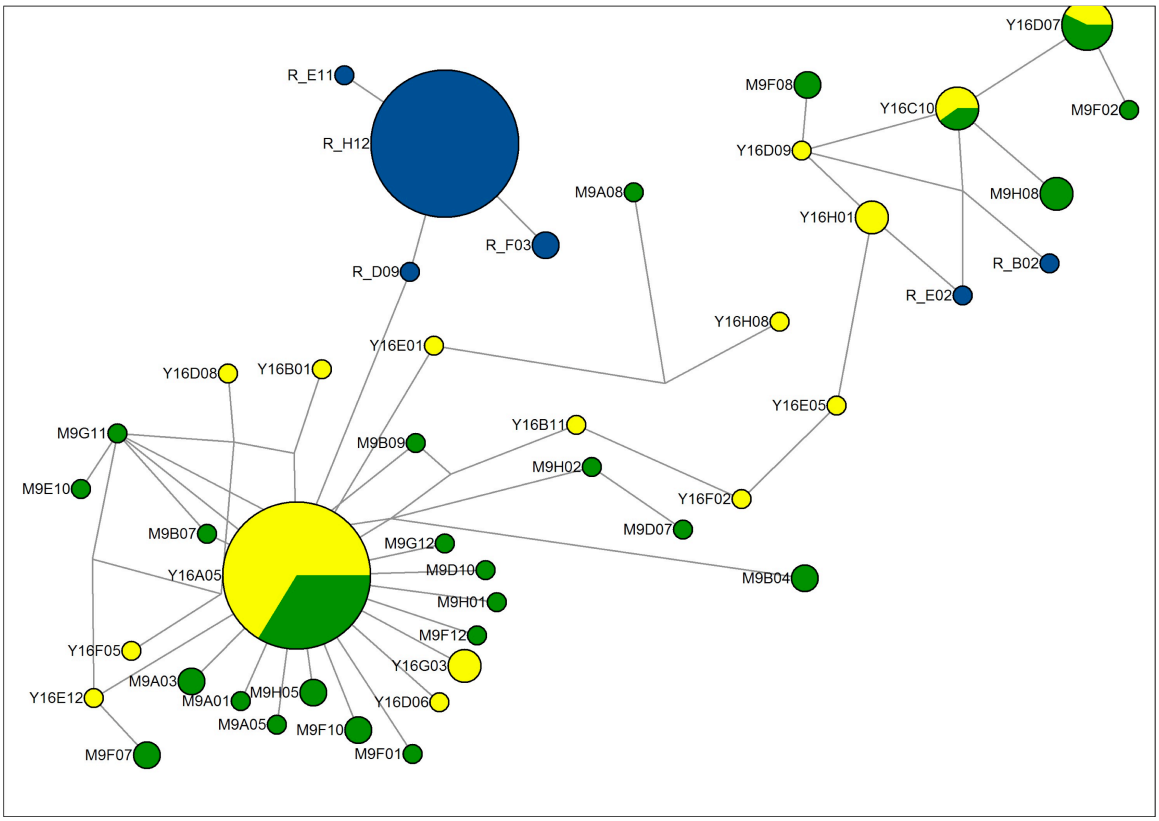

Figure 5-2. Minimum Spanning Network of the crab family's *nad2* DNA clone sequences. Different colors denote the different origins of sequences, blue as the male adult crab R, yellow as the female Y16, and green as the offspring megalope M9. Each circle stands for one distinct haplotype; the size of circles is proportional to the frequency of halotypes; and the branch length is proportional to the genetic distance. This minimum spanning network of all haplotypes was produced with Median Joining calculation (Bandelt, Forster et al. 1995).

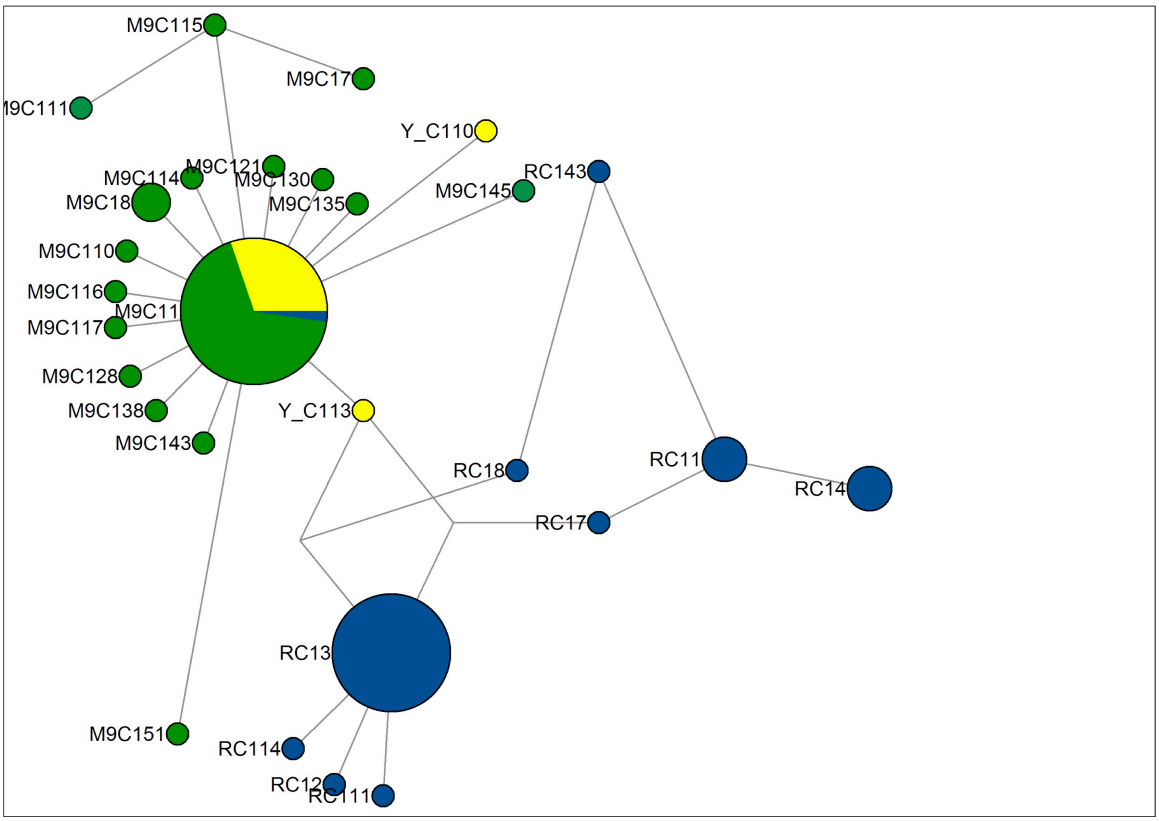

Figure 5-3. Minimum Spanning Network of the crab family's *cox1* DNA clone sequences. Different colors denote the different origins of sequences, blue as the male adult crab R, yellow as the female Y16, and green as the offspring megalope M9. Each circle stands for one distinct haplotype; the size of circles is proportional to the frequency of halotypes; and the branch length is proportional to the genetic distance. This minimum spanning network of all haplotypes was produced with Median Joining calculation (Bandelt, Forster et al. 1995).

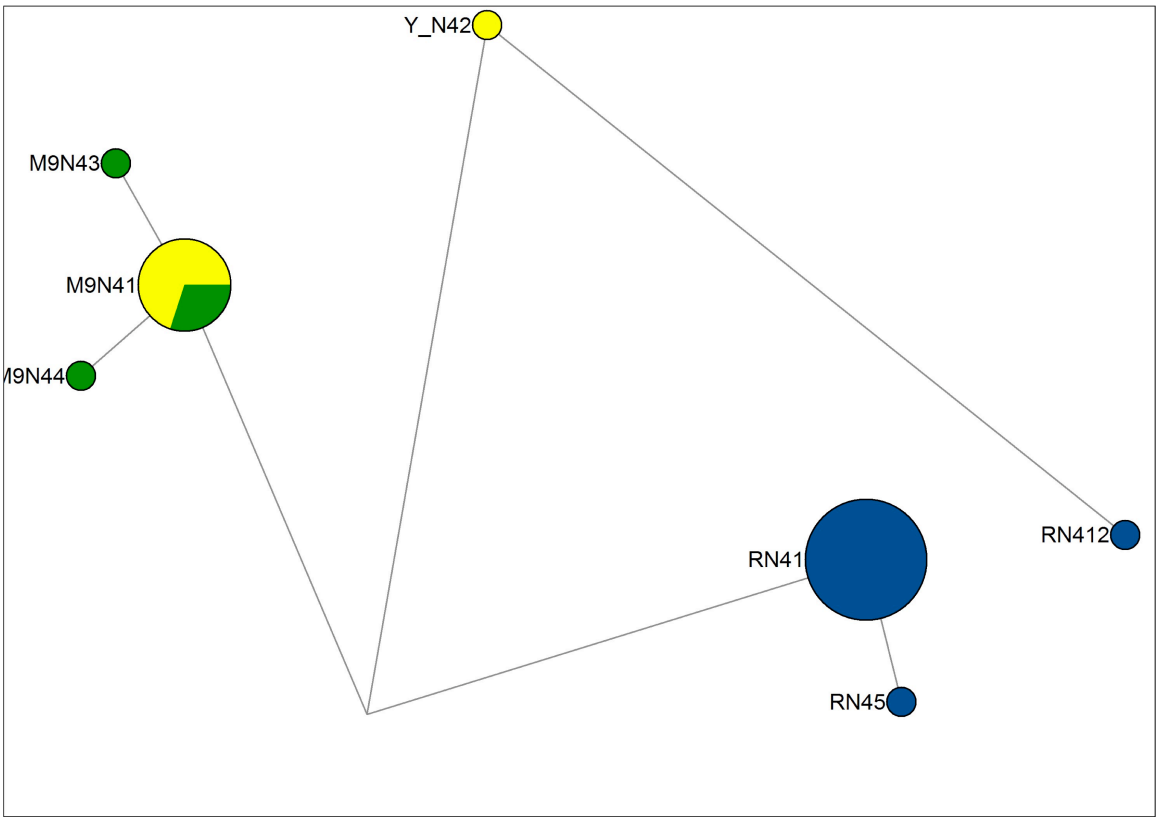

Figure 5-4. Minimum Spanning Network of the crab family's *nad4* DNA clone sequences. Different colors denote the different origins of sequences, blue as the male adult crab R, yellow as the female Y16, and green as the offspring megalope M9. Each circle stands for one distinct haplotype; the size of circles is proportional to the frequency of halotypes; and the branch length is proportional to the genetic distance. This minimum spanning network of all haplotypes was produced with Median Joining calculation (Bandelt, Forster et al. 1995).

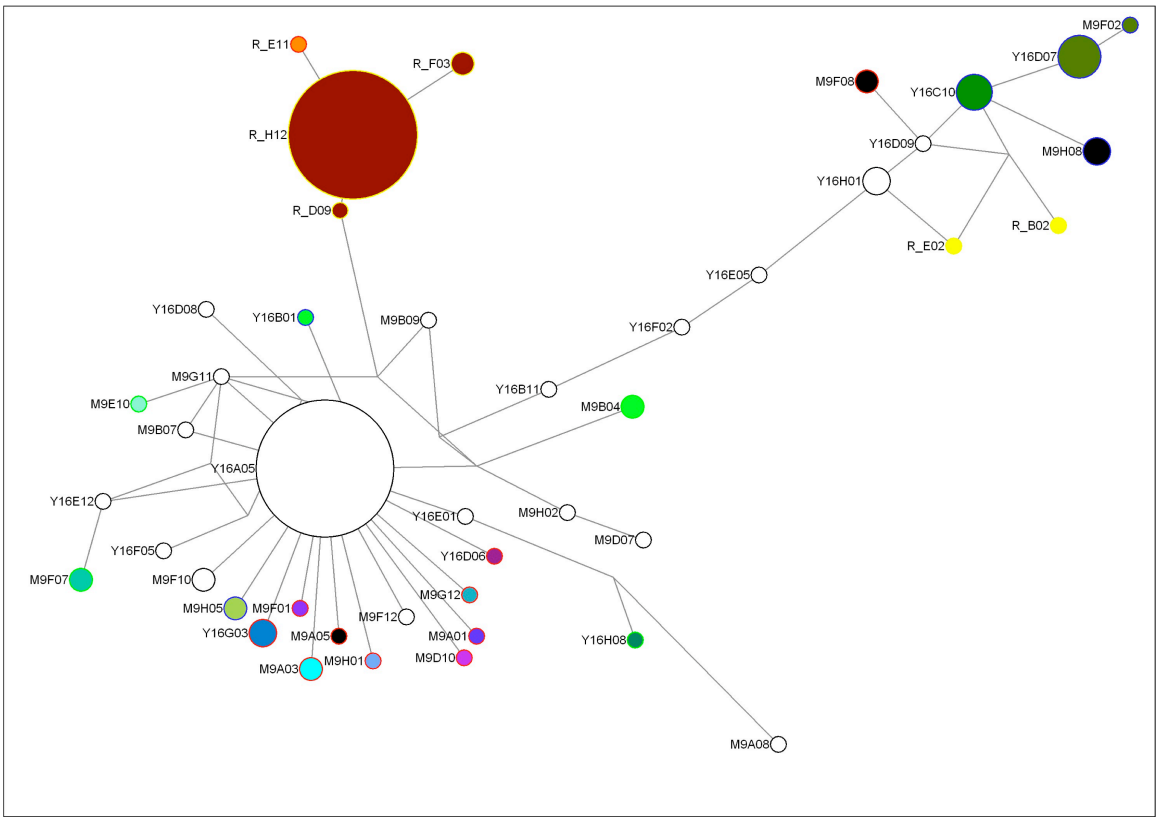

Figure 5-5. Minimum Spanning Network of the crab family's *nad2* DNA sequences. Different colors are used for distinct amino acid sequences. Black circles are the sequences including stop codons. Each circle stands for one distinct DNA haplotype; the size of circles is proportional to the frequency of halotypes; and the branch length is proportional to the genetic distance. This minimum spanning network of all haplotypes was produced with Median Joining calculation (Bandelt, Forster et al. 1995).

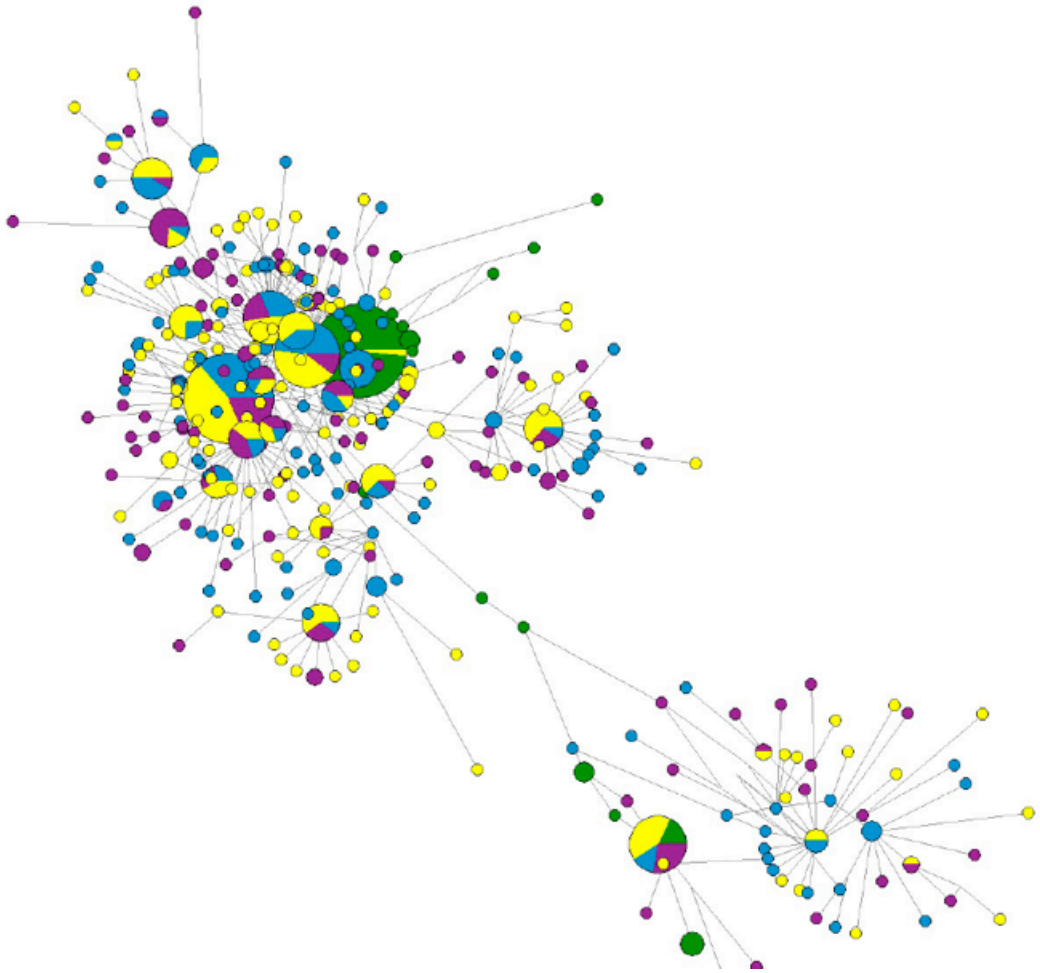

Figure 5-6. Minimum Spanning Network of *nad2* DNA sequences of Y16 and other wild crabs. Different colors denote the origin of sequences, blue as the wild crabs collected from Massachusetts, yellow as the wild crabs from Gulf of Mexico, purple as the wild crab from Chesapeake Bay, and green as the Y16's clones. Each circle stands for one distinct haplotype; the size of circles is proportional to the frequency of halotypes; and the branch length is proportional to the genetic distance. This minimum spanning network of all haplotypes was produced with Median Joining calculation (Bandelt, Forster et al. 1995).

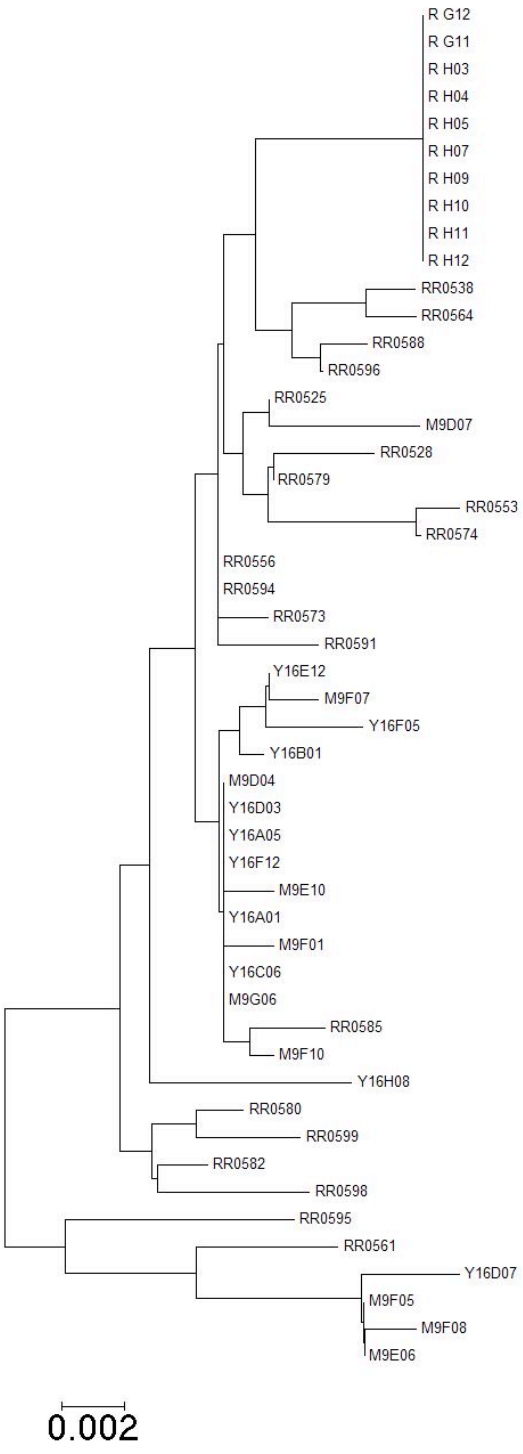

Figure 5-7. A Neighbour Joining tree generated on full *nad2* alignment. All positions weighted equally. 10 clone sequences were randomly selected for each of Y16, R and M9 in addition to 20 random RR wild crabs' PCR sequences to construct the tree.

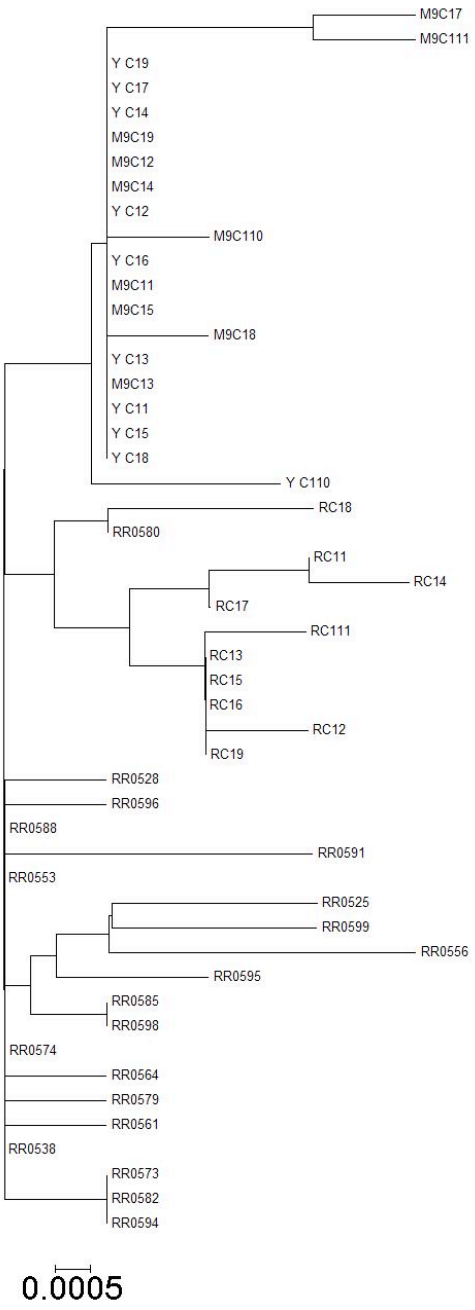

Figure 5-8. A Neighbour Joining tree generated on full *cox1* alignment. All positions weighted equally. 10 clone sequences were randomly selected for each of Y16, R and M9 in addition to 20 random RR wild crabs' PCR sequences to construct the tree.

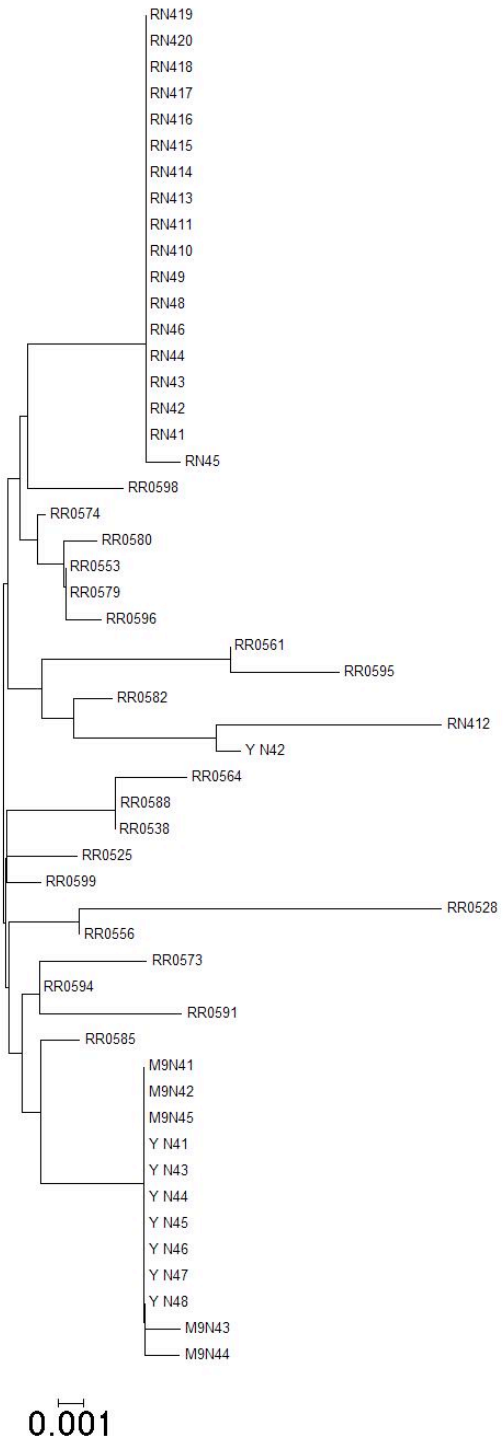

Figure 5-9. A Neighbour Joining tree generated on full *nad4* alignment. All positions weighted equally. 10 clone sequences were randomly selected for each of Y16, R and M9 in addition to 20 random RR wild crabs' PCR sequences to construct the tree.

Summary

The findings

This thesis presents data clearly indicating the unprecedented hyper-variability in the mitochondrial genome of *C. sapidus*. The mtDNA variation extended to multiple regions, including *cox1, nad4*, *12S* and particularly *nad2*. The haplotype diversity of *nad2* gene was always nearly 1, and the nucleotide diversity was about 1%. Even the conserved 12S RNA showed a considerable number of substitutions. The variability of *cox1* and *nad4* was between *nad2* and 12S regions. This hyper-variability in the mtDNA allows using a single mtDNA gene (*nad2*) to distinguish hatchery-produced crabs from wild crabs after release to the wild.

Using *nad2* sequence to study the genetic composition of wild crab populations, I found no dominant haplotypes in the population but instead a low-frequency recurrent haplotypes with a large number of singletons. Because of this high diversity within every sampled populations, the mtDNA lacks geographic structure in wild populations from the Atlantic to Gulf of Mexico.

In addition to the high genetic diversity in populations, mtDNA appears to be heteroplasmic within individual crabs. By cloning high fidelity PCR products, we confirmed that for *cox1, nad4,* and *nad2*, single individual crab/megalopa harbored dozens of copies of the same mitochondrial genes. And the copy number was not saturated with the possibility of inadequate sampling in clones. The heteroplasmy in the blue crab appears to under maternal inheritance without paternal contribution. Some

minor haplotypes were represented in wild populations, which indicated they were fully functional while other minor haplotypes contained stop codons and/or non-synonymous substitutions which may influence the viability of the mitochondria.

The thinkings

C. sapidus is not the only species with such a high diversity in mtDNA. For example, coconut crab (Lavery, Moritz et al. 1996), brown and white shrimps (McMillen-Jackson and Bert 2003; McMillen-Jackson and Bert 2004), the East China Sea mackerel (Shui, Han et al. 2009), and even the Balkan sheep (Cinkulov, Popovski et al. 2008), all have comparable haplotype diversities. However the Balkan sheep had a deep lineage structure that was not often observed in marine species including the animal of our interest. And more importantly, all these studies were done with the control region. In this sense the blue crab is unique with the extreme diversity in its *nad2* gene.

The feature of high genetic diversity plus the lack of deep phylogenetic structure could be in part due to the high fecundity of marine species. One female crab can produce millions of eggs and the effective population size of crabs is usually a very large number even though it is much smaller than the actual population size that is possibly in millions or even billions. Given these enormous numbers, during the reproduction process the hierarchical populations of crab mitochondrial DNA (within-mitochondria; within-oocytes; within organism; and among organisms within population) go through very rapid population expansions to an extreme large population. Thus as a consequence of rapid population expansions, it is reasonable that a number of mutations occur and some of them manage to survive in next generations. Apparently, if the mutations are neutral,

the genetic drift force can't remove them as effectively as in those low fecundity species. Therefore high diversity is not hard to achieve.

Moreover if an individual is heteroplasmic and the amount of its minor haplotypes accumulate to a certain percentage, the heteroplasmy may be amplified in her offspring and then more diversity can be added to the population.

The current model of animal mtDNA inheritance is simplistic with the mtDNA homoplasmy, general uniparental inheriting and no recombination. As more evidence of mtDNA heteroplasmy gets reported, a new model of mtDNA inheritance needs to be established. It will be interesting to study through computer simulations. When a dominant haplotype co-exists with multiple lower-frequency haplotypes, the process of inheritance and the consequences in the population will be much more complicated if there is no strong selection pressure. The heteroplasmic pattern (e.g. the ratio of distinct haplotypes) within individuals will certainly vary among organisms and generations. And another question is whether it will result in an expansion or even a fixation of that heteroplasmic group.

The sustained heteroplasmic mtDNA in blue crabs may have some biological significance. An interesting example is "A Thirty Million Year-Old Inherited Heteroplasmy" in oniscids (Doublet, Souty-Grosset et al. 2008). This G/A single nucleotide heteroplasmy consistently exists across multiple oniscidae species, which suggests this heteroplasmic DNA has survived millions of years in this group. It seems

the persistence of this heteroplasmic DNA is because it encodes two tRNAs ($tRNA^{Ala}$ and $tRNA^{Val}$) due to the G/A polymorphism and this fragment of sequence is the only endogenous source of these two tRNAs in *Armadillidium vulgare* mitochondria.

Not all heteroplasmic sequences are so significant. But heteroplasmy helps to keep the high diversity, which may help the animals when it is needed.

Reference

Abbott, C. L., M. C. Double, et al. (2005). "An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in Thalassarche albatrosses." Molecular Ecology **14**(11): 3605-3613.

Antunes, A., J. Pontius, et al. (2007). "Mitochondrial introgressions into the nuclear genome of the domestic cat." Journal Of Heredity **98**(5): 414-420.

Avise, J. C., J. Arnold, et al. (1987). "Intraspecific phylogeography - the mitochondrial-DNA bridge between population-genetics and systematics." Annual Review Of Ecology And Systematics **18**: 489-522.

Azuma, N., Y. Kunihiro, et al. (2008). "Genetic variation and population structure of hair crab (Erimacrus isenbeckii) in Japan inferred from mitochondrial DNA sequence analysis." Marine Biotechnology **10**(1): 39-48.

Bandelt, H. J., P. Forster, et al. (1995). "Mitochondrial portraits of human-populations using median networks." Genetics **141**(2): 743-753.

Billington, N. (2003). Mitochondrial DNA. Population genetics: principles and applications for fisheries scientists E. M. Hallerman. Bethesda, Maryland., American Fisheries Society. **chapter 4:** 59-100.

Birky, C. W., Jr. (1995). "Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution." Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences **92**(25): 11331-11338.

Birky, C. W., Jr., C. A. Demko, et al. (1978). "Uniparental inheritance of mitochondrial genes in yeast: dependence on input bias of mitochondrial DNA and preliminary investigations of the mechanism." Genetics **89**(4): 615-651.

Birky, C. W., Jr., T. Maruyama, et al. (1983). "An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results." Genetics **103**(3): 513-527.

Black, W. C. and R. L. Roehrdanz (1998). "Mitochondrial gene order is not conserved in arthropods: Prostriate and metastriate tick mitochondrial genomes." Molecular Biology And Evolution **15**(12): 1772-1785.

Bogenhagen, D. F. (1999). "Repair of mtDNA in vertebrates." American Journal Of Human Genetics **64**(5): 1276-1281.

Boore, J. L. (1999). "Animal mitochondrial genomes." Nucleic Acids Research **27**(8): 1767-1780.

Boursot, P., H. Yonekawa, et al. (1987). "Heteroplasmy in mice with deletion of a large coding region of mitochondrial-Dna." Molecular Biology And Evolution **4**(1): 46-55.

Boyce, T. M., M. E. Zwick, et al. (1989). "Mitochondrial-DNA in the bark weevils - size, structure and heteroplasmy." Genetics **123**(4): 825-836.

Branco, J. O. and S. Masunari (2000). "Reproductive ecology of the blue crab, *Callinectes danae* Smith, 1869 in the Conceicao Lagoon system, Santa Catarina Isle, Brazil." Revista Brasileira De Biologia **60**(1): 17-27.

Breton, S., H. D. Beaupre, et al. (2007). "The unusual system of doubly uniparental inheritance of mtDNA: isn't one enough?" Trends In Genetics **23**(9): 465-474.

Breton, S., G. Burger, et al. (2006). "Comparative analysis of gender-associated complete mitochondrial genomes in marine mussels (Mytilus spp.)." Genetics **172**(2): 1107-1119.

Brown, W. M., M. George, Jr., et al. (1979). "Rapid evolution of animal mitochondrial DNA." Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences **76**(4): 1967-1971.

Cao, L., B. S. Ort, et al. (2009). "The control region of maternally and paternally inherited mitochondrial genomes of three species of the sea mussel genus *mytilus*." Genetics **181**(3): 1045-1056.

Chatterjee, A., E. Mambo, et al. (2006). "Mitochondrial DNA mutations in human cancer." Oncogene **25**(34): 4663-4674.

Cinkulov, M., Z. Popovski, et al. (2008). "Genetic diversity and structure of the West Balkan Pramenka sheep types as revealed by microsatellite and mitochondrial DNA analysis." Journal Of Animal Breeding and Genetics **125**(6): 417-426.

Cole, M. A. and R. P. Morgan (1978). "Genetic-variation in 2 populations of blue-crab, *Callinectes-Sapidus*." Estuaries **1**(3): 202-205.

Cook, B. D., S. E. Bunn, et al. (2007). "A comparative analysis of population structuring and genetic diversity in sympatric lineages of freshwater shrimp (*Atyidae : Paratya*): concerted or independent responses to hydrographic factors?" Freshwater Biology **52**(11): 2156-2171.

Crochet, P. A. and E. Desmarais (2000). "Slow rate of evolution in the mitochondrial control region of gulls (*Aves : Laridae*)." Molecular Biology And Evolution **17**(12): 1797-1806.

Dawid, I. B. (1972). "Evolution of mitochondrial DNA sequences in Xenopus." Developmental Biolology **29**(2): 139-151.

Dawid, I. B. and A. W. Blackler (1972). "Maternal and cytoplasmic inheritance of mitochondrial DNA in Xenopus." Developmental Biolology **29**(2): 152-161.

Densmore, L. D., J. W. Wright, et al. (1985). "Length variation and heteroplasmy are frequent in mitochondrial-DNA from parthenogenetic and bisexual lizards (Genus *Cnemidophorus*)." Genetics **110**(4): 689-707.

Doublet, V., C. Souty-Grosset, et al. (2008). "A thirty million year-old inherited heteroplasmy." PLoS ONE **3**(8): e2938.

Egger, J. and J. Wilson (1983). "Mitochondrial inheritance in a mitochondrially mediated disease." New England Journal Of Medicine **309**(3): 142-146.

Excoffier, L., G. Laval, et al. (2005). "Arlequin ver. 3.0: An integrated software package for population genetics data analysis." Evolutionary Bioinformatics Online **1**: 47-50.

Fisher, C. and D. O. F. Skibinski (1990). "Sex-biased mitochondrial-DNA heteroplasmy in the marine mussel *Mytilus*." Proceedings Of The Royal Society Of London Series B-Biological Sciences **242**(1305): 149-156.

Fitz, H. C. and R. G. Wiegert (1991). "Tagging juvenile blue crabs, *Callinectes-sapidus,* with microwire tags - retention, survival, and growth through multiple molts." Journal Of Crustacean Biology **11**(2): 229-235.

Fontaine, K. M., J. R. Cooley, et al. (2007). "Evidence for paternal leakage in hybrid periodical cicadas (*Hemiptera: Magicicada spp.*)." PloS One **2**(9): e892.

Forward, R. B. and D. Rittschof (1994). "Photoresponses of crab megalopae in offshore and estuarine waters: implications for transport. ." Journal Of Experimental Marine Biology Ecology (182): 183–192.

Fu, Y. X. (1997). "Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection." Genetics **147**(2): 915-925.

Fu, Y. X. and W. H. Li (1993). "Statistical tests of neutrality of mutations." Genetics **133**(3): 693-709.

Gissi, C., F. Iannelli, et al. (2008). "Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species." Heredity **101**(4): 301-320.

Gold, J. R. and L. R. Richardson (1990). "Restriction site heteroplasmy in the mitochondrial-DNA of the marine fish *Sciaenops-ocellatus* (L)." Animal Genetics **21**(3): 313-316.

Gray, M. W. (1992). "The endosymbiont hypothesis revisited." International Review Of Cytology-A Survey Of Cell Biology **141**: 233-357.

Hauswirth, W. W. and P. J. Laipis (1982). "Mitochondrial-DNA polymorphism in a maternal lineage of holstein cows." Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences **79**(15): 4686-4690.

Hoarau, G., S. Holla, et al. (2002). "Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish Platichthys flesus." Molecular Biology And Evolution **19**(12): 2261-2264.

Holt, I. J., A. E. Harding, et al. (1990). "A new mitochondrial disease associated with mitochondrial-DNA heteroplasmy." American Journal Of Human Genetics **46**(3): 428-433.

Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogeny. ." Bioinformatics (17): 754-755.

Ingman, M., H. Kaessmann, et al. (2000). "Mitochondrial genome variation and the origin of modern humans." Nature **408**(6813): 708-13.

Kivisild, T., P. Shen, et al. (2006). "The role of selection in the evolution of human mitochondrial genomes." Genetics **172**(1): 373-87.

Kondo, R., Y. Satta, et al. (1990). "Incomplete maternal transmission of mitochondrial DNA in Drosophila." Genetics **126**(3): 657-63.

Kordos, L. M. and R. S. Burton (1993). "Genetic differentiation of Texas Gulf-coast populations of the blue-crab *Callinectes-sapidus*." Marine Biology **117**(2): 227-233.

Kvist, L., J. Martens, et al. (2003). "Paternal leakage of mitochondrial DNA in the great tit (*Parus major*)." Molecular Biology And Evolution **20**(2): 243-247.

Ladoukakis, E. D., C. Saavedra, et al. (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**(4): 755-769.

Ladoukakis, E. D. and E. Zouros (2001). "Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA." Molecular Biology And Evolution **18**(7): 1168-1175.

Lavery, S., C. Moritz, et al. (1996). "Indo-Pacific population structure and evolutionary history of the coconut crab *Birgus latro*." Molecular Ecology **5**(4): 557-570.

Lee, S. B. and J. W. Taylor (1993). "Uniparental inheritance and replacement of mitochondrial DNA in *Neurospora tetrasperma*." Genetics **134**(4): 1063-1075.

Librado, P. and J. Rozas (2009). "DnaSP v5: A software for comprehensive analysis of DNA polymorphism data." Bioinformatics **doi: 10.1093/bioinformatics/btp187.**

Lightowlers, R. N., P. F. Chinnery, et al. (1997). "Mammalian mitochondrial genetics: heredity, heteroplasmy and disease." Trends In Genetics **13**(11): 450-455.

Lin, L. Y., I. P. Cheng, et al. (1990). "Maternal transmission of mitochondrial DNA in ducks." Biochemical And Biophysical Research Communications **168**(1): 188-193.

Lipcius, R. N. and W. T. Stockhausen (2002). "Concurrent decline of the spawning stock, recruitment, larval abundance, and size of the blue crab *Callinectes sapidus* in Chesapeake Bay." Marine Ecology-Progress Series **226**: 45-61.

Liu, Y., R. L. Liu, et al. (2009). "Genetic differentiation between populations of swimming crab *Portunus trituberculatus* along the coastal waters of the East China Sea." Hydrobiologia **618**: 125-137.

Lopez, J. V., S. Cevario, et al. (1996). "Complete nucleotide sequences of the domestic cat (*Felis catus*) mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in the nuclear genome." Genomics **33**(2): 229-246.

Luttikhuizen, P. C., J. Campos, et al. (2008). "Phylogeography of the common shrimp, *Crangon crangon* (L.) across its distribution range." Molecular Phylogenetics and Evolution **46**(3): 1015-1030.

Marchington, D. R., M. S. Scott Brown, et al. (2002). "No evidence for paternal mtDNA transmission to offspring or extra-embryonic tissues after ICSI." Molecular Human Reproduction **8**(11): 1046-1049.

McMillen-Jackson, A. L. and T. M. Bert (2003). "Disparate patterns of population genetic structure and population history in two sympatric penaeid shrimp species (*Farfantepenaeus aztecus* and *Litopenaeus setiferus*) in the eastern United States." Molecular Ecology **12**(11): 2895-2905.

McMillen-Jackson, A. L. and T. M. Bert (2004). "Genetic diversity in the mtDNA control region and population structure in the pink shrimp *Farfantepenaeus duorarum*." Journal Of Crustacean Biology **24**(1): 101-109.

McMillen-Jackson, A. L. and T. M. Bert (2004). "Mitochondrial DNA variation and population genetic structure of the blue crab *Callinectes sapidus* in the eastern United States." Marine Biology **145**(4): 769-777.

McMillen-Jackson, A. L., T. M. Bert, et al. (1994). "Population-genetics of the blue-crab *Callinectes-sapidus* - modest population structuring in a background of high gene flow." Marine Biology **118**(1): 53-65.

Miller, T. J. and E. D. Houde (1998). Blue crab target setting, final repot to U.S. EPA living resources subcommittee. . [UMCES]CBL**:** 22.

Mishmar, D., E. Ruiz-Pesini, et al. (2003). "Natural selection shaped regional mtDNA variation in humans." Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences **100**(1): 171-176.

Moraes, C. T., E. A. Schon, et al. (1989). "Heteroplasmy of mitochondrial genomes in clonal cultures from patients with Kearns-Sayre Syndrome." Biochemical And Biophysical Research Communications **160**(2): 765-771.

Moritz, C., T. E. Dowling, et al. (1987). "Evolution of animal mitochondrial-DNA - relevance for population biology and Systematics." Annual Review of Ecology and Systematics **18**: 269-292.

Munoz, J., A. Gomez, et al. (2008). "Phylogeography and local endemism of the native Mediterranean brine shrimp Artemia salina (*Branchiopoda : Anostraca*)." Molecular Ecology **17**(13): 3160-3177.

Nabholz, B., S. Glemin, et al. (2009). "The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals." BMC Evolutionary Biology **9**: 54.

Nardi, F., A. Carapelli, et al. (2001). "The complete mitochondrial DNA sequence of the basal Hexapod *Tetrodontophora bielanensis*: Evidence for heteroplasmy and tRNA translocations." Molecular Biology And Evolution **18**(7): 1293-1304.

Nei, M. (1987). Molecular Evolutionary Genetics. New York, Columbia Univ. Press.

Nei, M. and F. Tajima (1981). "Genetic drift and estimation of effective population Size." Genetics **98**(3): 625-640.

Nelson, K. and D. Hedgecock (1980). "Enzyme polymorphism and adaptive strategy in the Decapod Crustacea." American Naturalist **116**(2): 238-280.

Nesbo, C. L., M. O. Arab, et al. (1998). "Heteroplasmy, length and sequence variation in the mtDNA control regions of three percid fish species (*Perca fluviatilis, Acerina cernua, Stizostedion lucioperca*)." Genetics **148**(4): 1907-1919.

Obata, M., C. Kamiya, et al. (2006). "Sperm mitochondrial DNA transmission to both male and female offspring in the blue mussel *Mytilus galloprovincialis*." Development, Growth Differentiation **48**(4): 253-261.

Okimoto, R., J. L. Macfarlane, et al. (1992). "The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*." Genetics **130**(3): 471-498.

Palmer, J. D. (1988). "Intraspecific variation and multicircularity in Brassica mitochondrial DNAs." Genetics **118**(2): 341-351.

Papetti, C., P. Lio, et al. (2007). "Antarctic fish mitochondrial genomes lack ND6 gene." Journal Of Molecular Evolution **65**(5): 519-528.

Parr, R. L., J. Maki, et al. (2006). "The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation." BMC Genomics **7**.

Pellegrin, G. J., V. Guillory, et al. (1999). Length-based estimates of total mortality for Gulf of Mexico blue crab. . Proceedings of the Blue Crab Mortality Symposium, Lafayette. Louisiana, Gulf States Marine Fisheries Commission.

Place, A. R., X. J. Feng, et al. (2005). "Genetic markers in blue crabs (*Callinectes sapidus*) II. Complete mitochondrial genome sequence and characterization of genetic variation." Journal Of Experimental Marine Biology And Ecology **319**(1-2): 15-27.

Quesada, H., H. Stuckas, et al. (2003). "Heteroplasmy suggests paternal co-transmission of multiple genomes and pervasive reversion of maternally into paternally transmitted genomes of mussel (Mytilus) mitochondrial DNA." Journal Of Molecular Evolution **57 Suppl 1**: S138-147.

Rogers, A. R. and H. Harpending (1992). "Population growth makes waves in the distribution of pairwise genetic differences." Molecular Biology And Evolution **9**: 552–569.

Ronquist, F. and J. P. Huelsenbeck (2003). "MRBAYES 3: Bayesian phylogenetic inference under mixed models. ." Bioinformatics **19**: 1572-1574.

Rosing, H. S., L. C. Hopkins, et al. (1985). "Maternally inherited mitochondrial myopathy and myoclonic epilepsy." Annals Of Neurology **17**(3): 228-237.

Rotig, A., J. L. Bessis, et al. (1992). "Maternally inherited duplication of the mitochondrial genome in a syndrome of proximal tubulopathy, diabetes mellitus, and cerebellar ataxia." American Journal Of Human Genetics **50**(2): 364-370.

Rugolo, L. J., K. S. Knotts, et al. (1998). "Stock assessment of Chesapeake Bay blue crab (*Callinectes sapidus Rathbun*)." Journal Of Shellfish Research **17**(2): 493-517.

Ryer, C. H., J. vanMontfrans, et al. (1997). "Cannibalism, refugia and the molting blue crab." Marine Ecology-Progress Series **147**(1-3): 77-85.

Safrany, E., V. Csongei, et al. (2007). "Mitochondrial DNA and its mutations: new advances in a new field." Orvosi Hetillap **148**(21): 971-978.

Shitara, H., J. I. Hayashi, et al. (1998). "Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage." Genetics **148**(2): 851-857.

Shoubridge, E. A. and T. Wai (2007). "Mitochondrial DNA and the mammalian oocyte." Current Topics In Developmental Biology **77**: 87-111.

Shui, B. N., Z. Q. Han, et al. (2009). "Mitochondrial DNA variation in the East China Sea and Yellow Sea populations of Japanese Spanish mackerel *Scomberomorus niphonius*." Fisheries Science **75**(3): 593-600.

Slatkin, M. and R. R. Hudson (1991). "Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations." Genetics **129**: 555–562.

Sligh, J. E., S. E. Levy, et al. (2000). "Maternal germ-line transmission of mutant mtDNAs from embryonic stem cell-derived chimeric mice." Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences **97**(26): 14461-14466.

Sodmergen, Q. Zhang, et al. (2002). "Reduction in amounts of mitochondrial DNA in the sperm cells as a mechanism for maternal inheritance in *Hordeum vulgare*." Planta **216**(2): 235-244.

Solignac, M., M. Monnerot, et al. (1983). "Mitochondrial-DNA heteroplasmy in *Drosophila-Mauritiana*." Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences **80**(22): 6942-6946.

Sotelo, G., P. Moran, et al. (2008). "Genetic variation of the spiny spider crab *Maja brachydactyla* in the northeastern Atlantic." Marine Ecology-Progress Series **362**: 211-223.

St John, J. C. and G. Schatten (2004). "Paternal mitochondrial DNA transmission during nonhuman primate nuclear transfer." Genetics **167**(2): 897-905.

Stewart, J. B., C. Freyer, et al. (2008). "Purifying selection of mtDNA and its implications for understanding evolution and mitochondrial disease." Nature Reviews Genetics 9: 657-662.

Tajima, F. (1983). "Evolutionary relationship of DNA sequences in finite populations." Genetics **105**(2): 437-460.

Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics **123**(3): 585-95.

Tajima, F. (1993). Measurement of DNA polymorphism. . Introduction to Molecular Paleopopulation Biology. N. Takahata and A. G. Clark. Tokyo, Japan Scientific Societies Press, Sinauer Associates**:** 37-59.

Tamura, K., J. Dudley, et al. (2007). "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. ." Molecular Biology and Evolution **24**: 1596-1599.

Tatarenkov, A. and J. C. Avise (2007). "Rapid concerted evolution in animal mitochondrial DNA." Proceedings Of The Royal Society B-Biological Sciences **274**(1619): 1795-1798.

Taylor, M. F. J., S. W. McKechnie, et al. (1993). "The Lepidopteran Mitochondrial control region - structure and evolution." Molecular Biology And Evolution **10**(6): 1259-1272.

Taylor, R. W., M. T. McDonnell, et al. (2003). "Genotypes from patients indicate no paternal mitochondrial DNA contribution." Annals Of Neurology **54**(4): 521-524.

Theologidis, I., S. Fodelianakis, et al. (2008). "Doubly uniparental inheritance (DUI) of mitochondrial DNA in *Donax trunculus* (Bivalvia: Donacidae) and the problem of its sporadic detection in Bivalvia." Evolution **62**(4): 959-970.

Van Hove, J. L. K., C. Freehauf, et al. (2008). "Infantile cardiomyopathy caused by the T14709C mutation in the mitochondrial tRNA glutamic acid gene." European Journal Of Pediatrics **167**(7): 771-776.

Vilkki, J., M. L. Savontaus, et al. (1990). "Segregation of mitochondrial genomes in a heteroplasmic lineage with Leber hereditary optic neuroretinopathy." American Journal Of Human Genetics **47**(1): 95-100.

Wai, T., D. Teoli, et al. (2008). "The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes." Nature Genetics **40**(12): 1484-1488.

Wang, J. Y., Y. S. Gu, et al. (2008). "MGB probe assay for rapid detection of mtDNA11778 mutation in the Chinese LHON patients by real-time PCR." Journal Of Zhejiang University-Science B **9**(8): 610-615.

Watterson, G. A. (1975). "Number of segregating sites in genetic models without recombination." Theoretical Population Biology **7**(2): 256-276.

White, D. J., J. N. Wolff, et al. (2008). "Revealing the hidden complexities of mtDNA inheritance." Mol Ecol **17**(23): 4925-4942.

Williams, A. B. (1974). "Swimming crabs of Genus *Callinectes* (Decapoda-Portunidae)." Fishery Bulletin **72**(3): 685-798.

Williams, A. B. (1984). Shrimps, lobsters, and crabs of the Atlantic coast of the eastern United States, Maine to Florida. . Washington, D.C., Smithsonian Institution Press.

Williams, S. T. and N. Knowlton (2001). "Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*." Molecular Biology And Evolution **18**(8): 1484-1493.

Williamson, D. (2002). "The curious history of yeast mitochondrial DNA." Nature Reviews Genetics **3**(6): 475-481.

Wolff, J. N., S. Gandre, et al. (2008). "Delimiting the frequency of paternal leakage of mitochondrial DNA in chinook salmon." Genetics **179**(2): 1029-1032.

Wu, J. M., R. K. Smith, et al. (2000). "Sequence heteroplasmy of D-loop and rRNA coding regions in mitochondrial DNA from Holstein cows of independent maternal lineages." Biochemical Genetics **38**(9-10): 323-335.

Ye, C. Z., Y. T. Gao, et al. (2008). "Association of mitochondrial DNA displacement loop (CA)(n) dinucleotide repeat polymorphism with breast cancer risk and survival among Chinese women." Cancer Epidemiology Biomarkers & Prevention **17**(8): 2117-2122.

Zeviani, M., C. Gellera, et al. (1990). "Tissue distribution and transmission of mitochondrial DNA deletions in mitochondrial myopathies." Annals of Neurology **28**(1): 94-97.

Zouros, E., A. Oberhauser Ball, et al. (1994). "An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*." Proceedings Of The National Academy Of Sciences Of The United States Of America-Biological Sciences **91**(16): 7463-7467.