ABSTRACT

Title of Dissertation: CRYPTIC DIVERSITY, ECOLOGICAL DIFFERENTIATION AND POPULATION GENETICS OF AN ESTUARINE COPEPOD, *ACARTIA TONSA* DANA 1849 (COPEPODA: CALANOIDA)

Gang Chen, Doctor of Philosophy, 2009

Dissertation directed by: Dr. Matthew P. Hare
Assistant Professor, Department of Biology, University of Maryland, College Park
Associate Professor, Department of Natural Resources, Cornell University, Ithaca (Current)

Surprising genetic diversity has been discovered in marine holoplankton, organisms that "drift" in water currents throughout their life cycle. This discovery challenges our assumptions and suggests that holoplankton species may have limited dispersal and/or have adapted to small-scale oceanographic features. In this study, I investigated population genetics of *Acartia tonsa*, a holoplanktonic estuarine copepod containing deeply-diverged mitochondrial lineages, on the United States Atlantic coast. The study goals include: 1) assessing its cryptic species/genetic diversity; 2) inferring evolutionary and geographic origins of its cryptic lineages; 3) testing environmental associations of cryptic lineages; 4) inferring evolutionary and ecological processes/mechanisms underlying population diversification of *A. tonsa*.

Phylogenetic analyses of DNA sequences from two gene loci, mitochondrial cytochrome *c* oxidase subunit I (mtCOI) and nuclear ribosomal internal transcribed spacer (nITS), resolved five morphologically cryptic, genetically diverged lineages that were reproductively isolated species based on genealogical concordance principle. Three co-distributed, deeply-diverged mtCOI lineages (*X*, *S*, *F*) showed significant population differentiation within lineages and contrasting phylogeographic patterns among lineages. Population structures and isolation by distance patterns detected for all lineages suggested that dispersal of *Acartia* lineages was more or less limited to adjacent estuaries; geographic isolation was a key mechanism underlying population diversification of *A. tonsa*. The highly diversified, relatively recent lineage *F* demonstrated a southern center of origin in Florida with northward stepwise diversification. Its distinct localized population structure and strong association with low-salinity environments suggested that environmental stressors (such as salinity) could act as physiological barriers to gene flow, facilitating diversification of *Acartia* populations.

Co-existing *Acartia* lineages were parapatrically distributed along the estuarine gradient across systems on the US Atlantic coast. Genetic, morphological and ecological evidence indicated niche partitioning and ecological differentiation of *A. tonsa* within estuaries. Multiple factors may have contributed to the observed parapatric distribution and niche partitioning, including selection by salinity, biological competition, and/or local adaptation. These findings in one of the best known estuarine copepods reinforce the general conclusion that marine biodiversity is substantially underestimated, not only in terms of species numbers, but also with respect to niche partitioning and the potential importance of ecological divergence in marine holoplankton.

CRYPTIC DIVERSITY, ECOLOGICAL DIFFERENTIATION AND POPULATION GENETICS OF AN ESTUARINE COPEPOD, *ACARTIA TONSA* DANA 1849 (COPEPODA: CALANOIDA)

by

Gang Chen

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:

Dr. Matthew P. Hare, Associate Profess, Chair
Dr. David Hawthorne, Associate Professor
Dr. Thomas D. Kocher, Professor
Dr. Michael R. Roman, Professor
Dr. Jeffrey W. Shultz, Associate Professor
Dr. Sara Via, Professor

Dedicated to my beloved family:

Lizhu Chen (**陈立柱**)

Shuying Chen (**陈淑英**)

Zhong Chen (**陈钟**)

and

Si Chen (**陈思**)

ACKNOWLEDGEMENTS

I owe my sincere gratitude to all the people and institutions involved in the development and completion of this dissertation research for their support and help.

I am most grateful to my advisor, Dr. Matthew P. Hare, former assistant professor of University of Maryland and current associate professor of Cornell University. He provided me the opportunity and lab facilities to initiate and finally finish this study. Without his broad research interests and knowledge, constructive and pertinent advice, and sincere friendship, this study could never be accomplished.

I wish to express my deep gratitude to all my current and past committee members, Dr. David Hawthorne, Dr. Thomas D. Kocher, Dr. Margaret Palmer, Dr. Michael R. Roman, Dr. Kerry L. Shaw, Dr. Jeffrey W. Shultz, and Dr. Sara Via for their constructive suggestions, valuable advice, and constant encouragement during the whole time course of this study.

Many colleagues and institutions helped with the collection of the geographic samples that are critical to this study. I would particularly thank the following people for recruiting me into their research cruises, helping me collect plankton samples, and sharing the environmental data: Drs. Michael R. Roman, David Kimmel, Edward Houde of University of Maryland Center for Environmental Science for sampling Choptank River, Chesapeake Bay in 2004; Dr. Eric K. Wommack of University of Delaware and Dr. Feng Chen of University of Maryland Center of Marine Biotechnology for sampling Delaware and Chesapeake Bay in 2005; Dr. Elizabeth Mann of Skidaway Institute of Oceanography for sampling Savannah estuary; Dr. Matthew P. Hare, Maria Murray, Dr.

Colin Rose for sampling Little Choptank River, Chesapeake Bay in 2002 and Florida lagoons in 2005. Dr. Edward Durbin of University of Rhode Island, Drs. Hans Dam and David Avery of University of Connecticut provided helpful suggestions and guidance on the proper onshore sites for samplings in Rhode Island and Connecticut. Dr. Edward Houde generously provided all the historical zooplankton samples in Chesapeake Bay between 1995–2000. Dr. Marko Reinikainen of University of Helsinki collected and shipped a Finland sample for this study. I would also thank Captains and crews of RVs *Cape Henlopen*, University of Delaware, *Aquarius*, Chesapeake Biological Lab, and *Gannet*, Skidaway Institute of Oceanography, for their help in collecting plankton samples and environmental data. William Connelly, Sukgeun Jung of Chesapeake Biological Lab and Leo Genyuk of University of Delaware spent their precious time answering my questions about cruise environmental databases.

I owe my warmest gratitude to all the old Hare lab members: Safra Altman, Sarah Kingston, Dr. Inken Kruse, Jennifer Murfree, Maria Murray, Dr. Colin Rose, Natasha Sherman, Kathryn Shultziski, Peter Thompson and Nihn Vu. In particular, Colin Rose, Peter Thompson, Maria Murray, Safra Altman and Sarah Kingston built an unforgettably joyful lab environment with their energetic, enthusiastic engagement in scientific research and humorous attitudes toward the PhD life. Their instantaneous assistances with the bench work and frequent discussions on lab research topics greatly helped me carry out and improve my study. I also deeply appreciate their warmest friendship and constant spiritual support.

During the past seven years, many colleagues and graduate students showed their interests in my study and spent their precious time reading my work, listening to my talk,

and providing their thoughtful opinions. Dr. Grace Wyngaard of James Madison University had very thoughtful input in early project development. She also generously shared her nuclear ITS primers designed for freshwater copepods with my study. Horn Point Lab colleagues, Drs. David Kimmel, Jamie Pierson and William Boicourt provided valuable comments and discussions on the first chapter of this dissertation that helped its later publication and improved the followed projects. Drs. Michael Cummings and Adam Bazinet helped with the genealogical sorting analysis, including both scientific discussions of my case study and technical instructions of computer-programmed computation. Graduate students and friends, Julie Byrd, James Penttengill, Michael Lloyd, Jamie Grace, Sky Lesnick read my early draft of dissertation and provided helpful suggestions for its improvement.

I also warmly thank current and past directors of the PhD programs in Behavior, Ecology, Evolution and Systematics (BEES), Drs. Michele Dudash and Gerald Wilkinson, for their help with financial support for this study. This study was partially supported by summer fellowships of BEES program, Department of Biology. The College of Chemical and Life Science provided me a travel award for presenting part of results in Evolution meeting in 2006. I sincerely thank graduate secretaries Lois Reid, Dana Campbell, Abby Kula, and Carrie Soltanoff for their administrative help at every stage from start to finish of my graduate study.

I wish to express my special thanks to a unique group of Chinese scholars and graduate students in Biology and Psychology Departments: Jiakun Song, Yezhong Tang, Xiaohong Deng, Zhen Shi, Kaiwen He, Feng Rong, Haiyan He, Youjun Wang, Huan Luo, Xin Tian, Juanjuan Xiang, Ping Du, Yadong Wang, Qinwen Liu, and Xiaojun Feng.

This is a group with deep cultural roots, wide open mind, bright intelligence, and wonderful success in pursuits of scientific careers with no boundaries. I am proud of living among them and appreciate the experience and lessons shared by each other.

No words can express my love and gratitude to my family: My parents, Mr. Lizhu Chen and Mrs. Shuying Chen, my elder brother Zhong Chen, and my younger sister Si Chen. Living separately on the other half of the planet, they have been trying all possible ways to show me their everlasting love and unconditional support, which always energize me to go through my tough time and pursue my life goals. I would attribute all my research successes to the love and support I received from my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1
Introduction to Dissertation

This dissertation research is centered with a marine copepod *Acartia tonsa* Dana 1849 (Copepoda, Calanoida), studying its recently raised ecological and evolutionary questions. Since this copepod represents an important group of marine organisms that share similar ecological, behavioral and evolutionary features, questions raised for this taxon have much broader implications. In this opening chapter, I attempt to introduce a broad background and general context for this research.

Cryptic diversity of marine holoplankton: a speciation paradox

To a marine ecologist or biological oceanographer, marine organisms are often classified into three groups: plankton, benthos, and nekton, distinguished by their living environment and functional roles in marine ecosystem (Lalli & Parsons 1993). The copepod *Acartia tonsa* belongs to the plankton group, more specifically, a holoplankton subgroup whose members by definition have two basic features: 1) weak swimming capability that only allows them to drift with ocean currents, in contrast to fishes (nekton) that can swim against water flows; 2) lifetime residence in the water column in contrast to benthic invertebrates (benthos) that only enter the planktonic system during early larval stages. Given the continuity of the oceans and their complex circulations, these two features grant holoplankton a high potential for dispersal and frequent population homogenization. As a logical consequence, high species diversity and rapid speciation

were thought to be rare in holoplankton taxa. Early empirical evidence seemed to support this view. For example, holoplankton has more cosmopolitan species with broader geographic ranges (van der Spoel & Pierrot-Bults 1979; van der Spoel & Heyman 1983; Palumbi 1992; Goetze 2003). In an extreme case, identical genotypes of Foraminifers are shared by populations from the two earth poles (de Vargas *et al.* 1999).

However, this traditional view has been increasingly challenged since the 1990s. Many studies using molecular techniques have revealed abundant cryptic species diversity in holoplankton taxa, such as algae (Saez *et al.* 2003), protists (Fenchel 2005; Šlapeta *et al.* 2006), cnidarians (Dawson & Jacobs 2001; Schroth *et al.* 2002), chaetognaths (Thuesen *et al.* 1993; Peijnenburg *et al.* 2004), copepods (Bucklin *et al.* 1996; Lee 2000; Caudill & Bucklin 2004; Goetze 2003, 2005), and euphausiids (Zane *et al.* 1998; Zane *et al.* 2000), etc. Among them, some species previously thought to be cosmopolitan were found to consist of multiple diversified genetic lineages. These findings counter the expected consequence of speciation under frequent gene flow, which here I called “the speciation paradox of holoplankton”.

The holoplankton speciation paradox reflects the gaps in our current knowledge of the taxonomy, systematics, behavior, ecology, evolution and oceanography of this group. It challenges basic assumptions we often made for holoplankton evolution. First, are the seas continuous or structured? Nowadays, there is no doubt about the heterogeneity of marine environment to its inhabitants, especially in estuarine systems (Day *et al.* 1989; Bilton *et al.* 2002). However, what remains to puzzle us is how the seas are structured to holoplankton in the absence of apparent physical barriers. Second, are holoplankton truly global travelers? Does limited dispersal capability play a role in

structuring holoplankton populations? This study on molecular ecology of the copepod *A. tonsa* was developed with an aim to address these interesting questions and seek insights on population diversification and speciation of not only the copepod *A. tonsa* but also marine holoplankton in general.

Recognition of cryptic species using the genealogical approach

It is well documented that genetically diverged marine sibling taxa often lack diagnostic morphological characters (Reviewed by Knowlton 1993), especially in marine holoplankton (Fenchel 2005; Šlapeta *et al.* 2006). Two causes of species crypsis have been proposed. One is called morphological stasis, which states that certain organisms have a stagnant evolution of morphological traits compared to a rapid molecular evolution (Knowlton 1993; Rocha-Olivares *et al.* 2001; Lee & Frost 2002). The other simply is lack of adequate morphological studies. No matter which one rules, cryptic species prevents us from accurately assessing marine biodiversity, discovering processes and mechanisms of marine speciation, and performing species-based ecological studies. Therefore, testing species status of cryptic genetic lineages of *A. tonsa* is one of the major goals of this dissertation research.

Given the long history of debates and controversies over species concepts (e.g. Endler 1989; Hey 2001; Coyne & Orr 2004), it is essential to define the species concept used in this study. In this study, I focused mainly on genealogical species concept (GSC) and used the genealogical approach to test species status of genetic lineages. Briefly, a genealogical species is a most recently derived, exclusive group of organisms all of whose genes coalesce more recently with each other than with those outside the group

(Baum & Shaw 1995). The GSC has two advantages over other species concepts. First, as phylogenetic units, genealogical species inform a speciation history in addition to merely species status. Second, genealogical species recognition does not require the assessment of reproductive compatibility, which is often difficult to perform for marine organisms but critical to biological species concept (BSC, a group of natural populations reproductively isolated from others, Mayr 1942).

Although the BSC and GSC emphasize different consequences of a speciation event, they are congruent with the genealogical concordance principle (Avise & Ball 1990). According to Avise & Ball (1990), a concordant genealogical cladistic structure among independently evolved neutral genes signals a substantially long time period of reproductive isolation between taxa. Thus, concordant genealogies across independent loci provide authoritative evidence for species recognition. The principle has been widely adopted to genetically test species status (e.g. Rising & Avise 1993; Bernardi *et al.* 1993; Koufopanou *et al.* 1997; Schizas *et al.* 1999; Gomez *et al.* 2002; Walker *et al.* 2006). It is conservative and powerful for identification of biological species status, in that complete lineage sorting is the ultimate consequence of speciation. However, it is too conservative to identify incipient species, in which a degree of reproductive isolation has occurred but genealogical sorting of ancient polymorphisms is incomplete. To extend the principle of genealogical concordance to this situation, Cummings *et al.*'s (2008) quantitative genealogical sorting approach was developed to quantify and test the extent of genealogical sorting among non-monophyletic populations. Statistically significant exclusivity of individuals from a paraphyletic clade against a null distribution of random cladistic assignments indicates a substantial divergence between this clade and others

(Cummings *et al.* 2008). Genealogical concordance between non-monophyletic clades with significant genealogical sorting indices also suggests substantial reproductive isolation and good species status of populations. In Chapters 2 and 3 of this dissertation, the genealogical concordance approach was applied to the copepod *Acartia tonsa* to assess the species status of its cryptic genetic lineages.

To test cryptic species in *A. tonsa*, I used neutral genetic markers of both mitochondrial DNA (mtDNA) and nuclear DNA (nDNA), including two mitochondrial gene loci, cytochrome *c* oxidase subunit I (mtCOI) and 16S ribosomal DNA (mt16S), and one nuclear internal transcribed spacer (nITS) gene locus. Mitochondrial COI and mt16S are well studied mtDNA markers for the phylogenetic study of animal populations. Due to maternal inheritance and absence of recombination, Mitochondrial DNA is generally considered a selectively neutral, rapidly evolved single-locus genetic marker (Harrison 1989; Avise 2004). Nuclear ITS is a subportion of the nuclear ribosomal RNA cistron, conventionally including the entire ITS1, 5.8S rRNA gene, and ITS2 region. It has been extensively used in phylogenetic studies of both plants and animals (Baldwin *et al.* 1995; Schulenburg *et al.* 1999; Coleman & Vacquier 2002; Young & Coleman 2004). Nuclear ITS undergoes concerted evolution, a process in which unequal crossing over, gene duplication, or gene conversion homogenizes variation among tandem copies (Zimmer *et al.* 1980; Dover 1982; Moritz & Hillis 1996). Concerted evolution results in less sequence variation within a random mating population compared with neutral expectations, but accentuates the sequence divergence level among rapidly isolated groups (Hershkovitz *et al.* 1999). According to gene coalescent theory (Kingman 1982; Hudson 1990), it is expected to take N_e (effective population size) generations for

mtDNA and $4N_e$ generations for nuclear DNA (nDNA) to coalesce two gene variants under the classic Wright-Fisher neutral mutation model (Ewens 1979). Thus, at least $4N_e$ generations are required to attain gene concordance between mtCOI and nITS loci on average after the onset of reproduction isolation.

Phylogeography study of *Acartia tonsa*

One evolutionary question raised by revelation of cryptic diversity within a species is where those cryptic taxa originated. This is especially important when cryptic species are found in sympatry or parapatry, such as in *A. tonsa*. Their evolutionary origins will inform us how they evolved over space and time. One goal of this study is to reconstruct population history of *A. tonsa* on the Atlantic coast of the United States, recognize possible origins of its genetic lineages, and infer processes and/or mechanisms underlying its population divergence using the phylogeographic approach.

Phylogeography studies geographic distribution of gene copies as well as genealogical lineages within and among closely related species, reconstruct population history and test evolutionary processes and mechanisms underlying the formation of species (Avise 2000, 2004). The approach has been broadly applied to all major groups of organisms, such as plants, fungi, virus, and many animals. It takes advantage of the maternal inheritance and rapid divergence rate of mtDNA to resolve geographic and phylogenetic relationships of mtDNA haplotypes at the population level. In many cases, the genealogical network informatively tells the historical change (origination, migration, expansion, etc.) of populations. The most influential example is the Out-of-Africa hypothesis for human history (Cann *et al.* 1987; Templeton 2002). Nevertheless, inferring

the speciation processes and mechanisms involved in population history often requires sophisticated analyses that integrate population genetics, phylogenetics and statistical methods (Knowles & Maddison 2002; Templeton 2004, 2008). In Chapter 3, I used mtCOI sequence data to construct phylogeographic patterns of *A. tonsa* genetic lineages, resolving their possible origins on the US east coast. Traditional population genetics approaches, such as Analysis of Molecular Variance (AMOVA), isolation by distance (IBD) test, etc. were resorted to test processes and mechanisms underlying observed phylogeographic patterns.

Marine holoplankton and the environment: Niche partitioning

The coexistence of cryptic taxa (e.g., *A. tonsa*) within a study system (e.g., the estuary of Chesapeake Bay) also raises an ecological question of whether they share the same ecological niche, or how they survive interspecific competition. A niche is a species-specific space defined by all the associated ecological factors (Hutchinson 1957). The principle of ecological competitive exclusion predicts that two species cannot occupy the same niche for a long time due to their fierce interspecific competition for resources (Gauze 1934; Hutchinson 1957; Hardin 1960). The way to reduce competition and maximize resource use is to diversify niches. Therefore, cryptic taxa identified in sympatry (or partially co-occurring) can be particularly informative about the niche diversification that has recently occurred within marine habitats.

Sympatric distribution of cryptic species could result from a failure to recognize their distinct niches due to inadequate sampling and inappropriate scale. Careful analysis is needed to understand how, and at what scale, niche partitioning has evolved in high-

dispersal holoplankton species. To recognize niches and niche partitioning, a first step is to describe environmental associations and ecological correlates of cryptic taxa. Highly expected environmental association is reinforced by the strong interactions found between marine organisms and environment. Marine holoplankton species are well known for their sensitive and instantaneous response to many physical, chemical and biological factors of their living environments, thus often regarded as indicators of environmental conditions (e.g. de Vargas *et al.* 1999; Beaugrand 2004; Bonnet & Frid 2004). Although numerous mechanisms could lead to these patterns, environmentally restricted distribution in high-dispersal holoplankton could well signal strong selection along marine environmental gradients. In Chapters 2 and 4, distribution of *A. tonsa* cryptic lineages in relation to environmental factors was investigated, with a goal to test for environmental association and niche partitioning developed in *A. tonsa*.

The study system: *Acartia tonsa* on the US Atlantic coast

Acartia tonsa Dana, 1849 is nominally a calanoid copepod species (Copepoda, Calanoida) with a global wide distribution along coast of Indo-Pacific and Atlantic (Mauchline 1998). Its worldwide colonization history reflects its global dispersal capacity, though the actual dispersal processes are largely unknown. On the east coast of the United States, *A. tonsa* occurs as a seasonally dominant species in estuarine and coastal environments (McAlice 1981; Paffenhöfer & Stearns 1988; Tester & Turner 1991). As a main food of many fish species and major grazer of phytoplankton, *A. tonsa* plays important trophodynamic roles in marine ecosystem, making it one of the best studied copepods in the world (Baird & Ulanowicz 1989; Mauchline 1998).

Acartia tonsa lives its whole life (12 developmental stages) in the water column, with a generation time of 7–25 days (Heinle 1966; Mauchline 1998) and adult longevity of 26 days (Paffenhöfer 1991). Mature males and females conduct sexual reproduction. A mature female has a lifetime fecundity of 435–718 eggs (Zillioux & Gonzalez 1972; Parrish & Wilson 1978). After mating, females release internally fertilized eggs directly to the water at a rate of 18–50 eggs/day (Mauchline 1998). Such reproductive potential along with short generation time of *A. tonsa* often yield natural populations with an enormous size. *A. tonsa* also reportedly produces diapause eggs (eggs that enter developmental dormancy as a physiological response to certain environmental cues), which help populations survive harsh environments (Zillioux & Gonzalez 1972; Marcus 1996). All these life-history features of *A. tonsa* provide ample opportunities for its population expansion, invasion and recolonization of its populations.

Acartia tonsa has been regarded as a single species in many ecological studies. It is thus a species with amazing capabilities of living in a wide range of salinity (0.3 to 36.9 PSU) and temperature (–1 to 35ºC) (Lance 1964; Gillespie 1971; Gonzalez 1974; Johnson & Allen 2005). However, recent genetic studies suggested this cosmopolitan species might instead be a species complex. Caudill & Bucklin (2004) reported four deep genetic lineages (lineages separated by large genetic distances) in *A. tonsa* populations from the US Atlantic coast based on mitochondrial 16S ribosomal gene locus. Chen & Hare (2008) found two lineages (including one novel to Caudill & Bucklin 2004) in Chesapeake Bay and confirmed their status of morphologically cryptic species by genealogical concordance between a mitochondrial DNA (mtCOI) and a nuclear gene locus (nITS). These genetic findings directly motivated this dissertation research for

investigation of cryptic diversity in *A. tonsa* and questions raised about its evolution and ecology.

My study was conducted in two focal geographic systems: 1) Chesapeake Bay, the largest estuary of the United States; and 2) the US east coast from Rhode Island down to Florida. Each system serves different goals of this study. In general, I focused the study on estuarine systems. As a transitional zone between seawater and freshwater, estuaries are known for their unstable heterogeneous environmental and ubiquitous ecological gradients (Day *et al*. 1990). *A. tonsa* living in estuaries has to evolve to deal with the environment. A careful scrutiny of *A. tonsa* population structure in a representative estuary will provide a valuable picture of the *Acartia* story at a smaller scale. Chesapeake Bay well serves this purpose, because it has abundant *A. tonsa* all year around (Kimmel & Roman 2004) and strong environmental gradients due to partially-mixing hydrographic process (Schubel & Pritchard 1987).

The US Atlantic coast has estuaries with different sizes, mixing types, climates, and other distinct physical features. Adequate samples from key estuaries will allow us to test the *A. tonsa* story across the diversity of estuaries. More important, the US east coast sampled in this study extends 1000+ km long, providing a large geographic scale to test dispersal capacity of *A. tonsa*. As I mentioned previously, dispersal seems a key factor determining *A. tonsa* genetic structure. At a scale of 1000+ km, dispersal of *A. tonsa* among estuaries could be restricted by both the physical processes (estuarine and coastal flow patterns) and the biological processes (physiological tolerance of physical stressors). This was discussed in Chapters 3 and 4.

Outline of this dissertation research

The general goals of this study include: 1) assessing cryptic genetic/species diversity of the copepod *A. tonsa*; 2) inferring evolutionary and geographic origins, and phylogeographic history of its deep lineages; 4) testing environmental associations of *A. tonsa* genetic lineages; 5) inferring evolutionary, ecological processes and/or mechanisms governing population diversification of *A. tonsa.*

In the following chapters of this dissertation, Chapter 2 reports a study conducted in Chesapeake Bay. Temporal and geographical populations of *A. tonsa* were intensively sampled to reveal its genetic structure and population divergence. Two gene loci, mtCOI and nITS, were sampled to test species status of two Chesapeake Bay lineages. Distribution of RFLP-genotyped lineages was correlated to environmental factors, testing for niche partitioning and its consistency over time and space.

Chapter 3 reports a phylogeography study conducted over a large geographic scale. Genetic diversity and species diversity were assessed using samples from east coast of the United States (between Rhode Island to Florida). Genealogical concordance was used to test species status for all genetic lineages surveyed. Phylogeographic origins of these lineages were inferred based on their mtCOI phylogeographic patterns. Processes and/or mechanisms governing the population diversification and speciation of *A. tonsa* were also tested.

Chapter 4 reports a study focusing on lineage distribution along ecological gradients within estuaries. Compositions of three common genetic lineages (*F*, *S* and *X*) were surveyed in multiple estuarine systems along the US eastern coast, testing for consistency and robustness of salinity association of genetic lineages. Physical and

biological features of multiple distinct estuaries were explored, looking for factors that contribute to niche partitioning and diversification of *A. tonsa* species complex. Body size of distinct genetic lineages was also assessed, adding evidence for paralleled morphological divergence of *A. tonsa.*

In the last chapter, I review all the results and conclusions of each single chapter and discussed the overall story of copepod *A. tonsa* and its implications to evolution and ecology of marine holoplankton. Future work is proposed to improve our understandings beyond this dissertation research.

Chapter 2
Cryptic ecological diversification of *Acartia tonsa* in Chesapeake Bay

Abstract

The recent discovery of cryptic species in marine holoplankton, organisms that "drift" in oceanic currents throughout their life cycle, contrasts with their potential for long distance passive dispersal and presumably high gene flow. These observations suggest that holoplankton species are adapting to surprisingly small-scale oceanographic features and imply either limited dispersal or strong selection gradients. *Acartia tonsa* is a widespread and numerically dominant estuarine copepod containing deep mitochondrial lineages within and among populations along the Northwestern Atlantic coast. In this study, I investigated *A. tonsa* populations in Chesapeake Bay with the goals of testing species status for the deeply diverged lineages and testing for their association with environmental features over space and time. Phylogenetic analyses of DNA sequences from mitochondrial cytochrome *c* oxidase I (mtCOI) and the nuclear ribosomal internal transcribed spacer (nITS) resolved two concordant monophyletic clades. Deep divergence between the two clades (13.7% uncorrected sequence divergence for mtCOI and 32.2% for nITS) and genealogical concordance within sympatric populations strongly suggest that the two clades represent reproductively isolated cryptic species. Based on restriction fragment length polymorphisms of mtCOI, representatives from the two clades were found consistently associated with contrasting salinity regimes (oligohaline *vs.* meso-polyhaline) with an overlap between 2 and 12 PSU in samples from 1995 to 2005. Finding these patterns in one of the best known estuarine copepods reinforces the

conclusion that marine biodiversity is underestimated, not only in terms of species numbers, but also with respect to niche partitioning and the potential importance of ecological divergence in marine holoplankton.

Introduction

Since the publication of Knowlton's (1993) review on sibling species in the sea it has become ever more apparent that many marine species contain cryptic lineages (e.g., benthic invertebrates, Lazoski *et al.* 2001; Sponer & Roy 2002; fishes, Ruzzante *et al.* 1996; Colborn *et al.* 2001; and mammals, Rosel *et al.* 1994; Dalebout *et al.* 2006). "Sibling" and "cryptic" both refer to taxa that are difficult to distinguish morphologically even if greater scrutiny ultimately reveals diagnostic phenotypic characters. Cryptic taxa identified in sympatry (at least partially co-occurring) can be particularly informative about the niche diversification that has recently occurred within marine habitats. These insights are expected based on the ecological competitive exclusion principle that predicts coexisting species must have partitioned their resource use so as to minimize overlap and reduce competition (Gauze 1934; Hutchinson 1957). Because species often modify their geographic range, niche diversification by competitive exclusion may or may not be associated with species origins, but it does require some measure of reproductive isolation. In marine environments reproductive isolation may be facilitated at small spatial scales by phenological differences (Ruzzante *et al.* 1996) or by differential tolerances/preferences along environmental gradients of depth (Kruse & Reise 2003), salinity (Bekkevold *et al.* 2005), temperature (Hilbish *et al.* 1994), or pollutants (Schizas *et al.* 2001; Rocha-Olivares *et al.* 2004). In principle, species with the

potential for broad dispersal (10~100+ km) are least likely to speciate across small scale (1~10km) habitat heterogeneities or adapt to small scale habitat features. Exceptions to this generalization are of particular interest because they indicate taxa with unappreciated dispersal capabilities (philopatry) or those experiencing particularly strong selection gradients.

Holoplankton, a group of small marine organisms that live in the water column throughout their entire life cycle, are by definition drifters or weak swimmers generally moving less than one cm/sec. Long distance dispersal is expected to be common in these species based on the potential for advection in oceanographic currents and the broad geographic species ranges commonly described (van der Spoel & Heyman 1983). Few non-fossil data have been available bearing on mechanisms of speciation for these taxa, many of which are pan-global (Palumbi 1992; Goetze 2003). More recently molecular studies have documented abundant cryptic species diversity in holoplankton taxa such as algae (Saez *et al.* 2003), protists (Fenchel 2005; Šlapeta *et al.* 2006), cnidarians (Dawson & Jacobs 2001; Schroth *et al.* 2002), chaetognaths (Thuesen *et al.* 1993; Peijnenburg *et al.* 2004), copepods (Bucklin *et al.* 1996; Lee 2000; Caudill & Bucklin 2004; Goetze 2003, 2005), and euphausiids (Zane *et al.* 1998; Zane *et al.* 2000). Many of these cases are consistent with allopatric divergence across recognized continental dispersal barriers or between major oceanic gyres. The permeability of these barriers, however, apparently depends on species-specific oceanographic habitat preferences (Goetze 2005). Even at smaller scales, when cryptic taxa co-occur geographically, their association with distinct environmental features suggests ecological specialization and local adaptation. In one of the most remarkable examples, some of the planktonic foraminiferal taxa whose fossil

record of tests (hard shells) have been instrumental for paleontological dating, are now known to consist of multiple species (de Vargas *et al.* 1999; de Vargas *et al.* 2001; Darling *et al.* 2000). In one of the most common globally distributed species, *Orbulina universa*, three cryptic species were associated with ecologically specialized niches defined by chlorophyll concentration and upwelling (de Vargas *et al.* 1999). These cryptic species each shared haplotypes across oceans but had patchy patterns of abundance within oceans suggesting high dispersal but species-specific, condition-dependent reproduction (de Vargas *et al.* 1999). Careful analysis of similar cases is needed to understand how, and at what scale, niche partitioning has evolved in high-dispersal holoplankton species.

Acartia tonsa Dana, 1849 is a calanoid copepod species widely distributed in the coastal waters of the world's oceans (Mauchline 1998). It is a seasonally dominant consumer near the base of the food web in many estuaries of the western North Atlantic (Paffenhöfer & Stearns 1988; Tester & Turner 1991; Johnson & Allen 2005). In contrast to morphological similarities uniting individuals of this species, *A. tonsa* are capable of living in a wide range of salinities (0.3 to 36.9 PSU) and temperatures (-1 to 35ºC), consistent with a generalist estuarine phenotype (Lance 1964; Gillespie 1971; Gonzalez 1974; Johnson & Allen 2005). In Chesapeake Bay, on the eastern coast of the United States, *A. tonsa* occurs year round as a major component of the zooplankton community and an important food base for fishes (Brownlee & Jacobs 1987; Kimmel & Roman 2004). Average density of *A. tonsa* in Chesapeake Bay varies from 1700 to 6200 individuals/m^3 and peaks in spring and summer (Kimmel & Roman 2004). Generation time in Chesapeake Bay ranges from 7 to 13 days (Heinle 1966).

A recent study revealed deep molecular divergence within *A. tonsa* populations along the northwestern Atlantic coast (Caudill & Bucklin 2004). Based on mitochondrial 16S ribosomal DNA sequence data, Caudill & Bucklin (2004) resolved four deeply diverged phylogenetic clades with uncorrected sequence divergences of 10–14%. These macrogeographic mtDNA results suggest that *A. tonsa*, one of the best known and intensively studied copepods in the world, may represent a cryptic species complex with multiple lineages co-occurring within estuaries. However, several processes can generate or assemble diverse mitochondrial lineages within a single biological species, such as retention of ancient polymorphism in large populations or hybridization (Ballard & Rand 2005). Therefore, a major goal of this study was to test if the diverged mitochondrial lineages in *A. tonsa* represent reproductively isolated species. If so, then the phenotypic diversity and ecological roles of *A. tonsa* may be better understood as a partitioning of niches among differently adapted forms rather than evolution of an ecological generalist. To test for reproductive isolation, we took the genealogical concordance approach that has been widely applied in other taxa (Rising & Avise 1993; Schizas *et al.* 1999; Bernardi *et al.* 1993; Koufopanou *et al.* 1997; Gomez *et al.* 2002; Walker *et al.* 2006). If *A. tonsa* mtDNA clades are also supported by independent nuclear markers, we can conclude that these lineages have been reproductively isolated for a sufficiently long time to consider them species under the biological species concept (Moore 1995; Palumbi *et al.* 2001; Turelli *et al.* 2001; Rosenberg & Nordborg 2002).

Another important finding on population divergence in *A. tonsa* was that the same deeply diverged mtDNA lineages coexisted in multiple estuaries (Caudill & Bucklin 2004). This suggests that, if the mitochondrial lineages are indeed reproductively isolated

species, sampling at a finer scale and testing for environmental correlates may reveal niche partitioning. It is worth noting that some processes, such as recent range expansion and anthropogenic introduction can also create sympatry of cryptic species. This is especially likely in temperate regions responding to a recent glacial retreat (Hewitt 2000; Hewitt 2001) or estuaries heavily impacted by human activities (Carlton & Geller 1993; Lavoie *et al.* 1999; Ruiz *et al.* 2000). However, expected patterns of evolutionary equilibrium, such as isolation by distance (Hutchison & Templeton 1999), or temporal stability of subdivision patterns, can provide evidence for *in situ* evolution and against recent non-equilibrium processes.

I had two major goals in this study. First, we examined both mitochondrial and nuclear gene loci to test for cryptic species in *A. tonsa* within Chesapeake Bay, the largest estuary in the United States. Second, we extensively surveyed the distribution of *A. tonsa* genetic lineages in Chesapeake Bay relative to environmental features to test for niche partitioning and its consistency over time and space.

Materials and Methods

Copepod samples

A total of 1649 *A. tonsa* individuals from 50 samples were examined in this study (Table 2-1). These samples were collected from Chesapeake Bay during the time period of 1995–2005. Multiple research cruises were involved, including two cruises for Atlantic Coast Environmental Indicators Consortium in Chesapeake Bay (http://www.ACEINC.org); eight cruises for Trophic Interaction in Estuarine Systems,

project of Chesapeake Bay Large Marine Ecosystem Research (LMER-TIES, http://www.chesapeake.org/ties/ties.html); and one cruise for Microbial Observatory of Virioplankton Ecology (MOVE, K. Eric Wommack, University of Delaware). The authors collected 8 additional samples from shore. Standard plankton nets with a mesh size of 125–280 μm were towed either horizontally or vertically to collect zooplankton samples from different water layers (surface, bottom, above or below pycnocline). Samples were immediately preserved in 95% ethanol.

DNA extraction, amplification, and sequencing

In the laboratory, *A. tonsa* and congeneric species *A. hudsonica* individuals (adult and later larval stages CIV, CV) were identified and sorted out on the basis of diagnostic morphological characters (Bradford-Grieve 1999; Sabatini 1990). To extract genomic DNA, single individuals were first rehydrated in de-ionized water for more than 2 hrs and then heated at 99°C in 100–150 μl of 5% chelex solution (Bio-Rad Laboratories) for 8 min. After centrifugation at 2288 *g* for 5 min, supernatant DNA solutions were collected and stored at 4°C for genetic analysis.

Two gene fragments, mitochondrial cytochrome *c* oxidase subunit I (mtCOI) and nuclear ribosomal DNA internal transcribed spacer (nITS), were amplified by polymerase chain reaction (PCR). Amplifications used mtCOI primers 1490 and 2198 (Folmer *et al.* 1994) and universal ITS primers ITS-4 and ITS-5 (White *et al.* 1990). In order to relate results here with a previous study (Caudill & Bucklin 2004), we also amplified and sequenced a portion of mitochondrial 16S ribosomal DNA (mt16S) using universal 16S ar/br primers (Palumbi *et al.* 1996). For all genes, PCR amplifications were performed in

a 25 μl reaction volume with 1× Invitrogen buffer, 2.5 mM $MgCl_2$, 125μM dNTPs, 0.2 μM of each primer, 0.2 μg/μl BSA, 0.3 units of Invitrogen *Taq* polymerase, and 0.5–1 μl of genomic DNA. The optimized PCR conditions started with DNA denaturing at 95°C for 1 min, followed by 35 cycles of 30 sec denaturing at 95°C, 30 sec annealing at 50°C, and 50 sec extension at 72°C, and then a final extension at 72°C for 7 min. PCR amplifications were confirmed in a 1.2% agarose gel with ethidium bromide staining. This protocol was effective for all genes and most DNA samples. For sequencing, PCR products were purified by incubation with shrimp alkaline phosphatase and exonuclease I (USB Biochemical Corp.) for 30 min, and cleaned by isopropanol precipitation. Sequencing reactions used Big Dye terminator chemistry as recommended by the manufacturer (Applied Biosystems) and the same primers as used for PCR. Sequencing products were analyzed on an ABI 3100 DNA sequencer (Applied Biosystems).

Sequence data analyses

I edited DNA sequences using SEQUENCHER (Genecode Corp.). Accuracy of base calls was always confirmed by comparing both strands. Mitochondrial COI and mt16S sequences had no heterozygous base calls as expected for mtDNA. Some double base calls were observed in both strands of nITS, and we used the criteria of Hare and Palumbi (1999) to infer heterozygosity. It is unknown to what degree observed heterozygosity was within a locus versus across ITS copies in the genome. Therefore haplotype inferences based on Hardy-Weinberg assumptions were not appropriate and sequences with two or more heterozygous bases were excluded from our ITS dataset due to their ambiguous haplotypes. Sequence alignments of mt16S and nITS were done using

Clustal X v1.83 (Thompson *et al.* 1997) with default parameter settings and subsequently optimized manually. No insertions or deletions (indels) were found in mtCOI so the alignment from SEQUENCHER was used.

For each gene, phylogenetic relationships among sequences were assessed using the neighbor-joining (NJ), maximum parsimony (MP), and maximum likelihood (ML) methods implemented in computer software PAUP*4.0b10 (Swofford 2003). Uncorrected distance, Kimura-2-parameter (for mtCOI), and Tamura-Nei (for nITS and mt16S) corrected distances were used to construct distance trees, as suggested by Kumar *et al*. (1993). Heuristic searches were used among phylogenetic trees to find the minimum total tree length (NJ), smallest number of character state changes (MP), or maximum likelihood (ML). These searches were started from trees made by random sequence addition with 10 replicates, continued by tree-bisection-reconnection (TBR) branch-swapping. To test the reliability of inferred topologies, we performed bootstrapping tests with 1000, 500, and 100 pseudoreplicates for NJ, MP and ML trees respectively. One mtCOI sequence of *A. hudsonica* collected from Delaware Bay was used as an outgroup to root the mtCOI gene trees. The alignment between *A. tonsa* and *A. hudsonica* nITS sequences was too uncertain to provide a useful outgroup root, so midpoint rooting was applied to the nITS trees. To compare the tree topologies for mtCOI and nITS, a subset of sequences derived from the same individuals were analyzed as above.

DNA polymorphism was examined for both mtCOI and nITS genes, the latter assuming diploidy for comparative purposes. Nucleotide diversity (π, Nei 1987), haplotype diversity (*H*, Nei 1987), and Tajima's *D* (Tajima 1989) were calculated for the

entire sequence data sets (excluding outgroup) and for two distinct clades separately using MEGA v4.0 (Tamura *et al.* 2007) and DnaSP v4.10.4 (Rozas *et al.* 2003). Fu & Li's *D* and *F* statistics (Fu & Li 1993) were also used to test neutrality within clades using a representative sequence from the alternate clade as an outgroup. MtCOI nucleotide and haplotype diversities between two clades were compared and tested using *t*-test (Nei 1987). A relative rate test (Tajima 1993) implemented in MEGA was conducted using a sequence from the congeneric species *A. hudsonica* as an outgroup to test for rate heterogeneity between sequences from the two distinct mtCOI clades.

RFLP analyses

Sequence differences between distinct *A. tonsa* mtCOI clades were examined for diagnostic restriction fragment length polymorphisms (RFLP). Five units of *Hae*III endonuclease (New England Biolabs) were used to digest 5–10 μl of mtCOI PCR amplicon in a total volume of 15 μl with 1× NEB 2 buffer at 37°C for 3 h. The digested DNA fragments were visualized in 2% agarose gels with ethidium bromide staining.

I correlated the relative frequency of the two distinct mtCOI lineages (expressed as the percentage composition of lineage *S* in a sample, %*S*) with major environmental parameters, including temperature, salinity, dissolved oxygen, turbidity, and chlorophyll. For shipboard samples, real time environmental parameters for each sampling station, including depth, temperature, salinity, chlorophyll, turbidity and dissolved oxygen, were recorded using CTD instruments and averages across copepod sampling depths are reported. For shore-based samples, single measures of salinity and temperature were

taken at one meter depth with an YSI EC300 Conductivity/Temperature instrument (YSI Environmental Co.). Both Pearson's and Spearman's correlation coefficients were calculated between %*S* and each environmental parameter. Logistic regression models were fit to describe the quantitative relationships between %*S* and salinity expressed in practical salinity units (PSU, numerically equivalent to parts per thousand of salt by weight). Vertical stratification is a notable physical feature of some sampling sites relative to others, characterized by a relatively sharp salinity gradient (halocline) between upper and bottom mixing layers. To characterize stratification among our samples two indices were considered: total vertical difference in salinity through the water column (ΔS) and maximum rate of salinity change per meter (max PSU/m). All correlation and regression analyses were done with SAS v6.12 (SAS Institute Inc., Cary, NC, USA).

Results

Mitochondrial DNA

A 564 bp mtCOI fragment was analyzed. No insertions or deletions (indels) were needed for alignment. Predicted amino acid sequences had no stop codons within the reading frame. Among 76 *A. tonsa* individuals we recorded 29 distinct mtCOI haplotypes (Table 2-2, GenBank accession nos. EU274436 to EU274464). Within *A. tonsa*, 101 sites (17.9%) were polymorphic. Of these polymorphisms, 15 occurred at 1st codon positions and 85 at 3rd codon positions, producing a total of 3 predicted amino acid polymorphisms. Eighty-nine polymorphisms occurred in two or more sequences and were therefore parsimony-informative, but most of these distinguished two major clades. One sequence

from *A. hudsonica* (GenBank accession no. EU274431) was obtained and used as an outgroup for phylogenetic analysis.

Two clades separated by long branches were supported by bootstrap values of 100 in trees from all phylogenetic methods (Fig. 2-1). Individuals belonging to each clade were associated with contrasting salinity regimes (see below), so we designated the two clades saline (*S*) and freshwater (*F*) lineages. Net sequence divergence between the two clades was 13.7% based on uncorrected distance. Within clades *S* and *F*, nucleotide diversity was low (0.001–0.012, Table 2-3). Both nucleotide and haplotype diversity were significantly higher in clade *F* than clade *S* (Table 2-3, one-tailed *t*-test, $P < 0.001$ for all tests). High mtCOI nucleotide diversity in the *F* lineage was largely caused by three haplotypes (H27–H29) that formed a distinct clade with 72–88% bootstrap support and occurred in some but not all tributaries. Tajima's *D* was significantly positive if *S* and *F* clades were tested together as if they were one species (Table 2-3). Within clades Tajima's *D* was negative but only significant for clade *S*. Additional within-clade tests, Fu and Li's *D* and *F*, were not significant ($P > 0.10$). No sequences from the two clades showed significantly different substitution rates relative to the outgroup *A. hudsonica* (χ^2 test, $P > 0.35$).

Mitochondrial 16S sequences were also obtained from 14 *F*-lineage and 2 *S*-lineage individuals. Four mt16S haplotypes were found and combined with 19 *A. tonsa* haplotypes and 4 congeneric outgroup sequences from GenBank (Table 2-4). Based on 134 bp of overlapping sequence and neighbor joining analysis (Fig. 2-2), one clade with 100% bootstrap support contained the *S*-lineage haplotypes from this study along with the common B-lineage haplotypes described by Caudill & Bucklin (2004). Haplotypes

corresponding to the *F* lineage in this study formed an exclusive clade supported by 84% bootstrap support. The *F*-lineage mt16S clade had a sister relationship to Caudill & Bucklin's C-lineage clade found only in Texas.

Nuclear ITS

The amplified nITS sequence lengths from 43 individuals were between 533 and 581 bp. No indels were required to align each of two sequence sets. When the two groups were aligned together, however, Clustal default alignment parameters resulted in insertion of 17 indels. A central 346 bp unambiguous alignment containing 11 indels was used for all phylogenetic analyses (Table 2-5).

There were 106 (30.7%) polymorphic sites in *A. tonsa*, of which 105 distinguished two major sequence sets. One sequence set had no variation and the second set had 3 polymorphic nucleotide positions after removal of six sequences with two or more heterozygous sites. All of the variable positions in the second set were observed in both heterozygous and homozygous condition. A total of five alleles (GenBank Accession nos. EU274426 to EU274430) were recorded among $2n = 74$ chromosomes (assuming diploidy). Recombination appeared minimal based on the 4-gamete test of Hudson & Kaplan (1985) even with consideration of the unphased sequences (results not shown).

All phylogenetic methods separated two divergent nITS clades with strong bootstrap support (only neighbor-joining shown; Fig. 2-3): one clade contained only haplotype ITS1 and the other had haplotypes ITS2–5. Net sequence divergence between the two clades was 32.2% based on uncorrected distance. Within clade ITS2-5 sequence

diversity was low (0.002, Table 2-3) but haplotype diversity was moderately high (Table 2-3). Both nucleotide and haplotype diversity were significantly higher in the ITS2-5 clade corresponding to the *F* lineage than in the ITS1 clade (one-tailed *t*-test, $P < 0.001$). Tajima's *D* was non-significant within each clade, and significantly positive if both clades were tested together (4.805, $P < 0.001$, Table 2-3). Fu & Li's tests did not reject neutrality ($P > 0.10$). Very similar results were obtained when the PHASE program (Stephens *et al*. 2001) was used to infer ITS haplotypes from the full data set (results not shown).

Gene Tree Concordance

For 21 individuals with data from both mtCOI and nITS, phylogenetic trees were constructed using the neighbor joining method. The two gene trees showed complete concordance with respect to the well-supported major clade partition: individuals within each mtCOI clade (*S* and *F*) were also grouped into reciprocally monophyletic clades on the nITS tree (Fig. 2-3). The nITS clade with haplotype ITS1 corresponded to mtCOI lineage *S*, and clade ITS2-5 to lineage *F*. Thus, the higher level of diversity seen in mtDNA clade *F* relative to *S* was also seen in the corresponding nITS clades.

RFLP of mtCOI

The *Hae*III endonuclease assayed a RFLP distinguishing haplotypes in the two mtCOI clades. Two cutting sites between positions 36–37 and 176–177 in the sequences of *F* clade yielded a big fragment of 466 bp and two small fragments of 140 bp and 104

bp from the amplicon. In contrast, no *Hae*III cutting sites were found in *S* clade sequences due to polymorphisms at positions 37 (A/C), 175 (A/G) and 178 (T/C) (Table 2-2). A total of 1649 *A. tonsa* individuals were PCR-RFLP genotyped using *Hae*III. Both lineages were widespread within Chesapeake Bay, but lineage *S* was found predominantly in downstream areas and lineage *F* in upstream areas of the Chesapeake mainstem (Fig. 2-4A) and Chesapeake tributaries (Fig. 2-4B). Lineage *S* proportion of each sample (%*S*) showed a statistically significant correlation with salinity (Fig. 2-5, Pearson's $r = 0.58$, $P < 0.001$; Spearman's $r = 0.77$, $P < 0.001$). No significant correlations were found between %*S* and depth, temperature, chlorophyll, turbidity or dissolved oxygen. A logistic regression model describing the relationship between %*S* and salinity for all samples was highly significant:

$$\%S = \frac{100}{1+e^{1.0335-0.3874Sal}}$$

($N = 42$; Max-rescaled $R^2 = 0.65$; Likelihood ratio test, $P = 0.0011$).

The total vertical salinity difference, ΔS, and maximum rate of vertical salinity change were correlated across all samples (Pearson's $r = 0.86$, $P < 0.001$). The frequency distribution of ΔS was bimodal with only one sample between 1.4 and 3.7 PSU difference (data not shown). Thus, we classified samples with $\Delta S < 2$ as unstratified and > 2 as stratified. Twenty-two unstratified samples were evenly divided between the Chesapeake mainstem and tributary sites and had maximum rates of change from 0 to 1.34 PSU/m (mean = 0.39). In contrast, stratified sites ($\Delta S > 2$) were almost all in mainstem waters and had significantly sharper haloclines (maximum rate of change: 1.1 to 3.3 PSU/m, mean = 1.86, one-tailed *t*-test, $P < 0.001$). Unstratified and stratified sample sets each

showed logistic relationships distinct from the total data set but only the unstratified model was significant (Fig. 2-5)

For unstratified samples only:

$$\%S = \frac{100}{1+e^{2.4990-0.5159Sal}}$$

(N = 22; Max-rescaled R^2 = 0.82; Likelihood ratio test, P = 0.0014)

For stratified samples only:

$$\%S = \frac{100}{1+e^{-0.0976-0.3043Sal}}$$

(N = 20; Max-rescaled R^2 = 0.57; Likelihood ratio test, P = 0.1417)

Discussion

Cryptic species within Acartia tonsa in Chesapeake Bay

Based on extensive sampling of nominal *A. tonsa* within Chesapeake Bay I have found two deeply diverged mitochondrial lineages, *F* and *S*, in populations inhabiting relatively fresh and saline waters, respectively. Only the *S* lineage corresponds with previously described cryptic *Acartia* diversity. Mid-Atlantic estuaries were not sampled by Caudill & Bucklin (2004), but they reported the *S* lineage (their lineage “B”) in *A. tonsa* from as far away as New Hampshire and Texas and found it sympatric with additional deep mt16S lineages unobserved here (Fig. 2-2). Based on mt16S the *F* lineage is reciprocally monophyletic with a clade containing haplotypes found only in Texas by Caudill & Bucklin (2004; "C" lineage). Without geographically intermediate samples it is

impossible to confirm the evolutionary independence of the Chesapeake *F* and Texas C clades.

Two results support our inference that *F* and *S Acartia* lineages are reproductively isolated cryptic species rather than deep ancestral polymorphisms within a species with large effective population size. First, independently evolving mtCOI and nITS loci showed identical genealogical partitioning of *F* and *S* lineages with strong bootstrap support. Second, the existence of this multilocus concordance in co-occurring *Acartia* populations suggests that reproductive barriers are at least partly intrinsic and on the whole keep reproductive exchange, if any, minimal or nonexistent.

The deep mtCOI and nITS divergences separating the *F* and *S* clades are also consistent with species status. The mtCOI divergence is comparable to those reported between recognized species in other calanoid copepods (9–25%, Bucklin *et al.* 2003). However, based on mt16S sequences, percent sequence differences among *A. tonsa* lineages (10–14%) were only half that measured among other nominal *Acartia* species (19–28%, Caudill & Bucklin 2004).

To conclude that *A. tonsa* lineages *F* and *S* are good species based on genetic data we must rely on assumptions of orthology and neutrality for the assayed markers, aspects that are frequently uncertain (e.g. Nguyen *et al.* 2002; Antunes & Ramos 2005; Ballard & Rand 2005; Schmitz *et al.* 2005; Bazin *et al.* 2006). Inadvertent amplification of nuclear pseudogenes of mitochondrial origin can generate unexpectedly high divergence levels within nominal species (Zhang & Hewitt 1996). In this study, each mtCOI sequence was without stop codons or indels and no heterozygous nucleotide bases were recorded, all results consistent with orthologous amplifications. Genealogical concordance between

loci is also unexpected if either gene tree is based on paralogous comparisons. Neutrality of mtCOI within clades (the appropriate intraspecific testing level) was not rejected by any test except Tajima's *D* in lineage *S*. Tajima's *D* was negative in both lineages, a trend consistent with population expansion or recent genetic hitchhiking (Tajima 1989; Simonsen *et al.* 1995). Fu & Li's tests use outgroup sequences to improve the power to detect recent genetic hitchhiking (Fu 1997) but were not significant for *S* lineage. Thus, we tentatively conclude that non-neutral patterns in mtCOI reflect population expansion. Selective neutrality is also consistent with the lack of variation in mtCOI evolutionary rates between lineages *F* and *S*. The ITS locus is multicopy in eukaryotic genomes, widely believed to be evolving neutrally, and typically evolves under concerted evolution (Gerbi 1985; Hillis & Dixon 1991). Concerted evolution is expected to reduce variation within species and accelerate divergence between them, possibly contributing to the exceptionally deep divergence between and low polymorphism within lineages in ITS relative to mtCOI. These features of ITS make it difficult to interpret patterns of polymorphism but do not invalidate the inference of longstanding reproductive isolation based on genealogical concordance.

I did not find hybrids between lineages *F* and *S* using the nuclear ITS marker, but neither the absence of hybrids nor the presence of multilocus genealogical distinctions between *F* and *S* rule out hybridization. This is because hybrids are easily missed if they are rare or restricted temporally or spatially (Barton & Hewitt 1985). More intensive sampling in lineage overlap zones and laboratory cross-lineage breeding experiments are required to test whether reproduction isolation between lineage *F* and *S* is complete and due to intrinsic barriers.

Inadequate study may have contributed to the morphological crypsis of these *Acartia* species. Individuals from lineages *F* and *S* all had a suite of diagnostic characters previously used to separate *A. tonsa* from other *Acartia* species, including presence of rostral filament, spinule pattern of the reproductive segment, gender-specific shape and structure of the specified fifth legs, etc. (Bradford-Grieve 1999). Because cryptic *Acartia* species co-occur within some estuarine waters, previous morphological studies could easily have missed important distinctions within mixed samples (Garmew *et al.* 1994) or failed to sample *Acartia* lineages with restricted environmental distributions. Careful morphological comparison and description of *S* and *F* lineages is currently underway with a goal to find characters that distinguish them.

Morphological stasis may also be a general feature of copepod diversification as it has been reported in other copepod species including *Cletocamptus deitersi* (Gomez *et al.* 2004; Rocha-Olivares *et al.* 2001), *Eurytemora affinis* (Lee & Frost 2002), and *Acanthocyclops vernalis* (Dodson *et al.* 2003). The diversity of habitats used by these taxa suggests that morphological stasis in copepods does not depend entirely on habitat features (Hebert 1998). However, environmental stresses experienced by these taxa (e.g., gradients in temperature, salinity, and dissolved oxygen) may focus their adaptive diversification on physiological more than morphological traits (Knowlton 1993). If this is the case, traditional taxonomic systems largely based on morphological traits may have substantially underestimated marine biodiversity in spatially heterogeneous estuarine and intertidal environments (Bilton *et al.* 2002).

Ecological divergence and niche partitioning of A. tonsa *cryptic species*

A striking finding of this study was the strong correlation between salinity and *A. tonsa* lineage frequency (Fig. 2-5), indicating ecological divergence and niche partitioning of *A. tonsa* lineages. Lineage *F* was primarily associated with oligohaline waters (0.3–12 PSU) and lineage *S* with meso-polyhaline waters (2–26 PSU, up to the highest salinity examined here). The association of lineage *F* with lower salinity waters was observed in multiple Chesapeake tributaries and the mainstem of the bay, as well as in the spring and late summer of one year (Table 2-1, Fig. 2-4). The association was also observed as early as 1996 in the Chesapeake mainstem whereas drought conditions in 2005 elevated upper mainstem salinity relative to 1996. Accordingly, upper mainstem sites were dominated by *S* lineages in 2005 (compare salinity for samples M1–M3 versus T3–T11, Fig. 2-4). The temporal and spatial consistency of the salinity association suggests that it is neither ephemeral nor due to a single point-source introduction of a non-indigenous *F* lineage.

Salinity is likely to be a strong selective agent maintaining observed spatial distributions of *Acartia* lineages *S* and *F*, although it is possible that unmeasured covariates also are important (Lakkis 1994). Previous ecological studies have shown that salinity largely restricts the distribution of *A. tonsa* to estuaries (Tester & Turner 1991). Naupliar survival was found to be optimal at salinities of 20–25 PSU but much lower at salinities greater than 25 PSU (Tester & Turner 1991). Also, for *A. tonsa* collected from mesohaline waters, laboratory egg production at 15 PSU was significantly lower than at 20–35 PSU (Castro-Longoria 2003). I hypothesize that the salinity-associated niche

partitioning in Chesapeake *A. tonsa* results from adaptation by *S* and *F* lineages to different optimal salinities. To test this hypothesis, common garden experiments will be needed to measure fecundity and survivorship of each lineage across a range of salinities. A corollary to this hypothesis is that associations are driven by differential recruitment and selection in lineages *S* and *F* across salinity gradients, with little ability for individuals to track preferred salinity. If, on the other hand, copepod swimming ability is at all important for tracking preferred salinity then the pattern of association and the degree of spatial overlap between lineages might depend on hydrographic characteristics of an estuary.

In estuaries where tidal mixing is too weak to completely homogenize fresh and salt water, vertical stratification is generated by low density fresh water moving down estuary over higher density saltwater moving up estuary (Day *et al.* 1989). The large Chesapeake Bay system often contains, simultaneously, regions where salinity primarily varies horizontally (in tributaries with low to moderate fresh water input) while other regions are vertically stratified and have both horizontal and vertically salinity gradients (Schubel & Pritchard 1987). How well do weakly swimming *Acartia* regulate their distribution relative to vertical salinity gradients? Paired vertical comparisons were made at three strongly stratified sites in the Chesapeake mainstem and in every case the trend in *S*% was in the expected direction, higher in deeper more saline waters (Fig. 2-4). Diel vertical migrations (DVM) of ~15 m have been observed in *A. tonsa* (Cuker & Watson 2002) and most studies on copepod DVM attribute it primarily to trade-offs between predator avoidance and feeding (Hays 2003). Even though copepod swimming is too slow to directly counter currents, many larval planktons use vertical positioning within

layered currents of partially mixed estuaries to dramatically affect their spatial position prior to metamorphosis and settlement (Naylor 2006). If stratification, associated with opposing shear flows, is important for *A. tonsa* to track preferred salinities via vertical migration (Wooldridge & Erasmus 1980; Kimmerer & Mckinnon 1987), then *S* and *F* lineage distributions should be more discrete in stratified waters. Alternatively, if vertical stratification presents too sharp or dynamic a salinity gradient for *A. tonsa* lineages to track their preferred salinity then a stronger salinity association should be found along the more gradual, uni-dimensional, horizontal salinity gradients in unstratified waters. Observational data on *Eurytemora affinis* in a strongly tidal, unstratified estuary suggests that opposing shear flows are not necessary for copepods to track preferred salinities (Hough & Naylor 1991). However, comparisons of copepod distribution between stratified and unstratified waters are necessary to determine their relative capabilities in the range of habitats they will encounter.

In the logistic model describing the dependence of %*S* on salinity there is a large variance within intermediate salinities when all data are examined together (Fig. 2-5). At intermediate salinities of 2–10 PSU the *Acartia* samples from stratified waters, typically drawn from the less saline waters above a halocline, had higher %*S* than samples from unstratified sites of comparable salinity (Fig. 2-5). Assuming that salinity-associated fitness optima are constant across these two environments, one interpretation is that copepods in stratified waters were having more difficulty maintaining their preferred position within salinity gradients. This effect may be visible in Fig. 2-5 by examining sample pairs above and below haloclines at three sites. The difference in %*S* for each pair is much less than predicted by any of the logistic curves. Also, the variance around the

logistic model for unstratified samples is low relative to stratified samples or all samples combined (Fig. 2-5; rescaled R^2: 0.82 for unstratified vs. 0.57 for stratified and 0.65 for all sites). It may not be stratification *per se* that constrains the tracking of water masses by copepods, but the fact that conditions producing stratification are also likely to create more dynamic salinity gradients in space and time. Unlike the seasonally stable horizontal salinity gradient along the Choptank river axis (Fig. 2-4B, spring vs. summer, paired *t*-test, $P = 0.435$), the horizontal salinity gradient along the mainstem axis was not monotonic. For example, sites T4/5 and T6/7 had relatively lower surface salinity than the more upstream site T3 in 1996. Salinity patchiness over time and space in the mainstem is expected because of seasonally varied and spatially complex freshwater inputs from multiple tributaries (Schubel & Pritchard 1987). Thus, we tentatively conclude that the mechanisms used by *A. tonsa* lineages *S* and *F* to track preferred salinities are more efficient in vertically unstratified waters, suggesting that vertical migration across shear zones is not a primary determinant of *Acartia* position in the estuary.

Origins of A. tonsa *cryptic species in Chesapeake Bay*

Historical climate change has been hypothesized as a major force shaping current biogeographic patterns of *A. tonsa* along the U.S. Atlantic coast (McAlice 1981; Caudill & Bucklin 2004). Caudill & Bucklin (2004) associated the existence of deeply diverged *A. tonsa* lineages to Pleistocene glacial cycles, hypothesizing that populations diverged within isolated glacial refugia and subsequently expanded to co-inhabit many Atlantic estuaries. Similarly, diversification in a roughly co-distributed planktonic copepod,

Eurytemora affinis, seems to include allopatric mechanisms acting over broad regions (Lee 2000) coupled with recent osmoregulatory adaptive divergence within watersheds (Lee 1999).

In this report I have focused on the temporal and meso-spatial analyses of previously unreported associations between *Acartia* lineages and salinity. Determining how niche partitioning evolved in these taxa will require additional samples and experiments. Ongoing efforts to expand the geographic scope of our analyses suggest that the *F* lineage is not restricted to Chesapeake Bay. Furthermore, the non-sister relationship of *F* and *S* clades in the mt16S tree suggests that neither lineage originated in Chesapeake Bay. In principle, application of a molecular clock could help reject a Chesapeake origin for the *F* and *S* lineage split because the Chesapeake Bay formed as Pleistocene glaciers melted only ~10,000 years ago (Hobbs 2004). Unfortunately no copepod-specific clock calibration has been reported. In some well studied crustacean taxa the evolutionary rate of mtCOI divergence has been estimated to be around 2% MY^{-1} for example 1.4–2.6% for snapping shrimp *Alpheus* (Knowlton & Weigt 1998; Knowlton *et al.* 1993) and 1.66–2.33% for Jamaican crabs Grapsidae (Schubart *et al.* 1998). Applying a rate of 2% to *A. tonsa* mtCOI, the splitting of lineages *F* and *S* is estimated to have occurred 3.7 MYA, equivalent to at least 18.5 million copepod generations (5–6 gen/yr at Delaware Bay, Jeffries 1962). At this time the balance of evidence indicates that *F* and *S* lineages did not diverge *in situ* in Chesapeake Bay and provisionally dates their origin before Pleistocene glacial cycles.

The *S* lineage was the numerically and geographically dominant species in our samples. Even so, its genetic diversity was low relative to expectations for a species with

enormous census numbers and significantly lower than in the *F* lineage. When mt16S haplotypes from New England, Georgia and Texas were analyzed together with Chesapeake representatives total within-clade *S*-lineage diversity remained surprisingly low (Fig. 2-2). One possibility is that the *S* lineage experienced a genetic bottleneck during recent introduction to the western North Atlantic. This hypothesis requires a single introduction and subsequent spread through the western North Atlantic or multiple introductions from a single low-diversity source population. Both scenarios are plausible given that planktonic copepods are often transported in ballast water (Carlton & Geller 1993; Cordell & Morrison 1996; Lavoie *et al.* 1999). Broader geographic sampling is required to identify potential source populations.

Table 2-1 Sampling information of *Acartia tonsa* populations in Chesapeake Bay, including geographic location, maximum depth of a sampled site (D_{max}), sampled depth range (D_s), average salinity across sampled depth (PSU), RFLP sample size (n) and estimated proportion of *S* lineage (%*S*), and sequence sample sizes for *F* and *S* lineages. Salinity measurement details are in Materials and Methods. n.d.- no data.

								n COI seq		2*n* ITS seq	
Sample name	Latitude	Longitude	Sampling date	D_{max} (m)	D_s (m)	Salinity (PSU)	n RFLP (%*S*)	*S*	*F*	*S*	*F*
Bay Mainstem											
T1	39.33	76.20	July 22, 1996	10.0	0-9	0.34	16 (6.3)				
T2	39.23	76.24	July 22, 1996	9.0	0-6	2.46	34 (47.1)		2		
T3	39.18	76.28	July 22, 1996	10.1	0-6	3.96	36 (75.0)				
T4	39.11	76.30	May 6, 1996	9.0	0-3	1.54	33 (81.8)				
T5	39.11	76.30	May 6, 1996	9.0	6-9	6.88	36 (91.7)				
T6	39.06	76.33	May 5, 1996	10.0	0-3	2.46	36 (69.4)				
T7	39.06	76.33	May 5, 1996	10.0	7-10	6.20	37 (73.0)				
T8	39.00	76.36	May 6, 1996	19.0	0-3	3.33	38 (78.9)				
T9	39.00	76.36	May 6, 1996	19.0	16-19	10.12	36 (88.9)				
T10	39.07	76.40	July 23, 1996	12.0	0-5	5.91	43 (90.7)				
T11	38.83	76.41	July 20, 1996	16.1	0-5	7.66	30 (96.7)				
T12	38.00	76.17	July 17, 1996	6.0	0-6	11.04	19 (100.0)				
T13	37.67	75.95	July 18, 1996	6.0	0-6	14.80	59 (100.0)				
T14	37.67	76.20	Oct. 25, 1999	9.0	6-9	19.97	16 (100.0)				
T15	37.67	76.20	Oct. 19, 2000	10.0	7-10	18.99	17 (100.0)				
T16	37.66	76.19	Nov. 3, 1995	12.1	0-5	21.00	21 (100.0)				
T17	37.50	76.08	Nov. 1, 1997	11.7	0-3	24.08	17 (100.0)				
T18	37.50	76.08	Oct. 20, 1998	12.0	9-12	23.80	23 (100.0)				
T19	37.33	76.15	Oct. 30, 1996	8.0	0-5	17.31	21 (100.0)				
T20	37.05	76.03	Oct. 21, 1998	9.0	0-3	26.42	10 (100.0)				
M1	39.17	76.33	May 25, 2005	6.3	0-5	8.36	46 (100.0)				
M2	38.96	76.38	May 25, 2005	26.1	0-5	8.29	29 (100.0)				
M3	38.75	76.43	May 25, 2005	26.0	0-5	9.38	44 (100.0)				
M4	38.30	76.28	May 27, 2005	21.9	0-5	10.26	19 (100.0)				
M5	38.06	76.22	May 27, 2005	25.0	0-5	11.32	40 (100.0)				
A1	38.57	76.44	Mar. 29, 2004	21.2	0-5	11.44	48 (81.3)				
A2	38.57	76.50	Mar. 29, 2004	8.6	0-5	10.70	19 (94.7)				
A3	38.57	76.44	Aug. 9, 2004	19.6	0-5	8.03	43 (100.0)				

Table 2-1 *continued*

Sample name	Latitude	Longitude	Sampling date	D_{max} (m)	D_s (m)	Salinity (PSU)	n RFLP (%*S*)	*n* COI seq		2*n* ITS seq	
								S	*F*	*S*	*F*
A4	38.57	76.50	Aug. 9, 2004	8.4	0-5	8.47	38 (97.1)				
Choptank River											
C1	38.72	76.01	Mar. 30, 2004	9.7	0-5	1.63	21 (20.0)				
C2	38.66	75.96	Mar. 30, 2004	5.9	0-3	6.04	63 (36.5)		4	4	8
C3	38.58	76.02	Mar. 30, 2004	8.2	0-5	9.08	52 (88.7)				
C4	38.65	76.17	Mar. 30, 2004	8.2	0-5	10.50	40 (90.0)				
C5	38.64	76.33	Mar. 29, 2004	9.6	0-5	10.97	38 (94.7)				
C6	38.72	76.01	Aug. 10, 2004	9.7	0-5	3.33	22 (22.7)		1		2
C7	38.66	75.96	Aug. 10, 2004	6.4	0-5	5.61	36 (22.2)				
C8	38.58	76.02	Aug. 10, 2004	12.3	0-5	8.96	27 (74.1)				
C9	38.65	76.17	Aug. 9, 2004	7.1	0-5	10.79	47 (100.0)				
C10	38.64	76.33	Aug. 9, 2004	9.4	0-5	11.12	43 (100.0)				
G1	38.61	75.95	Aug. 8, 2003	< 3.0	0-2	n.d.	6 (33.3)		4		4
G2	38.57	76.06	July 13, 2003	n.d.	0-3	n.d.	5 (100.0)	3		6	
G3	38.59	76.13	July 12, 2003	< 3.0	0-2	n.d.	55 (78.2)	18	4	4	4
Patuxent River											
PA1	38.50	76.67	Aug. 4, 2004	< 3.0	0-1	n.d.	23 (34.8)	4	9	6	2
PA2	38.41	76.57	Aug. 4, 2004	< 3.0	0-1	n.d.	12 (100.0)				
PA3	38.40	76.53	May 27, 2005	8.6	0-5	7.76	34 (100.0)				
Potomac River											
PO1	38.19	76.71	May 26, 2005	6.4	0-5	4.20	44 (84.1)	5	5		
PO2	38.11	76.54	May 26, 2005	5.1	0-5	6.67	33 (84.8)				
Baltimore Inner Harbor											
BIH	39.28	76.61	July 31, 2003	< 3.0	0-2	n.d.	88 (44.3)	4	5	8	18
Little Choptank River											
L1	38.52	76.27	Apr. 8, 2002	n.d.	0-2	n.d.	3 (100.0)	3		2	
L2	38.52	76.27	Aug. 15, 2002	n.d.	0-2	n.d.	53 (100.0)	5		6	

Table 2-2 Mitochondrial COI sequence alignment of 29 *Acartia tonsa* haplotypes recorded in Chesapeake Bay. Only polymorphic sites are listed. Site numbers are labeled based on 564 bp fragment. Two groups (H1–8 and H9–29) are separated and shown relative to different reference sequences (H1 and H9) by using dots to denote matching nucleotides. Polymorphic sites between the two reference sequences are shaded.

Haplotype	Frequency	00000000000000000000000111111111111111112222222222222222223333333333333333334444444444444555555555555
		00001111233334445677889122445566777789990122223455566778990012233444555667890122224567789000122233345
		14782346212473690439254847281506025840368113672536902140281732817036589179219812482165676256403625849
H1	33	AAATGAAACATAATAGCTTAGTTGATATTTTATAATTATTTACTATATCGACAATTCCCACCTCAATTCGCTTTAAATTTATTAGATGAATTTATAAAGCT
H2	2	..................C..
H3	2	......G.....................C...................T..
H4	1	........................G..
H5	1	............................C..
H6	1	....................................C..
H7	1	....T.....................................C..............
H8	1	....C...
H9	5	ACGTGGAGTGCTCAGATTTTATCAGACACCAACGGCTTCACGTAGTCTTATTAGAAACTGTATATTCCTTTGCAAGTACCTGCAAATATAACATGTTGGAC
H10	3	G.........................A..
H11	2	..........................A...A..
H12	2	..........................A..
H13	1	.................C........A..
H14	1	.....................A....A..............................T.........................G.................
H15	1	.....................A....A..........................A.............................G.................
H16	1	...C...
H17	1	....C..........................
H18	1	....C............A.........................
H19	1	...C......................A....G..A............................
H20	1	..........................A...A............................
H21	1	.T........................A..A.........................
H22	1	.....................A....A..G.............A..T
H23	1	G...A.....................A..
H24	1	.................C........A..................C.......................................G...............
H25	1	.....................A....A..G......C..........
H26	1	.......A..................A..
H27	5	.......A.............A....A..T...A.............C....GA.................AA........T.G.....G...........
H28	2	.......A.............A..A.A..T...A.............C....GA.................AA........T.G..C..............
H29	1	.....................A....A..TG..A.............C....GA.................AT........T.G.................

Table 2-3 Summary statistics of DNA polymorphism for mtCOI and nITS genes in *Acartia tonsa* populations in Chesapeake Bay. Fu & Li's tests were only applied within clades. Statistical significance: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. n.a.- not applicable.

	Sample size	Haplotype Number	Haplotype diversity (*H*)	Nucleotide diversity (p)	Tajima'*D*	Fu & Li's test	
						D	*F*
mtCOI							
Clade *S*	42	8	0.384	0.001	**-1.840***	-0.530	-1.056
Clade *F*	34	21	0.954	0.012	-0.539	0.036	-0.168
all samples	76	29	0.805	0.074	**3.321*****		
nITS							
Clade *S*	36	1	0.000	0.000	0.000	n.a.	n.a.
Clade *F*	38	4	0.619	0.002	0.141	0.559	0.380
all samples	74	6	0.668	0.164	**4.805*****		

Table 2-4. *Acartia tonsa* mitochondrial 16S rDNA sequences used for phylogenetic analysis in the Chesapeake Bay study. Four Chesapeake Bay haplotypes (CB_F1-3, CB_S1) are from this study and the rest are from GenBank. The alignment only shows the polymorphic sites of a 134 bp overlapping fragment. CB_F1 was used as reference sequence. Matching nucleotides in other sequences were denoted by dots and gaps by dashes.

Haplotype	Geographic location	GenBank accession no.	0000000000000000000000000000000000000000000000000000000000000000000000000111111111111 0000000001111111111222222222233333333334444444444555555555566666689999999000011111112 1234567890123456789012345678901234567890123456789012345678923467801234568235803456780
Acartia tonsa			
CB_F1	Chesapeake Bay, USA	EU274433	TTTGGGGTAAATAAAAATAAAATTATT-TTGAATTAAATTGATCTTAATAAAACTAGATAAAGGTAAATTATACCATAATAATTT
CB_F2	Chesapeake Bay, USA	EU274435	..............G............-...
CB_F3	Chesapeake Bay, USA	EU274434	..............G............-...T...........
CB_S1	Chesapeake Bay, USA	EU274432	......................A....-....T...............CT......A............C.....T....G....
A1	NW Atlantic	AF502368	.......A....T..............-..A.TAA...........T.C.......A..........G.........TG...C..
A2	NW Atlantic	AF502369	.......A....T..............-..A.TAA...........T.C.......A..........G......T..TG...C..
A3	NW Atlantic	AF502371	.......A....T..............-..A.TAA...........T.C.......A..........G.........TG...C..
A7	NW Atlantic	AF502385	.......A....T..............-..A.TAA...........T.C.......A..........G.........TG...C..
A8	NW Atlantic	AF502370	.......A....T..............-..A.TAA...........T.C.......A..........G.........T....C..
B1	NW Atlantic	AF502376	......................A....-....T...............CT......A............C.....T....G....
B2	NW Atlantic	AF502377	......................A....-....T...............CT......A............C.....T....G....
B3	NW Atlantic	AF502380	......................A....-....T...............CT......A............C.....T....G....
B6	NW Atlantic	AF502378	......................A....-....T...............CT.-....A..--....T...C.....T....G....
B7	NW Atlantic	AF502374	......................A....-....T...............CT......A............C.....T....G....
B8	NW Atlantic	AF502379	......................A....-....T...............CG......A............C.....T....G....
B9	NW Atlantic	AF502375	......................A....-....T.....-.........CT......A............C.....T....G....
C1	NW Atlantic	AF502383	......................G....-.....C....C.......G..........G...........................
C2	NW Atlantic	AF502381	......................G....-.....C....C.......G..........G..................-.....-..
C3	NW Atlantic	AF502382	......................G....-.....C....C.......G..........G...................-.......
D1	NW Atlantic	AF502373	............T....A....A....-..T......T........G.C....T..AT...............AT..T...G...
D2	NW Atlantic	AF502372	............T....A....A....-..T......T........G......T..AT...............AT..T...G...
P1	NE Pacific	AF502384	............TT.....GG.C....-.AT.T.A.C.A.......T..T...T....C.T...........T..T.........
P2	NE Pacific	AF502386	............TT.....GG.C....-.AT.T.A.C.A.......T..T...T....C.T...........T..T.........
A. hudsonica	Narragansett Bay, USA	AF502389	--------------...C..C.---------.C.....C------------..T.---.TTG.........A.TT..GT......
A. clausi	Plymouth, UK	AF502388	..........G...T....TT.G....A..A.T......A......T......TA.A....T....TG.....ATT.......A.
A. longiremis	Håkoybotn Fjord, Norway	AF502387	---------------------------------------.......T..T...AATTG..G......G...A..TT.T....C..
A. pacifica	Seto Inland Sea, Japan	DQ071175	.......A...G.T....C.GT.....T.AATTGA.C........A.T.TT..T.CA.G...ATW.C.C.T..TTT.TTG.G.GG

Table 2-5 Nuclear ITS sequence alignment of 5 haplotypes recorded in *Acartia tonsa* in Chesapeake Bay. Sequences from two clades (ITS1 and ITS2–5) are separated with polymorphic sites between them shaded. Haplotypes ITS3–5 are shown relative to reference sequence ITS2 by using dots to denote matching nucleotides. Haplotype frequencies assume diploidy (2*n* = 74).

Haplotype	Frequency	Segment 1–90
ITS1	36	AAAGAATTCGCAG-GCAAGGCTAGCGAATTCAACAATCAAGCTAGCAAGCTTGCTATAGGGAGAGAGAGAGTGTGTGGGAAAGGCGGTAA
ITS2	19	AAAGGAATGTCGGCTCATGG-AAGCAAG--CATATGTGCAGCTACCA---TTAAT-CAGGTTGTGGGGACCTGCAATAGCAAGCAAGCAA
ITS3	14	.....-.....--....C.....---.....-.....
ITS4	4	.....-.G.....--....C.....---.....-.....
ITS5	1	.....-G.....--....C.....---.....-.....
		Segment 91–180
ITS1	36	GCTAGTGGACAAGGGAACACTGGGCTTGTAGCTTGAAGAAGATAGTAGCTATATACTAGAGTGAATGTGAATCGCTGGTCACATG-AACA
ITS2	19	GCTAGTAGGCAAGGGAACACTAGACTTGCAGCTTGAGGAAGATAGTAGCTAAATACTTGAGTGAATGTGAATCGCTGGTCACATGAAACG
ITS3	14	.....
ITS4	4	.....
ITS5	1	.....
		Segment 181–270
ITS1	36	TAGCTGGATTGAACGCAAGTGTAGGAACCAATGTCTTTGACTTTGGTTTCATATTCCTTGTCAAAACGAGGGCAGTGTGTGGG-AGCCCG
ITS2	19	TAGCTGGGTTGAACGCAAGGGTGAGAACCAAAGCC--AGGCTTTGGGTCTCAATTCTTTGTCAGGACATGCCTTTGGAGAAGGCAGTTAA
ITS3	14	.....--.....
ITS4	4	.....--.....
ITS5	1	.....--.....
		Segment 271–346
ITS1	36	TCTGTTCAAGGACAGATCAATTCCACCCTTCTTAAGCAGAACGAGGGAAGCTGTACGCCAGGCAAGGTCGCCTAGA
ITS2	19	GCCGTCTTAGAACGGATC-GTTCTACCCTCTGCAAA------GAGGTAGAATAGAAGGCAA--AAGGACTAATAGA
ITS3	14	.....-.....------.....--.....
ITS4	4	.....-.....------.....--.....
ITS5	1	.....-.....------.....--.....

Figure legends

Figure 2-1 Neighbor-joining gene tree of 29 mtCOI haplotypes based on uncorrected *p*-distance showing two mitochondrial lineages (*F*, *S*) of *Acartia tonsa* in Chesapeake Bay. Congeneric species *A. hudsonica* was used as the outgroup to root the tree. Bootstrapping support values greater than 70% from all three methods (NJ/MP/ML) are shown on the branches. Haplotype labels and frequencies (in parentheses) as in Table 2-2.

Figure 2-2 Neighbor-joining gene tree of *Acartia tonsa* mitochondrial 16S rDNA haplotypes based on Tamura-Nei distance (α=2), showing that the mitochondrial lineage *S* haplotype (CB_S1, boxed) in this study forms a clade with lineage B (B1–3 , B6–9) from Caudill & Bucklin (2004), whereas *F*-lineage haplotypes here (CB_F1–3, boxed) formed a sister clade to Caudill & Bucklin's lineage C (C1–3). Bootstrapping support values greater than 70% are shown on the branches. Four congeneric species *A. clausi*, *A. hudsonica*, *A. pacifica*, and *A. longiremis* were used as outgroups. Haplotype labels as in Table 2-4.

Figure 2-3 Neighbor-joining gene trees based on uncorrected *p*-distance demonstrating genealogical concordance between mtCOI (left) and nITS (right) in *Acartia tonsa* of Chesapeake Bay. Both trees only include a subset of individuals in which both genes had been sequenced. Bootstrapping support values greater than 70% from all three methods (NJ/MP/ML) are shown on the branches. The mtCOI tree was rooted by congeneric species *A. hudsonica* (not shown), while the nITS tree was rooted by mid-point rooting method.

Figure 2-4 Geographic distribution of *Acartia tonsa* mitochondrial lineages *F* and *S* in Chesapeake Bay mainstem (A) and tributaries (B). Pie shading depicts %*S* and numbers

on the pies are average salinity (PSU) of sampled water. Boxed pies in 2-4A represent samples from above- (upper pie) and below- (lower pie) a vertical cline in salinity at single localities. Circles in 2-4B identify sample groups from single tributaries. Three samples (G1, G2, and L1) are not shown due to small sample size.

Figure 2-5 Correlation between lineage *S* composition (%*S*) of *Acartia tonsa* and salinity in Chesapeake Bay. The bold solid line represents a general logistic regression model that fits all the data (both open and closed circles). The thin solid line represents a specific model that fits data from unstratified localities (closed circles). The dashed line represents a model that fits data from vertically stratified localities (open circles). Grey lines identify vertical sample pairs from three stratified sites (T4/5, T6/7, T8/9). Mathematical expressions for the three models are presented in Results.

Figure 2-1

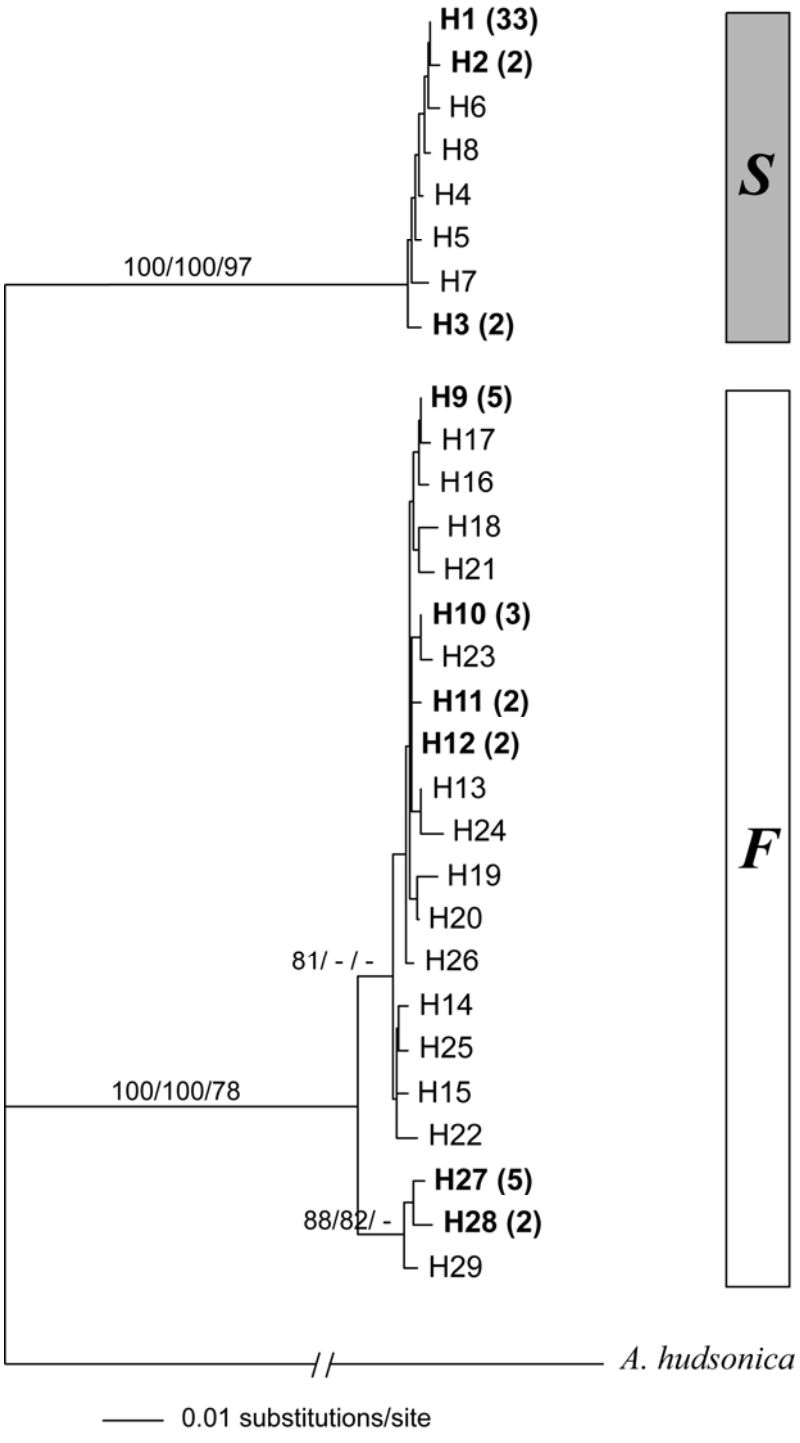

H1 (33)
H2 (2)
H6
H8
H4
H5
H7
H3 (2)
S
100/100/97
H9 (5)
H17
H16
H18
H21
H10 (3)
H23
H11 (2)
H12 (2)
H13
H24
H19
H20
H26
F
81/ - / -
H14
H25
H15
H22
100/100/78
H27 (5)
88/82/ -
H28 (2)
H29
A. hudsonica
0.01 substitutions/site

Figure 2-2

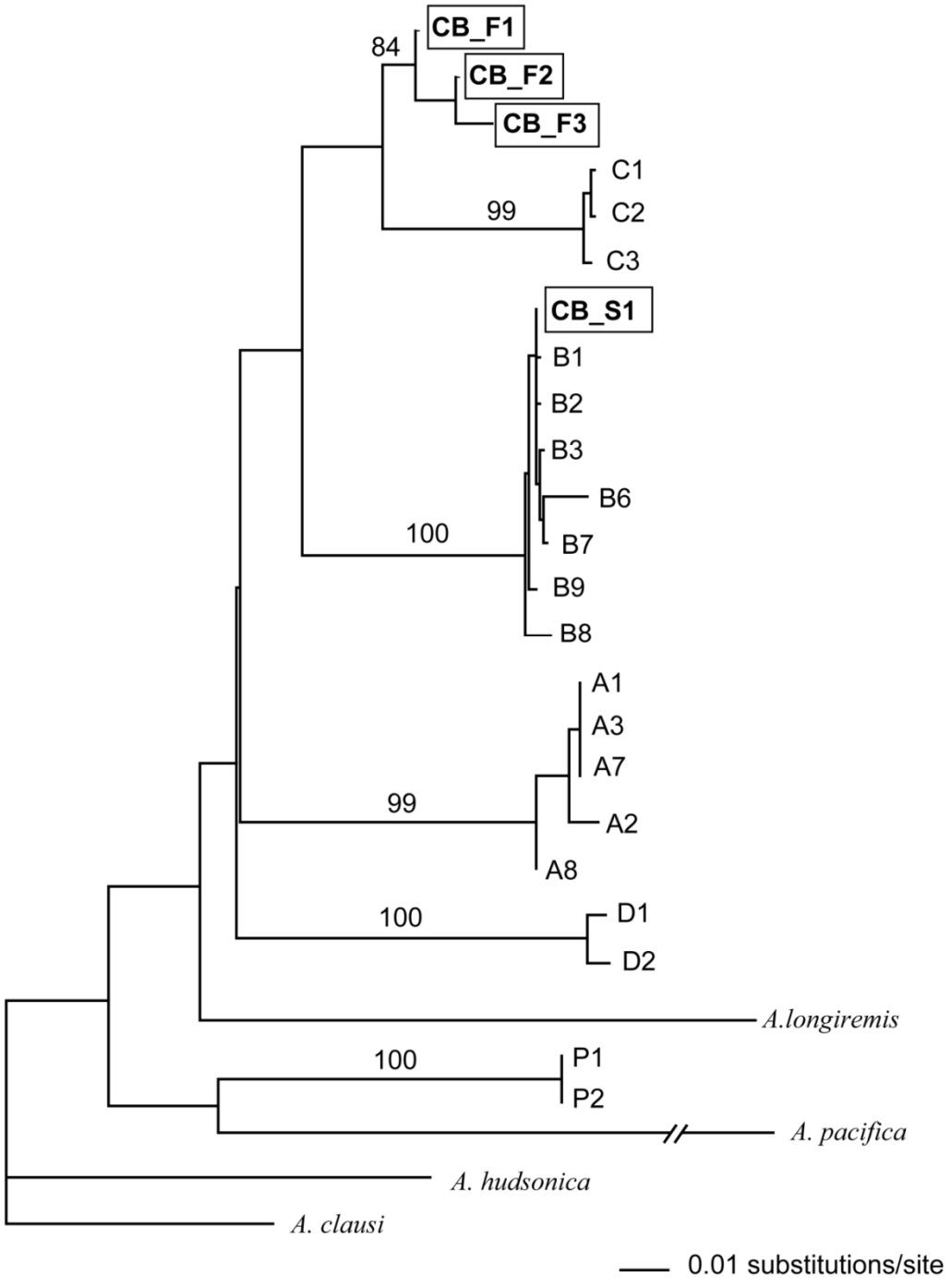

CB_F1
84
CB_F2
CB_F3
C1
99
C2
C3
CB_S1
B1
B2
B3
B6
100
B7
B9
B8
A1
A3
A7
99
A2
A8
100
D1
D2
A.longiremis
100
P1
P2
A. pacifica
A. hudsonica
A. clausi
0.01 substitutions/site

Figure 2-3

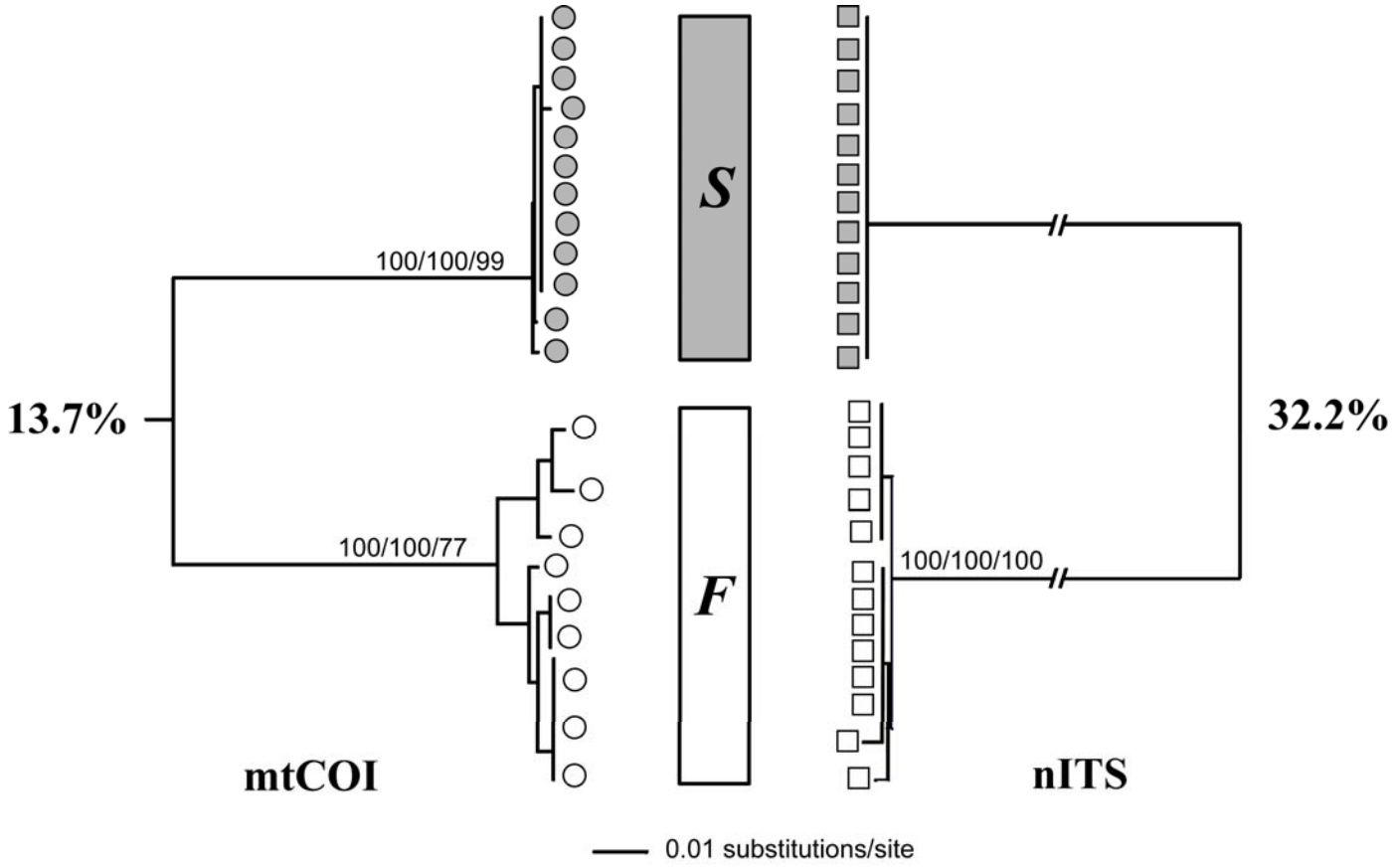

S
F
100/100/99
100/100/77
100/100/100
13.7%
32.2%
mtCOI
nITS
0.01 substitutions/site

Figure 2-4

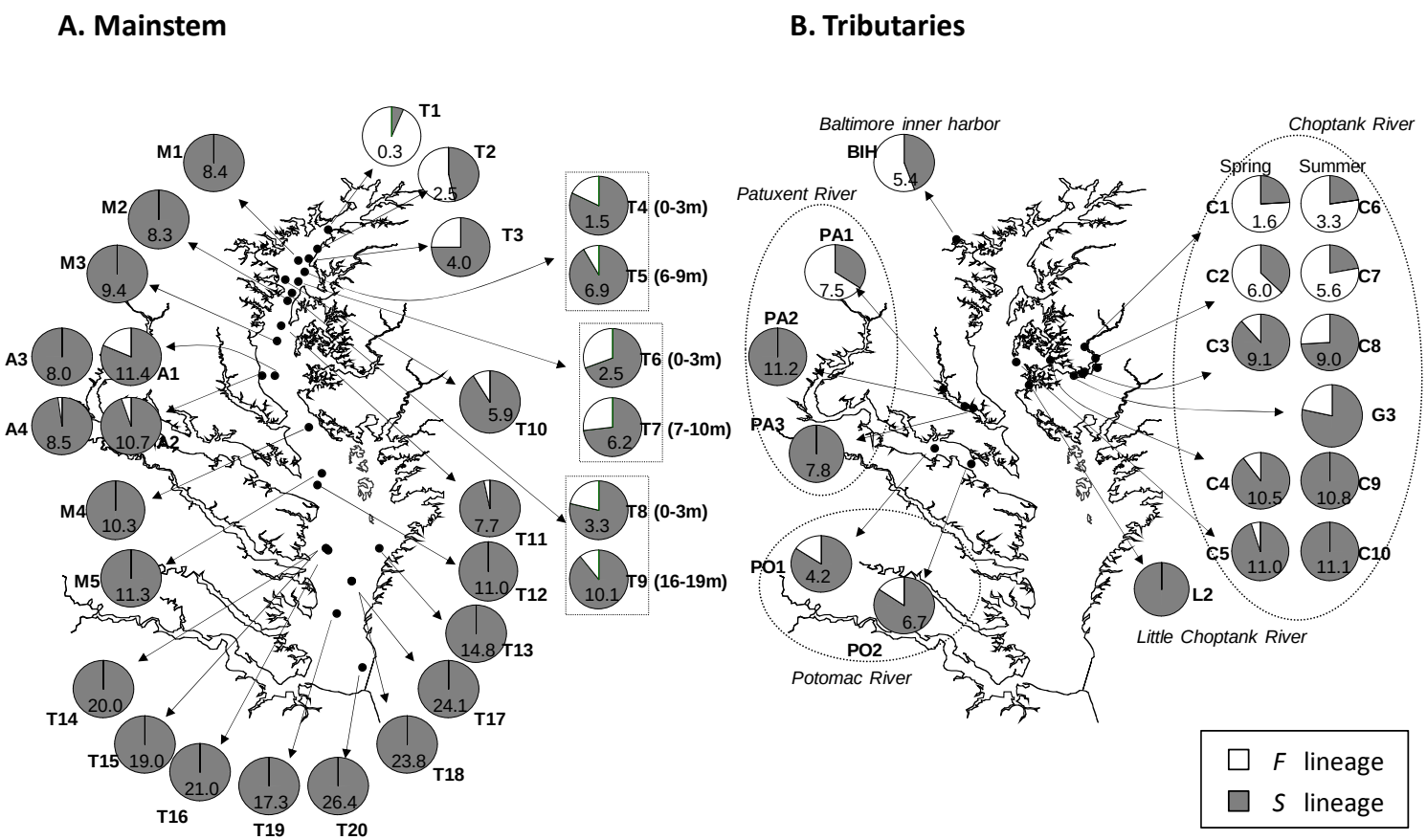
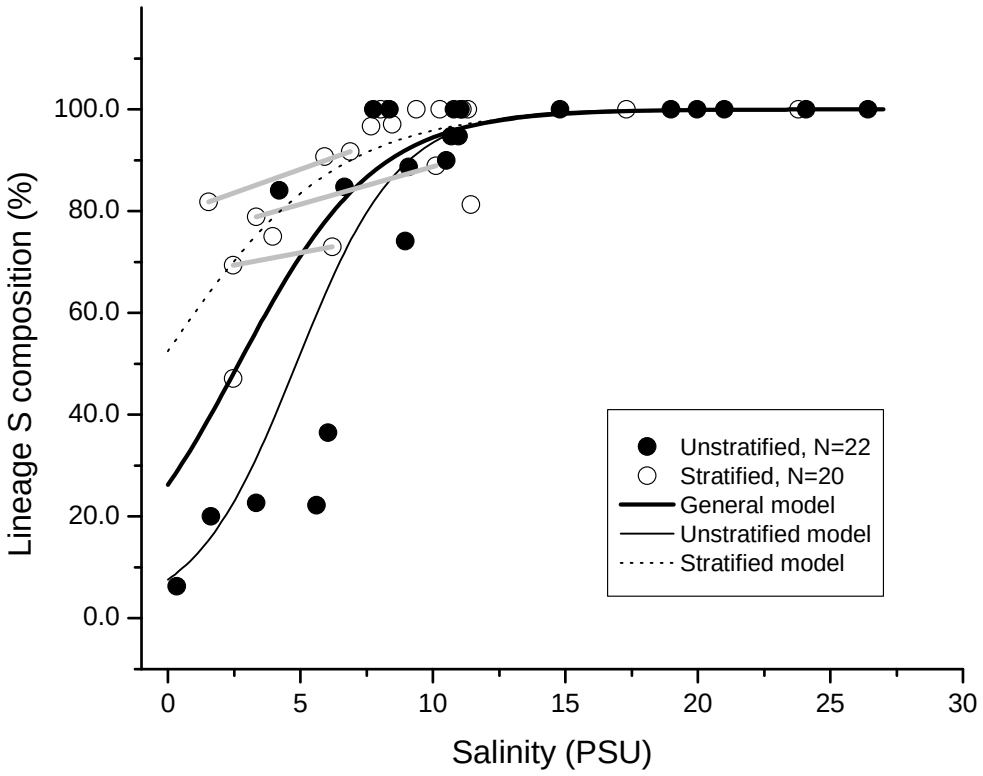

A. Mainstem
B. Tributaries
T1
0.3
M1
8.4
T2
2.5
M2
8.3
T3
4.0
M3
9.4
T4 (0-3m)
1.5
T5 (6-9m)
6.9
T6 (0-3m)
2.5
T7 (7-10m)
6.2
T8 (0-3m)
3.3
T9 (16-19m)
10.1
A3
8.0
11.4
A1
A4
8.5
10.7
A2
5.9
T10
M4
10.3
7.7
T11
M5
11.3
11.0
T12
14.8
T13
20.0
T14
24.1
T17
T15
19.0
21.0
T16
17.3
T19
26.4
T20
23.8
T18
Baltimore inner harbor
BIH
5.4
Patuxent River
PA1
7.5
PA2
11.2
PA3
7.8
PO1
4.2
6.7
PO2
Potomac River
L2
Choptank River
Spring
Summer
C1
1.6
C6
3.3
C2
6.0
C7
5.6
C3
9.1
C8
9.0
G3
C4
10.5
C9
10.8
C5
11.0
11.1
C10
Little Choptank River
F lineage
S lineage

Figure 2-5


Lineage S composition (%)
100.0
80.0
60.0
40.0
20.0
0.0
0
5
10
15
20
25
30
Salinity (PSU)
Unstratified, N=22
Stratified, N=20
General model
Unstratified model
Stratified model

Chapter 3
Population genetics and comparative phylogeography of *Acartia tonsa* on the US Atlantic coast

Abstract

The biogeography of marine holoplankton is more complex than previously thought. Unexpectedly high genetic diversity and small-scale geographic structures in many cosmopolitan species challenges the view of unrestrained dispersal of holoplankton and high gene flow in marine systems. Marine holoplankton thus provide opportunities to look into processes and mechanisms governing marine speciation under potential gene flow.

I investigated population genetics and phylogeography of a common estuarine copepod, *Acartia tonsa*, on the US Atlantic coast, using mitochondrial (mtCOI) and nuclear (nITS) gene markers. Phylogenetic analyses revealed five morphologically cryptic, genetically diverged lineages that were recognized as reproductively isolated species based on concordance between two gene genealogies. These lineages showed sympatric distribution (though patchy) in estuaries on the US Atlantic coast, but contrasting phylogeographic patterns: 1) Geographic structure were found at large scales (1000–2000 km) in two deep lineages (*X*, *S*; separated by large genetic distances) while at small scales (~100 km) within recently diversified lineage (*F*); 2) Highly diversified lineage *F* had a southern center of origin showing a stepwise northward colonization history since Pleistocene. 3) Isolation by distance (IBD) patterns were significant for all

lineages, suggesting that dispersal of *A. tonsa* was more or less limited to adjacent estuaries and distant gene flow followed the stepping-stone model.

In conclusion, geographic isolation due to limited dispersal played a major role in population differentiation and speciation of *A. tonsa*. Physiological adaptation to low-salinity environments with increased residence times might have facilitated geographic isolation of lineage *F* and sped up its diversification.

Introduction

The biogeography of marine holoplankton (organisms drifting along with currents all lifetime) is often speculated to be less hindered by geographic barriers due to their vast populations and high dispersal potential. In the extreme case, planktonic microbes were once believed to be "everything everywhere" on our planet (Whitfield 2005). Across a wide range of invertebrate taxa, holoplankton often show global or regional wide distributions associated with large-scale circulation patterns and were expected to have abundant cosmopolitan species (van der Spoel & Heyman 1983; van der Spoel & Pierrot-Bults 1979). In recent years, however, accumulated evidence revealed unexpectedly high genetic diversity in marine holoplankton, including many previously thought to be cosmopolitan (e.g. Bucklin *et al.* 1996; Zane *et al.* 1998; Goetze 2003; Peijnenburg *et al.* 2004; Fenchel 2005; Šlapeta *et al.* 2006; Chen & Hare 2008). These findings created a paradox about holoplankton evolution, i.e., the contradictory link between high genetic divergence and presumably strong gene flow. Meanwhile, marine holoplankton provide good opportunities to discover possible processes and mechanisms behind the paradox, which will eventually deepen our insights about marine speciation

and the generation of biodiversity in marine system. In this study I used DNA sequence analysis to examine the evolutionary diversification of an estuarine copepod species complex *Acartia tonsa* on the Atlantic coast of the United States.

Acartia tonsa Dana, 1849 is nominally a calanoid copepod species (Copepoda, Calanoida). It is a seasonally dominant and prolific member of the estuarine holoplankton and has been intensively studied in the field and laboratory (Mauchline 1998). *Acartia tonsa* lives its whole life (12 developmental stages) in the water column with a generation time of 7–25 days (Mauchline 1998) and adult longevity of 26 days (Paffenhöfer 1991). A mature female has a lifetime fecundity of 435–718 eggs (Zillioux & Gonzalez 1972; Parrish & Wilson 1978). After mating with males, internally fertilized eggs are spawned directly into the water at a rate of 18–50 eggs/day (Mauchline 1998). Such high reproductive potential along with short generation time often make *A. tonsa* populations astronomically huge. *Acartia tonsa* also reportedly produces diapause eggs (eggs that enter developmental dormancy as a physiological response to certain environmental cues), which help populations survive harsh environments (Zillioux & Gonzalez 1972; Marcus 1996). All these life-history features of *A. tonsa* provide ample opportunities for its population expansion, invasion and recolonization.

The taxonomic status of *A. tonsa* was well established around one and half century ago. Since then, it has been found in a wide range of Indo-Pacific and Atlantic coastal regions (Mauchline 1998). On the Northwest Atlantic coast, *A. tonsa* often occurs as a seasonally dominant species in estuarine and coastal environments (McAlice 1981; Paffenhöfer & Stearns 1988; Tester & Turner 1991). It has been regarded as a single species in many ecological studies; however, recent genetic studies suggested this

cosmopolitan species might be a species complex. Caudill & Bucklin (2004) reported four deeply-diverged genetic lineages in *A. tonsa* populations from the US Atlantic coast based on mitochondrial 16S ribosomal gene locus. Chen & Hare (2008) found two lineages (including one novel to Caudill & Bucklin 2004) in Chesapeake Bay and confirmed their status of morphologically cryptic species by genealogical concordance between a mitochondrial DNA (mtCOI) and a nuclear gene locus (nITS). These genetic findings set *A. tonsa* as a novel example reflecting the above-mentioned holoplankton evolution paradox. Resolving its population diversification history will certainly help us solve the holoplankton puzzle. To do so, I extensively sampled *A. tonsa* populations in major estuaries on the US Atlantic coast and investigated its population genetics and large-scale comparative phylogeography. The major goals were 1) to estimate genetic and species diversity of *A. tonsa* on the US east coast; 2) to recognize possible geographic or evolutionary origins of *A. tonsa* genetic lineages; 3) to infer possible evolutionary processes and mechanisms creating cryptic diversity of *A. tonsa.*

To assess species status within *A. tonsa* taxa, I applied the principle of genealogical concordance. Concordant genealogical structure among independently evolved neutral genes signals a substantial period of reproductive isolation between taxa (Avise & Ball 1990; Chen & Hare 2008). The genealogical concordance approach is conservative and powerful for identification of biological species status, in that complete lineage sorting is the ultimate consequence of reproductive isolation. However, to learn about diversification processes the focus is often on closely related taxa or populations, in which genealogies are typically non-exclusive, complicated by the sharing of ancestral polymorphisms, and discordant across loci. To extend the principle of genealogical

concordance to this situation, Cummings *et al.* (2008) developed a quantitative genealogical sorting approach to quantify and test the extent of genealogical sorting among non-monophyletic lineages. Statistically significant exclusivity of individuals from a paraphyletic clade against a null distribution of random cladistic assignments indicates a substantial divergence between this clade and others (Cummings *et al.* 2008). Genealogical concordance between non-monophyletic clades with significant genealogical sorting indices also suggests substantial reproductive isolation and good species status of lineages.

A previous phylogeography study of *A. tonsa* on the Northwest Atlantic coast revealed strong genealogical divergence, however weak geographic structure within and among deeply diverged lineages along the US east coast (Caudill & Bucklin 2004). Two common lineages appeared to coexist sympatrically along the coast. This could be a fact of no regional phylogeographic structure for *A. tonsa* due to its strong dispersal potential, or simply an artifact due to insufficient samples (4 samples over ~2000 km long coast). In this study, I investigated *Acartia* phylogeography on the US Atlantic coast from Rhode Island to Florida with intensive sampling of 20 estuarine systems. One key phylogeography hypothesis is the existence of geographic structure in *A. tonsa* genetic lineages. The expectation has two major reasons. First, the entire US east coast from Maine to Florida spans 20 latitude degrees (25–45°N) with biophysically diverse estuarine habitats. Latitudinal gradients are a well known factor affecting distribution, community composition and diversity of marine organisms along the Atlantic coast of North America (Hedgpeth 1957; Hayden & Dolan 1976; Kendall & Aschan 1993). Within the range, four coastal marine biogeographic provinces were well established

based on observed distribution of other organisms, including Acadian, Virginian, Carolinian and West Indian (Hedgpeth 1957; Gosner 1971; Briggs 1974; Engle & Summers 1999; Wares 2002). It is reasonable to expect phylogeographic structure of *A. tonsa* as a response to the environmental heterogeneity at this large geographical scale. Second, the US east coast experienced a history of climatic changes during the Pleistocene that induced dramatic environmental changes, such as sea level drop and rise (~200 m), displacement of isotherms and isohalines, and contraction of temperate and subtemperate coastal zones (Cronin 1988; Lambeck & Chappell 2001). These events are likely to have had an impact on the historical distribution of *Acartia* and its genetic lineages (McAlice 1981; Caudill & Bucklin 2004).

Another hypothesis for *Acartia* phylogeography is about the role of biological features. As we found two genetic lineages with distinct salinity affinities in Chesapeake Bay (low-salinity lineage *F* and high-salinity *S*), I hypothesize that contrasting phylogeographic patterns may exist among distinct *Acartia* lineages. More specifically, a physiological barrier may play a role as important as the physical barrier, if any, to *Acartia* gene flow. This could be tested through comparative phylogeography among coexisting *Acartia* lineages. Agreements in phylogeographic patterns among co-distributed lineages/species should reflect the forcing effect of shared demographic events and physical processes on evolution of the organisms, whereas disagreements highlight the consequential effect of their differential biological responses to shared environments (Bermingham & Moritz 1998). Within *A. tonsa*, we expect to see a higher level of genetic structure in low-salinity lineage than in high-salinity lineage, if salinity acts as a physiological barrier to *Acartia* dispersal.

Overall, this study aims to reveal phylogeographic history of *Acartia tonsa* on the US Atlantic coast, testing its genetic diversities, evolutionary or geographic origins and possible gene flow constraints at the large geographic scale. High species diversity at the studied scale and higher genetic diversity in low-salinity genetic lineages are predicted. Geographic isolation by limited dispersal and gene flow of *Acartia* may exist and have contributed to generation of its genetic diversity at the large scale. We should be able to identify allopatric origins of *Acartia* genetic lineages within the studied northwestern Atlantic region.

Materials and Methods

Sampling sites and copepod collection

Geographic samples from 21 estuarine systems were examined in this study, including Chesapeake Bay samples reported previously in Chen & Hare (2008) and samples from 18 other estuarine systems on the US eastern coast (between Rhode Island and Florida), and one European sample from Finland (Table 3-1). Most samples were collected during April–September, 2005, while Chesapeake Bay samples had a time span from 1995 to 2005 and two Delaware samples (DB1 and DB2) were collected in summer 2003.

Detailed sampling information for Chesapeake Bay samples was described in Chen & Hare (2008). The author collected most other samples from shore by independent road trips. Delaware Bay samples were collected from a research cruise for Microbial Observatory of Virioplankton Ecology (MOVE, E. K. Wommack, University of

Delaware) and Savannah River samples from one for marine phytoplankton research (E. Mann, Skidaway Institute of Oceanography). The European sample was obtained from M. Reinikainen, University of Helsinki, Finland. In general, zooplankton samples were towed horizontally or vertically using standard plankton nets. Sampling depths varied in different sampling events with most shore samples collected from upper water layer (1–3 m). All samples were immediately preserved in 95% ethanol.

DNA extraction, amplification, sequencing and RFLP genotyping

In the laboratory, adults and later larval stages (copepodids CIV, CV) of *Acartia tonsa* were identified under microscope based on their diagnostic morphological characters (Bradford-Grieve 1999; Sabatini 1990; see a brief summary in Methods of Chapter 4 of this thesis). Individuals with intact bodies were selected to extract genomic DNA. Single copepods were first rehydrated in de-ionized water for more than 2 h and then heated at 99 °C in 100–150 μl of 5% chelex solution (Bio-Rad Laboratories) for 8 min. After centrifugation at 2288 *g* for 5 min, supernatant DNA solutions were collected and stored at 4°C for genetic analysis.

We amplified and sequenced three gene fragments: mitochondrial cytochrome *c* oxidase subunit I (mtCOI), mitochondrial 16S ribosomal DNA (mt16S), and nuclear ribosomal DNA internal transcribed spacer (nITS). Amplifications used mtCOI primers 1490 and 2198 (Folmer *et al.* 1994), universal ITS primers ITS-4 and ITS-5 (White *et al.* 1990), and universal 16S ar/br primers (Palumbi *et al.* 1996). For all genes, PCR amplifications were performed in a 25 μl reaction volume with 1× Invitrogen buffer, 2.5

mM $MgCl_2$, 125 μM of each dNTP, 0.2 μM of each primer, 0.2 μg/μL BSA, 0.3 U of Invitrogen *Taq* polymerase, and 0.5–1 μL of genomic DNA. The optimized PCR conditions started with DNA denaturing at 95°C for 1 min, followed by 35 cycles of 30 sec denaturing at 95°C, 30 sec annealing at 50°C, and 50 sec extension at 72°C and then a final extension for 7 min. PCR amplifications were confirmed in a 1.2% agarose gel with ethidium bromide staining. This protocol was effective for all genes and most DNA samples. For sequencing, PCR products were purified by incubation with shrimp alkaline phosphatase and exonuclease I (USB Biochemical) for 30 min, and cleaned by isopropanol precipitation. Sequencing reactions used Big Dye terminator chemistry as recommended by the manufacturer (Applied Biosystems) and the same primers as used for PCR. Sequencing products were analyzed on an ABI 3100 DNA sequencer (Applied Biosystems). Both strands were sequenced for each gene to warrant the accuracy of base calls.

Based on 298 mtCOI sequences, distinct restriction fragment length polymorphism (RFLP) patterns were deduced to occur among three deeply-diverged mtCOI lineages (*F, X, S*) from digestion with endonuclease *Ase* I. For lineage *X,* the 6-base recognition sequence is found at positions 235–240, yielding a two-fragment pattern (304/406 bp based on 710 bp amplicon). Lineage *S* had two patterns, no-cut (710 bp, rare pattern) and a two-fragment (167/543 bp) pattern with the restriction site located at positions 475–480. Three *Ase* I RFLP patterns were found in lineage *F*: a two-fragment (118/592 bp) pattern and two three-fragment patterns (118/167/425 bp, common, 114/118/478 bp, rare). Five units of *Ase* I endonuclease (New England Biolabs) were used to digest 5–10 μL of mtCOI PCR amplicon in a total volume of 15 μL with 1× NEB

3 buffer at 37°C for 3 h. The digested DNA fragments were visualized in 2.5% agarose gels with ethidium bromide staining.

Data analysis

All DNA sequences were edited using SEQUENCHER (Genecode Corp.). Accuracy of base calls was always confirmed by comparing both strands. No heterozygous base calls were found in mtCOI and mt16S sequences as expected for haploid mtDNA. Coding gene mtCOI showed no insertions or deletions (indels), thus its sequences were aligned using SEQUENCHER. Sequence alignments of mt16S and nITS were done using Clustal X v1.83 (Thompson *et al.* 1997) with default parameter settings and subsequently optimized manually. Double base calls were observed in some nITS sequences as expected for this multicopy nuclear gene, reflecting heterozygosity and/or variations among different chromosomes. Heterozygous sites were determined based on the criteria of Hare & Palumbi (1999). It is unknown how many copies of nITS genes across genomes of studied individuals, which make haplotype inferences based on Hardy-Weinberg assumptions inappropriate. For nITS sequences with more than 1 ambiguous heterozygous sites, alleles were determined based on known alleles from homozygotes and its frequencies, assuming diploidy (Clark 1990). For one individual with four unresolved ambiguous heterozygous sites, two pseudo alleles were inferred based on two known alleles. They were included as OTUs (operational taxon unit) in phylogenetic analysis, however, excluded from all other population genetic analyses.

For each gene, phylogenetic relationships among allelic sequences were examined using the neighbor-joining (NJ), maximum parsimony (MP), and maximum likelihood

(ML) methods. Phylogenetic trees were constructed based on both uncorrected genetic distance and gene-specific corrected distance using computer software PAUP*4.0b10 (Swofford 2003). Substitution models of Kimura-2-parameter (K2P, for mtCOI) and Tamura-Nei (TrN, for nITS) were selected for the correction of genetic distance in neighbor-joining method and modeling nucleotide substitution in maximum likelihood analyses, as suggested by Kumar *et al*. (1993). For mt16S, the model of Tamura-Nei with $\alpha = 2.0$ (TrN-G) used in Caudill & Bucklin (2004) were applied for consistency and comparison purpose. In parsimony analysis, gaps in nITS and mt16S sequences were coded as separate characters following coding rules of Simmons & Ochoterena (2000). Best MP and ML trees were obtained by heuristic searches initiated with random sequence addition (10 replicates) and continued by tree-bisection-reconnection (TBR) branch-swapping. Statistical tests for the reliability of tree topologies were implemented by bootstrap sampling with 1000, 500, and 100 pseudoreplicates for NJ, MP, and ML trees respectively. Mitochondrial COI gene trees were rooted by a sequence of congeneric species *A. hudsonica* collected from Delaware Bay, and mt16S gene trees by *A. longiremis* mt16S sequence from GenBank (Accession no. AF502387). Nuclear ITS gene trees were rooted by midpoint rooting method due to lack of a useful outgroup. To test for mtCOI rate heterogeneity among distinct clades, relative rate tests were conducted using computer program RRTree (Robinson-Rechavi & Huchon 2000). Default Kimura-2-parameter distance was used to estimate the rate differences among tested clades relative to outgroup. Tree topologies (neighbor-joining) were taken into account to weigh the sequences so as to minimize the effect of unbalanced sequence sampling (Robinson *et al.* 1998). Representative sequences from adjacent clades or a

sequence from congeneric species *A. hudsonica* (for testing entire *A. tonsa* data set) were used as outgroups for relative rate tests.

To test for tree topology concordance between mtCOI and nITS, sequences obtained from the same subset of individuals were analyzed as above. For lineage *F1-3* with discordant COI and ITS topologies, genealogical concordance between the gene trees was quantified and tested independently for the trees derived from the three methods (NJ, MP, and ML respectively) using Cummings *et al*.'s (2008) genealogical sorting approach. The genealogical sorting index, *gsi*, measures the degree of exclusive ancestry for labeled groups within a phylogenetic tree. The normalized index varies from zero to one, zero indicating a random mixture of group labels relative to any clade structure and one resulting from perfect agreement between group labels and clade structure. In the latter case, assuming that there is some bootstrap support for the clades, the result supports an inference of low gene flow between groups and complete sorting of ancestral polymorphisms. To quantify genealogical concordance I labeled individuals according to their clade membership in the mtCOI tree. The *gsi* was calculated from the same individuals and their mtCOI clade membership labels in the nITS tree. In order to test the null hypothesis that labels are randomly distributed among clades, individual labels were randomized and assigned to the branch tips of the fixed gene tree 1,000 times. For each clade/group, the statistical distribution of *gsi* under the null hypothesis was obtained by calculating the *gsi* for each label-permuted tree. When the observed *gsi* value fell into the top 5% of the null *gsi* distribution, then the null hypothesis was rejected in favor of significant clade structure (not necessarily exclusivity). Three major geographic population sources involved in the *F1-3* clade (Pamlico River, NC, Indian River Bay, DE

and Chesapeake Bay) were also tested on both mtCOI and nITS trees. Individuals were labeled by geographic location and *gsi* calculated for 3 different locations. All the computation and permutation tests were completed using the R Package genealogicalSorting 2.0 (Bazinet *et al.* 2008).

Minimum spanning networks of mtCOI haplotypes were constructed for three deeply diverged clades (*X*, *S*, *F* respectively) showing phylogenetic relationships among haplotypes in a 2-dimensional space. The networks were calculated based on statistical parsimony with 95% probability limit (Templeton *et al.* 1992), using computer program TCS v1.21 (Clement *et al.* 2000). An ITS network for sublineage *F1-3* was constructed using nITS sequence data and same method in TCS program.

DNA polymorphism was examined for mtCOI and nITS genes, with the latter assuming diploidy for comparative purposes. Nucleotide diversity (π, Nei 1987) and haplotype diversity (*H*, Nei 1987) were calculated for both the entire sequence data set (excluding outgroup) and distinct clades separately using MEGA v4.0 (Tamura *et al.* 2007) and DnaSP v4.10.4 (Rozas *et al.* 2003). Comparisons of π and *H* between clades of interest were conducted using *t*-test (Nei 1987). Within distinct clades and/or populations, allelic diversity (R_d) was estimated based on the rarefactioned measure of allelic richness, which is by definition the expected number of alleles in a clade/population for a specified sample size (Kalinowski 2004, n = 5 was selected in this study). This rarefactioned allelic diversity allows unbiased comparisons among unequally sampled clades/populations. Rarefaction analyses were done using the computer program RAREFAC (Petit *et al.* 1998). To test neutrality for studied genes, Tajima's *D* (Tajima 1989) and Fu & Li's statistics (*D*, *F*, *D** and *F*, Fu & Li 1993) were computed for the clades of interest and the

entire data set. Representative sequences from adjacent clades or a sequence from congeneric species *A. hudsonica* (for testing entire *A. tonsa* data set) were used as outgroups for testing Fu & Li's *D* and *F*.

Geographic population structure within lineages was tested using Weir & Cockerham's *F* statistics (Weir & Cockerham 1984) and Raymond & Rousset's exact test (Raymond & Rousset 1995a) based on mtCOI haplotype frequency data. Analysis of molecular variance (AMOVA, Excoffier *et al.* 1992) was conducted for lineage *X*, *S*, and *F1-3* to partition the genetic variance (including both nucleotide difference and haplotype frequency) within and among geographic samples and between regions to test for all possible hierarchical geographic structures. Statistical significance of F_{ST} and Φ statistics was tested via permutation with 1000 replicates. All computations and permutation tests were implemented using computer program ARLEQUIN 3.1 (Excoffier *et al.* 2005). Isolation by distance (IBD, Wright 1943) was tested in ARLEQUIN by performing a Mantel test between two pairwise distance matrices, Rousset's (1997) genetic distances, $F_{ST}/(1-F_{ST})$, and the shortest open-water geographic distances between populations. Pairwise Rousset distances were estimated using computer program GENEPOP v4.0 (Raymond & Rousset 1995b). Geographic distance was measured using ArcGIS Explorer and its imagery base map (http://www.esri.com/software/arcgis/explorer/index.html). Significant positive correlation suggests that the equilibrium between genetic drift and dispersal has been reached in a population following the stepping-stone model of gene flow (Hutchison & Templeton 1999).

Results

Mitochondrial DNA

I analyzed a 564 bp fragment of mtCOI gene. A total of 98 distinct haplotypes were recorded from 298 sequenced *Acartia tonsa* individuals, including 29 haplotypes previously reported from Chesapeake Bay (Chen & Hare 2008). Within the alignment, 150 sites were polymorphic, of which 128 (85.3%), 20 (13.3%) and 2 (1.3%) were found at 3rd, 1st and 2nd codon positions respectively. Nucleotide substitutions of variable sites yielded 7 amino acid substitutions. No stop codons were found within reading frame of predicted protein sequences. Among all variable sites, 131 (87.3%) were parsimony-informative.

Three deeply diverged clades (*F, S,* and *X*) were resolved with strong bootstrap support by all three phylogenetic analysis methods (Fig. 3-1). Uncorrected net sequence differences among the three clades were between 10.3% and 12.6%, and within-clade sequence difference averages ranged from 0.2% to 2.9%. Five subclades (*F1 – 5*) were further recognized within *F* lineage (Fig. 3-1), although only the two most basal subclades, *F5* and *F4*, had consistently strong bootstrap support. Among these subclades net sequence differences ranged from 1.3% to 6.5% and within-clade differences were between 0.4% and 1.8%. *F* lineage as a whole had significantly greater diversity than *X* or *S* (π and *H*, *t*-tests: $P < 0.001$). *X* lineage showed the lowest levels of both nucleotide and haplotype diversity (Table 3-2). Diversity levels of *S* lineage were higher in magnitude than those of *X,* though only marginally significant (π, *t*-test: $P = 0.062$; *H*, *t*-test: $P = 0.087$). Within *F* lineage, no statistically significant differences were found among the 5 subclades.

In general, mtCOI data showed neutral variations within lineages, though significantly non-neutral for the entire data set (Table 3-2). No significant differences in substitution rate were detected among the three deeply diverged lineages (*X*, *S and F*) or among the subclades within *F* (relative rate test: dK/sd = -1.7145~0.8867, $P > 0.05$).

Mitochondrial 16S gene sequence data were obtained from 26 representative individuals for different COI lineages. A 134 bp fragment was overlapping with Caudill & Bucklin (2004)'s sequence data in GenBank and was used to examine clade correspondence between the two studies. A total of eight mt16S haplotypes were recorded in this study, including four (16S1–4) reported previously for Chesapeake Bay (Chen & Hare 2008). Two haplotypes of lineage *X* in this study were grouped with all 5 A-type haplotypes from Caudill & Bucklin (2004). One haplotype of lineage *S* was closely related to all the B-type haplotypes from Caudill & Bucklin (2004). Six haplotypes of lineage *F* and 3 Texas haplotypes (C1–3) found in Caudill & Bucklin (2004) were grouped into a clade distant from all other clades, though not strongly supported by bootstrap test (Fig. 3-2).

Nuclear ITS and Genealogical Concordance

A 378 bp fragment of nITS gene was selected for genetic analysis. A total of 23 distinct alleles were recorded from 138 individuals, including 5 previously reported from Chesapeake Bay (Chen & Hare 2008). Doubly heterozygous sequences were observed in 21 individuals. No sequences used in this study have more than 2-base ambiguous sites. Haplotype phase resolution was trivial in 20 cases by assuming that the most common haplotypes were involved. One additional case with four ambiguous sites was unresolved.

Twenty-three insertions or deletions (indels) with lengths ranging from 1 to 29 bp were present in the final alignment. There were 169 variable sites (44.7%) within the fragment with 153 parsimony-informative.

Three deeply diverged clades (*F*, *S* and *X*) defined by the mtCOI gene tree were also resolved in all nITS trees. Based on a subset of 128 individuals in which both mtCOI and nITS sequences were obtained, comparison of mtCOI and nITS tree topologies showed genealogical concordance in three deeply diverged mtCOI haplotype groups (*F1-5*, *S*, and *X*). Individuals from each mtCOI haplotype group formed an exclusive nITS group, and corresponding clades were well-supported in both trees by bootstrap tests (Fig. 3-3). One way of quantifying concordance is with the Genealogical Sorting Index (*gsi*), using clade membership in one gene tree to identify specimens in the other gene tree and calculate the *gsi*. Strict concordance in clade membership between trees will yield a *gsi* value of 1.0 and this was the result at the level of major lineages *F*, *S* and *X*. Within the *F* lineage group, the strongly supported subclades *F4* and *F5* also showed strict concordance between mtCOI and nITS (Fig. 3-3). Like mtCOI, the nITS tree had three recognizable subclades exclusive of *F4* and *F5* with bootstrap values varying widely among methods but each subclade was 90 or above for NJ bootstrap. Genealogical concordance between mtCOI subclades *F1–3* and the corresponding nITS subclades was low as indicated by clade-specific *gsi* values between 0.20–0.54. Permutation tests rejected the null hypothesis that the mtCOI clade membership labels on each individual (numbers in Fig. 3-3) formed one mixed group within the nITS tree (all $P < 0.001$), indicating that some of the mtCOI clade structure was present in the nITS tree. Interestingly, whereas the mtCOI subclades *F1–3* did not correspond very closely to

geography, the nITS subclades showed strong phylogeographic structure. If we ignored the source locations with trivially small sample sizes, three geographic source populations (Pamlico River, NC, Indian River Bay, DE and Chesapeake Bay) showed complete lineage sorting (*gsi* = 1.0) for the nITS gene tree (permutation tests: $P < 0.001$), but incomplete sorting in the mtCOI tree (*gsi* = 0.464–0.812, permutation tests: $P < 0.001$; Fig. 3-3).

Net uncorrected ITS sequence differences among the three deeply diverged clades (*F*, *S* and *X*) were between 26.3% and 30.1%, with average within-clade differences between 0.1% and 3.7%. Within the *F* lineage, net uncorrected sequence differences among lineages *F1-3*, *F4*, and *F5* ranged from 2.7% to 12.9%, with average within-lineage differences less than 2.2%. Consistent with mtCOI data, *F* lineage as a whole showed much higher diversities than both *X* and *S* (π and *H*, *t*-test: $P < 0.001$; Table 3-2). Nucleotide diversity of *X* lineage is significantly lower than *S* (*t*-test: $P < 0.05$; Table 3-2), whereas no significant difference was found in haplotype diversity (*t*-test: $P = 0.127$).

Nuclear ITS sequence variation in lineage *S*, *X*, and *F* is apparently neutral (Table 3-2), though the entire dataset including all lineages together and the subclade *F4* shows deviation from neutrality. No significant differences in substitution rate occurred among three deeply diverged lineages (*X, S* and *F*) and among the subclades within *F* (relative rate test: dK/sd = -0.7245~1.3221, $P > 0.05$).

Lineage distribution and mitochondrial phylogeography

Overall, all three ancient lineages (*F*, *S*, *X*) showed a broad geographic distribution along the US east coast (Fig. 3-4). Seventeen (84%) out of 20 river systems accommodated at least two different lineages, and 9 (45%) of the 20 contained all three (Table 3-1, Fig. 3-4). However, all lineages are clearly patchy at both the large and small (within an estuary) scales, as indicated by varying relative abundance and discontinuous presence/absence among geographic samples. Lineage *S* was the most frequent lineage found in northern estuarine systems between Neuse River, NC and Pettaquamscutt Lake, RI (average 50.8% *vs.*8.4% average for all more southerly sites). In contrast, lineage *F* showed roughly the opposite trend with a slightly different mid-Atlantic transition point (average 9.2% between Chesapeake Bay and Pettaquamscutt Lake *vs.* 44.4% average across more southern localities) (Fig. 3-4). In neither case was the average salinity sampled between north and south regions significantly different (*t*-tests: $P = 0.266–0.630$). No similar macrogeographic trends were found for lineage *X*, although it was overwhelmingly dominant in the Hudson-Raritan River System and completely missing from the Chesapeake Bay and Albemarle-Pamlico Sound (Fig. 3-4). The only European sample, from Finland, was exclusively *X* individuals ($n = 28$ based on RFLPs).

The three deeply-diverged lineages showed contrasting parsimony networks of mtCOI haplotypes. In lineage *X*, all haplotypes were closely related and arranged like a hub and spokes of a wheel. One numerically dominant haplotype (x_1) was central in the network and occurred in all geographic samples, including Finland (Fig. 3-5). Most intermediate frequency haplotypes were either unique to single localities (e.g. x_3) or shared by adjacent drainages (e.g. x_2, x_4, x_7). Only two intermediate frequency haplotypes

(x_5 and x_8) were shared between non-adjacent drainages, being found in both South Daytona Beach, FL and Winyah Bay, SC (Fig. 3-5).

In lineage *S*, the mtCOI haplotype network showed a clear regional structure. Most haplotypes were recorded from the northern region (from Pettaquamscutt Lake, RI to Neuse River Estuary, NC) forming a northern haplotype group, while four southern haplotypes differed with the northern group by several substitution steps (Fig. 3-6). Two closely-related dominant haplotypes (s_1 and s_2) were not only central to northern haplotypes, but also shared by southern populations (s_1 by Savannah and s_2 by South Daytona Beach; Fig. 3-6). Lineage *S* was uncommon in the southern region and therefore the 6 haplotypes found among 22 *S* individuals may be a poor representation of southern diversity.

In contrast to *S* and *X*, lineage *F* had the most complex haplotype network containing no common and widely shared haplotypes. A majority of haplotypes (50 of 64 total) were singletons. Only two haplotypes were found in more than one location, in each case occurring in two or three adjacent geographic samples (Fig. 3-7). The full haplotype set formed three parsimony sub-networks, each separated by greater than 10 substitution steps. The three sub-networks corresponded to the three well-supported major subclades *F1-3*, *F4* and *F5* resolved in the mtCOI gene tree. The *F1-3* sub-network exclusively contained samples from North Carolina northward and showed loose phylogeographic structure within this northern region. Florida haplotypes occur only in one of the two southern sub-networks (Fig. 3-7). The nITS network clearly showed three haplotype groups separated by a number of substitution steps. These groups well

correspond to three geographic sources (Indian River Bay, Chesapeake Bay, and Pamlico River (Fig. 3-7).

Geographical allelic diversity, population differentiation and isolation by distance

After controlling for sample size variation with rarefaction, allelic diversity (R_d) showed the greatest variation among different geographical locations in the *X* lineage (Fig. 3-8). This is partly a consequence of the very low diversity found in lineage *X* from Finland. Among North American samples no clear latitudinal trend was evident in any lineage. Lineage *F* had an average diversity level of 3.414 ($n = 4$), which is significantly higher than those of *X* (1.563, $n = 8$) and *S* (1.834, $n = 7$) (*t*-tests: $P < 0.001$). No significant difference was found between lineage *S* and *X* (1.834 *vs.* 1.563, *t*-test: $P = 0.491$).

To statistically test for geographic structure I pooled samples into distinct estuarine and/or tributary populations within large estuarine systems such as Chesapeake Bay (detailed pooling designs shown in Fig. 3-9). Isolated samples with sizes of $n \leq 5$ were excluded from analyses. In lineage *X*, six US estuarine populations from Connecticut River, CT to South Daytona Beach, FL were used. Geographic structure at a scale of ~2,000 km was significant based on Weir & Cokerham's F_{ST}, AMOVA's Φ_{ST} and exact test of population differentiation (all tests: $P < 0.001$; Table 3-3). No hierarchical partitions had significant between-group variance (all non-significant Φ_{CT}). Thus, 79% of the total variance accounted for within-populations and 21% among-populations.

Analysis on *S* lineage was based on 8 estuarine populations from Connecticut River, CT to South Daytona Beach, FL (Fig. 3-9). *S* lineage showed significant F_{ST}, Φ_{ST}, and overall population differentiation (all tests: $P < 0.001$; Table 3-3). The 2-region hierarchical partition with maximized between-group variance was found with two regions separating 6 northern populations, including North Carolina, from the remaining two southern populations ($\Phi_{CT} = 0.694$, $P < 0.05$), accounting for 69% of total variance (Table 3-3).

In *F* lineage, analyses were only conducted for the combined subclade *F1-3*, with 5 estuarine populations between Thames River, CT and Pamlico River, NC (Fig. 3-9). Significant F_{ST} was found ($P < 0.001$; Table 3-3), although it was substantially lower than those for lineages *X* and *S*. The AMOVA's Φ_{ST} and the exact test of population differentiation were also significant, while no hierarchical structure was detected among the populations (Table 3-3).

Genetic pattern of isolation by distance (IBD) was tested at the continental scale for lineage *X* (excluding Finland), and in comparable northern regions including North Carolina for the *S* and *F* (subclade *F1-3*) lineages. Lineage *X* showed a significant IBD pattern ($r = 0.749$, Mantel test: $P = 0.017$; Fig. 3-9) at the continental scale but much of the association between genetic and geographic distance across the large mid-Atlantic gap in the occurrence of the lineage. Significant IBD patterns ware also found for 6 estuarine populations of northern *S* group ($r = 0.521$, Mantel test: $P = 0.014$; Fig. 3-9) and five estuarine populations of lineage *F1-3* ($r = 0.874$, Mantel test: $P = 0.008$; Fig. 3-9) compared at the same geographic scale.

Discussion

Cryptic species complex and biodiversity of Acartia tonsa *along the US Atlantic coast*

High species diversity was revealed within the nominal species *Acartia tonsa* using genetic tools in this study. Based on genealogical concordance between two independently evolved genes mtCOI and nITS, *A. tonsa* is a species complex consisting of 5 morphologically cryptic and often sympatrically mixed species (*X, S, F1-3, F4* and *F5*).

Conclusion of species status based on genealogical concordance assumes that examined gene markers are orthologous and neutral. However, the assumptions may not hold in many cases confounded often by pseudogenes or non-neutral processes (e.g. Antunes & Ramos 2005; Ballard & Rand 2005; Bazin *et al.* 2006). In this study, neutrality tests for mtCOI and nITS were overall non-significant within distinct lineages of interest for a suite of test statistics (Table 3-2). For mtCOI, only the *S* lineage showed significant negative Fu & Li's D^* and F^*, which may be caused by population demographic effect, such as population expansion or recent genetic hitchhiking (Simonsen *et al.* 1995; Fu 1997). Since Tajima's *D*, Fu & Li's *D* and *F* statistics did not reject the neutrality. I tentatively conclude that mtCOI is neutral in this study, which is also consistent with no heterogeneity of evolutionary rates among lineages (Relative rate tests: $P > 0.5$). However, given the huge census size of *Acartia* populations, deep lineages and low polymorphisms within lineages are still unexpected for a selectively neutral locus. Neutrality of mtDNA has long been questioned (see William *et al.* 1995). Bazin *et al.*'s (2006) provided solid evidence showing constant mtDNA diversity among animal groups and poor reflection of effective population size. According to Bazin *et al.*'s (2006),

mtDNA is subject to more frequent selective sweeps for invertebrates with larger population size, which effectively wipe the diversity for the non-recombining genome. We should be very cautious to use mtDNA alone to infer population history, even neutrality were not rejected by traditional tests. In this study, mtCOI and nITS were combined to test genealogical concordance and infer evolutionary history of *Acartia* populations.

Divergence of nITS among *Acartia* lineages shows more rapid evolutionary rate than mtCOI, creating deep divergence between lineages and low polymorphism within lineages (Fig. 3-3). More strikingly, genealogical sorting of nITS was also found rapid in three geographic populations with *gsi* = 1.0 (Fig. 3-3). This pattern could be best explained by concerted evolution of nITS. The multicopy ITS locus is often assumed neutral under concerted evolution (Gerbi 1985; Hillis & Dixon 1991). As expected from concerted evolution, gene duplication and homogenization reduce genetic variation within populations and increase divergence between isolated populations. Rapid genealogical sorting of nITS in geographically separated populations provided strong evidence for geographic isolation and gene flow constrains between *Acartia* populations, a process essential to allopatric speciation (Coyne & Orr 2004). With regards to this aspect, nITS serves as a good genetic marker for detection of geographic structure and gene flow between populations. Although neutrality tests were rejected by some testing statistics for some lineages, such as *F4* (Table 3-2), nITS is still valid for inference of species status based on genealogical concordance, because it does not confound longstanding reproductive isolation. When genealogical sorting of mtCOI was completed

for certain lineages (*X*, *S*, *F4*, *F5*), concordant patterns arose, providing evidence for reproduction isolation between these lineages.

As a multi-copy nuclear gene, ITS sequence data often has multiple ambiguous sites that make the phase determination and haplotype inference difficult using traditional methods with an assumption of Hardy-Weinberg equilibrium. The accuracy of allele determination should be concerned because it directly influences evolutionary conclusions made on the basis ITS sequence divergence. In this study, I am confident of the accuracy of my ITS sequence data and its validation in phylogenetic and population genetic inference in *A. tonsa* for three reasons. First, all the nITS sequences have good qualities with clean base signals for both homozygous and heterozygous sites. Only a small amount of cases (21 out of 138 total) have ambiguous sites that need phase determination. For the selected 378 bp fragment, all the ambiguous sites have only 2-base ambiguities, allowing an assumption of diploidy when determining their phase. The majority of homozygous cases provided solid reference alleles for inference of alleles in heterozygotes. Second, although the nITS sequences have significant variation in length ranging from 472 bp to 619 bp. Within groups with the same length, sequences are conveniently aligned with no indels (insertions and deletions) or just a single-base indel. The selected 378 bp fragment is located in the central region of the full nITS alignment with a large amount of indels. It is a region with fewer indels, more conservative sites across lineages and presumably containing the 5.8s rDNA protion (Appendix A, Table A-1). Therefore, the observed large variation in nITS sequence data are primarily attributed to sequence divergence across lineages and less likely affected by inaccuracies in allele calls. Third, alleles determined in this study are also similar to the results obtained using

computer program PHASE (Stephens *et al.* 2001). In addition, the Hudson & Kaplan (1985)'s four-gamete tests showed no significant recombination events with consideration of both phased and unphased sequences.

Two of the geographically widespread and phylogenetically deep lineages analyzed here (*X*, *S*) were initially described by Caudill & Bucklin (2004) based on samples from three New England estuaries, Georgia and Texas. The single European sample examined here consisted of a single lineage, *X*, consistent with founder effects during the anthropogenic introduction of *Acartia* (Mauchline 1998; David *et al.* 2007). Lineage *X* also had the most regionally discontinuous distribution along the U.S. coast, being entirely absent from very well sampled mid-Atlantic estuaries, suggesting that it may be a recent nonindigenous arrival on both sides of the Atlantic.

By sampling multiple habitats within each US estuary, our continental survey also revealed a broadly distributed and rapidly diversifying lineage (*F*) that was previously known only from the Gulf of Mexico (Caudill & Bucklin 2004, their haplotype group C; Fig. 3-2). The reason that *F* lineage was not prevailing in Caudill & Bucklin's (2004) study is very likely due to their insufficient sampling. Their samples were mainly from the New England area where *F* lineage is less abundant and sample size was relative small. Nevertheless, Caudill & Bucklin (2004) found a very distinct but rare haplotype group in Rhode Island and Massachusetts estuaries that was not recorded in this study, and another highly divergent form in the Northeast Pacific. Together, these two studies suggest that diversification within this species complex is occurring in both temperate and subtropical waters with more lineages yet to be found at a global scale.

Because Caudill & Bucklin (2004) found representatives of lineage *F* (their haplogroup C) only in Texas, they interpreted its deep divergence with other lineages (*S*, *X* of this study) as evidence supporting the genetic distinctiveness between *Acartia* populations in the Atlantic and Gulf of Mexico (GOM). With the benefit of our broader sampling of the *F* lineage from the US east coast it tentatively appears that the GOM haplotypes are nested within Atlantic *F* sublineage diversity. This observation requires further study because the 16S data are not sufficient to provide robust hypotheses about the clade structure within the diverse *F* lineage. However, we note that the basal position of sublineage *F5* was found with both mtDNA and nITS and is also reflected in the 16S gene tree (albeit with weak bootstrap support), suggesting some degree of accurate 16S clade structure. Regardless of which *F* sublineage proves to be the closest relative to the Texas C haplotypes from Caudill & Bucklin (2004), the relatively long branch for the latter compared to Atlantic-Atlantic divergences still supports the contention (Caudill & Bucklin 2004) that the Florida peninsula has acted as a gene flow barrier as seen with other taxa (Avise 2004).

Molecular data also suggested that diversification of species within *A. tonsa* might have a long evolutionary history. When applying a molecular clock calibrated for the mtCOI gene of marine crustaceans (~2% divergence per million years, e.g. Knowlton *et al.* 1993; Knowlton & Weigt 1998; Schubart *et al.* 1998), three deep Acartia lineages (*F*, *S*, *X*) shared most recent common ancestors 3.2–2.6 MYA during late Pliocene. Within the diversifying lineage *F*, three well-supported monophyletic clades (*F1-3, F4*, *F5*) were derived 1.6–0.3 MYA during and since the early Pleistocene. While these dates are very tentative, the general distinction between three ancient lineages (*F*, *S*, *X*) and relatively

recent diversification within the *F* lineage is apparent in all three genes, mtCOI, 16S and nITS.

Combined all above-discussed evidence together, I believe that at least three deeply diverged lineages (*F, X, S*) are reproductively isolated species, rather than structured populations within a single species. The key distinction between species and structured populations lies in the signals of reproductive isolation. These signals are stronger in deeply diverged lineages *F, X, S*, given their sympatric distribution where deep divergence has to be maintained by reproductive isolation, if not strong selection. Strong genealogical concordance was also observed for three *F* sublineages (*F1-3*, *F4*, *F5*), which allow us to hypothesize that these lineages are reproductively isolated species too. However, evidence of reproductive isolation is less obvious for 3 *F* sublineages, given their geographically isolated distribution with relatively short genetic distances among them. The genealogical concordance principle does not distinguish scenarios causing reproductive isolation, such as geographic isolation, premating or postmating barriers, etc. In the case of geographic isolation, genealogical divergence and concordance can reasonably be found among geographically isolated populations with unknown status of reproductive incompatibilities. It is thus important to further test the species status of *F* sublineages through laboratory crossing experiments. Three interbreeding attempts between geographically isolated populations have been reported for three *Acartia* nominal species. Crosses were successful in *A. californiensis* between populations from Estero de Punta Banda, Baja California and Mission Bay, California (Trujillo-Ortiz *et al.* 1999), but failed in *A. clausi* between Atlantic and Pacific populations (Enrique Carrillo *et al.* 1974). Hill (2004) conducted crosses among four *A.*

tonsa populations from Great Bay, NH, Buzzards Bay, MA, Narragansett Bay, RI, and Beaufort Inlet, NC (Hill 2004), with successes among the latter three and failure between Great Bay with others. If sympatry of cryptic lineages is common in all *Acartia* taxa, it is hard to compare and explain these results unless we know both the genetic structure and lineage identities in above tests. Given the known *Acartia* cryptic lineages and geographic distribution from the present study, it is more meaningful to test reproductive isolations among lineages rather than geographic populations.

Southern origin and stepwise northward diversification of Acartia *lineage F*

The three ancient *Acartia* lineages (*F*, *S*, *X*) showed different phylogeographic patterns based on mtCOI DNA variation. The most striking and novel finding was the stepwise northward divergence of lineage *F* subclades. Three strongly supported subclades (*F5*, *F4*, *F1-3*) of lineage *F* were associated with three geographic regions along the US east coast: basal clade *F5* in Florida, middle subclade *F4* in South Atlantic Bight, and recent subclade *F1-3* in the mid-Atlantic Bight (Fig. 3-1, 3-7). This latitudinal polarity indicates a southern center-of-origin for lineage *F*. Both branch lengths and genealogical concordance suggest that sublineages *F4* and *F5* are relatively old and probably pre-Pleistocene. Two points argue for a post-Pleistocene origin for the population structure found within sublineage *F1-3* in the mid-Atlantic Bight. First, in this region the spatial configuration of low salinity habitat occupied by lineage *F* in the upper reaches of estuarine tributaries was too dynamic during Pleistocene glacial cycles to leave a simple allopatric signature among contemporary estuaries (Cronin 1988; Hobbs 2004). Second, allopatry among post-Pleistocene estuaries seems to explain much of the

diversification within sublineage *F1-3* as evidenced by the large Φ_{ST} for mtCOI (48% of variation explained) and near-perfect phylogeographic structure found with nITS. Post-Pleistocene environmental changes have been associated with latitudinal genetic gradients in other marine species, but mostly with respect to levels of intraspecific population structure or intra-population diversity (Castric *et al.* 2001; Wares 2002; Marko 2004; McMillen-Jackson & Bert 2004). Few studies, terrestrial or marine, have found evidence for this level of population structure due to post-Pleistocene neutral drift (i.e., without positing Pleistocene refugia). Thus, under this interpretation the rate of diversification appears to have been very high for *F* lineage populations in the mid-Atlantic and New England.

In this study we provided the first evidence of a northward stepwise diversification pattern in *Acartia* along the US east coast. The directional colonization, if associated with post-Pleistocene climate change, should be seen in all the co-distributed *Acartia* lineages. It is however only manifest in lineage *F* as a result of interaction between several ecological and evolutionary processes pertinent to *F*. First, the stepwise diversification pattern indicates a restriction of post-Pleistocene gene flow among estuaries in lineage *F*. This restriction is reasonably attributed to the ecological affinity of lineage *F* for low salinity waters (Chen & Hare 2008; Chapter 4 of this thesis). Geographic distance, salinity tolerance, or both could limit gene exchange between estuaries. Second, if genetic drift among allopatric populations was the diversification mechanism in lineage *F* and all *F* sublineages have had the same low-salinity habitat restriction leading to isolated populations, then comparable rapid diversification also would be expected within sublineage *F4* in the South Atlantic Bight and *F5* in Florida.

Our geographic sampling in the south is not sufficient to rigorously reject phylogeographic structure within these sublineages, but at least for *F*5 there is none evident even though the distribution of this clade spans a widely recognized biogeographic marine province boundary at Cape Canaveral (Saunders *et al.* 1986; Reeb & Avise 1990; Hare & Avise 1998; Herke & Foltz 2002; Baker *et al.* 2008). The contrast in phylogeographic structure between sublineages *F4* and *F5* and *F1-3* suggests that additional mechanisms besides post-Pleistocene allopatry may have been important for diversification in the north, but our current data are equivocal.

It is worth noting that the northward stepwise diversification of *F* lineage counters the direction of flows along the US Atlantic coast, which in average go southward due to the Coriolis effect (Epifanio & Garvine 2001). Behind the conflicting biological and physical patterns, there are some reconciling points that may deserve much attention. First, the general southward coast flows did not have a strong impact on distribution and dispersal of *Acartia F* lineages. Otherwise, we have little chance to find the northward stepwise diversification pattern. This further reinforces the conclusion of geographic restrictions of *F* lineage by either physical or physiological barriers. Second, the northward stepwise diversification of *F* lineage is possibly caused by specific historical events with irregular physical processes than well recognized regular physical processes at present days. It is worth looking into paleontological oceanography of Atlantic coast for evidence of historical change of flow patterns and salinity structures. Third, possibilities also exist for plenty of sporadic physical events favoring northward transport of *Acartia*, including wind-induced northward surface flow during summer, occasional

northward flows caused by frontal eddies or meanders of the Gulf Stream front (reviewed by Epifanio & Garvine 2001; Werner *et al.* 1993).

Contrasting phylogeography of Acartia *lineages*

Lineage *X* appeared to be a relatively well-mixed population, with a common haplotype shared across all geographic samples and low divergence among haplotypes (Fig. 3-5). In strong contrast, lineage *F* showed multiple relatively isolated regions and population structure even within regions (Fig. 3-7). Lineage *S* had an intermediate pattern between those of *X* and *F*, with both widely-shared haplotypes and heavily regionalized ones roughly corresponding to zoogeographic provinces (Fig. 3-6). These contrasting phylogeographic patterns suggest that the three major lineages have among-estuary gene flow propensities that are rank ordered $X > S > F$.

The strongest gene flow barriers for lineages *F* and *S* seem to be spatially associated with long-recognized zoogeographic province boundaries at Cape Hatteras and Northeast Florida (Briggs 1974, 1995). These province boundaries have been recognized based on many different kinds of taxa (Briggs 1974; Engle & Summers 1999; Wares 2002; Briggs 2007), but have not been considered with respect to holoplankton. The similar biogeographic patterns across taxa suggest common mechanisms, possibly including both hydrographic barriers and latitudinal adaptive constraints. The absence of significant regional AMOVA partitioning of *X* lineage diversity, despite a large gap in its continental distribution, suggests that this lineage may be more capable than the others of dispersing through oceanic shelf waters. However, even in the *X* lineage most haplotypes besides the most common are found at only one or a few regionally proximate sites and

overall F_{ST} is moderate indicating that only adjacent estuaries are typically accessed via along-shelf dispersal. This is consistent with the physical features of the US Atlantic coast. Hydrographic processes and flow fields along the US east coast evidently support regional movement of zooplankton between adjacent estuaries, while a long-range displacement requires series of stepping-stone movements (Scheltema 1975; Epifanio & Garvine 2001).

Sharing the same physical system with lineage *X*, lineage *F* however showed strong population substructure and significant IBD at smaller geographic scales, suggesting that gene flow was substantially reduced among estuaries. What is particularly interesting for this lineage is its strong affinity to low-salinity waters (Chen & Hare 2008; Chapter 4 of this thesis). Higher gene flow restriction in lineage *F* populations at smaller scales might be associated with its physiological preference and/or tolerance of salinity. If *F* is a lineage that has adapted to low-salinity environment, high salinity may act as a physiological barrier for lineage *F* to disperse across estuaries.

Phylogeographic structure of lineage *S* is comparable to the two traditional zoogeographic provinces, Virginian in north and Carolinian province in south, except the boundary in this case is south of Cape Hatteras and the Albemarle-Pamlico Sound behind the Cape (Briggs 1974; Engle & Summers 1999). Estuaries in the Albemarle-Pamlico Sound have similar patterns of *Acartia* lineage composition and distribution with those in Chesapeake Bay. Both systems were dominated by lineage *F* and *S* with the absence of *X* (Fig. 3-4) and shared closely related haplotypes (Fig. 3-6). The protrusion of Cape Hatteras is well recognized as a zoogeographic break due to sharp temperature gradients and convergence of currents (Palumbi 1994; Engle & Summers 1999). However, large

amounts of estuarine outflow from Chesapeake Bay may be entering Albemarle-Pamlico Sound and crossing Cape Hatteras southward (Epifanio & Garvine 2001), which provides mixing processes for the two systems. Interestingly, similar zoological affinities also were found in the benthic communities of these large estuaries. Engle & Summers (1999) examined the latitudinal gradients in benthic community composition along the entire US east coast. Benthic macroinvertebrate composition of Albemarle-Pamlico Sound was more similar to Virginian fauna than to the fauna of South Carolina and Georgia (Engle & Summers 1999). One parsimonious explanation could be that benthic species distributions were largely controlled by their pelagic larval dispersal and recruitment, subject to the same hydrographic processes for as copepods.

Large scale phylogeography studies also provide us an opportunity to test anthropogenic introduction of *Acartia*. *Acartia* are subject to anthropogenic transport in ballast waters, a dispersal and invasion path for many marine zooplanktons (Carlton & Geller 1993; Cordell & Morrison 1996; Lavoie *et al.* 1999). In this study, several commercial harbors were involved, including New Haven, CT, New York, NY, Baltimore, MD, and Savannah, GA. If anthropogenic transport is an important process for *Acartia* dispersal, we should see no significant geographic structure and IBD patterns in either lineage. Thus the positive phylogeographic structures of this study basically rejected the anthropogenic transport hypothesis along the US Atlantic coast, or at least this process did not play an important role in *Acartia*. Nevertheless, I would leave this hypothesis for a further test for two reasons. First, at a global scale, anthropogenic transport has to play an important role in *Acartia* dispersal and global wide colonization (Mauchline 1998; David *et al.* 2007). In highly dispersed lineage *X*, the finding that

European sample shared a common mtCOI haplotype with Atlantic samples could indicate the trans-Atlantic anthropogenic introduction event. Therefore, testing anthropogenic transport and its role in *Acartia* phylogeography might be more effective at the global scale with more continental geographic samples. Second, even on the US Atlantic coast, the possibility of anthropogenic transport of *Acartia* may not be completely ruled out. In lineage *S*, two common haplotypes were shared by northern and southern haplotype groups, even though a strong structure was found between them. The common haplotypes carried by a few southern individuals could be ancient polymorphisms retained by southern populations. Alternatively, they may possibly be transported through ballast waters from northern estuaries. Unfortunately the southern samples were too few to address this question.

Conclusions

The estuarine copepod *Acartia tonsa* has multiple cryptic genetic lineages co-distributed along the US east coast. Genealogical concordance between mtCOI and nITS genes supports reproductively isolated species status for at least five major lineages. The three major lineages (*X*, *F* and *S*) originated prior to the Pleistocene, while some sublineages within *F* showed a post-Pleistocene diversification with a southern center of origin. Our results are consistent with geographic isolation and limited dispersal generating *Acartia* population divergence, with variation among lineages in their sensitivity to this genetic drift mechanism. Gene flow barriers may include physical factors (e.g. distance between estuaries, latitudinal gradient, hydrodynamics) and physiological adaptation to environmental factors (e.g. Salinity, temperature). The *A.*

tonsa species complex provides a new window for studying marine biogeography and evolutionary history of holoplankton in the western North Atlantic.

Table 3-1 Information of geographic and genetic sampling of *Acartia tonsa* on the US Atlantic coast. Three major lineages (*F, S, and X*) were classified based on mtCOI gene marker and its concordance with nITS gene. Sequence sample sizes are listed for mtCOI (*n*), nITS (2*n*), and mt16S (*n*). RFLP sample sizes were listed for pooled populations (*N*, in bold).

					mtCOI (*n*)			nITS (2*n*)			mt16S (*n*)		
Sample name & locality		Date	Latitude	Longitude	*F*	*S*	*X*	*F*	*S*	*X*	*F*	*S*	*X*
Rhode Island													
PQL	Pettaquamscutt Lake	12 August 2005	41.49°N	71.45°W		4			4				
	Pettaquamscutt Lake (*N* = 19)				**10%**	**90%**	**0%**						
Connecticut													
TR1	Thames River Estuary	11 August 2005	41.35°N	72.09°W	1								
TR2	Thames River Estuary	11 August 2005	41.43°N	72.10°W	10			1					
	Thames River Estuary(*N* = 26)				**42%**	**39%**	**19%**						
CTR1	Connecticut River Estuary	12 August 2005	41.35°N	72.38°W		3	6		2	2			
CTR2	Connecticut River Estuary	12 August 2005	41.32°N	72.35°W		5							
CTR3	Connecticut River Estuary	13 August 2005	41.32°N	72.35°W		3	3						
CTR4	Connecticut River Estuary	13 August 2005	41.29°N	72.35°W		3	2						
CTR5	Connecticut River Estuary	13 August 2005	41.35°N	72.38°W		1							
	Connecticut River Estuary (*N* = 125)				**1%**	**85%**	**14%**						
	Housatonic River Estuary (*N* = 22)				**0%**	**100%**	**0%**						
New York – New Jersey													
HUR1	Hudson River Estuary	13 August 2005	40.83°N	73.97°W			2						
HUR2	Hudson River Estuary	14 August 2005	40.70°N	74.07°W		1							
HAR	Hackensack River	14 August 2005	40.85°N	74.03°W			6			6			
NYB	New York Bay	14 August 2005	40.54°N	74.13°W		4	7		1				
	Hudson-Raritan River System (*N* = 171)				**1%**	**6%**	**93%**						
	Great Egg Harbor (*N* = 19)				**0%**	**0%**	**100%**						
Delaware													
DB1	Indian River Bay	8 August 2003	38.60°N	75.17°W	7	2	4	4	3	2	1		1
DB2	Indian River Bay	8 August 2003	38.63°N	75.01°W	2			3			1		
	Indian River Bay (*N* = 24)				**54%**	**17%**	**29%**						
DB4	Delaware Bay	28 May 2005	39.22°N	75.29°W	1	2	4		2	5			
DB5	Delaware Bay	28 May 2005	39.08°N	75.19°W	1	4		1	3				
	Delaware Bay (*N* = 21)				**10%**	**52%**	**38%**						

Table 3-1 *continued*

Sample name & locality		Date	Latitude	Longitude	mtCOI (*n*)			nITS (2*n*)			mt16S (*n*)		
					F	*S*	*X*	*F*	*S*	*X*	*F*	*S*	*X*
Chesapeake Bay, MD-VA													
BIH	Baltimore Inner Harbor	31 July 2003	39.28°N	76.61°W	5	4		12	4		3	1	
PA1	Patuxent River	4 August 2004	38.50°N	76.67°W	9	4		3	3		4		
PO1	Potomac River	26 May 2005	38.19°N	76.71°W	5	5							
C1	Choptank River	30 March 2004	38.72°N	76.01°W	1						2		
C2	Choptank River	30 March 2004	38.66°N	75.96°W	4			5	2				
C6	Choptank River	10 August 2004	38.72°N	76.01°W	1			1			1		
G1	Choptank River	8 August 2003	38.61°N	75.95°W	4			2			3		
G2	Choptank River	13 July 2003	38.57°N	76.06°W		3			3				
G3	Choptank River	12 July 2003	38.59°N	76.13°W	4	18		2	2		1		
L1	Little Choptank River	8 April 2002	38.52°N	76.27°W		3			1				
L2	Little Choptank River	15 August 2002	38.52°N	76.27°W		5			4			1	
T1	Mainstem	22 July 1996	39.33°N	76.20°W	3								
T2	Mainstem	22 July 1996	39.23°N	76.24°W	4								
T4	Mainstem	6 May 1996	39.11°N	76.30°W		6							
T6	Mainstem	5 May 1996	39.06°N	76.33°W		3							
T8	Mainstem	6 May1996	39.00°N	76.36°W		5							
T11	Mainstem	20 July 1996	38.83°N	76.41°W		5							
M3	Mainstem	25 May 2005	38.75°N	76.43°W		5							
	Chesapeake Bay (*N* = 1649)				**16%**	**84%**	**0%**						
North Carolina													
PML1	Pamlico River Estuary	18 September 2005	35.54°N	77.04°W	2			2					
PML2	Pamlico River Estuary	18 September 2005	35.38°N	76.75°W	8			5					
PML3	Pamlico River Estuary	18 September 2005	35.25°N	76.59°W	2			1					
	Pamlico River Estuary (*N* = 68)				**69%**	**31%**	**0%**						
NSR	Neuse River Estuary	18 September 2005	34.97°N	76.81°W		12							
	Neuse River Estuary (*N*=46)				**39%**	**61%**	**0%**						
	Albemarle Sound (*N* = 93)				**90%**	**10%**	**0%**						
	Cape Fear River Estuary (*N* = 22)				**91%**	**0%**	**9%**						

Table 3-1 *continued*

Sample name & locality		Date	Latitude	Longitude	mtCOI (*n*)			nITS (2*n*)			mt16S (*n*)		
					F	*S*	*X*	*F*	*S*	*X*	*F*	*S*	*X*
South Carolina													
GTL	Winyah Bay	18 September 2005	33.37°N	79.27°W			11			4			
	Winyah Bay (*N* = 30)				**0%**	**0%**	**100%**						
Georgia													
SVR2	Savannah River Estuary	20 September 2005	32.04°N	80.87°W		13	5		1	5			
SVR3	Savannah River Estuary	20 September 2005	32.04°N	80.89°W			4			4			
SVR4	Savannah River Estuary	20 September 2005	32.04°N	80.91°W	1		1			1			
SVR5	Savannah River Estuary	20 September 2005	32.05°N	80.93°W			4			4			
SVR7	Savannah River Estuary	20 September 2005	32.09°N	80.99°W			1			1			
SVR8	Savannah River Estuary0	20 September 2005	32.08°N	81.05°W	2		1	2					
	Savannah River Estuary (*N* = 176)				**16%**	**36%**	**48%**						
	South Georgia (*N* = 61)				**31%**	**5%**	**64%**						
Florida													
JAX	Jacksonville	29 April 2005	30.33°N	81.62°W	12		3	3		3	1		2
	Jacksonville (*N* = 24)				**71%**	**0%**	**29%**						
FD	Faver-Dykes Stake Park	30 April 2005	29.67°N	81.27°W	11		2	3			3		1
	Faver-Dykes (*N* = 17)				**88%**	**0%**	**12%**						
SDB	South Daytona Beach	1 May 2005	29.21°N	81.01°W	1	7	12		4	3			
	South Daytona Beach (*N* = 63)				**2%**	**13%**	**85%**						
SEB	Sebastian	1 May 2005	27.83°N	80.51°W	3	2		2	2		1	1	
	Sebastian (*N* = 16)				**19%**	**12%**	**69%**						
European coast													
FIN	Poiju, Finland	18 August 2005	59.86°N	23.27°E			10			4			
	Finland (*N* = 28)				**0%**	**0%**	**100%**						

Table 3-2 Summary statistics of *Acartia tonsa* populations on the US Atlantic coast, including sample size (*n* for mtCOI, 2*n* for nITS), nucleotide diversity (π), haplotype diversity (*H*), Allelic diversity based on rarefactioned allelic richness (R_d), Tajima's *D*, and Fu & Li's statistics. Significant values are highlighted in bold face. Statistical significance: $0.01 < P < 0.05$ (*), $0.001 < P < 0.01$ (**), $P < 0.001$ (***). n/a: not applicable.

Lineage	Sample size	Number of alleles	R_d	π	*H*	Tajima's *D*	Fu & Li's statistics			
							D	*F*	*D**	*F**
mtCOI										
S	132	22	3.028	0.0055	0.738	-1.138	-1.879	-1.919	-3.308*	-2.965*
X	88	12	3.006	0.0024	0.620	-1.460	0.794	-0.010	0.445	-0.282
F overall	104	64	n/a	0.0290	0.974	0.378	-0.274	-0.119	-1.070	-0.560
F1	40	25	n/a	0.0059	0.942	-1.704	-1.153	-1.664	-2.296	-2.481
F2	26	14	n/a	0.0071	0.926	-0.545	0.302	0.087	-1.093	-1.082
F3	8	7	n/a	0.0101	0.964	-0.063	0.712	-0.542	-0.327	-0.295
F1-3	74	46	4.751	0.0174	0.974	-0.265	-0.272	-0.372	-1.758	-1.414
F4	25	13	n/a	0.0042	0.777	-0.967	-2.182	-2.101	-2.076	-2.032
F5	5	5	n/a	0.0043	1.000	0.000	1.448	1.515	0.000	0.000
Entire data	324	98	n/a	0.0904	0.926	3.436**	1.298	2.926*	-0.193	1.925*
nITS										
S	82	2	n/a	0.0066	0.287	1.068	1.064	1.209	1.272	1.423
X	86	2	n/a	0.0011	0.396	1.174	0.498	0.813	0.502	0.817
F overall	108	19	n/a	0.0371	0.908	-0.276	2.029*	1.263	1.786*	1.106
F1-3	82	15	n/a	0.0200	0.864	0.814	0.967	1.176	0.644	0.844
F4	20	3	n/a	0.0181	0.653	2.647**	1.641*	2.329*	1.501*	2.129*
F5	6	1	n/a	0.0000	0.000	n/a	n/a	n/a	n/a	n/a
Entire data	276	23	n/a	0.1998	0.865	5.163***	n/a	n/a	2.706*	4.674*

Table 3-3 Summary statistics and estimates of population differentiation for major lineages of *Acartia tonsa* on the US Atlantic coast. Within each lineage, samples were pooled into estuaries and/or tributary populations (see Fig. 3-9); populations with small sample size (< 5) were excluded from analysis. Statistically significant estimates were highlighted in bold face and classified as $P < 0.05$ (*), $P < 0.01$ (**), $P < 0.001$ (***). n/a: not applicable.

Lineage	*X*	*S*	*F1-3*
Sample size (*n* individuals)	63	106	55
Population number (*N*)	6	8	5
Weir & Cockerham's F_{ST}	0.296***	0.212***	0.037***
Exact test of population differentiation	$P < 0.001$	$P < 0.001$	$P < 0.001$
AMOVA			
Φ_{ST} (%variance)	0.210*** (21%)	0.712*** (29%)	0.476*** (48%)
Φ_{CT} (%variance)	n/a	0.694* (69%)	n/a
Φ_{SC} (%variance)	n/a	0.086* (3%)	n/a

Figure legends

Figure 3-1 Neighbor-joining gene tree of 98 mitochondrial COI haplotypes based on uncorrected *p*-distance for *Acartia tonsa* on the US Atlantic coast. Three major lineages (*S*, *F*, and *X*) separated by long branches and five subclades (*F1–5*) are labeled. Bootstrap support values (>70%) from all phylogenetic analysis methods (NJ/MP/ML) are shown for supported clades. A sequence of the congeneric species *A. hudsonica* was used as the outgroup to root the gene tree.

Figure 3-2 Neighbor-joining gene tree of mitochondrial 16S haplotypes of *Acartia tonsa* based on uncorrected *p*-distance. Bootstrap support values (> 70%) from all phylogenetic analysis methods (NJ/MP/ML) are shown for supported clades. Boxed labels represent sequences collected in this study and others are from Caudill & Bucklin (2004). Lineage identity of the sequenced individuals was determined based on their mtCOI sequences. Three deep mtCOI lineages (*F*, *X*, *S*) were coded with color shade of boxes (white for *F*, gray for *S*, and black for *X*). Parenthetical labels next to boxed labels are also used to indicate lineage identity including three *F* sub-lineages (*F1-3*, *F4*, and *F5*).

Figure 3-3 Genealogical concordance between mtCOI (left) and nITS (right) gene trees for *Acartia tonsa* on the US Atlantic coast. Both are neighbor-joining trees based on uncorrected *p*-distance. The same individuals were used to collect sequences from both genes and are represented with circles (mtDNA) or squares (nITS) next to the sequence obtained. Bootstrap support values (>70%) from all phylogenetic analysis methods (NJ/MP/ML) are shown for supported clades. Concordant clades are highlighted with dotted boxes. Individuals of subclades *F1–3* were labeled with numbers (1, 2, 3 respectively) corresponding to clade membership. For the top non-concordant portion of

the figure, geographic sources were distinguished for individuals with filled colors, including Chesapeake Bay (green), Indian River Bay, DE (cyan), Delaware Bay (gray), Pamlico River, NC (red), and Thames River, CT (black). Genealogical sorting index (*gsi*) and its statistical significance are listed for clades *F1–3* to summarize two tests: genealogical concordance between mtDNA and nITS (all samples included) and phylogeographic sorting of mtDNA haplotypes from Chesapeake Bay (green), Indian River Bay (cyan) and Pamlico River (red).

Figure 3-4 Overall geographic distribution of *Acartia tonsa* lineages on the US Atlantic coast. Acartia tonsa individuals are genotyped based on restriction fragment length polymorphisms. If an estuary was represented by a single sample, it is represented using an open circle while systems with multiple pooled samples are enclosed in a rectangle. Number of individuals, *n*, is listed under each location name.

Figure 3-5 Mitochondrial COI minimum spanning network for *Acartia X*. Sampling localities are shown with boxes and the fill color distinguishes nine major estuarine systems from the US Atlantic coast and the European coast. Network balloons represent mtCOI haplotypes recorded in *X* lineage and labeled from *x*1 to *x*12. Balloons are sized in proportion to the overall frequency of the haplotypes. For each haplotype, numerical compositions of different geographic sources are showed as a pie chart inside the balloon, with geographic origins color-coded according to the color codes used for geographic locations on the left map. Small open balloons represent hypothetical, unobserved haplotypes and balloon connections all represent a single nucleotide substitution.

Figure 3-6 Mitochondrial COI minimum spanning network for *Acartia S* lineage. Sampling localities are shown with boxes and the fill color distinguishes nine major

estuarine systems from the US Atlantic coast. Network balloons represent mtCOI haplotypes recorded in *S* lineage with only eight common haplotypes labeled from *s*1 to *s*8. Balloons are sized in proportion to the overall frequency of the haplotypes. For each haplotype, numerical compositions of different geographic sources are showed as a pie chart inside the balloon, with geographic origins color-coded according to the color codes used for geographic locations on the left map. Small open balloons represent hypothetical, unobserved haplotypes and balloon connections all represent a single nucleotide substitution.

Figure 3-7 Mitochondrial COI (left) and nuclear ITS (right) minimum spanning networks for *Acartia F* lineage. Three circled sub-networks on the left are based on mtCOI gene locus. The right square-boxed one is nITS network for lineage *F1-3*. Sampling localities are shown with boxes and the fill color distinguishes nine major estuarine systems from the US Atlantic coast. Network balloons represent mtCOI or nITS haplotypes recorded in *F* lineage. Balloons are sized in proportion to the overall frequency of the haplotypes. For each haplotype, numerical compositions of different geographic sources are showed as a pie chart inside the balloon, with geographic origins color-coded according to the color codes used for geographic locations on the left map. Small open balloons represent hypothetical, unobserved haplotypes and balloon connections all represent a single nucleotide substitution. Genetic distances (substitution steps) among three circled sub-networks are too large to construct a reliable full network based on statistical parsimony. Therefore, the three sub-networks are connected with dashed lines. The three sub-networks also correspond to three *F* sub-clades (*F1-3, F4, F5*) in the mtCOI gene tree (Fig. 3-1). Based on the cladogenesis of three clades in the mtCOI tree, a northward

stepwise diversification of *A. tonsa* along the US east coast is indicated by an arrow-headed curve near the coastline.

Figure 3-8 Rarefactioned allelic diversity and its geographic variation in *Acartia* lineages *X*, *S*, and *F1-3*. Allelic diversity was rarefactioned to a sample size of 5. The x-axis from left to right shows a list of geographic populations in the southward order. Data points are shown in scattered circles with filled color coding different lineages (white for *F1-3,* gray for *S*, and black for *X*). Adjacent data points are connected with solid lines while gapped by dotted lines. Population abbreviations: FIN – Finland; TR – Thames R. Estuary, CT; CTR – Connecticut R. Estuary, CT; HUR – Hudson R. Estuary, NJ-NY; DB – Delaware Bay; IRB – Indian R. Bay, DE; CB – Chesapeake Bay; PML – Pamlico R. Estuary, NC; NSR – Neuse R. Estuary, NC; GTL – Winyah Bay, SC; SVR – Savannah R. Estuary; SDB – South Daytona Beach, FL.

Figure 3-9 Isolation by distance (IBD) tests for *Acartia* lineages *X*, *S* and *F* on the US Atlantic coast. The figure also shows sample lumping designs for AMOVA analyses and IBD tests. Two square-boxed southern populations (Savannah R. Estuary and S. Daytona Beach) were excluded from IBD test of lineage *S* for comparison with IBD of sublineage *F1-3*.

Figure 3-1

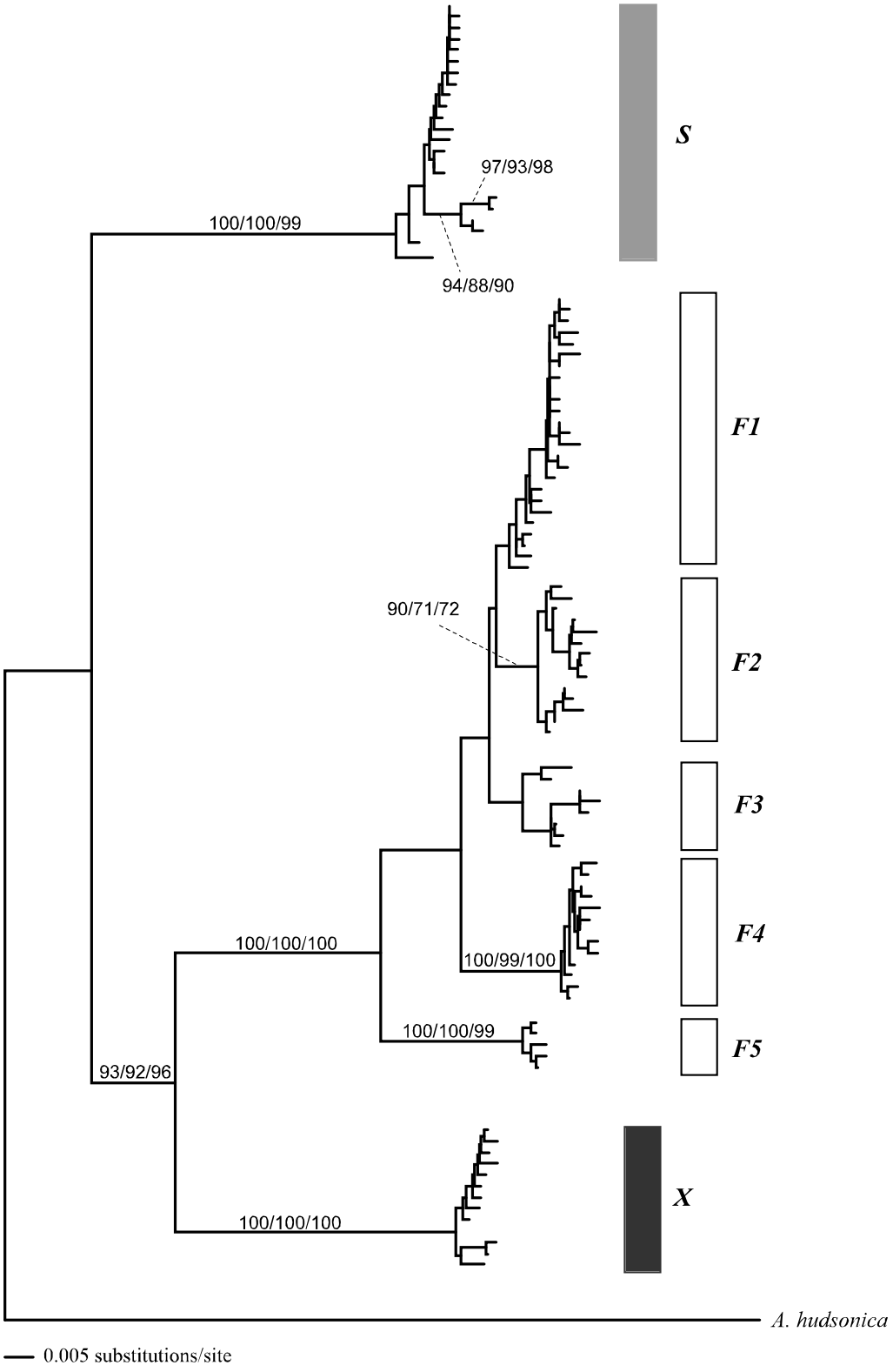

S
97/93/98
100/100/99
94/88/90
F1
90/71/72
F2
F3
F4
100/100/100
100/99/100
100/100/99
F5
93/92/96
X
100/100/100
A. hudsonica
0.005 substitutions/site

Figure 3-2

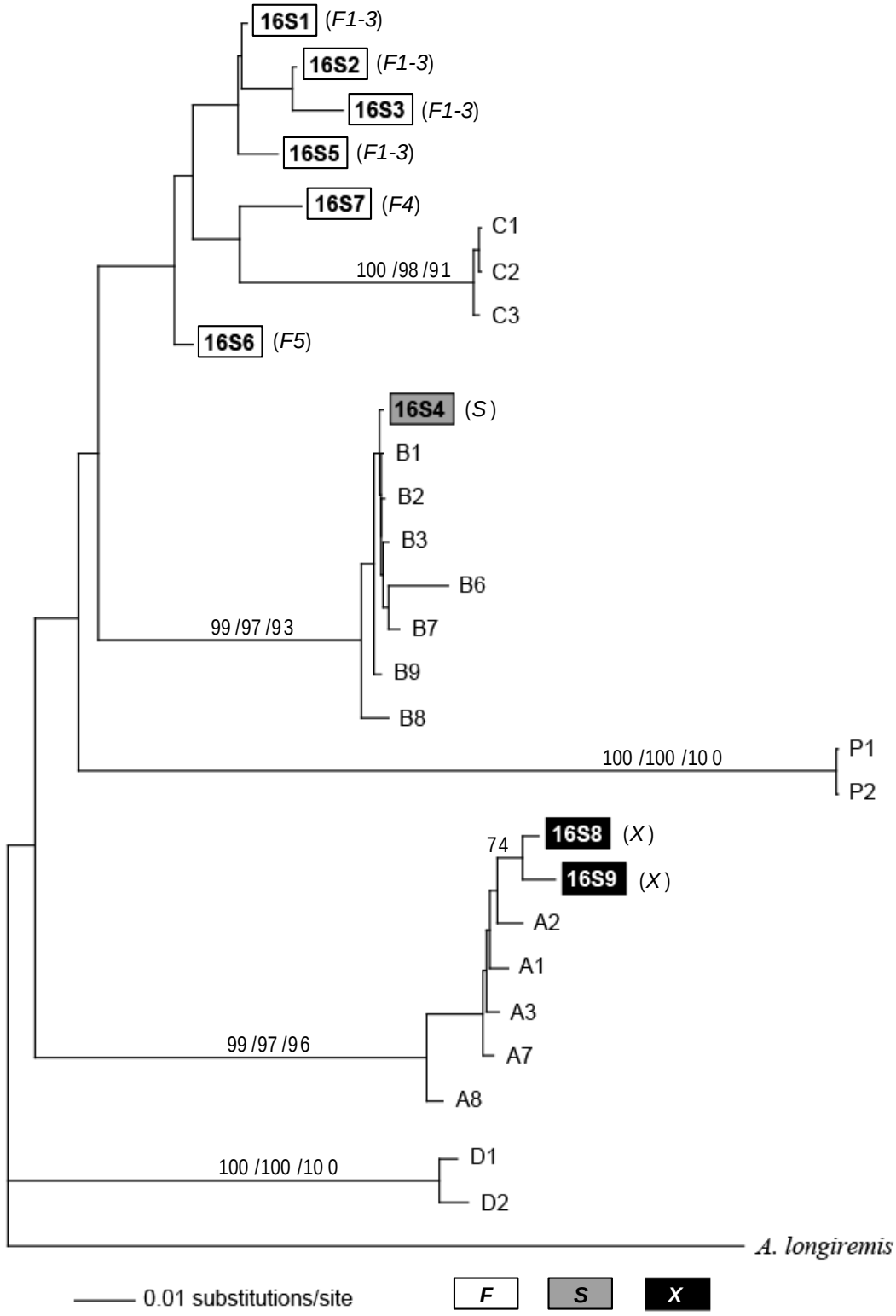

16S1 (F1-3)
16S2 (F1-3)
16S3 (F1-3)
16S5 (F1-3)
16S7 (F4)
C1
C2
C3
100 /98 /91
16S6 (F5)
16S4 (S)
B1
B2
B3
B6
B7
B9
B8
99 /97 /93
100 /100 /100
P1
P2
16S8 (X)
74
16S9 (X)
A2
A1
A3
A7
A8
99 /97 /96
100 /100 /100
D1
D2
A. longiremis
0.01 substitutions/site
F
S
X

Figure 3-3

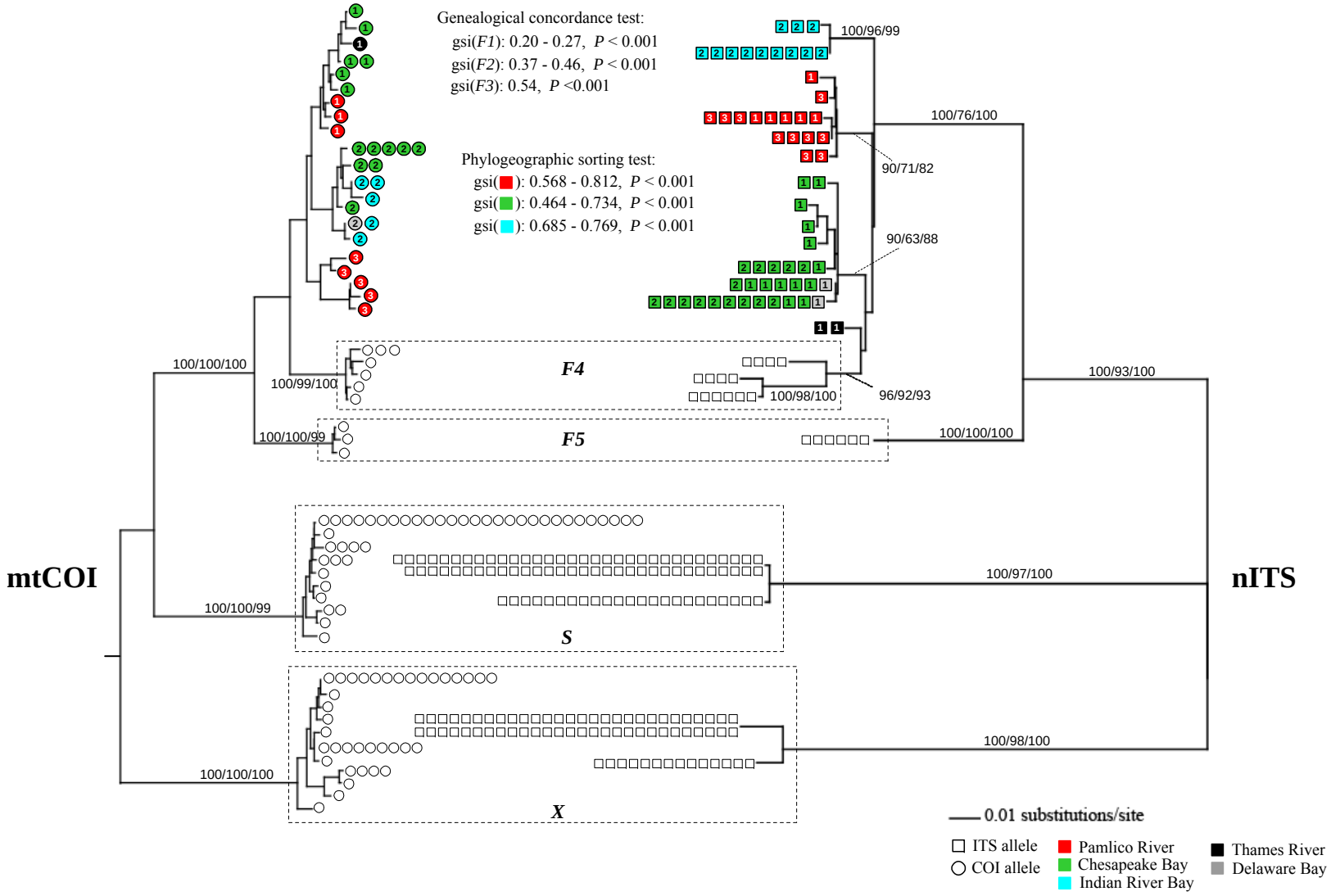

Genealogical concordance test:
gsi(F1): 0.20 - 0.27, P < 0.001
gsi(F2): 0.37 - 0.46, P < 0.001
gsi(F3): 0.54, P <0.001
Phylogeographic sorting test:
gsi(■): 0.568 - 0.812, P < 0.001
gsi(■): 0.464 - 0.734, P < 0.001
gsi(■): 0.685 - 0.769, P < 0.001
100/96/99
100/76/100
90/71/82
90/63/88
96/92/93
100/98/100
100/100/100
100/99/100
100/100/99
100/93/100
100/100/100
F4
F5
S
X
mtCOI
nITS
100/100/99
100/97/100
100/98/100
100/100/100
0.01 substitutions/site
ITS allele
COI allele
Pamlico River
Chesapeake Bay
Indian River Bay
Thames River
Delaware Bay

Figure 3-4

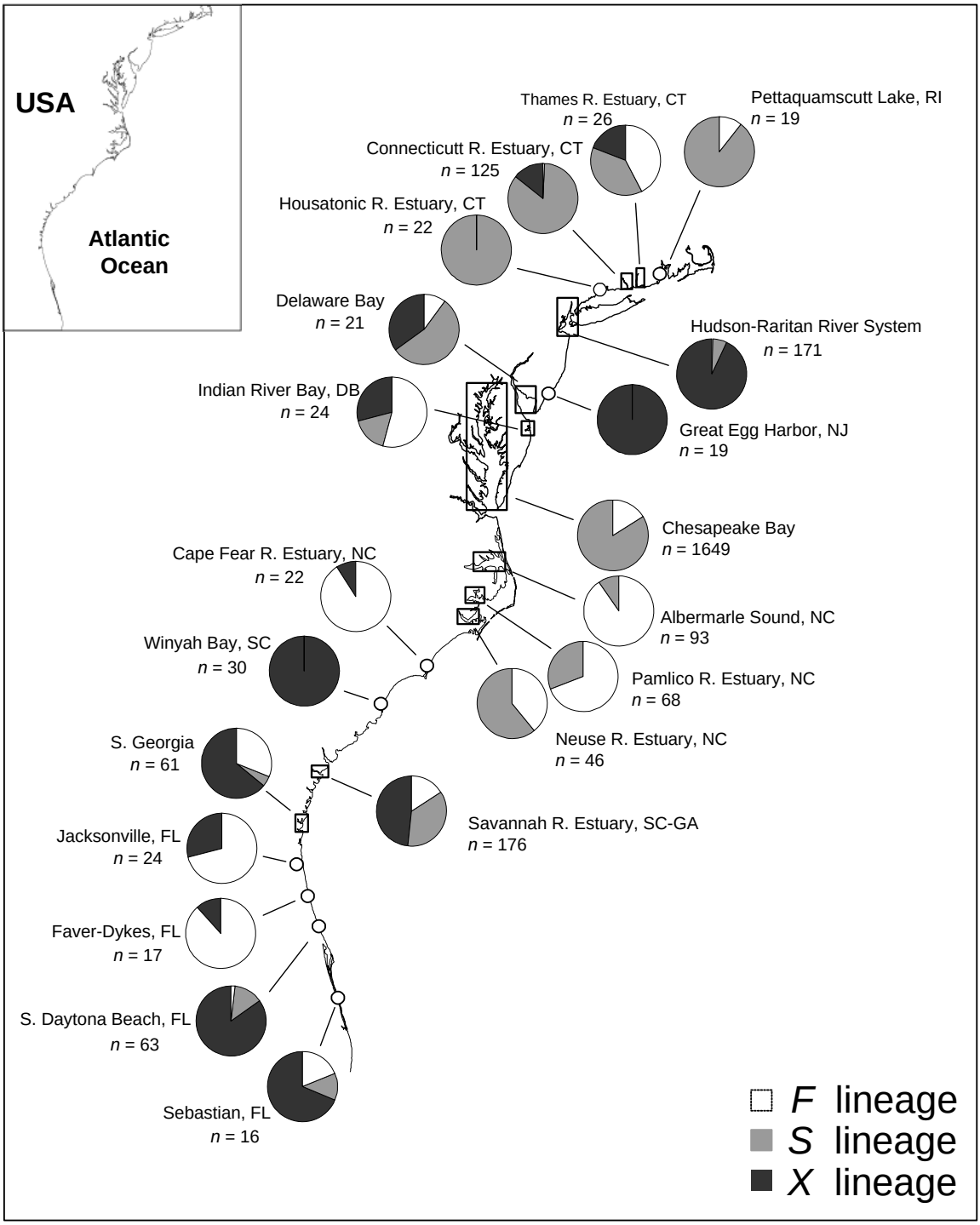

USA
Atlantic Ocean
Thames R. Estuary, CT
n = 26
Pettaquamscutt Lake, RI
n = 19
Connecticutt R. Estuary, CT
n = 125
Housatonic R. Estuary, CT
n = 22
Delaware Bay
n = 21
Hudson-Raritan River System
n = 171
Indian River Bay, DB
n = 24
Great Egg Harbor, NJ
n = 19
Chesapeake Bay
n = 1649
Cape Fear R. Estuary, NC
n = 22
Albermarle Sound, NC
n = 93
Winyah Bay, SC
n = 30
Pamlico R. Estuary, NC
n = 68
Neuse R. Estuary, NC
n = 46
S. Georgia
n = 61
Savannah R. Estuary, SC-GA
n = 176
Jacksonville, FL
n = 24
Faver-Dykes, FL
n = 17
S. Daytona Beach, FL
n = 63
Sebastian, FL
n = 16
F lineage
S lineage
X lineage

Figure 3-5

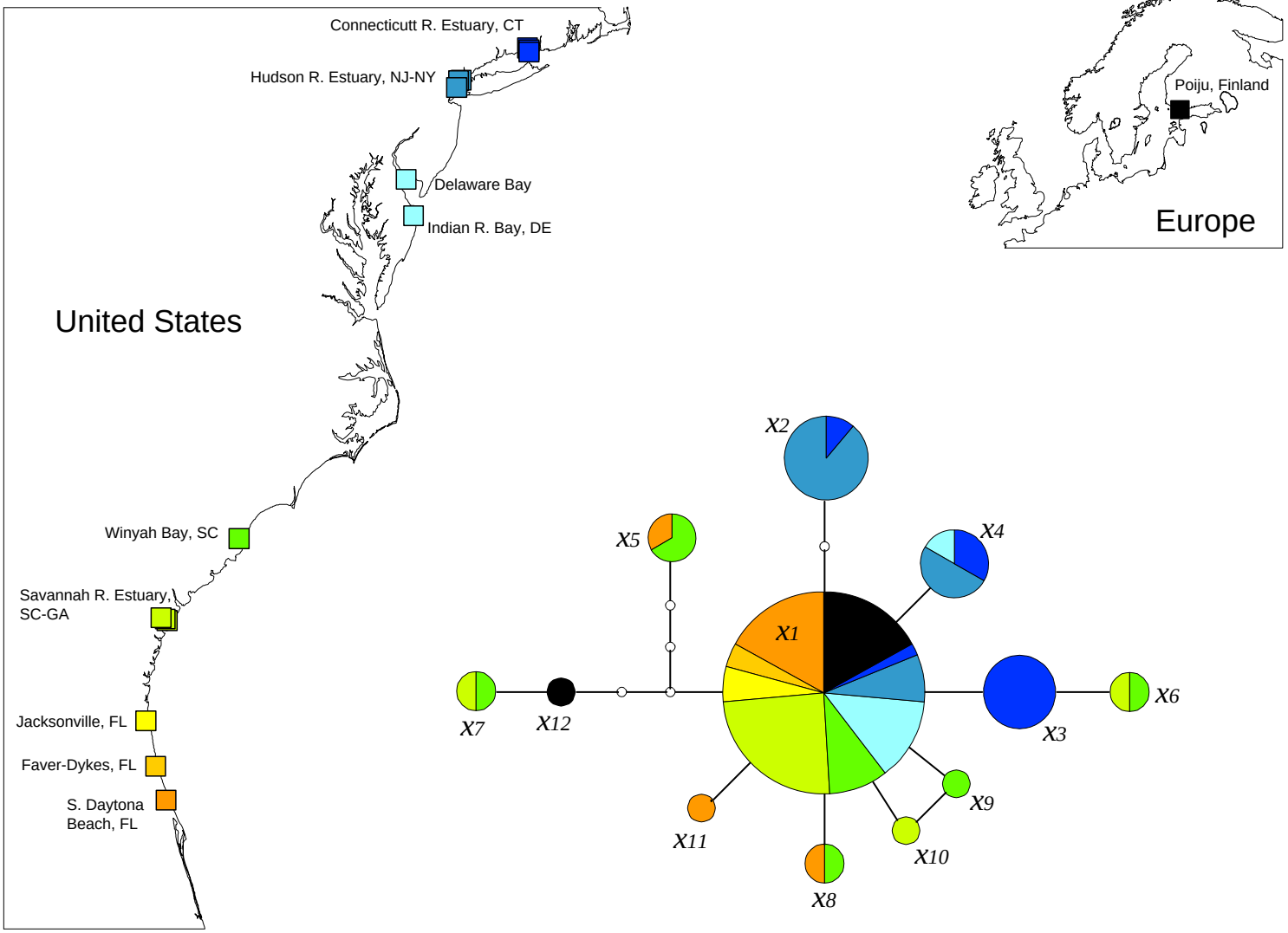

Connecticutt R. Estuary, CT
Hudson R. Estuary, NJ-NY
Delaware Bay
Indian R. Bay, DE
United States
Winyah Bay, SC
Savannah R. Estuary, SC-GA
Jacksonville, FL
Faver-Dykes, FL
S. Daytona Beach, FL
Poiju, Finland
Europe
X1
X2
X3
X4
X5
X6
X7
X8
X9
X10
X11
X12

Figure 3-6

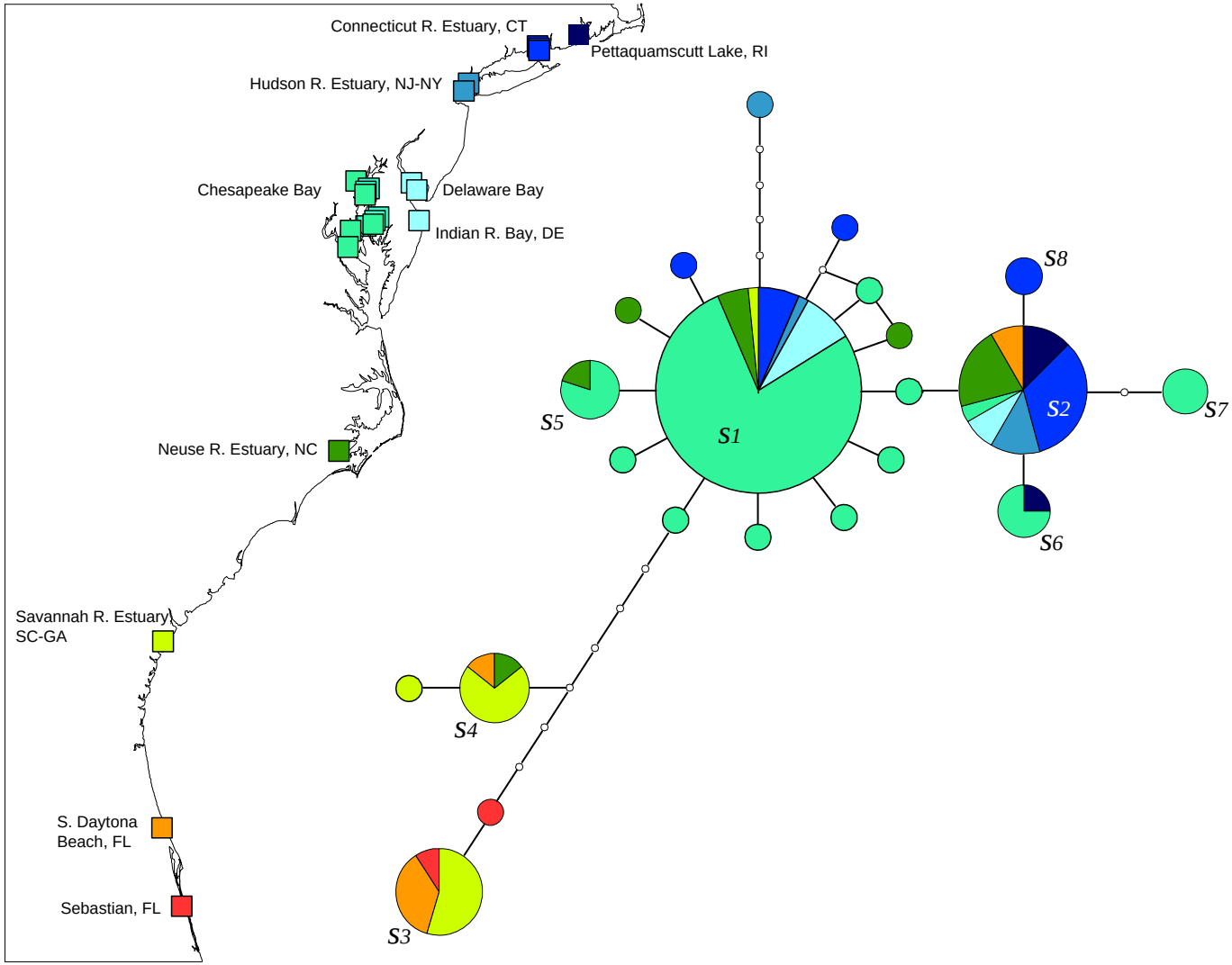

Connecticut R. Estuary, CT
Pettaquamscutt Lake, RI
Hudson R. Estuary, NJ-NY
Chesapeake Bay
Delaware Bay
Indian R. Bay, DE
Neuse R. Estuary, NC
Savannah R. Estuary
SC-GA
S. Daytona
Beach, FL
Sebastian, FL
S1
S2
S3
S4
S5
S6
S7
S8

Figure 3-7

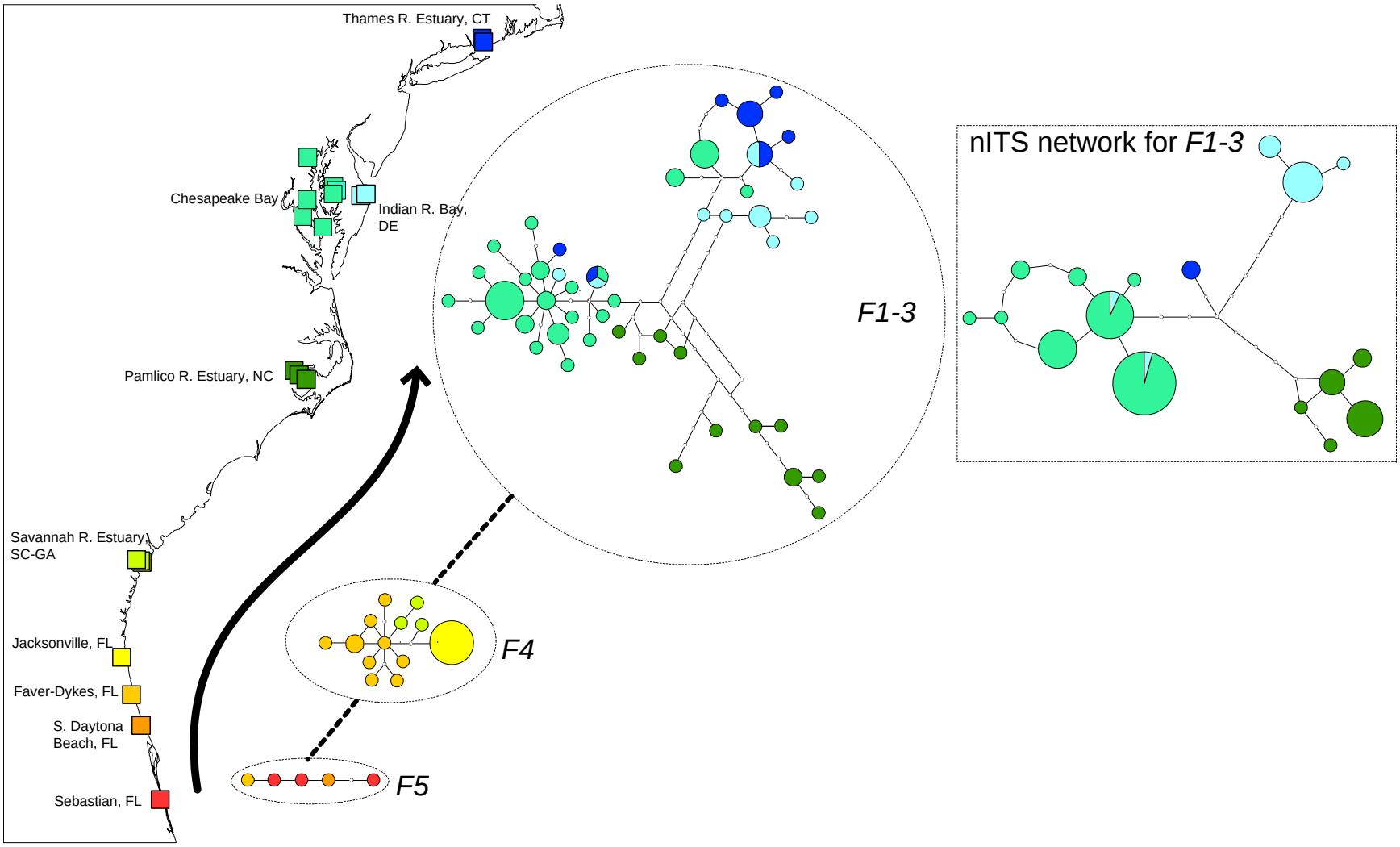

Thames R. Estuary, CT
Chesapeake Bay
Indian R. Bay, DE
Pamlico R. Estuary, NC
Savannah R. Estuary, SC-GA
Jacksonville, FL
Faver-Dykes, FL
S. Daytona Beach, FL
Sebastian, FL
F1-3
F4
F5
nITS network for F1-3

Figure 3-8

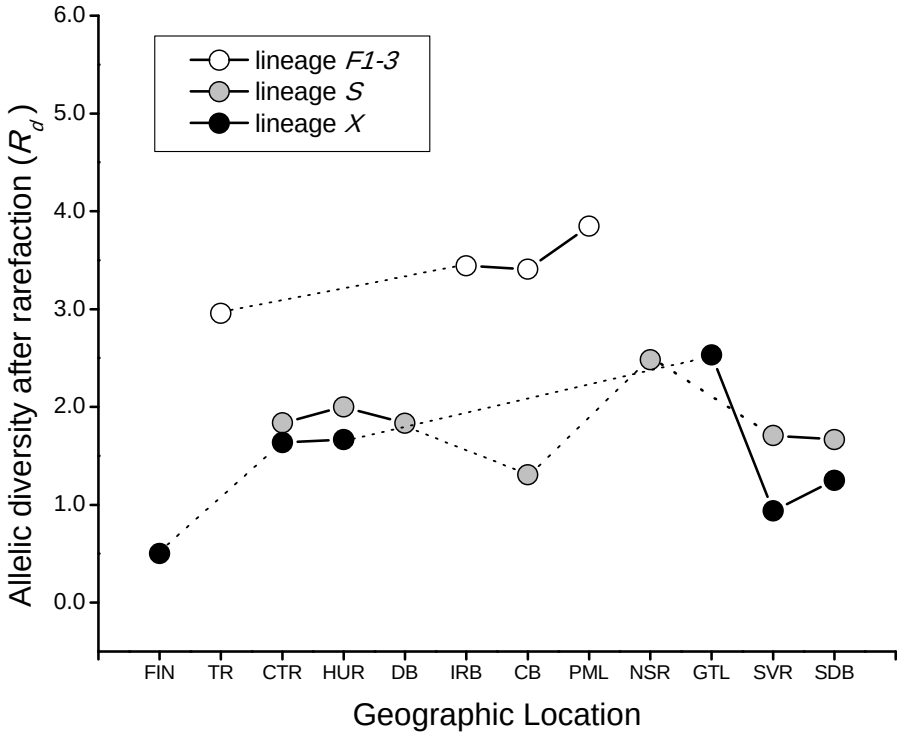
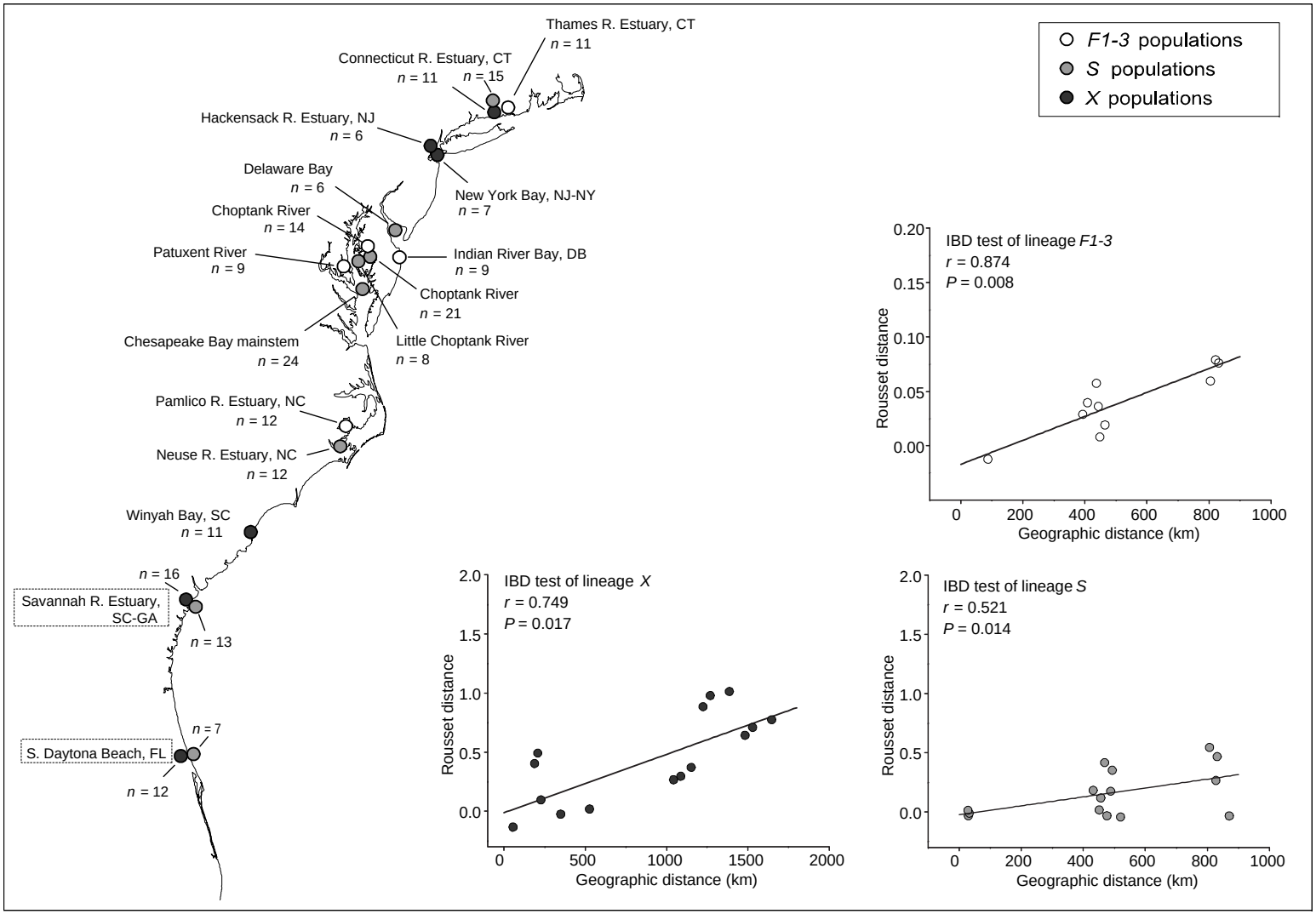

lineage F1-3
lineage S
lineage X
Allelic diversity after rarefaction (R_d)
6.0
5.0
4.0
3.0
2.0
1.0
0.0
FIN
TR
CTR
HUR
DB
IRB
CB
PML
NSR
GTL
SVR
SDB
Geographic Location

Figure 3-9


F1-3 populations
S populations
X populations
Thames R. Estuary, CT
n = 11
Connecticut R. Estuary, CT
n = 11
n = 15
Hackensack R. Estuary, NJ
n = 6
Delaware Bay
n = 6
New York Bay, NJ-NY
n = 7
Choptank River
n = 14
Patuxent River
n = 9
Indian River Bay, DB
n = 9
Choptank River
n = 21
Chesapeake Bay mainstem
n = 24
Little Choptank River
n = 8
Pamlico R. Estuary, NC
n = 12
Neuse R. Estuary, NC
n = 12
Winyah Bay, SC
n = 11
n = 16
Savannah R. Estuary, SC-GA
n = 13
n = 7
S. Daytona Beach, FL
n = 12
IBD test of lineage F1-3
r = 0.874
P = 0.008
Rousset distance
Geographic distance (km)
IBD test of lineage X
r = 0.749
P = 0.017
IBD test of lineage S
r = 0.521
P = 0.014

Chapter 4
Ecological differentiation of cryptic genetic lineages in *Acartia tonsa*

Abstract

Recognition of ecologically mediated divergence in diversifying populations and/or coexisting sibling species is an important step towards understanding the relative role of ecological differentiation in generation and maintenance of biological diversity. This is particularly important for broadly-dispersed marine holoplankton with mixed cryptic taxa. In this study, I investigated ecological differentiation of *Acartia tonsa*, a cosmopolitan estuarine copepod consisting of multiple co-existing deep genetic lineages (lineages separated by large genetic distances). Composition and distribution of three *Acartia* lineages (*F*, *S* and *X*) was examined within and across estuaries on the US eastern coast, testing the generality of salinity association patterns found in a previous study. The results showed that lineage *F* was strongly associated with low-salinities (0~5 PSU), while Lineage *S* and *X* were euryhaline for a wide range of salinities (5~30 PSU). Coexisting lineages were spatially separated with overlaps along salinity gradient, suggesting salinity associated niche partitioning in several estuaries. However, unusual regional distribution and inconsistent salinity delineation for niches suggest that factors other than salinity, including competition, may also play important roles in niche partitioning of *Acartia*. Salinity restriction might be relaxed as a consequence of adaptation in more diversified lineage *F*. The three cryptic genetic lineages also displayed significant difference in body size, with *S* larger than *F* and *X* individuals, providing evidence in favor of morphological divergence and against morphostasis of *Acartia*.

Introduction

In general, with spatially random dispersal, the spatial scale over which populations become adapted along an environmental gradient is positively correlated with the spatial scale of gene flow (Slatkin 1973; Endler 1977). Marine environments have few impenetrable physical barriers, but sharp physical and biological gradients are common, especially within estuaries (Bilton *et al.* 2002). In many cases marine taxa adapted to different portions of an environmental gradient have evolved some measure of reproductive isolation through differences in phenology (Ruzzante *et al.* 1999) or contrasting tolerances/preferences of depth (Kruse & Reise 2003), salinity (Bekkevold *et al.* 2005), temperature (Hilbish *et al.* 1994), or pollutants (Schizas *et al.* 2001; Rocha-Olivares *et al.* 2004). Surprising abilities of broadly-dispersing taxa to find and/or choose desirable habitat have been demonstrated (Bierne *et al.* 2003; Kinlan *et al.* 2005; Bradbury *et al.* 2008), but strong viability selection can also generate local 'adaptation' across a physical gradient after dispersal each generation (Williams 1975; Koehn *et al.* 1980; Bertness & Gaines 1993; Wilhelm & Hilbish 1998; Schmidt & Rand 2001). In addition to the response by each species to environmental gradients, species interactions will further shape environmental associations of taxa. Thus, there are numerous mechanisms that can lead to juxtaposed distributions of related taxa along an environmental gradient. Nonetheless, in general, the greater the dispersal potential of an organism, the more interesting a pattern of restricted distribution becomes because it suggests unappreciated philopatric capabilities or an example of particularly strong selection.

Holoplankton, a group of small marine organisms that live in the water column throughout their entire life cycle, are by definition drifters or weak swimmers generally moving less than 1 cm/s. Long-distance dispersal is expected to be common in these species based on the potential for advection in oceanographic currents and the broad geographical species ranges commonly described (van der Spoel & Pierrot-Bults 1979; van der Spoel & Heyman 1983). For example, the biogeography of many holoplankton species suggests allopatric divergence across continental dispersal barriers or between major oceanic gyres (van der Spoel & Pierrot-Bults 1979; Bucklin *et al.* 1996; Norris 1999; Beaugrand *et al.* 2002; Darling *et al.* 2007). However, with the benefit of molecular studies to identify closely related cryptic taxa, these have sometimes been found to co-occur geographically and be associated with distinct environmental features (e.g. de Vargas *et al.* 1999; Schizas *et al.* 2001; Ortells *et al.* 2003; Chen & Hare 2008). In one example, cryptic species each shared haplotypes across oceans but had patchy patterns of abundance within oceans, suggesting high dispersal but species-specific, condition-dependent reproduction (de Vargas *et al.* 1999). Careful analysis of similar cases is needed to understand how, and at what scale, niche partitioning has evolved in high-dispersal holoplankton species.

Acartia tonsa Dana, 1849 (Copepoda, Calanoida) is a well-studied planktonic copepod that seasonally dominates the zooplankton community in estuaries along the US east coast (Mauchline 1998; Johnson & Allen 2005). Recent studies have demonstrated that this cosmopolitan copepod is composed of multiple genetically distinct but morphologically cryptic species, which commonly coexist within estuaries (Caudill & Bucklin 2004; Chen & Hare 2008; Chapter 3 of this study). A previous study in

Chesapeake Bay, the largest estuary in the United States, revealed a strong association between two genetic lineages (*F* and *S*) and environmental salinities, with lineage *F* showing strong affinity to fresher water, while *S* lineage was meso-polyhaline (Chen & Hare 2008). The salinity association of *Acartia* lineages in Chesapeake Bay was temporally stable over years and was found across both horizontal and vertical gradients. These observations were interpreted as evidence of niche partitioning by *A. tonsa* populations; however, it was unclear whether salinity tolerance was a primary factor driving niche partitioning, or if community interactions or other factors might be important as well. For example, in Chesapeake Bay the association between these two lineages and salinity was tighter along relatively stable salinity gradients compared to times and locations with more hydrodynamic mixing (Chen & Hare 2008), possibly indicating limits to the behavioral ability to track preferred water masses.

Using an observational approach to explore the primacy of salinity gradients for structuring *Acartia* cryptic species, the most valuable extension of previous results will come from tests of association between *Acartia* lineages and environmental variables under a greater diversity of physical, geographic, and ecological contexts. Macrogeographic analysis of the *Acartia* species complex along the US eastern coast revealed a third major cryptic species, lineage *X*, in addition to the *F* and *S* lineages in Chesapeake Bay (Caudill & Bucklin 2004; Chapter 3 of this thesis). Furthermore, the *F* lineage was found to have geographically discrete monophyletic subclades only one of which was present in Chesapeake Bay. A major goal of this study was to measure environmental associations for lineage *X* and the non-Chesapeake *F* sublineages. All three major lineages were broadly distributed, providing the opportunity to compare

lineage associations with salinity in temperate and subtemperate latitudes and in two major zoogeographic provinces. Some estuaries had all three lineages co-occurring whereas in the majority of estuaries only two lineages were found. Limited sampling certainly contributed to this pattern, but it was largely the result of a large gap in the distribution of lineage *X* in Chesapeake Bay and Albemarle-Pamlico Sound, North Carolina, as well as regional prevalence of *S* lineage in the north and *X* lineage in the south. At any rate, the diversity of lineage compositions observed across 20 estuarine systems provided here for a rigorous continental-level test of the salinity associations previously measured for *F* and *S* lineages in Chesapeake Bay (Chen & Hare 2008).

As I mentioned above, observed environmental association may be a consequence of different stories. As a stressor for nearly all the estuarine organisms, salinity could act as a selective force on survival of *A. tonsa.* As a consequence, salinity-associated distribution of *Acartia* lineages reflects their salinity tolerance/preference and their fundamental niches largely defined by salinity. Alternatively, salinity-associated distribution may not be necessarily related to salinity. *A. tonsa*, a well-known estuarine zooplankton, may have evolved to tolerate a wide range of salinity. The observed niche partitioning may result from a response to other physical factors or biological competition for food resources that happens to vary along salinity gradient. As matter of fact, a number of physical factors often correlate with salinity to some extent (Day *et al.* 1989). Phytoplankton composition varying along estuarine gradients has been well documented, including distinct riverine/coastal assemblages (e.g. Ahel *et al.* 1996; Muylaert *et al.* 2000; Quinlan & Phlips 2007) and/or size-fractioned groups (e.g. Sin *et al.* 2000). Zooplankton specialized on different phytoplankton food may partition their food niches

following phytoplankton distributions along estuarine gradients. Hoffmeyer & Figueroa (1997) found two copepods, *Eurytemora affinis* and *A. tonsa*, had specialized oral integument structures associated with their different feeding habits, which well explained the coexistence of two species in Bahia Blanca estuary, Argentina.

The crypsis of *Acartia* lineages may link to morphological stasis, a phenomenon of stagnant morphological evolution relative to molecular evolution often found in insects and copepods (Knowlton 1993; Lee & Frost 2002; Rocha-Olivares *et al.* 2001). However, inadequate morphological scrutiny may be a simple reason causing our failure to distinguish them. In this study, morphometric variations of *Acartia* body size were investigated with an aim to discern possible morphological divergence paralleled to genetic divergence. Organismal body size (specifically body length, volume, weight or mass) is one of the most easily measured morphological traits, yet has complex evolutionary trajectories affected by nearly all the biological processes, including taxonomic affiliation, life history, physiology and ecology (Blackburn & Gaston 1994; Roy 2008). In marine planktonic copepods, temporal and spatial variations in body size among populations have been widely reported (Mauchline 1998; Hirst *et al.* 1999; Kobari *et al.* 2003; Gaudy & Verriopoulos 2004). These variations have been related to a variety of ecological factors, such as temperature, salinity, oxygen concentration, food level, predation, etc., suggesting that ecological factors may play an important role in regulating copepod body size. Previous studies on *A. tonsa* have also shown substantial variations in its body size. Garmew *et al*. (1994) reported different prosome body length of *A. tonsa* from geographic populations of Chesapeake Bay, Montauk Bay, USA and coast of Peru, however, no environmental association were studied. Gaudy & Verriopoulos (2004)

reported huge spatial and seasonal variations in Berre Lagoon, France, and attributed them to multiple environmental factors (temperature, salinity, chlorophyll and particulate seston). Since cryptic genetic lineages have been found in *A. tonsa* (Caudill & Bucklin 2004; Chen & Hare 2008), it is essential to test if observed body size variations are accounted by lineage discrepancy. Significant lineage-specific variations of morphometric measures would reject morphostasis of *Acartia*, providing a new window to look into its population differentiation.

Overall, two major goals of this study are 1) test the persistence of salinity associations with *Acartia* lineages across diverse contexts; 2) test the morphometrical variations among *Acartia* lineages. I hypothesize that salinity association of *Acartia* lineage is common across estuaries and body size variations exist among three *Acartia* lineages.

Materials and Methods

Study sites, collection of samples and environmental data

A total of 85 geographic samples were examined in this study from estuaries distributed along ca. 2000 km of US east coast between Rhode Island and Florida (Table 4-1). Nearly half of them were from the largest estuary in the United States, Chesapeake Bay, including the main stem of the bay and its tributaries over an 11-year time span from 1995–2005 (for detailed sampling information see Chen & Hare 2008). Most other samples were collected during August–September 2005 with a plankton net from shore. I sampled from research vessels in Delaware Bay (hosted by Microbial Observatory of

Virioplankton Ecology, K. Eric Wommack, University of Delaware) and Savannah River (hosted by Elizabeth Mann, Skidaway Institute of Oceanography). In general, zooplankton samples were towed horizontally or vertically using standard plankton nets (mesh size of 125–250 μm). Sampling depths varied but most shore-based samples were from surface waters (1–3 m). Zooplankton samples were immediately preserved in 95% ethanol.

For shipboard sampling, real time temperature (in Celsius degree) and salinity (in PSU, practical salinity unit) associated with samples were recorded using CTD instruments and averaged across sampling depths. For shore-based samples, single measures of salinity and temperature were taken at a depth of 1 m with an YSI EC300 Conductivity/Temperature instrument (YSI Environmental).

The physical characteristics of 12 estuarine systems involved in this study were summarized using data from the literature and from multiple environmental monitoring programs. Thirteen parameters were selected to compare basic physical and ecological features of the estuaries that may affect zooplankton distribution, including vertical stratification, estuary size (length, width, depth, volume, and area), areal size of salinity zone (tidal fresh, mixing, seawater zone), freshwater input rate, tidal range, flushing time, temperature (range and duration for *Acartia* development), and copepod food abundance. Among them, temperature duration (D_T) is a parameter designed to measure the environmental quality of an estuary for *Acartia* development. McAlice (1981) reviewed temperature required for *Acartia tonsa* to hatch from egg, grow, develop and reproduce, proposing that a minimum of 20°C is required for *Acartia* to develop large populations in temperate estuaries. In this study, D_T refers to the number of days per year in an estuary

with temperature > 20°C, averaged using available data of surface water temperature from marine monitoring stations (See Appendix B for details).

Food level for *Acartia* was estimated as carbon unit (mgC/m^3), converted from chlorophyll *a* concentration in Bricker et al. (2007) using a moderate factor of 50 gC/gChl.a (Geider 1987; Veldhuis & Kraay 2004; Llewellyn *et al.* 2005). For gross comparisons at a macrogeographic scale, three categorical food levels were determined: 0~250 mgC/m^3 (low), 250~1000 mgC/m^3 (moderate) and > 1000 mgC/m^3 (high). Paffenhofer & Stearns (1988) found from feeding experiments that food capturing efficiency of *Acartia tonsa* decreases at food levels below 22 mgC/m^3. Detailed specifications and estimation methods of other environmental parameters are described in Appendix B.

Copepod classification and morphological measurement

In the laboratory, *Acartia tonsa* individuals (adult and later larval stages CIV, CV) were identified based on their diagnostic morphological characters (Bradford-Grieve 1999; Sabatini 1990). Briefly, key characters include the presence of rostral filaments and postdorsal spinules on the first two urosomites in both sexes. In females, the fifth leg (P5) has a bulbous terminal spine with a denticulate tip and an equally long plumose seta on its internally projected segment B2. For male, the right fifth leg (P5) has a conspicuous inner projection on segment Re2 and a large proximal rounded expansion on segment B2, while the left fifth leg has spines on segment Re2+3 and a rounded inner expansion on segment B2. Before proceeding to DNA extraction, four morphological traits that characterize copepod body size and shape were measured to the nearest 0.1 μm for each

individual under a microscope: prosome length (PL), urosome length (UL), full body length (BL = PL + UL), and the ratio of prosome over urosome length (P/U) (Fig. 4-1). Only adult body size data were used in this study.

DNA extraction, amplification, and RFLP genotyping

Genomic DNA was extracted using 5% chelex solution (Bio-Rad Laboratories) for identified *A. tonsa* individuals. Intact copepod bodies were first rehydrated in de-ionized water for more than 2 h and then heated at 99°C in 5% chelex solution for 8 min. After centrifugation at 2288 *g* for 5 min, supernatant DNA solutions were collected and stored at 4°C for subsequent genetic analysis.

Mitochondrial cytochrome *c* oxidase subunit I (mtCOI) gene was PCR-amplified using Folmer's primers 1490 and 2198 (Folmer *et al*. 1994). PCR reactions were performed in a total volume of 25 uL with 1X Invitrogen buffer, 2.5mM $MgCl_2$, 125 μM dNTPs, 0.2 μM of each primer, 0.2 ug/uL BSA, 0.3 U of Invitrogen *Taq* polymerase, and 0.5 – 1 uL of template DNA. PCR thermocycles started with DNA denaturation at 95°C for 1 min, followed by 35 cycles of 30 sec denaturation at 95°C, 30 sec annealing at 50°C, and continued with 50 sec extension at 72°C, and then a final extension at 72°C for 7 min. Amplified PCR products were visually confirmed in 1.2% agarose gels with ethidium bromide staining.

Distinct restriction fragment length polymorphisms (RFLPs) were generated with digestion of endonuclease *Ase* I for deeply diverged mtCOI lineages based on 298 mtCOI sequences (Chapter 3 of this thesis). The RFLP patterns were used to classify *A. tonsa* individuals into three genetic lineages (*F, X, S*). PCR-amplified mtCOI products (5 – 10

μL each) were digested using 5 U of *Ase* I (New England Biolabs) in 15 μL reaction volumes with 1x manufacturer buffer at 37°C for 3 h. The digested DNA fragments were visualized in 2.5% agarose gels with ethidium bromide staining. Three lineages differ in both fragment length and numbers as described in Chapter 3 of this study (Fig. 4-2).

Statistical analysis

Percentage compositions of *Acartia* lineage *F*, *S*, and *X* (expressed as %*F*, %*S*, and %*X* respectively) were calculated for each sample. The full data set of these proportions was correlated with environmental factors including temperature, salinity, and geographic location. Pearson's correlation coefficient was calculated between lineage compositions and each environment factor. For lineages showing significant correlations, logistic regression models were fit to describe the quantitative relationships between lineage compositions and salinities.

Lineage proportions were compared along salinity gradients within a single estuary (Savannah River Estuary) and across estuaries (the U.S. eastern coast as a whole). Nonparametric permutation tests modified from Perry & Smith (1994) and Syrjala (1996) were applied to test the null hypothesis that all lineages were evenly distributed along the salinity gradient. Briefly, empirical cumulative distribution function (CDF) of lineage composition for the *i*th lineage at the *k*th salinity level s_k (in ascending order) were defined as

$$\Gamma_i(s_k) = \sum_{\forall s \le s_k} \frac{d_i(s)}{D_i} \quad (1)$$

where d_i denotes percentage composition of the ith lineage and D_i the summed d_i used for normalization of the distribution. A Cramér-von Mises type statistic (ψ) designed by Syrjala (1996) was used to test the difference between CDFs of two lineages, as summed over all the salinity levels (k = 1, 2, 3,…, K) :

$$\psi = \sum_{k=1}^{K}[\Gamma_1(s_k) - \Gamma_2(s_k)]^2 \quad (2)$$

Significant ψ indicates that the observed difference between two lineage distributions along the salinity gradient was not by chance ($\alpha = 0.05$). Statistical significance of observed ψ was determined based on an empirical probability distribution of ψ created by permuting observed percentage compositions between lineage pairs and across salinity levels. Exhaustive permutation with all possible pseudoreplicates was done for testing salinity levels ≤ 6 (total permutation number = $K!{\cdot}2^K$; 40680 for K = 6), whereas 10,000 randomly selected pseudoreplicates for salinity levels > 6. Nonparametric permutation tests were programmed and performed using computer statistical package R version 2.8.0 (R Development Core Team 2008). All R scripts were available on request.

For pooled systems, salinity-associated distribution patterns, 6 bin levels of salinity (0~5, 5~10, 10~15, 15~20, 20~25, and 25~30 PSU) were adopted with pooled point data falling into the respective bins, looking for optimal ranges for each lineage. Mean percentage compositions of each bin were used as d_i in Eq.(1) to calculate Γ distributions and ψ statistics. For each permutation replicate, randomization of lineage composition across lineage pairs and salinity level was conducted on the point data level. Mean values of 6 salinity bins were then recalculated for each replicate and used to calculate empirical distribution of ψ statistics.

Morphometric variations were analyzed using all the data from adult individuals with known genetic identity (mtCOI lineage *F*, *S*, *X*). Correlations between four measures (prosome length, urosome length, total body length, and prosome-urosome ratio) were assessed using Pearson's coefficients. Variations in body measures among genetic lineages, between sexes and seasons were tested using Kruskal-Wallis test and Wilcoxon signed-rank test (nonparametric counterpart of ANOVA and *t*-test respectively).

To test spatial variations of body size and possible geographic partitions, a nested ANOVA taking in account of unbalanced sample sizes (Kuehl 2000) was written in R script and performed using computer program R version 2.8.0 (R Development Core Team 2008). For females of three lineages sampled during August-September 2005, individuals were first pooled into arbitrarily defined estuaries, which were next grouped into a higher level of regions. Partitions of body size variation within estuaries, among estuaries and among regions were computed using nested ANOVA. Statistical significance was determined analytically based on probability distribution of *F* statistics. Regional grouping was guided by sample data distribution and *a priori* knowledge of geographic system at different scale. Body size variation over latitudinal gradient was tested for each lineage using Pearson's correlation coefficient for geographic samples and their corresponding latitudes.

Results

Lineage sympatry and regional distributions

My 85 collections were from 20 different estuarine systems. All three lineages co-occurred within 7 estuaries and only two lineages were found in the rest. Two broad

patterns contributed to the distinctiveness of three macrogeographic regions. First, lineage *X* was completely absent from Chesapeake and Albemarle-Pamlico estuarine systems despite extensive sampling. Curiously, the other mid-Atlantic estuary sampled (much more sparsely), Delaware Bay and Indian River Bay near the mouth of the Delaware Bay, yielded all three lineages. Second, the regional proportional dominance of one or two lineages over others shifted across the latitudinal range that we sampled despite the fact that collections from each region were drawn from a similar distribution of salinities. Lineage *S* and *X* dominated in the three northern estuarine systems, the Connecticut River Estuary, Thames River Estuary and Hudson River System (Fig. 4-3, 4-4). In the south-Atlantic Bight, the *S* lineage typically had a minority presence in estuaries relative to *X* and *F*.

Association between lineages and salinity

No statistical correlations were found within or among estuaries between *Acartia* lineage proportions and latitude or environmental factors other than salinity (data not shown). *Acartia* lineage *F* showed a strong association with environmental salinity, particularly in Chesapeake-Albemarle-Pamlico System (CAPS) where only lineage *F* and *S* were found. Lineage composition in a sample (%*F*) was negatively correlated with salinity (Pearson's r = -0.614, $P < 0.001$). The quantitative relationship between %*F* and salinity (*Sal*) can be best fit by a logistic regression model (Fig. 4-5A):

$$\%F = \frac{100}{1+e^{-1.4825+0.3804\,Sal}}$$ (Likelihood ratio test: $P < 0.001$).

When all the data were considered, association between lineage *F* and low salinities was still significant (Pearson's r = -0.467, $P < 0.001$). The logistic model for the whole data set can be expressed as follows:

$$\%F = \frac{100}{1+e^{-0.3595+0.1669\,Sal}}$$ (Likelihood ratio test: $P < 0.001$).

One sample (MNH) located between Albemarle Sound and Pamlico Sound had a relatively high proportion of *F* individuals in high-salinity water (22.0 PSU, Fig. 4-6). This sample was excluded from the logistic regression and association tests because its source waters are known to experience sharp mixing between low-salinity Albemarle water (2 – 5 PSU) and high-salinity Pamlico water (> 25 PSU) (Lin *et al.* 2007).

Lineage *S* showed the opposite composition to *F* in the CAPS system, and statistical significance of the correlation and logistic model were the same as for *F*, with the opposite sign for the correlation coefficient (Fig. 4-7B). However, association between lineage *S* and high salinities was not significant for all the data (Pearson's r = 0.048, P = 0.664; Fig. 4-7B).

Salinity association of lineage *X* was tested with a full data set including 50 nulls (0%) of CAPS samples due to the complete absence of *X* individual and a subset excluding all null data from CAPS. Neither case showed significant association between its percentage composition (%*X*) and salinities (Pearson's r = 0.322, P = 0.214 for all samples; r = 0.246, P = 0.155 for non-CAPS samples).

Habitat separation of coexisting *Acartia* lineages along estuarine salinity gradients was seen both within single estuaries and in general pooling patterns across estuaries. Strikingly in Savannah River estuary, three coexisting lineages (*F*, *S*, and *X*) also showed

significant separation along salinity gradients. Lineage *F*, *S* and *X* appeared to dominate in low- (4.9~10 PSU), medium- (10~25 PSU) and high-salinity zone (> 25 PSU) respectively (Fig. 4-8). Differences in these salinity-associated distributions in Savannah estuary were statistically significant between *F* and *S* (nonparametric permutation test: $P = 0.04$), while marginally between *F* and *X* ($P = 0.08$), and between *X* and *S* ($P = 0.12$).

When samples across estuaries were pooled together, lineage distribution patterns were different among three geographic regions: North of Chesapeake Bay, Chesapeake-Albemarle-Pamlico Systems (CAPS), and South of Pamlico Sound. In Chesapeake-Albemarle-Pamlico systems where two lineages coexisted, lineage *F* and *S* were clearly separated with *F* relatively dominating low-salinities (0~5 PSU) and *S* medium-high salinities (> 5 PSU), though the two overlapped between 0~15 PSU (Fig. 4-5B). This pattern was statistically significant by a nonparametric permutation test ($P < 0.001$). In southern estuaries where three lineages involved, the general pattern was similar to that in Savannah River Estuary: lineage *F, X, S* occupied 0~15, 5~30, 20~30 PSU salinity zone respectively (Fig. 4-5C). The separations of *F* and *S*, *F* and *X* along salinity gradient were statistically significant (nonparametric permutation test, $P < 0.01$). This pattern was not solely a result of the distribution in Savannah River; all three lineages were found in Florida shallow lagoons, with *F* relatively abundant in low-salinity waters (10.1 – 10.9 PSU) and the opposite for *X* (24.1 – 27.5 PSU) (Fig. 4-9). Lineage *S* was relative scarce, as it is along the entire South Atlantic Bight.

Patterns in northern estuaries were unusual for the presence of lineage *F* in 10~30 PSU zone and widespread of *X* and *S*. Two dominating lineages appeared to be antagonistic to each other, where *X* aggregated more in the middle of salinity range

"driving" *S* to both ends (Fig. 4-5A). No statistical significance was found for this slight separation between *X* and *S*.

Morphometrical variability among A. tonsa *lineages*

A high level of morphometrical variability was found among 1633 *A. tonsa* individuals for all the examined measures, including prosome length (PL: 546.1~1287.0 μm), urosome length (UL: 111.6~306.0 μm), total body length (BL: 675.8~1573.0 μm), and prosome-urosome ratio (P/U: 3.2~6.6). The three body length measures were strongly correlated to each other (Pearson's $r = 0.80$~0.99, $P < 0.001$), while weakly correlated to P/U ratio (Pearson's $r = -0.52$~0.09, $P < 0.001$).

Variations in body size were most conspicuously accounted by discrepancies between two sexes, among the three major genetic lineages and seasons (Table 4-2; Fig. 4-10). Overall, the females had significantly greater mean body length (PL, UL, and BL) and prosome-urosome ratio (P/U) than those of males within each lineage, except no significant difference in UL for all lineages in summer and lineages *S*, *X* in fall (Table 4-2). Within sexes and seasons, among-lineage variations were strongly significant for body length measures (BL, PL, UL; Table 4-2). Lineage *S* showed the largest mean body lengths and prosome-urosome ratio (P/U). One only exception was the UL in spring showing relative smaller than those of lineage *F*. No consistent rank of size was found for lineage *F* and *X*. Lineage *F* had greater TBL and PL than *X* in spring, while the trend reversed in summer and fall. Variations among lineages, sexes and seasons were particularly strong in Choptank River, Chesapeake Bay (Fig. 4-11). Two lineages (*F* females, *S* females and males) in the Choptank River showed significantly larger body

sizes in spring than in summer 2004 (Wilcoxon test, $P < 0.001$ for all three comparisons between spring and summer). In both seasons, *S* females were significant larger than males (Wilcoxon test, $P < 0.001$ for all tests). In spring, *S* females were significant larger than *F* females at all geographic locations (paired wilcoxon test, $P < 0.05$), whereas data were not sufficient to test difference in summer.

Spatial variations of body size among geographical samples were also substantial when lineage, sex and season were controlled. In the Choptank River, both *F* and *S* lineages showed a body size zonation along the river with downstream individuals significantly larger than upstream ones (Fig. 4-11; Pooled upstream samples (0 – 30 km) vs. downstream samples (30–70 km), Wilcoxon test, $P < 0.001$ for both female lineages). At a large geographic scale, geographic differences among pooled estuary samples were strongly significant (Krustal-Wallis test, $P < 0.001$ for all lineages). However, no significant hierarchical geographic structures were detected when using nested ANOVA to test possible regional difference (Fig. 4-12). Neither latitudinal gradients were detected for each lineage (Fig. 4-12; Pearson's $r = 0.06–0.23$, $P = 0.554–0.876$ for all three lineages).

Discussion

Our main goal in this study was to test the generality of salinity associations originally found for the *Acartia* lineages *F* and *S* in Chesapeake Bay (Chen & Hare 2008), and test environmental associations of a third lineage, *X*, that is broadly distributed outside of Chesapeake Bay (Chapter 3 of this thesis). Association tests previously reported for Chesapeake Bay *Acartia* involved a geographically restricted monophyletic

subclade of the *F* lineage (Chapter 3 of this thesis). By broadening my scope to the continental scale I was able to test whether the oligohaline restriction found for an *F* sublineage in Chesapeake Bay also occurred with the same sublineage in other mid-Atlantic estuaries as well as in other allopatric and monophyletic *F* sublineages. If restriction to oligohaline waters is a basal characteristic found in all the *F* lineage taxa then it strongly suggests that a narrowing of salinity tolerance evolved once and has persisted during the south to north radiation of *F* lineage subclades (Chapter 3 of this thesis). For the broadly distributed *S* lineage, if the salinity association found in Chesapeake Bay was unique to that context, it might suggest that other factors specific to local environment (e.g., food composition and distribution, migration behavior, estuarine hydrodynamics, competition) may be more important for community structuring than salinity.

Salinity associations of Acartia lineages

Lineage *F* showed a uniformly strong response to environmental salinity gradient in two out of three regions examined. In Chesapeake Bay, lineage *F* originally was named a freshwater lineage because of its restriction to mostly oligohaline waters (0~5 PSU) and it proportionately declined in abundance relative to lineage *S* as salinity increased (Chen & Hare 2008). I found this pattern generally repeated in three southeastern estuaries: Albemarle-Pamlico Estuary, Neuse Estuary and Savannah Estuary, where deeply diverged subclades of the *F* lineage occur.

In contrast to these patterns in the mid-Atlantic and south-Atlantic Bight, however, lineage *F* individuals were sporadically found in high-salinity areas (17.0~26.1

PSU) in Florida and in three northern estuaries: Thames River Estuary, Connecticut River Estuary, and Hudson-Raritan Estuary (Fig. 4-3, 4-4). Several factors may have contributed to these contrasts. First, oligohaline waters were poorly sampled in Florida and northern estuaries. Thus, the generally low abundance of *F* lineage in these estuaries and its occurrence at higher salinities could be an artifact caused by sampling during an environmentally marginal period. I also may have missed the local sources of low-salinity *F*, because a vast majority of low-salinity upstream areas of Hudson River Estuary, Connecticut River, Thames River and St. Johns River were not sampled (Fig. 4-3, 4-4, 4-9). Therefore, more extensive sampling must be conducted to adequately measure the salinity associations of lineage F in these regions.

A second possibility is that the less restrictive salinity association patterns of lineage *F* in the north and in Florida provide an accurate reflection of broader salinity tolerances. Compared with *S* and *X*, lineage *F* is a highly diversified lineage consisting of multiple allopatric sublineages (Chapter 3 of this thesis). New England and Florida were both at the geographic limits of our sampling where we found genetically distinct *F* sublineages predominating. The Florida sublineage *F5* was phylogenetically most basal and appeared to be relatively old, whereas north of Cape Hatteras *F* sub-lineages were relatively recently evolved (Chapter 3 of this thesis). It is reasonable to expect that life history features that vary latitudinally in *Acartia* could have led to distinct patterns of adaptation in New England populations of lineage *F* relative to its progenitor sublineages to the south. For example, the seasonal demography of *A. tonsa* populations varies across latitudes. Generally, *A. tonsa* populations were reported to develop in late spring and thrive through summer. During winter at temperate latitudes they enter into dormancy in

the form of diapause eggs that sink into the benthic sediment. These diapause eggs get resuspended and build new populations the next spring (Jeffries 1962; Sullivan & McManus 1986; Sullivan *et al.* 2007). In this life cycle, diapause eggs are believed to play an important role in retaining genetic diversity through temporal gene flow, providing ample chances for populations to survive various extreme living conditions (Grice & Marcus 1981; Marcus 1996). The southern limit of a diapause life history for *Acartia* in the western North Atlantic is not known. However, there may be an interaction between the extent of seasonal diapause, the seasonal timing of precipitation expanding oligohaline habitat and the adaptive benefits of a relatively narrow salinity tolerance envelope.

Latitudinal gradients are less likely to be important for explaining the weak oligohaline association of lineage *F* in Florida relative to the South Atlantic Bight. My sampling in Florida was sparse with respect to environmental gradients and geography, and 2005 was a relatively dry spring potentially limiting preferred conditions for the genetically distinct *F* sublineage that occurs there. I will simply note that in eastern Florida, except for the St. Johns River, estuarine habitat primarily consists of shallow lagoons with small watersheds and dispersed inlets allowing tidal exchange with continental shelf waters. This type of estuary, while present along 36% of the continental coastline from Cape Cod to Miami (Chen, unpublished data) and therefore quite common geographically, is not the spatially dominant configuration of estuarine waters. The only other lagoon sampled in this study was in Delaware and *F*-lineage *Acartia* were abundant (unfortunately salinity was not determined).

An interesting question raised by the distribution of lineage *F* across estuaries is how it successfully colonized and prevailed in the Chesapeake-Albemarle-Pamlico System (CAPS) but less in others. We compared physical environment and ecological features of nine estuaries on the eastern coast to determine whether there are any gross correlates with the distinct regional patterns of lineage occurrence, abundance, and salinity association (Table 4-3). Apparently two important features provided ample chances for lineage *F* to sustain in their preferred low-salinity waters; temporally and spatially stable oligohaline water of sufficient size, and a long duration of favorable temperature for growth and development. Chesapeake Bay and Albemarle-Pamlico Sound are the top two largest estuarine systems of the United States, with estuarine surface areas greater than 5,000km^2 (Schubel & Pritchard 1987; Paerl *et al.* 2001; Table 4-3). The oligohaline zone of these systems is proportionally large and often stabilized by their long flushing time (1~3 months, Table 4-3). In both systems, *A. tonsa* populations were supported by abundant food resources and longer period of favorable temperature (> 20°C, 125~191 days/yr) for *Acartia* (Schubel & Pritchard 1987; Sellner 1987; Lin *et al.* 2007). Lack of these two features also partially explains the colonization failure of lineage *F* in other estuaries. For example, absence of lineage *F* in northern estuaries may be related to short temperature duration for sustained population growth (73~93 days). Both the Connecticut River Estuary and Thames River Estuary are small salt-wedge estuaries characterized by strong seasonal stratifications (Garvine 1975). During seasons with high river discharge, a thin layer of freshwater often rides on top of intruded salient water all over the estuaries. Such unstable environment may have prevented *Acartia* lineage *F* from colonizing in northern estuaries.

A comparison between the Savannah River Estuary and Connecticut River Estuary is illuminating because of their similarities. These two riverine estuaries share many physical features: similar size (0.01–0.09 km^3 volume, 36–42 km^2 surface area), shape (50–100 km long, 0.4–0.6 km wide), salinity structure (21–28 km^2 mixing area), and freshwater input rate (226–322 m^3/s). However, the abundance and distribution of lineage *F* was quite different in these two systems, possibly due to their contrasting features of mixing (mainly forced by tides, salt-wedge *vs*. partially mixing), flushing time (1 *vs*. 5.6 days), and temperature duration (85 *vs*. 209 days). Lineage *F* occurrence in Savannah River appears to take advantages of its relative high mixing, longer flushing time and long temperature duration. In an estuary as small as Connecticut River and Savannah River, resident copepods of low salinity water have to face the risk being flushed out of the estuary. Thus how they regulate their local distribution may possibly make the difference we observed in the two estuaries as well. *A. tonsa* is reportedly capable of migrating vertically in a diel rhythm (Cuker & Watson 2002). As we previously studied in Chesapeake Bay, *Acartia F* lineages appeared to regulate well in unstratified tributaries than strongly stratified mainstem areas (Chen & Hare 2008). This is consistent with the findings of Savannah estuary and Connecticut estuary, with the latter has more stratified environment and less tight salinity association of *F*. However, more extensive samplings are required for the argument.

In contrast to lineage *F*'s affinity for low-salinities, lineages *S* and *X* were both euryhaline (5~30 PSU, this study). Their variable distributions within estuaries may have been determined by other factors than salinity. Interestingly, the three major lineages were significantly partitioned along the salinity gradient of Savannah River Estuary, with

lineages *F*, *X* and *S* aggregating in the oligohaline, meso-polyhaline and polyhaline zone respectively (Fig. 4-8). These salinity associations were not found elsewhere where *S* and *X* lineages occur together, suggesting that niche partitioning among *Acartia* genetic lineages is partially driven by, or accentuated by community interactions (e.g., competition) or other local environmental factors.

A suite of factors (abiotic or biotic) other than salinity may vary along estuarine gradient and contribute to niche partitioning of *Acartia* lineages. Most of the geographic samples used in this study were collected from the upper layer of seawater and the summer season, with less variation in other abiotic factors (e.g. temperature, dissolved oxygen) than salinity. The observed niche partitioning patterns are therefore more likely affected by biotic factors untested in this study, if not salinity. Among them, food (e.g. phytoplankton) composition, interspecific competition and selective predation deserve the focal attentions. First of all, lineages may compete for food where food availability including both quantity and quality is limited. Food availability plays an important role in regulating abundance, egg production and population development of *A. tonsa* (Durbin *et al.* 1983; Kleppel & Burkart 1995; Kleppel *et al.* 1998). Although *A. tonsa* has been reported as an omnivorous copepod capable of feeding on diverse diets, such as phytoplankton (Turner 1984), detritus (Roman 1984), protist (Ederington *et al.* 1995), etc. Given the cryptic lineages hidden in *A. tonsa*, it might need a re-examination of its lineage-specific feeding habit, preferences and possible specialization. If *Acartia* lineages have their own preferences on food diets, it is conceivable that food composition varied along estuarine gradient may cause niche partitioning among *Acartia* lineages. If *Acartia* lineages are truly omnivorous and feeding generalists, food limitation may not be a factor

leading to niche partitioning because there is abundant food in estuarine environments (Paffenhöfer & Stearns 1988; Reaugh 2005; Lloyd 2006).

Interspecific competitions for food and space are not limited to cryptic *Acartia* lineages studied. In many estuaries, *A. tonsa* is also living with various other marine organisms at the same trophic level, including two well-studied estuarine copepods *Eurytemora affinis*, *Acartia hudsonica*, and many cladocerans (e.g. Brownlee & Jacobs 1987; Sullivan & McManus 1986; Lakkis 1994; Kimmel & Roman 2004; Sullivan *et al.* 2007). Therefore the observed niche patterns might result from interactions between all coexisting and competing species components in the system.

Besides competitors, predators of *Acartia* may also have a contributing role in determining distribution and niche partitioning of *Acartia. Acartia* are a major food of larval fishes and invertebrate carnivores (e.g. jellyfishes, chaetognaths) (Mauchline 1998). In principle, if the predators have food preferences by, for example, body size, swimming speed etc. They might have differential predation on different *Acartia* lineages. Long-lasting selective predation pressure and predator-prey relationship may drive niche diversification of both predators and preys. *Acartia* are also hosts to many ectosymbionts (Ho & Perkins 1985; Weissman *et al.* 1993). In this study, I found many low-salinity individuals from Choptank River, Chesapeake Bay were infected by peritrich ciliates (personal observation). These ciliates appear to only occur in low-salinity area. If they have preferences on different *Acartia* lineage hosts, it would be another factor contributing to niche partitioning of *Acartia*. Unfortunately, studies on these factors are rare.

Body size variations of Acartia *lineages*

Unlike many other adaptive morphological traits, body size is well known for its extreme plasticity in response to a variety of environmental factors, especially food availability and temperature (Stearns & Koella 1986; Atkinson 1994; Blanckenhorn 2000). In this study, we found tremendous variation in body lengths of *Acartia tonsa*, of which we are particularly interested in those within and among the three distinct genetic lineages. As well recognized for marine calanoid copepods (Mauchline 1998), we found significant sexual dimorphism in each genetic lineage with females consistently larger. Thus, my morphometrical analyses were all controlled for sex.

Variability of body size observed within each lineage was high. Seasonal comparisons were only possible with Chesapeake Bay samples and showed that water temperature had a strong influence on body size in agreement with the "temperature-size rule" that predicts smaller sizes at higher temperature (Atkinson 1994). This further requires a control for season when testing other variation components of body size. Fortunately, most geographic samples were collected during August–September 2005 when water temperatures were between 25 and 30°C.

After controlling for sex and season, the three genetic lineages showed significant differences in body size and shape: *S* was generally bigger than *F* and *X*, while the latter two have no consistent rank for each other. While the overall comparison of females between lineages did not control for the effects of food availability across sample sites, the fact that *S* was consistently bigger than *F* within the same samples (Fig. 4-10) and within estuaries (Fig. 4-11) suggests an evolutionary divergence of body size in what are otherwise cryptic species.

Geographic variations in mean female body size within lineages did not show any trend with latitude following the "Bergmann rule" (Bergmann 1847) that predicts smaller sizes in southern populations (at low latitudes) if temperature rules (Fig. 4-12). According to Bergmann rule and temperature-size rule, we failed to see larger body size in Finland (11.0 °C), yet we did see larger body sizes from Chesapeake Bay and Delaware Bay that included spring samples from lower temperature. An important reason for this failure is the insufficient samples that could not effectively tease apart other confounding factors, for example, food availability. Gaudy & Verriopoulos (2004) also reported huge variability in *A. tonsa* body size in Berre Lagoon, France, which depends on local environmental conditions including temperature, salinity, food abundant and quality. To test for latitudinal gradients of body size, we need a large sample size that should effectively control for all local environmental factors. Meanwhile, most samples were from summer 2005 with a narrow temperature range of 25–30 °C across geographic regions (Table 4-1), which may have weakened the temperature effect. Therefore, Bergmann's rule may not be completely falsified for *Acartia tonsa.* We believe that temperature plays the strongest role in regulating *Acartia* body size and thermal plasticity of *A. tonsa* was the cause of high variations within lineages.

Despite large variances found in body size of *A. tonsa*, significant differences across lineages provided evidence for morphological divergence paralleled to genetic divergence. This finding strengthens our belief that there should be morphological traits signaling *Acartia* population divergence and useful for distinguishing genetic lineages. Gaudy & Verriopoulos (2004) proposed to use body length and body proportion (prosome-urosome ratio) to identify individuals to local populations that experienced

unique local conditions. Garmew *et al.* (1994) found prosome length helpful in distinguishing distinct geographic populations. In my study, I found high correlations among prosome length, urosome length and total body length, which partially caused the prosome-urosome ratio less unique across lineages in some cases. Again, insufficient sampling maybe another reason for loosing statistical powers to test variations of P/U ratio. However, prosome length is a good indicator showing variations among lineages, sexes, and geographic samples. More thorough studies on morphological variation in *Acartia* lineages should be an important step taken from this study.

Body size variations among *Acartia* lineages also bear adaptive implications. Particularly, smaller size of low-salinity lineage *F* relative to high-salinity *S* in Chesapeake Bay may hint an adaptation of *F* to low-salinity environment. Most free living holoplanktonic copepods have denser bodies than seawaters including the fully saline oceanic waters (Mauchline 1998). They have to spend energy swimming around to maintain their positions in water column, otherwise would sink down to the bottom. The sinking rates of *A. tonsa* in still water were empirically recorded between 0.6–1.0 mm/s (Jacobs 1961; Jonsson & Tiselius 1990). The sinking rate is expected to vary with salinities and body sizes. First of all, it is well-known that saltier, denser waters have more buoyancy than fresher waters. Copepods in fresher waters would have faster sinking rates and spend more energy staying in water column than those in saltier waters. Second of all, the larger the body, the faster it sinks given the same body density. Weissman *et al.* (1993) reported that sinking rates of a congeneric copepod, *A. hudsonica*, linearly increased with its prosome length. Given these two facts, smaller

body sizes would be favored in copepods living in fresher waters, because they would lower the sinking rates and allow copepods to save energy for other living expenses.

Conclusions

In this study, I found complex patterns of distribution and co-occurrence for three genetic lineages (*F*, *S*, *X*) of an estuarine copepod *Acartia tonsa* along salinity gradients within estuaries. Lineage *F* showed specialization to low-salinity waters while lineage *S* and *X* were more euryhaline. The differential distributions along salinity gradients of Savannah River Estuary and Chesapeake-Albmemarle-Pamlico system strongly support salinity associated niche partitioning among *Acartia* lineages. However, the inconsistent distribution patterns found in additional estuaries suggest that (1) some distinct sublineages of *F* may have evolved less restrictive salinity preferences/tolerances, and (2) observed niche partitioning is also affected by other unmeasured ecological factors. Although this diversification of *Acartia* species is likely to have involved physiological or behavioral adaptations that increased fitness at different salinities, morphostasis kept these lineages hidden despite intensive research on their ecology. Now, with the benefit of molecular identification of lineages, body size differences between the lineages are apparent when environmental plasticity is accounted for. Further investigation is likely to reveal additional adaptive differentiation that facilitates the co-occurrence of these lineages in estuaries.

Table 4-1 Sampling information of three *Acartia* genetic lineages (*F, S,* and *X*), including sampling time, geographic locations, lineage compositions, RFLP sample size (*N*), sample sizes for morphometric measures (*N*f – female, *N*m – male) and environmental salinity (S) and temperature (T). n.d. – no data. / - data not used for lineage association analysis due to small sample size.

							Lineage composition (%)				*N*f			*N*m		
Sample name & locality		Date	Latitude	Longitude	S (psu)	T (ºC)	*F*	*S*	*X*	*N*	*F*	*S*	*X*	*F*	*S*	*X*
Rhode Island																
PQL	Pettaquamscutt Lake	12 Aug. 2005	41.49°N	71.45°W	15.1	28.4	10.0	90.0	0.0	19	1	15		1		
Connecticut																
TR1	Thames River Estuary	11 Aug. 2005	41.35°N	72.09°W	25.2	25.3	10.0	40.0	50.0	10	1	1			1	
TR2	Thames River Estuary	11 Aug. 2005	41.43°N	72.10°W	17.0	28.3	100.0	0.0	0.0	10	10					
TR3	Thames River Estuary	11 Aug. 2005	41.52°N	72.08°W	5.2	28.1	0.0	100.0	0.0	6		6				
CTR1	Connecticut River Estuary	12 Aug. 2005	41.35°N	72.38°W	8.2	27.3	0.0	74.2	25.8	31		4	2		3	2
CTR2	Connecticut River Estuary	12 Aug. 2005	41.32°N	72.35°W	19.4	24.8	0.0	100.0	0.0	30		7			8	
CTR3	Connecticut River Estuary	12 Aug. 2005	41.32°N	72.35°W	11.7	26.2	0.0	82.8	17.2	29		8			8	1
CTR4	Connecticut River Estuary	12 Aug. 2005	41.29°N	72.35°W	26.1	22.0	5.3	78.9	15.8	19	1	3	2		5	1
CTR5	Connecticut River Estuary	12 Aug. 2005	41.35°N	72.38°W	4.2	27.8	0.0	87.5	12.5	16		2			3	
HOR	Housatonic River Estuary	13 Aug. 2005	41.20°N	73.11°W	10.0	27.1	0.0	100.0	0.0	26		7			12	
New York – New Jersey																
HUR1	Hudson River Estuary	13 Aug. 2005	40.83°N	73.97°W	13.2	28.3	0.0	0.0	100.0	27			14			3
HUR2	Hudson River Estuary	14 Aug. 2005	40.70°N	74.07°W	20.4	29.6	3.6	3.6	92.9	28	1	1	21			3
HAR1	Hackensack River	14 Aug. 2005	40.85°N	74.03°W	5.2	30.2	0.0	11.1	88.9	27			11			2
HAR2	Hackensack River	14 Aug. 2005	40.73°N	74.10°W	18.3	30.0	0.0	0.0	100.0	20			11			2
NWB	Newark Bay	14 Aug. 2005	40.68°N	74.13°W	21.2	28.1	0.0	0.0	100.0	27			5			1
RAR	Raritan River	14 Aug. 2005	40.47°N	74.36°W	11.1	28.8	0.0	21.4	78.6	14			2		2	7
NYB	New York Bay	14 Aug. 2005	40.54°N	74.13°W	24.2	29.2	0.0	14.3	85.7	28		1	16		1	2
GEH	Great Egg Harbor	15 Aug. 2005	39.29°N	74.63°W	19.7	30.0	0.0	0.0	100.0	19			2			3
Delaware																
IRB1	Indian River Bay	8 Aug. 2003	38.60°N	75.17°W	n.d.	n.d.	60.0	13.3	26.7	15	1					
DB4	Delaware Bay	28 May 2005	39.22°N	75.29°W	15.0	16.8	7.7	30.8	61.5	13	1	4	7			1
DB5	Delaware Bay	28 May 2005	39.08°N	75.19°W	23.2	15.8	12.5	87.5	0.0	8	1	4			3	
Chesapeake Bay, MD-VA																
PA2	Patuxent River	3 Aug. 2004	38.41°N	76.57°W	n.d.	n.d.	0.0	100.0	0.0	12		1				
PA3	Patuxent River	27 May 2005	38.40°N	76.53°W	7.8	17.9	0.0	100.0	0.0	42		12			14	
PO1	Potomac River	26 May 2005	38.19°N	76.71°W	4.2	17.9	15.9	84.1	0.0	37	5	30		2	5	
PO2	Potomac River	26 May 2005	38.11°N	76.54°W	6.7	17.2	13.2	86.8	0.0	33	2	19		1	8	
C1	Choptank River	30 Mar 2004	38.72°N	76.01°W	1.6	10.6	80.0	20.0	0.0	20	11	3				
C2	Choptank River	30 Mar 2004	38.66°N	75.96°W	6.0	9.8	61.3	38.7	0.0	62	39	23				

Table 4-1 *continued*

Sample name & locality		Date	Latitude	Longitude	S (psu)	T (ºC)	Lineage composition (%)				*N*f			*N*m		
							F	*S*	*X*	*N*	*F*	*S*	*X*	*F*	*S*	*X*
C3	Choptank River	30 Mar 2004	38.58°N	76.02°W	9.1	9.3	11.8	88.2	0.0	51	6	39			5	
C4	Choptank River	30 Mar 2004	38.65°N	76.17°W	10.5	9.2	10.0	90.0	0.0	40	2	28		1	7	
C5	Choptank River	29 Mar 2004	38.64°N	76.33°W	11.0	8.7	5.3	94.7	0.0	38	1	19		1	7	
C6	Choptank River	10 Aug. 2004	38.72°N	76.01°W	3.3	25.7	77.3	22.7	0.0	22						
C7	Choptank River	10 Aug. 2004	38.66°N	75.96°W	5.6	25.7	77.5	22.5	0.0	40	29	7		2		
C8	Choptank River	10 Aug. 2004	38.58°N	76.02°W	9.0	25.4	25.9	74.1	0.0	27	5	11		1	5	
C9	Choptank River	9 Aug. 2004	38.65°N	76.17°W	10.8	26.1	0.0	100.0	0.0	47		18			8	
C10	Choptank River	9 Aug. 2004	38.64°N	76.33°W	11.1	24.8	0.0	100.0	0.0	46		19			9	
A1	Mainstem	29 Mar 2004	38.57°N	76.44°W	11.4	8.1	18.8	81.2	0.0	48	4	25		5	12	
A2	Mainstem	29 Mar 2004	38.57°N	76.50°W	10.7	8.4	5.3	94.7	0.0	19						
A3	Mainstem	9 Aug. 2004	38.57°N	76.44°W	8.0	25.3	0.0	100.0	0.0	43		43				
A4	Mainstem	9 Aug. 2004	38.57°N	76.50°W	8.5	25.0	2.6	97.4	0.0	38	1	29			6	
M1	Mainstem	25 May 2005	39.17°N	76.33°W	8.4	14.9	0.0	100.0	0.0	46		21			11	
M2	Mainstem	25 May 2005	38.96°N	76.38°W	8.3	15.4	0.0	100.0	0.0	29		20			3	
M3	Mainstem	25 May 2005	38.75°N	76.43°W	9.4	15.6	0.0	100.0	0.0	44		28			14	
M4	Mainstem	27 May 2005	38.30°N	76.28°W	10.3	16.3	0.0	100.0	0.0	30		25			1	
M5	Mainstem	27 May 2005	38.06°N	76.22°W	11.3	16.5	0.0	100.0	0.0	40		32			7	
T1	Mainstem	22 July 1996	39.33°N	76.20°W	0.3	25.7	93.8	6.3	0.0	16						
T2	Mainstem	22 July 1996	39.23°N	76.24°W	2.5	25.1	52.9	47.1	0.0	34						
T3	Mainstem	22 July 1996	39.18°N	76.28°W	4.0	24.9	25.0	75.0	0.0	36						
T4	Mainstem	6 May 1996	39.11°N	76.30°W	1.5	16.1	21.6	78.4	0.0	37	6	1		2	27	
T5	Mainstem	6 May1996	39.11°N	76.30°W	6.9	13.3	9.3	90.7	0.0	43	2	2		1	23	
T6	Mainstem	5 May 1996	39.06°N	76.33°W	2.5	16.7	31.7	68.3	0.0	41	12	9		1	19	
T7	Mainstem	5 May 1996	39.06°N	76.33°W	6.2	13.9	27.0	73.0	0.0	37	7	5		2	20	
T8	Mainstem	6 May1996	39.00°N	76.36°W	3.3	15.8	20.5	79.5	0.0	39	6	18		2	13	
T9	Mainstem	6 May1996	39.00°N	76.36°W	10.1	12.1	8.3	91.7	0.0	36	1	7		1	20	
T10	Mainstem	23 July 1996	39.07°N	76.40°W	5.9	24.4	9.3	90.7	0.0	43						
T11	Mainstem	20 July 1996	38.83°N	76.41°W	7.7	25.3	3.3	96.7	0.0	30						
T12	Mainstem	17 July 1996	38.00°N	76.17°W	11.0	25.8	0.0	100.0	0.0	19						
T13	Mainstem	18 July 1996	37.67°N	75.95°W	14.8	26.7	0.0	100.0	0.0	59						
T14	Mainstem	25 Oct. 1999	37.67°N	76.20°W	20.0	16.5	0.0	100.0	0.0	16		15			1	
T15	Mainstem	19 Oct. 2000	37.67°N	76.20°W	19.0	18.0	0.0	100.0	0.0	17		17				
T16	Mainstem	3 Nov. 1995	37.66°N	76.19°W	21.0	17.9	0.0	100.0	0.0	21		20			1	
T17	Mainstem	1 Nov. 1997	37.50°N	76.08°W	24.1	15.6	0.0	100.0	0.0	17		5			4	
T18	Mainstem	20 Oct. 1998	37.50°N	76.08°W	23.8	20.2	0.0	100.0	0.0	23		16			4	

Table 4-1 *continued*

Sample name & locality		Date	Latitude	Longitude	S (psu)	T (ºC)	Lineage composition (%)			*N*	*N*f			*N*m		
							F	*S*	*X*		*F*	*S*	*X*	*F*	*S*	*X*
T19	Mainstem	30 Oct. 1996	37.33°N	76.15°W	17.3	17.3	0.0	100.0	0.0	21		15			4	
T20	Mainstem	21 Oct. 1998	37.05°N	76.03°W	26.4	19.5	0.0	100.0	0.0	10		7			3	
T21	Mainstem	2 Nov. 1997	37.67°N	76.18°W	15.5	20.1	/	/	/	/		3				
T22	Mainstem	19 Oct. 1998	37.67°N	76.17°W	18.4	19.7	/	/	/	/		1				
Albemarle-Pamlico Sound, NC																
SCU	Albemarle Sound	17 Sept 2005	35.92°N	76.25°W	2.0	29.5	96.9	3.1	0.0	32	28	1		3		
ALG	Albemarle Sound	17 Sept 2005	35.90°N	76.03°W	2.5	26.7	100.0	0.0	0.0	20	11					
MNH	Albemarle Sound	17 Sept 2005	35.91°N	75.77°W	22.0	27.3	80.5	19.5	0.0	41	10	6		14	2	
PML1	Pamlico River	18 Sept 2005	35.54°N	77.04°W	1.2	26.8	100.0	0.0	0.0	18						
PML2	Pamlico River	18 Sept 2005	35.38°N	76.75°W	6.9	26.7	100.0	0.0	0.0	14						
PML3	Pamlico River	18 Sept 2005	35.25°N	76.59°W	11.4	27.1	41.7	58.3	0.0	36	4	8		6	4	
NSR1	Neuse River	18 Sept 2005	34.97°N	76.81°W	10.2	26.7	30.6	69.4	0.0	36	2	13		3	8	
NSR2	Neuse River	18 Sept 2005	35.10°N	77.04°W	2.1	29.1	70.0	30.0	0.0	10	3	1		1	1	
CFR	Cape Fear River Estuary	18 Sept 2005	34.23°N	77.95°W	2.2	26.9	91.0	0.0	9.0	22	6	2		2		
South Carolina																
GTL	Winyah Bay	18 Sept. 2005	33.37°N	79.27°W	8.4	26.9	0.0	0.0	100.0	30			25			2
Georgia																
SVR2	Savannah River	20 Sept 2005	32.04°N	80.87°W	29.7	28.0	0.0	22.5	22.5	40		5	3		3	
SVR3	Savannah River	20 Sept 2005	32.04°N	80.89°W	27.1	28.4	0.0	77.5	64.3	28		1	11			1
SVR4	Savannah River	20 Sept 2005	32.04°N	80.91°W	25.6	28.5	2.9	35.7	51.4	35		1	2			
SVR5	Savannah River	20 Sept 2005	32.05°N	80.93°W	20.7	28.3	0.0	45.7	86.7	15		1	1		1	
SVR6	Savannah River	20 Sept 2005	32.05°N	80.93°W	16.3	28.4	/	/	/	1		1				
SVR7	Savannah River	20 Sept 2005	32.09°N	80.99°W	10.6	28.3	18.2	13.6	68.2	22			5	1	2	3
SVR8	Savannah River	20 Sept 2005	32.08°N	81.05°W	5.4	28.2	63.9	2.8	33.3	36	18	1	12	5		
BWK	Brunswick, GA	19 Sept 2005	31.12°N	81.48°W	22.8	30.1	0.0	7.1	92.9	42		1	5			3
DRN	Darien, GA	19 Sept 2005	31.33°N	81.45°W	0.6	29.2	100.0	0.0	0.0	19	15			2		
Florida																
JAX	Jacksonville	29 Apr. 2005	30.33°N	81.62°W	10.9	n.d.	71.0	0.0	29.0	24	4		2	3		
FD	Faver-Dykes Stake Park	30 Apr. 2005	29.67°N	81.27°W	10.1	n.d.	88.0	0.0	12.0	17	5		1		1	
SDB	South Daytona Beach	1 May 2005	29.21°N	81.01°W	27.5	n.d.	3.2	12.7	84.1	63	1	2	35	1	5	11
SEB	Sebastian	1 May 2005	27.83°N	80.51°W	24.1	n.d.	19.0	12.0	69.0	16	3	1	11		1	
European coast																
FIN	Poiju, Finland	18 Aug. 2005	59.86°N	23.27°E	5.7	11.0	0.0	0.0	100.0	28			7			7

Table 4-2 Summary of morphometrical variations of *Acartia tonsa* among sexes, genetic lineages and seasons. Kruskall-Wallis tests (K-W test) and post-hoc comparisons were conducted to test for difference among lineages within sex. K-W results were marked by letters *a*, *b*, *c* at significance level $\alpha = 0.05$ after Bonferroni adjustment. Wilcoxon tests (W-test) were conducted within lineages to compare difference between sexes. Statistical significance: $P > 0.5$ (n.s.), $P < 0.5$ (*), $P < 0.01$ (**), $P < 0.001$ (***). std – standard deviation.

		Total body length (μm)			Prosome length (μm)			Urosome length (μm)			Prosome-urosome ratio		
Lineage/season		♀	♂	W-test	♀	♂	W-test	♀	♂	W-test	♀	♂	W-test
Spring (Mar – May)													
F	mean	1151.7^{b}	990.4^{b}	***	934.7^{b}	794.7^{b}	***	217.0^{a}	195.7^{a}	*	4.38^{b}	4.10^{b}	**
	std	(142.8)	(170.8)		(112.1)	(133.6)		(38.9)	(40.4)		(0.52)	(0.32)	
	n	119	23		119	23		119	23		119	23	
S	mean	1228.2^{a}	1053.6^{a}	***	1019.9^{a}	856.8^{a}	***	208.3^{b}	196.8^{a}	**	4.95^{a}	4.38^{a}	***
	std	(127.6)	(112.1)		(100.0)	(88.7)		(32.0)	(26.2)		(0.44)	(0.31)	
	n	377	225		377	225		377	225		377	225	
X	mean	929.9^{c}	821.2^{c}	***	761.0^{c}	661.7^{c}	***	168.9^{c}	159.6^{b}	*	4.55^{b}	4.15^{b}	**
	std	(57.7)	(49.3)		(52.1)	(41.4)		(17.8)	(11.0)		(0.49)	(0.23)	
	n	56	13		56	13		56	13		56	13	
K-W test		***	***		***	***		***	***		***	***	
Summer (Aug)													
F	mean	857.1^{c}	773.0^{b}	**	707.2^{c}	622.3^{b}	***	149.9^{b}	149.2^{ab}	n.s.	4.74^{b}	4.18	**
	std	(66.9)	(16.9)		(53.9)	(10.4)		(15.5)	(12.2)		(0.30)	(0.30)	
	n	50	4		50	4		50	4		50	4	
S	mean	973.3^{a}	885.4^{a}	***	809.4^{a}	720.7^{a}	***	163.6^{a}	164.7^{a}	n.s.	4.96^{a}	4.39	***
	std	(76.6)	(91.2)		(64.3)	(74.1)		(15.5)	(19.7)		(0.31)	(0.31)	
	n	183	71		183	71		183	71		183	71	
X	mean	875.5^{b}	800.2^{b}	***	724.6^{b}	648.9^{b}	**	150.9^{b}	151.3^{b}	n.s.	4.82^{b}	4.30	***
	std	(56.6)	(65.0)		(48.1)	(52.9)		(12.7)	(15.1)		(0.34)	(0.28)	
	n	93	34		93	34		93	34		93	34	
K-W test		***	***		***	***		***	**		***	n.s.	
Fall (Sept – Nov)													
F	mean	813.8^{c}	746.1^{b}	***	673.4^{c}	604.2^{b}	***	140.3^{c}	141.9^{b}	*	4.82^{b}	4.27	***
	std	(74.8)	(42.8)		(63.6)	(35.9)		(14.7)	(11.3)		(0.38)	(0.30)	
	n	97	37		97	37		97	37		97	37	
S	mean	1020.1^{a}	877.5^{a}	***	847.7^{a}	712.2^{a}	***	172.4^{a}	165.3^{a}	n.s.	4.95^{a}	4.32	***
	std	(115.4)	(97.4)		(93.7)	(80.5)		(24.1)	(19.8)		(0.35)	(0.32)	
	n	139	38		139	38		139	38		139	38	
X	mean	939.4^{b}	839.2^{a}	***	779.5^{b}	680.2^{a}	***	159.9^{b}	159.1^{a}	n.s.	4.89^{ab}	4.28	***
	std	(55.6)	(39.2)		(44.1)	(33.9)		(13.7)	(9.6)		(0.26)	(0.25)	
	n	66	9		66	9		66	9		66	9	
K-W test		***	***		***	***		***	***		*	n.s.	

Table 4-3 Comparative physical and ecological features of major estuaries on the US Atlantic coast, including estuary type, estuary size (length, width, depth, volume and area), salinity structure and zone areas, freshwater input (FWI), flushing time (FT), annual temperature range (Temp), Temperature duration (D_T for T > 20°C) favorable to *Acartia* populations, and food level. See Appendix B for details of definition and estimation methods. Numbers without specified literature sources are estimated by the author. ? – data not available; / - not applicable

	Type[1]	Length (km)	Width (km)	Depth (m)	Volume (km^3)	Area (km^2)	Salinity zone area (km^2)[2]			FWI[3] (m^3/s)	Tidal range (m)	FT (days)	Temp (°C)	D_T (d/yr)	Food level[4] (mgC/m^3)
							Tidal fresh	Mixing	Seawater						
Long Island sound															
Thames Estuary	SW	26[a]	0.8	4.1	0.09	34[h]	?	?	?	12	0.8	3.8 – 5.6	1.3 – 22.6	68	?
Connecticut R. Estuary	SW	100[b]	0.6	2.2[g]	0.09[g]	42[g]	21[g]	21[g]	0[g]	226	0.8[g]	1[g]	-0.1 – 25.1	90	Low
Long Island Sound	WM	150[c]	20[c]	19.5[g]	63.45[g]	3,259[g]	0[g]	196[g]	3,063[g]	272	1.9[g]	56[g]	1.9 – 24.4	86	High
Hudson-Raritan System															
Hudson R. Estuary	PM	250[d]	1.5	8.3	2.45	295	109	186	0	499	1.5[d]	0.1 – 4[i]	1.7 – 25.9	113	Moderate-High
Entire system	PM	/	/	6.1[g]	4.90[g]	799[g]	109[g]	411[g]	278[g]	511	1.4[g]	9[g]	0.4 – 25.7	110	Moderate-High
Chesapeake Bay															
Mainstem	PM	320[e]	24.3	7.3[g]	51.12[g]	6,974[g]	328[g]	5,496[g]	1,151[g]	2,280[e]	0.5[g]	105[g]	1 – 28[e]	125	Moderate-High
Choptank R. Tributary	PM	85	4.4	3.1[g]	1.27[g]	411[g]	5[g]	406[g]	0[g]	2 – 4	0.5[g]	19[g]	1.8 – 27.4	127	High
Albemarle-Pamlico System															
Pamlico/Pungo Estuary	PM	63	5.2	1.6[g]	0.73[g]	452[g]	0[g]	452[g]	0[g]	9	0.2[g]	39[g]	4.1 – 30.2	169	High
Neuse Estuary	PM	86	6.2	2.9[g]	1.30[g]	456[g]	5[g]	451[g]	0[g]	24	0.2[g]	73[g]	3.4 – 29.7	174	High
Albemarle Sound	WM	110	15	2.5[g]	6.24[g]	2,497[g]	599[g]	1,898[g]	0[g]	102	0.6[g]	9[g]	5.1 – 29.7	140	High
Pamlico Sound	WM	129	30	2.9[g]	13.71[g]	4,680[g]	0[g]	4,418[g]	262[g]	33	0.4[g]	34[g]	4.9 – 31.4	191	High
Savannah R. Estuary	PM	50[f]	0.4	3.3[g]	0.01	36	8	28	0	322	1.9	5.6[j]	9.5 – 29.3	184	Moderate

[1] Based on salinity structure (Dyer 1997). SW: Salt-wedge, PM: Partially mixed, WM: Well-mixed

[2] Based on Bricker et al. (2007). Tidal fresh: 0~0.5 PSU, Mixing: 0.5~25 PSU, Seawater: > 25 PSU

[3] Except Chesapeake mainstem and Choptank R. estuary, all values were daily average estimates for 30 days prior to sampling date

[4] Estimates were converted from Chl.a level in Bricker *et al*. (2007). Low: 0 ~250 mgC/m^3, Moderate: 250~1000 mgC/m^3, High: > 1000 mgC/m^3

Literature sources*: a*. Trench (2005); *b*. Garvine (1975); *c*. http://www.longislandsoundstudy.net/fact.html; *d*. Geyer & Chant (2006); *e*. Schubel & Pritchard (1986); *f*. Reinert & Peterson (2008); *g*. Bricker *et al*. (2007); *h*. McCarthy (2006); *i*. Howarth *et al*. (2000); *j*. Alber & Sheldon (1999).

Figure legends

Figure 4-1 Illustration of morphometric measures for *Acartia tonsa* body size and shape. The figure shows posterior view of *A. tonsa* body trunks after Bradford-Grieve (1999). Abbreviations: PL – prosome length, UL – urosome length, BL – total body length.

Figure 4-2 Agarose gel picture showing restriction fragment length polymorphism (RFLP) patterns of mtCOI gene recognized in *Acartia* lineages. Lineage *X* has a single 2-fragment pattern (304, 406 bp). Lineage *S* has two patterns: 2-fragment *S* type I (167, 543 bp) and no-cut *S* type II (710 bp; not shown). Lineage *F* has three patterns: 3-fragment *F* type I (118, 167, 425 bp), 2-fragment *F* type II (118, 592 bp) and a rare 3-fragment *F* type III (114, 118,478 bp; not shown). The first lane on the left shows 1 kb standard DNA ladder. The first lane on the right shows RFLP standard prepared from known DNA samples.

Figure 4-3 Lineage compositions and distribution of *Acartia tonsa* in Thames River Estuary and Connecticut River Estuary of Long Island Sound. Numbers on the pies are salinities associated with the samples. Three lineages are color-coded with white (*F*), light gray (*S*) and dark gray (*X*).

Figure 4-4 Lineage compositions and distribution of *Acartia tonsa* in Hudson River Estuary System. Numbers on the pies are salinities associated with the samples. Three lineages are color-coded with white (*F*), light gray (*S*) and dark gray (*X*).

Figure 4-5 Pooled distributions of *Acartia* lineages along the binned salinity gradient in three geographic regions (lineage composition: mean plus standard error bar). (A) Estuaries north of Chesapeake Bay; (B) Chesapeake-Albemarle-Pamlico System (CAPS);

(B) Estuaries south of CAPS. Three lineages are color-coded with white (*F*), light gray (*S*) and dark gray (*X*).

Figure 4-6 Lineage compositions and distribution of *Acartia tonsa* in Pamlico River Estuary, Neuse River Estuary of Albemarle-Pamlico Sound. Numbers on the pies are salinities associated with the samples. Three lineages are color-coded with white (*F*), light gray (*S*) and dark gray (*X*).

Figure 4-7 Association plots between composition of *Acartia tonsa* lineages and salinity. (A) Lineage *F*; (B) Lineage *S*. One outlier was excluded from regression models for data from Chesapeake-Albemarle-Pamlico System (CAPS).

Figure 4-8 Lineage compositions and distribution of *Acartia tonsa* in Savannah River Estuary. Numbers on the pies are salinities associated with the samples. Three lineages are color-coded with white (*F*), light gray (*S*) and dark gray (*X*). Logistic regression models were fit for lineage compositions of *F* and *S* against salinities, showing trends of change (dotted lines). A polynomial regression model is fit for lineage *X*, yielding a maximal composition of 89.6% at the salinity of 17.4 PSU.

Figure 4-9 Lineage compositions and distribution of *Acartia tonsa* in Florida lagoons. Numbers on the pies are salinities associated with the samples. Three lineages are color-coded with white (*F*), light gray (*S*) and dark gray (*X*).

Figure 4-10 Body size variations among female *Acartia* lineages in summer. (A) Scatter plot of all samples with regression models for *F*, *S*, and *X* lineage; (B) Sex- and lineage-specific differences of mean prosome length in two different seasons. Three lineages are

color-coded with white (*F*), light gray (*S*) and dark gray (*X*). Error bar: standard deviation.

Figure 4-11 Temporal and spatial variations in prosome length of *Acartia* lineages in Choptank River, Chesapeake Bay. Error bar: standard deviation.

Figure 4-12 Latitudinal variations in female prosome length of *Acartia* lineages for samples collected during August–September 2005. Geographic samples were pooled into estuarine systems (labeled on graph). Three lineages are color-coded with white (*F*), light gray (*S*) and dark gray (*X*). Error bar: standard error of mean.

Figure 4-1

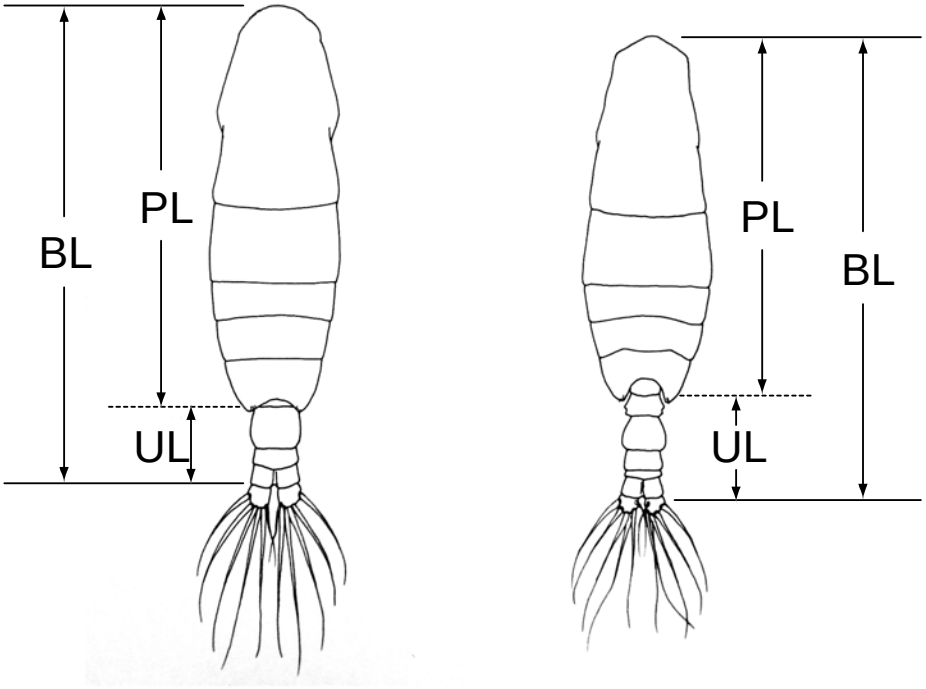


Figure 4-2

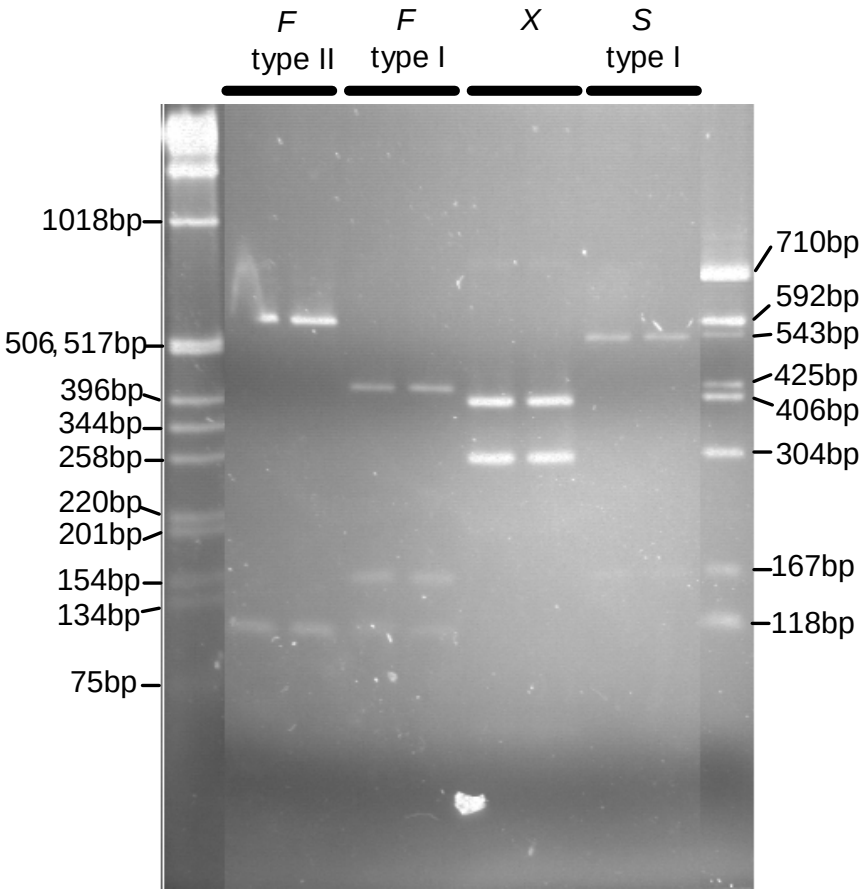

F
type II
F
type I
X
S
type I
1018bp
506, 517bp
396bp
344bp
258bp
220bp
201bp
154bp
134bp
75bp
710bp
592bp
543bp
425bp
406bp
304bp
167bp
118bp

Figure 4-3

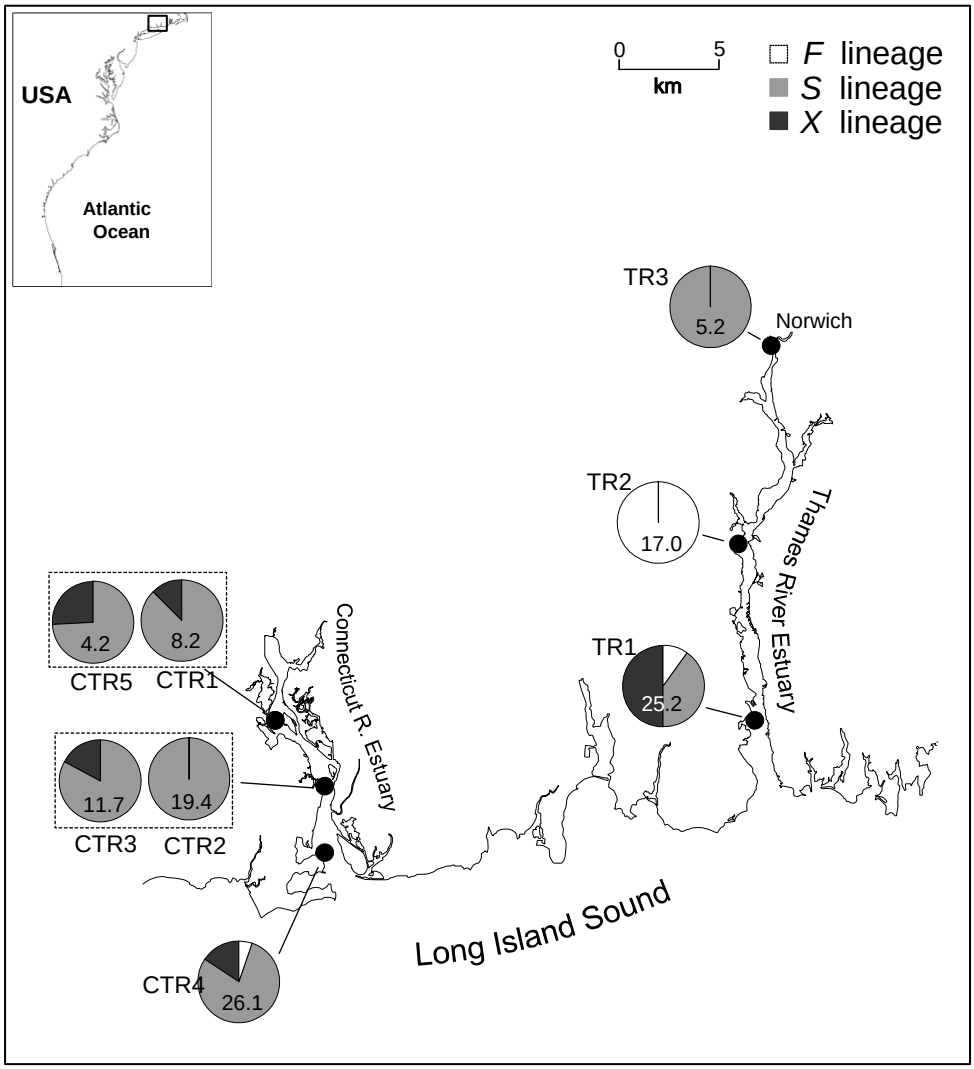

USA
Atlantic Ocean
0
5
km
F lineage
S lineage
X lineage
TR3
5.2
Norwich
TR2
17.0
Thames River Estuary
TR1
25.2
4.2
8.2
CTR5
CTR1
Connecticut R. Estuary
11.7
19.4
CTR3
CTR2
Long Island Sound
CTR4
26.1

Figure 4-4

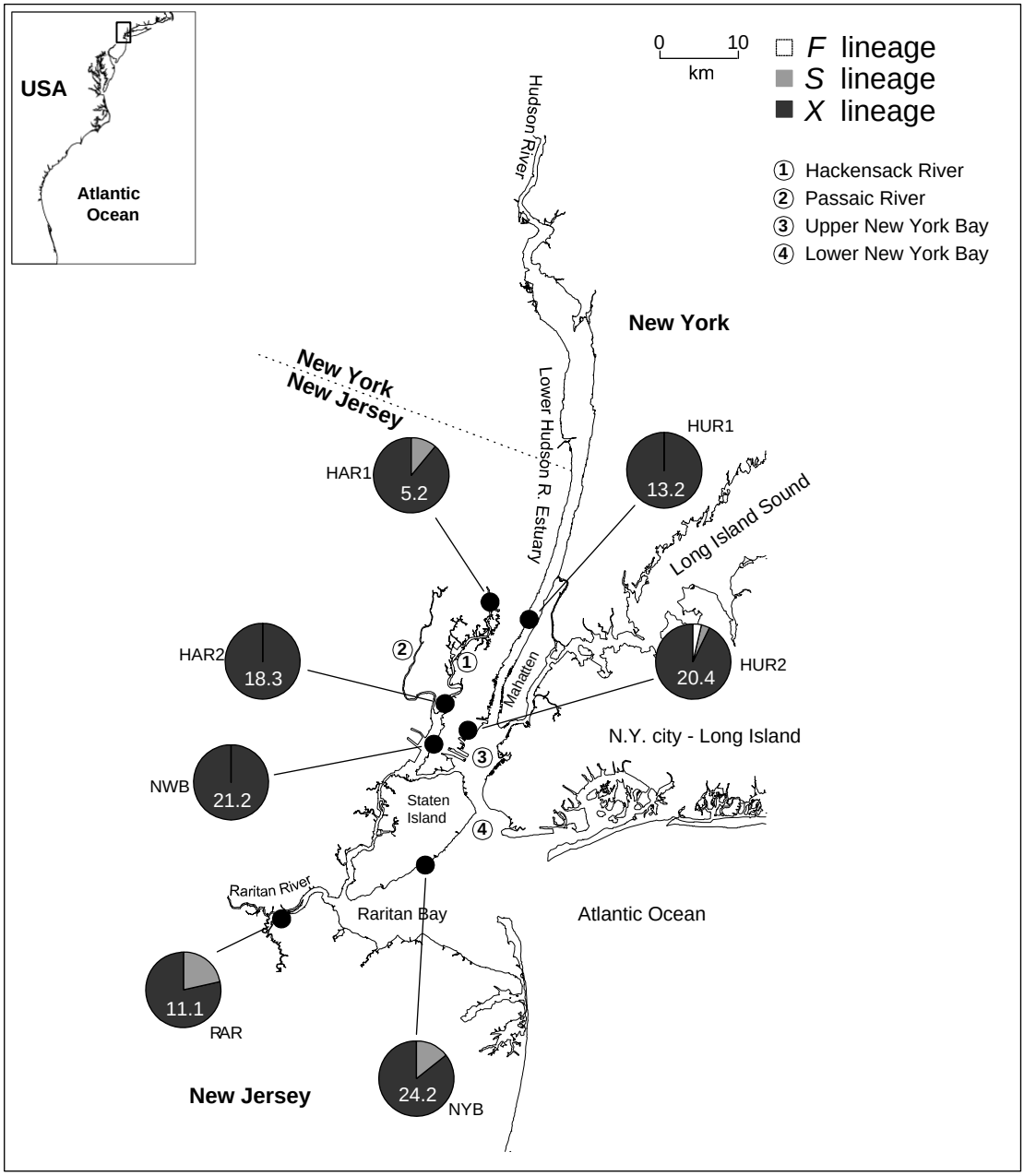

USA
Atlantic Ocean
0
10
km
F lineage
S lineage
X lineage
① Hackensack River
② Passaic River
③ Upper New York Bay
④ Lower New York Bay
Hudson River
New York
New York
New Jersey
Lower Hudson R. Estuary
HUR1
13.2
HAR1
5.2
Long Island Sound
HAR2
18.3
HUR2
20.4
Manhatten
N.Y. city - Long Island
NWB
21.2
Staten Island
Raritan River
Raritan Bay
Atlantic Ocean
11.1
RAR
New Jersey
24.2
NYB

Figure 4-5

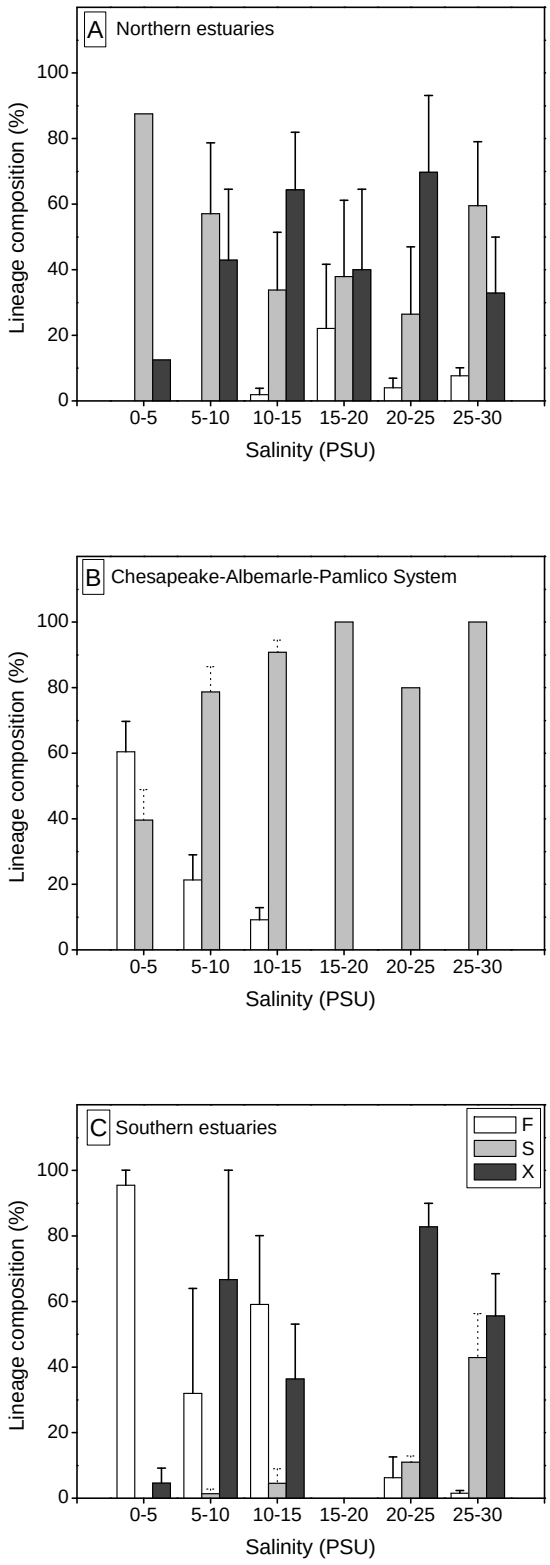

A Northern estuaries
B Chesapeake-Albemarle-Pamlico System
C Southern estuaries
F
S
X
Lineage composition (%)
0
20
40
60
80
100
0-5
5-10
10-15
15-20
20-25
25-30
Salinity (PSU)

Figure 4-6

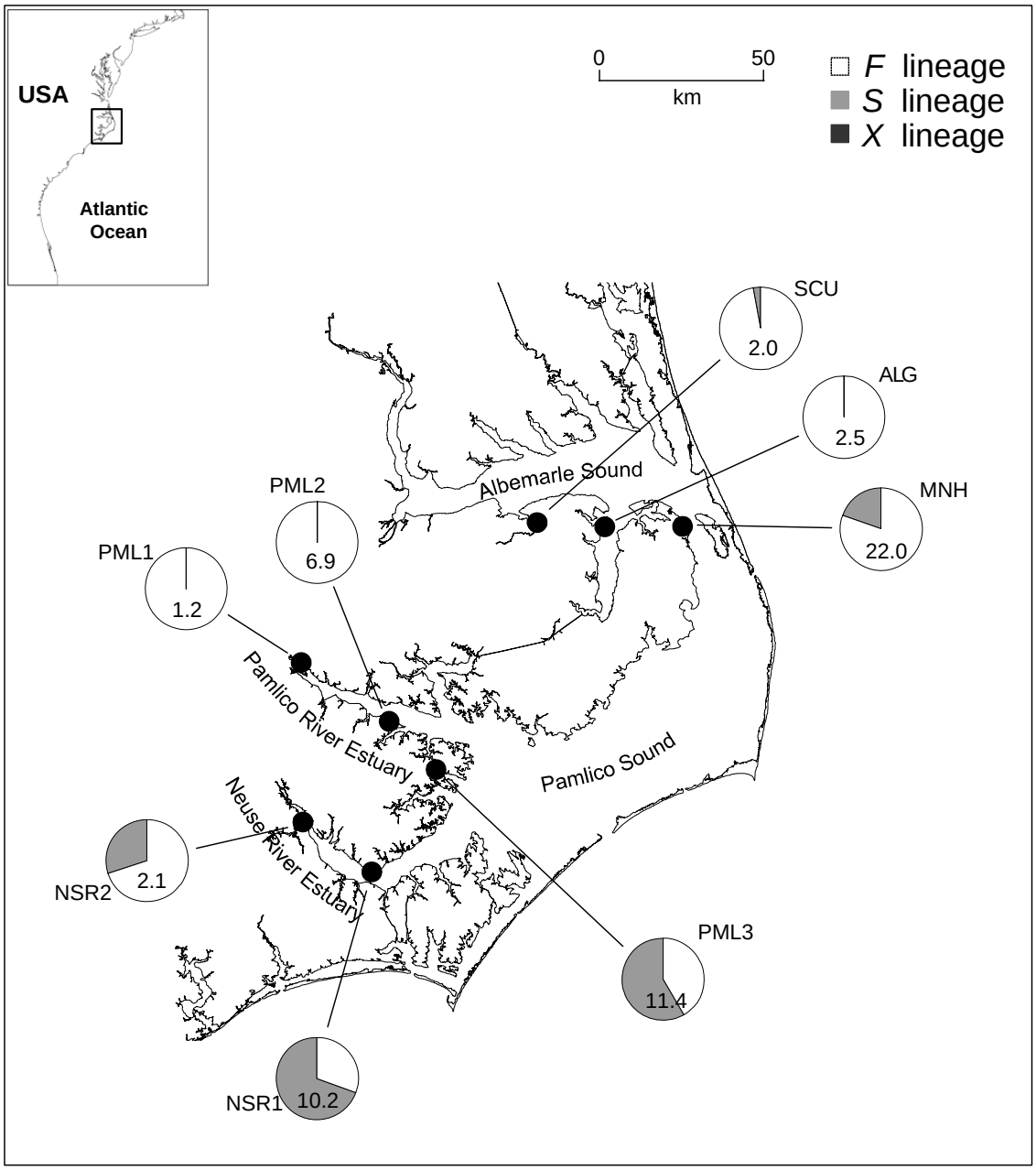
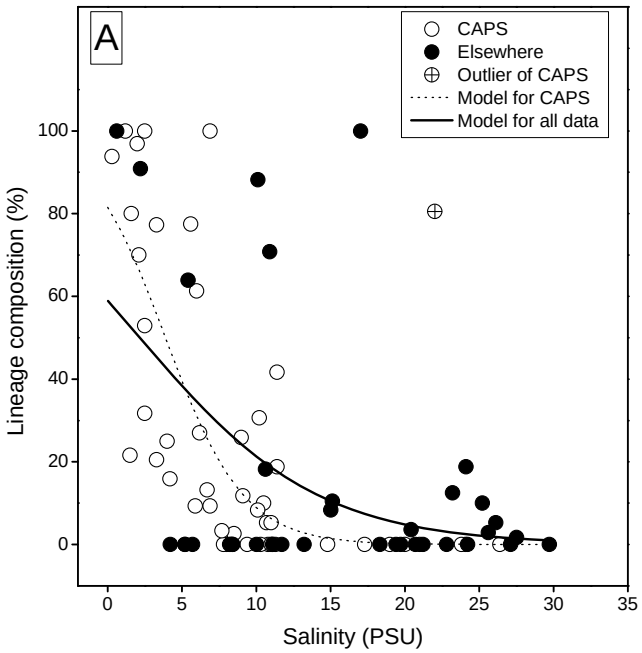

USA
Atlantic Ocean
0
50
km
F lineage
S lineage
X lineage
SCU
2.0
ALG
2.5
MNH
22.0
Albemarle Sound
PML2
6.9
PML1
1.2
Pamlico River Estuary
Pamlico Sound
Neuse River Estuary
NSR2
2.1
PML3
11.4
NSR1
10.2

Figure 4-7


A
CAPS
Elsewhere
Outlier of CAPS
Model for CAPS
Model for all data
Lineage composition (%)
Salinity (PSU)
0
20
40
60
80
100
0
5
10
15
20
25
30
35

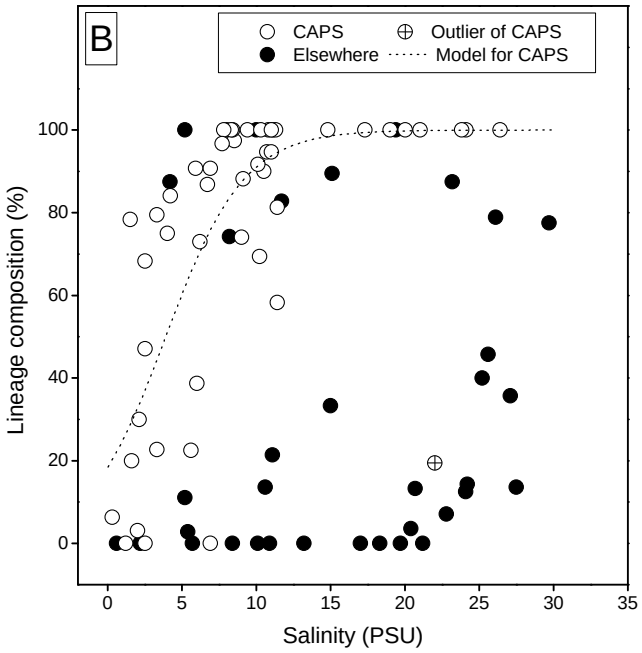

B
CAPS
Outlier of CAPS
Elsewhere
Model for CAPS
Lineage composition (%)
Salinity (PSU)
0
20
40
60
80
100
0
5
10
15
20
25
30
35

Figure 4-8

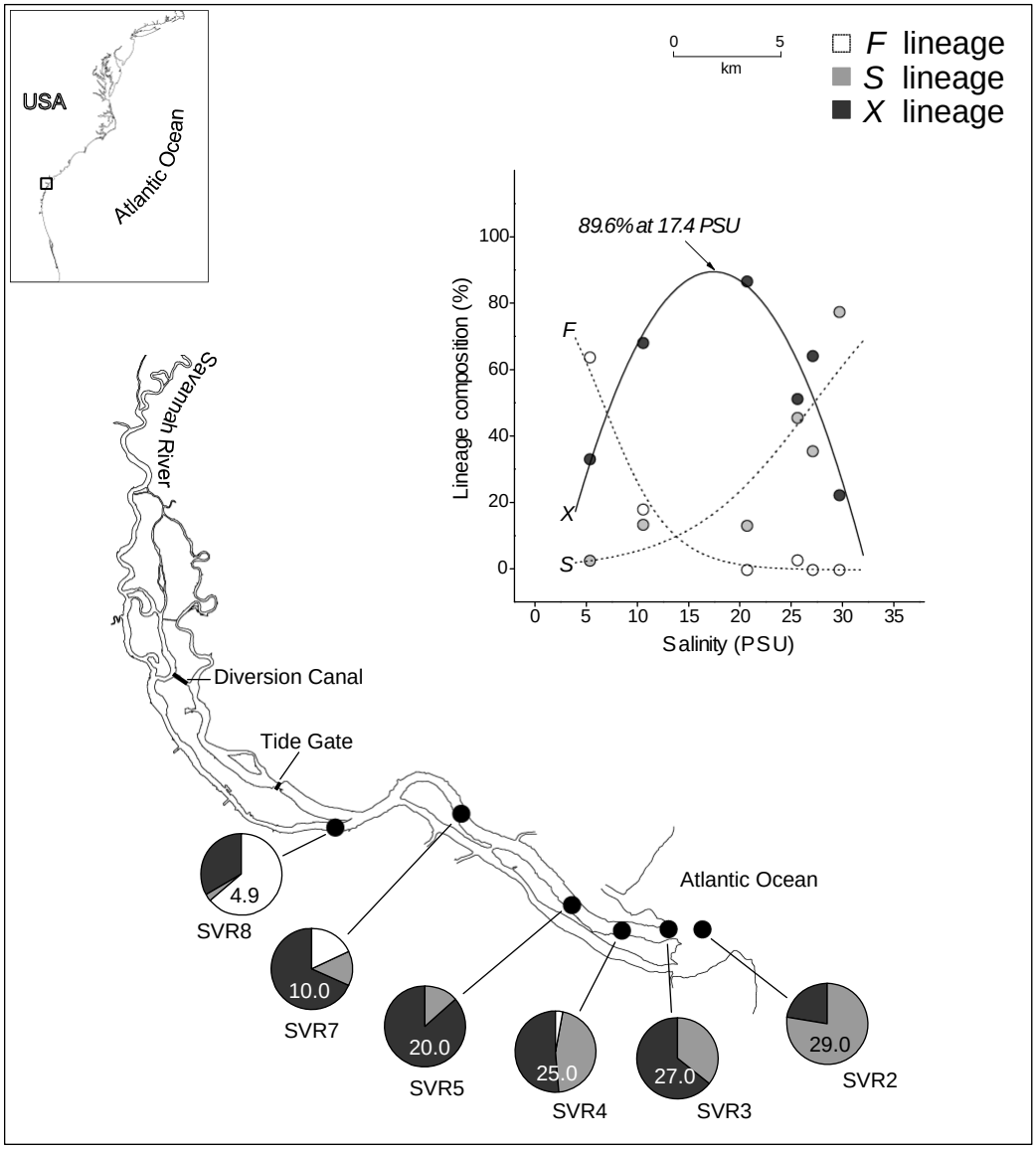

USA
Atlantic Ocean
0
5
km
F lineage
S lineage
X lineage
89.6% at 17.4 PSU
Lineage composition (%)
Salinity (PSU)
F
X
S
Savannah River
Diversion Canal
Tide Gate
Atlantic Ocean
4.9
SVR8
10.0
SVR7
20.0
SVR5
25.0
SVR4
27.0
SVR3
29.0
SVR2

Figure 4-9

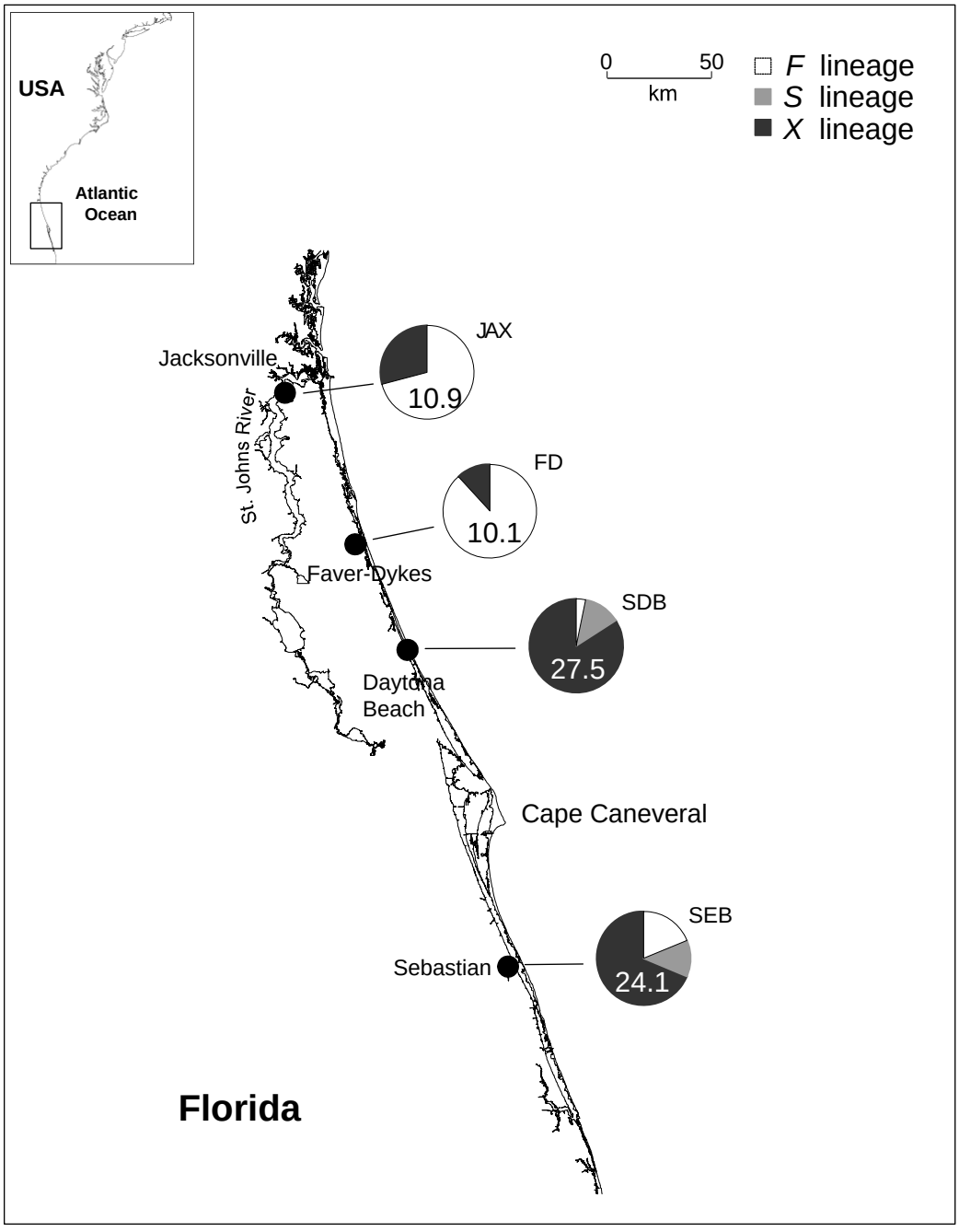
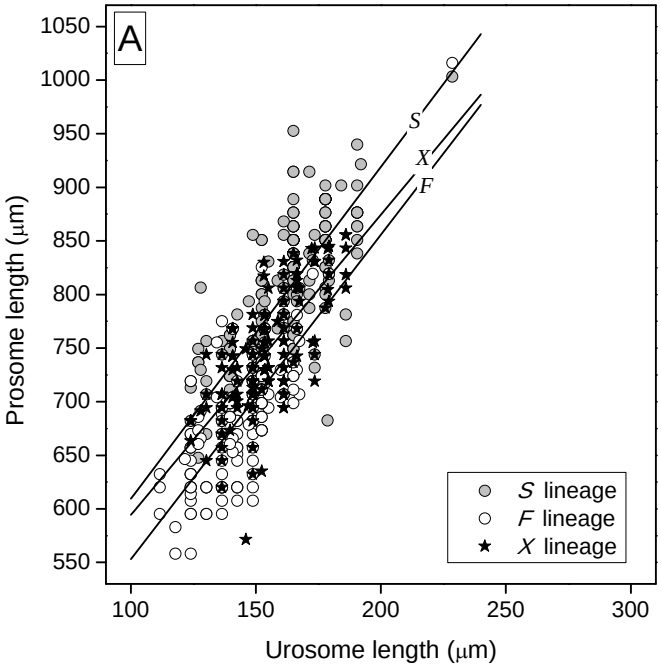

USA
Atlantic Ocean
0
50
km
F lineage
S lineage
X lineage
Jacksonville
JAX
10.9
St. Johns River
FD
10.1
Faver-Dykes
SDB
27.5
Daytona Beach
Cape Caneveral
SEB
Sebastian
24.1
Florida

Figure 4-10


A
Prosome length (μm)
Urosome length (μm)
S
X
F
S lineage
F lineage
X lineage

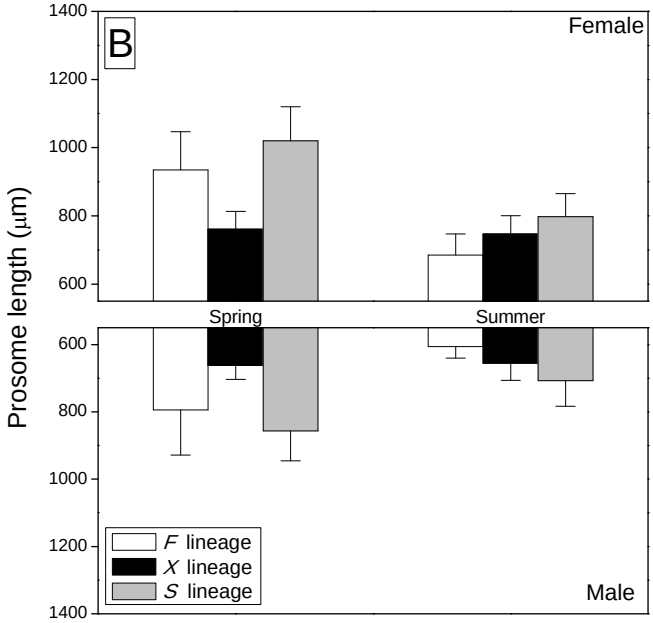

B
Female
Male
Spring
Summer
Prosome length (μm)
F lineage
X lineage
S lineage

Figure 4-11

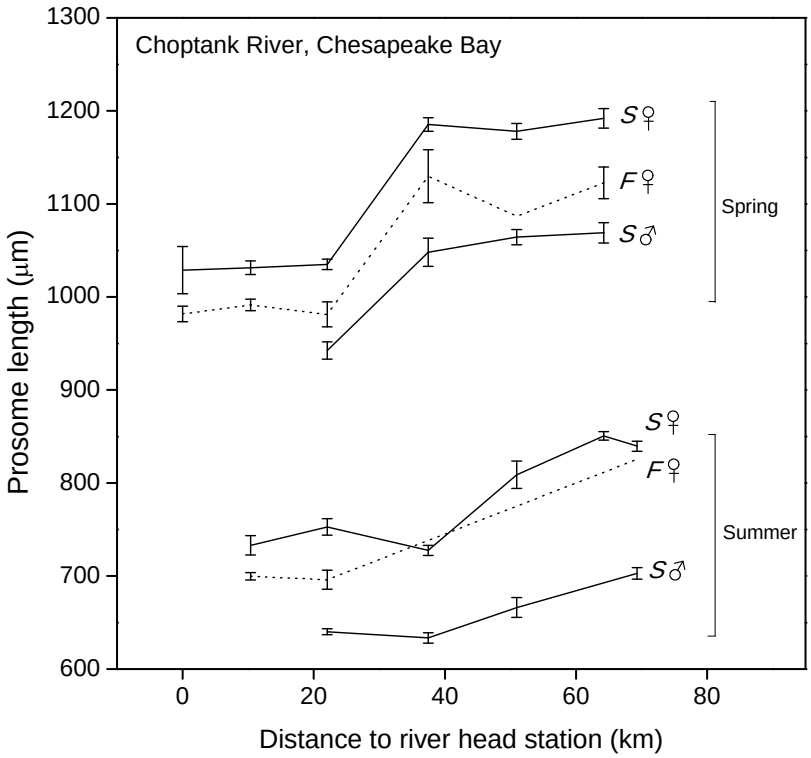
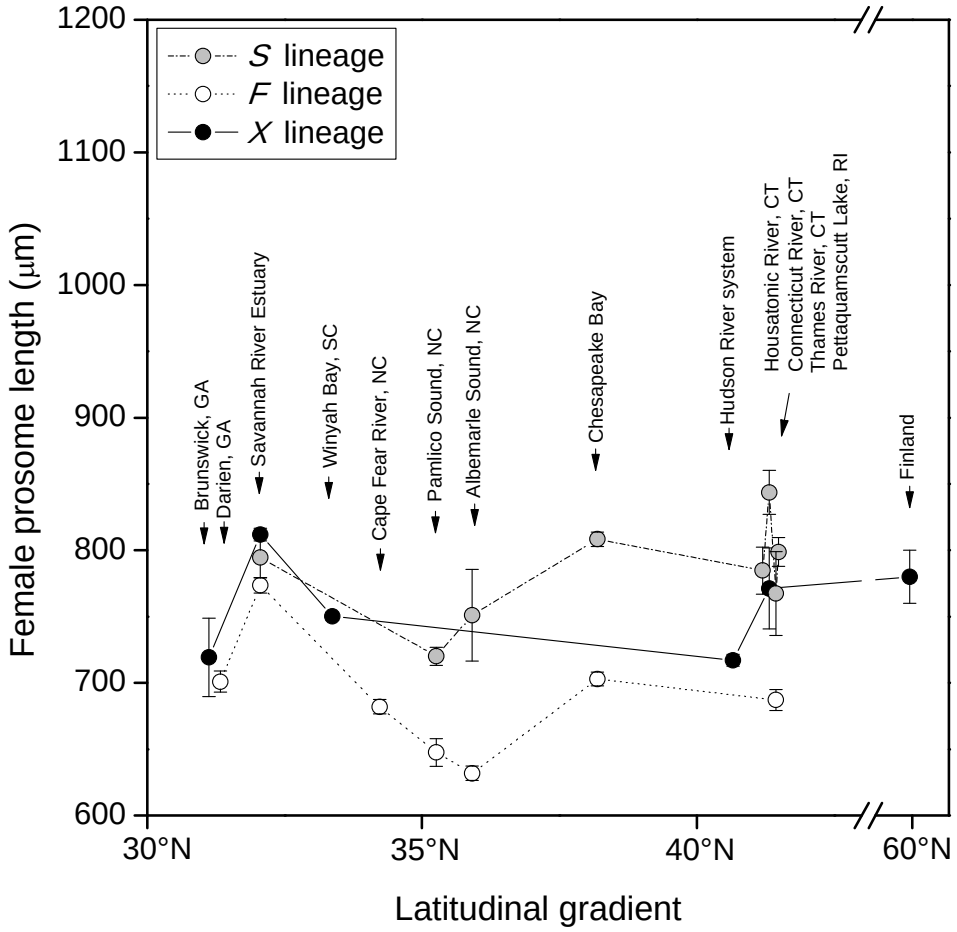

Choptank River, Chesapeake Bay
Prosome length (μm)
Distance to river head station (km)
1300
1200
1100
1000
900
800
700
600
0
20
40
60
80
S♀
F♀
S♂
Spring
S♀
F♀
S♂
Summer

Figure 4-12


S lineage
F lineage
X lineage
Female prosome length (μm)
1200
1100
1000
900
800
700
600
30°N
35°N
40°N
60°N
Latitudinal gradient
Brunswick, GA
Darien, GA
Savannah River Estuary
Winyah Bay, SC
Cape Fear River, NC
Pamlico Sound, NC
Albemarle Sound, NC
Chesapeake Bay
Hudson River system
Housatonic River, CT
Connecticut River, CT
Thames River, CT
Pettaquamscutt Lake, RI
Finland

Chapter 5
General concluding remarks and future work

My dissertation study has produced quite a few interesting results about ecology and evolution of copepod *Acartia tonsa*. Conclusions on different aspects of *A. tonsa* ecology and evolution have been stated in Chapters 2–4 respectively. In this chapter, I would highlight the most significant ones to make some general concluding remarks and propose important future work to extend this study.

Cryptic species diversity

The first big discovery of this study is cryptic species of *Acartia tonsa*. I believe that the cosmopolitan copepod *A. tonsa* is a species complex consisting of at least 5 morphologically cryptic species worldwide, supported by genetic, ecological and partially morphological evidence. Large genetic distances among deeply diverged *Acartia* lineages and genealogical concordance between mitochondrial and nuclear DNA markers could not be explained by any evolutionary processes or mechanisms other than reproductive barriers of gene flows. Cohesion of genetic lineages to ecological factors (salinity in this case) and morphometrical differentiation are consistent evidence for distinctiveness of species.

The finding has two profound implications. First, the cryptic diversity of *A. tonsa* forces us to carefully re-inspect all biological and ecological knowledge on the nominal species *A. tonsa*. Particular caution is called for when extrapolating results and

conclusions from previous studies on local populations. Second, processes and mechanisms underlying *A. tonsa* cryptic diversification allow us to suggest that marine holoplankton biodiversity has been substantially underestiamted. One of the key messages from this study is that the seas are structured rather than continuous even to marine holoplankton. The structure of oceans is scale-dependent, determined by both physical barriers (such as distance, hydrodynamic processes) and physiological barriers from environmental stressors (Temperature, Salinity, etc.). This was illustrated by comparative phylogeographic structures among three *Acartia* lineages (Chapter 3). The structure of the seas is certainly the mechanism generating and maintaining biodiversity of marine organisms. From the case of *A. tonsa*, I believe that many other cosmopolitan species of marine holoplankton should be re-examined using genetic tools, and I expect that cryptic species diversity would be found as in *A. tonsa*.

Limited dispersal in marine holoplankton

Another important finding of this study is that holoplankton may have strongly limited disposal. In *A. tonsa*, effective dispersal was limited between adjacent estuaries on the US Atlantic coast. The limitation however is scale-dependent and limiting factors varied with taxa. Both physical barriers and physiological barriers could limit dispersal of marine holoplankton. For the low-salinity adapted *Acartia* lineage *F*, a physiological barrier (salinity tolerance) strongly constrained its gene flow at a smaller geographic scale, and facilitated its population diversification.

The limited dispersal and constrained gene flow in a representative marine holoplankton suggest that geographic isolation is an important mechanism for

holoplankton speciation. This is particularly demonstrated by the *gsi* study in Chapter 3. Rapid sorting of nITS lineages in three geographically separated populations at a smaller scale (Chesapeake Bay, Indian River Bay, DE, and Pamlico River, NC) reflect the geographically-limited gene flow in *Acartia*. As to the holoplankton paradox of speciation (high genetic diversity under strong gene flow), an answer from my study would be that we were wrong about the dispersal capacity of most holoplankton taxa.

Ecological differentiation and niche diversification

One key finding of this study is the association of *Acartia* lineages with particular salinities and their niche diversification along estuarine salinity gradient. However, niche partitioning shows complex patterns across lineages and estuaries on the US east coast. Overall what I learned is 1) Salinity-directed niche partitioning is significant between a low-salinity specialized lineage (*F*) and others. The Chesapeake Bay study made a strong case for this. 2) Salinity is not the only factor distinguishing niches among studied genetic lineages, especially the euryhaline ones (*X* and *S*). 3) Interspecific competition appears to play an important role in partitioning realized niches among studied lineages. It is important for us to further characterize their niches with more factors, testing which of them (environmental, biological or both) determine niche diversification of *A. tonsa*.

Niche partitioning often signals the ecologically mediated differentiation. Although it requires more tests to conclude the process, *A. tonsa* species complex is a valuable system for those tests. Critical evidence includes adaptive divergence in response to ecological factors that parallels with genetic divergence. I found the body size variation among *Acartia* lineages is promising. A bigger-sized euryhaline lineage (*S*) and

small-sized freshwater lineage (*F*) is likely an adaptive response of *A. tonsa* to their habitat salinities, where size may grant a motility advantage to euryhaline lineage to overcome salinity-associated water viscosity.

Evolutionary history of *Acartia tonsa*

In this study, phylogeography of *A. tonsa* on east coast of the United States did not resolve the history of three deeply diverged lineages due to their old ages (originated likely before Pleistocene). However, sublineages in *F* showed a recent post-Pleistocene history. More interestingly, the diversification of lineage *F* had a southern center of origin (Florida) and showed northward stepwise cladogenesis.

The directional colonization may be associated with historical climate change events, however, it failed to exhibit in other co-distributed lineages. The working explanation is geographic isolation and neutral genetic drift. Evidence of geographic structure was found in all lineages. Since the lineage *F* has strong ecological affinity to restricted low-salinity waters, it is likely more sensitive to geographic isolation and neutral drift mechanism. As a consequence, stronger phylogeographic structure at smaller scales was found in *F* than in euryhaline lineages. Processes and mechanisms underlying the unidirectional stepwise diversification could not be inferred based on current data. More southern samples are required to address this question.

Some other messages may reasonably be taken from the phylogeography study. First, it has been hypothesized that *Acartia tonsa* was introduced from Europe to North America (Mauchline 1998). In my study, a Finland sample consists uniformly of individuals from lineage *X*, the high dispersal euryhaline lineage, which possibly signals

a founder effect of introduction from North America to Europe. Thus, we may hypothesize that European *A. tonsa* may be introduced from North America. Of course, to test this hypothesis, we need survey the European coast with more geographic samples. Second, *Acartia* is reasonably subject to anthropogenic transport via ballast waters for its abundance in estuaries. However, the significant isolation by distance patterns in all lineages tend to reject this hypothesis, or at least the process does not significantly affect the genetic patterns. This seems to counter the knowledge about colonization history of *A. tonsa.* More specific studies and tests are required to address this question.

Future work

Along with interesting results and conclusions, this study also posed many questions awaiting further investigations to answer. I have discussed these questions more or less in previous chapters. Here I would list the most urgent work to follow up in near future.

Morphological study

Morphological study of *Acartia tonsa* should be the immediate work to follow this dissertation study. It should include careful comparison and description of representative individuals from different genetic lineages with goals of 1) finding diagnostic characters or metrics (e.g. genital structure) to distinguish them; 2) finding more informative traits (e.g. body size) that parallel to genetic divergence of *Acartia*. Advanced techniques such as scanning electronic microscopy (SEM), morphometrical analysis, and image analysis will have to be applied in order to attain the goals. The work

should be able to provide marine ecologists the key for identification of *Acartia* species so as to develop species-based research. It also assesses the relative role of morphological change in evolution of *Acartia.*

Fine genetic structure within lineage and global phylogeography

The mtDNA markers did a good job resolving recent diversification of lineage *F*, but provided low resolution for other lineages, and especially at smaller geographic scales (within an estuary). I could not quantify the fine population structure in local systems, like the Chesapeake Bay. It is not clear whether there is no structure for these populations or simply no power to detect the structure within lineages. The study using mtDNA and nITS markers mostly addresses the consequences of genetic drift or genealogical sorting after a long evolutionary history. It has poor power to show fine genetic structure associated with recent events in local systems. Fine genetic structure within lineages in relation to environment is important to provide a better picture about population differentiation of *Acartia*. To study fine genetic structure, we need hypervariable neutral genetic markers instead of mtDNA markers. Microsatellite markers (short tandem repeat sequences present in eukaryotic genomes) are powerful candidates for their high levels of length polymorphism and high mutation rates (Li *et al.* 2002; Zhang & Hewitt 2003; Edmands & Harrison 2003).

The present phylogeography study does not resolve origins of all deep *Acartia* lineages because of the limited geographic scale studied. To fully recover its evolutionary history, we need study the global phylogeography of *A. tonsa*, including Indian-Pacific, northwestern Atlantic, Gulf of Mexico, Northeastern Atlantic (European coast), and

South Atlantic regions as well. In the present study, I missed some genetic lineages reported in Caudill & Bucklin's (2004) study, including their rare haplotype group D from New England (Rhode Island and Massachusetts) and group P from Northeast Pacific region. I highly expect more novel genetic lineages from different geographic regions. Their phylogeographic relationship should help to reconstruct natural history of *A. tonsa*. Furthermore, additional gene loci should be considered in future studies of phylogeography, given the increasingly mounted controversies over neutrality of mtDNA and possible misrepresentation of population history (William *et al.* 1995; Zhang & Hewitt 1996; Bazin *et al.* 2006).

Characterization of niches and species boundary

In this study, salinity association and partitioned distribution of *Acartia* lineages along the estuarine gradient strongly suggested niche partitioning in *A. tonsa*. However, I was unable to describe a niche using all its associated ecological factors. It is essential to characterize and compare niches of different species for fully understanding how niches develop, diversify, and how species occupy or shift niches during early speciation. For a further study, more environmental data should be collected for describing fundamental niches (determining factors, niche breadth, and niche overlap between sympatric species) of different *Acartia* lineages. Multivariate analysis will be a powerful tool for screening key factors that distinguish fundamental niches.

Niche characterization also helps to delineate boundaries between sympatric species and possible hybrid zones. I did not find hybrids between *Acartia* lineages using the nuclear (nITS) marker, however, we may not rule out hybridization of largely

overlapped *Acartia* lineages because hybrids are easily missed if they are rare or restricted temporally or spatially (Barton & Hewitt 1985). More intensive sampling with guidance of hybrid zone will provide chances to test the existence of hybrids, i.e., whether reproduction isolation between lineages is complete.

In addition to sampling natural populations and testing their environmental associations, laboratory experiments are another important approach to discovery of lineage-specific niches and their diversification. Physiological tolerances to environmental factors, reproductive isolation and hybridization can be tested in controlled breeding and crossing experiments using both lab-reared populations and natural populations. Combining experimental data and observed natural patterns would provide more precise pictures of *Acartia* distribution in relation to its environment and deeper insights of its regulatory mechanisms.

APPENDICES

Appendix A

Nuclear ITS gene sequences and the alignment used for genetic analysis of *Acartia tonsa*

This appendix provides supplementary information on nITS sequences obtained from *A. tonsa* and its alignments. Twenty five alleles were recorded/inferred based on a 378 bp sequence fragment. The full alignment of 25 sequences was shown in Table A-1.

Table A-1 The full alignment of 378 bp nuclear ITS sequences of 25 alleles in *Acartia tonsa* . The alignment was performed by Clustal X program with default parameter settings and optimized manually. Lineage identity and frequency (freq., 2*n*) of each allele was listed. The most frequent alleles from five major lineages (*F1-3*, *F4*, *F5*, *S*, *X*) were used as references. Identical nucleotides compared to reference sequences were denoted by dots. Variable sites among reference sequences were marked in bold. Gaps were denoted by dashes. Four pseudo-alleles determined for one individual with unresolved heterozygous sites were also included in this alignment (FN01, FN02, FN16 and FN17; see text).

Alleles	Lineage/freq.(2*n*)	Sequence alignment: Nucleotide positions no.1–95
		000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
		000000000111111111122222222223333333333444444444455555555556666666666777777777788888888889999999
		123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
FN01	*F1-3* (24)	TTTCCGTAGGTGAACCTTCGGAAGGATCATATAGTAGAGGAAAGGAATGTCGGC----------TCATGGAAGCAAG
FN02	*F1-3* (14)	.....----------.....
FN03	*F1-3* (10)	.....A.....A----------.....
FN04	*F1-3* (9)	.....----------.....G.....
FN05	*F1-3* (8)	.....T..A.....A----------.....G.
FN06	*F1-3* (4)	.....T..A.....A----------.....G.
FN07	*F1-3* (3)	.....A.....A----------..T.....
FN08	*F1-3* (2)	.....----------...GGGA.....
FN09	*F1-3* (2)	.....----------...GGGA.....
FN10	*F1-3* (2)	.....T..A.....----------.....
FN11	*F1-3* (2)	.....T..A.....A----------.....G.
FN12	*F1-3* (1)	.....T..A.....A----------.....G.
FN13	*F1-3* (1)	.....----------.....TA.
FN14	*F1-3* (1)	.....T..A.....A----------.....G.
FN15	*F1-3* (1)	.....A.....A----------...AGGA....G.
FN16	*F1-3* (1)	.....----------.....G.....
FN17	*F1-3* (1)	.....----------.....G.....
FN18	*F4* (10)	TTTCCGTAGGTGAACCTTCGGAAGGATCATATAGAAGAGTAAAGTAAAGTCTGCTCTCATGGAAGAGTAAAGTAAAGTCGGCTCATGGAAGCAAG
FN19	*F4* (6)	.....T.A..G.....G..----------.....
FN20	*F4* (4)	.....T.A..G.....G..----------.....
FN21	*F5* (6)	TTTCCGTAGGTGAACCTTCGGAAGGATCATAAAGTC--GGAAAGGAAAGTCGGAA----------CCATGGTAGCAGG
SN01	*S* (68)	TTTCCGTAGGTGAACCTTCGGAAGGATCATATAGTA--GAAAGGTCGGCCCAAAGAATTCGCAGGCAA----------GGCTAGCGAATTCAA
SN02	*S* (14)	.....--.....C.....---------.....
XN01	*X* (63)	TTTCCGTAGGTGAACCTTCGGAAGGATCATAGTG----GAGTTGTATTGTCGGGAACAAGGAATCGTAGA----------GCTAGTCTTAACGA-
XN02	*X* (23)	.....----.....---------.....-

Table A-1 *Continued*

Alleles	Lineage/freq.(2*n*)	Sequence alignment: Nucleotide position nos.96–190
		000011111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
		999900000000001111111112222222222333333333344444444455555555556666666666777777777788888888889
		678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
FN01	*F1-3* (24)	CATATGTGC-AGCTACCATTAATCAGGTTGTGGGGACCTGCA-AT----AGCAAGCAAGCAAG----CTAGTAGGCAAGGGAACACTAGACTTGC
FN02	*F1-3* (14)	....C....-.....-..----.....----.....
FN03	*F1-3* (10)	....C....-.....C.....-.ATGCA...G.....----.....
FN04	*F1-3* (9)	....C....-.....-..----.....----.....
FN05	*F1-3* (8)	....C....-.....-..AGCA..G...T.....---...A.....
FN06	*F1-3* (4)	....C....-.....-..AGCA......T.....---...A.....
FN07	*F1-3* (3)	....C....-.....C.....-.ATGCA...G.....----.....
FN08	*F1-3* (2)	....C....-.....-..----.....----.....
FN09	*F1-3* (2)	....C....-.....-..-----.G.....----.....
FN10	*F1-3* (2)	..C.C....-.....C.....-..----.....----.....
FN11	*F1-3* (2)	....C....-.....-..AGCA...G..T.....---...A.....
FN12	*F1-3* (1)	....C....-.....-.A----.....T.....---...A.....
FN13	*F1-3* (1)	....C....-.....-..----.....----.....
FN14	*F1-3* (1)	....C....-.....-..AGAA.....T.....---...A.....
FN15	*F1-3* (1)	....C....-.....C.....-.ATGCA...G.....----.....
FN16	*F1-3* (1)	.....-.....-..----...G.....----.....
FN17	*F1-3* (1)	....C....-.....-..----...G.....----.....
FN18	*F4* (10)	CATTCGTGC-AGCTACCATTAATCAGGTTGTGGGGACCTGCA-AT----AGCAAGCAAGCAAG----CTAGTTGGCAAGGGAACACTAGACTTGC
FN19	*F4* (6)	.....-.....C.....-..----.....----.....
FN20	*F4* (4)	.....-.....-..----.....----.....
FN21	*F5* (6)	CATTCGTGC-AGCAACCATTAACCAGGC-AAGGGTGCCTGCC-GAAGCAAGCAAGCAAGCAAGTGAGCAAGTAGACAAGGGAACACTTGGCTTGC
SN01	*S* (68)	CAATCAAGCTAGCAAGCTTGCTATAGGGAGAGAGAGAGAGTGTG-TG----GGAAAGGCGGTAAG----CTAGTGGACAAGGGAACACTGGGCTTGT
SN02	*S* (14)	.....A...-..-----.TG.....C...----.....A.....A.....
XN01	*X* (63)	-ATCCATACCAGGCCCAAGGCGCCTGCCAAAGAGAAAGCGCA-GGTGAAGGCGCTAAAGCAAA---GCTAGTGGACAAGGGAACACTAGGCTTGT
XN02	*X* (23)	-.....A..A.....---.....A.....

Table A-1 *Continued*

Alleles	Lineage/freq.(2*n*)	Sequence alignment: Nucleotide position nos.191–285
		11111111122222222222222222222222222222222222222222222222222222222222222222222222222222222222222
		99999999900000000001111111111222222222233333333334444444444555555555566666666667777777777888888
		12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
FN01	*F1-3* (24)	AGCTTGAGGAAGATAGTAGCTAAATACTTGAGTGAATGTGAATCGCTGGTCACATGAAACGTAGCTGGGTTGAACGCAAG-GGTGAGAACCAAAG
FN02	*F1-3* (14)	....-..............
FN03	*F1-3* (10)	....T.-..............
FN04	*F1-3* (9)	....-..............
FN05	*F1-3* (8)	....-A.............
FN06	*F1-3* (4)	....-A.............
FN07	*F1-3* (3)	....T.-..............
FN08	*F1-3* (2)	....-..............
FN09	*F1-3* (2)	....A...........-..............
FN10	*F1-3* (2)	....-..............
FN11	*F1-3* (2)	....-A.............
FN12	*F1-3* (1)	....-A.............
FN13	*F1-3* (1)	....-..............
FN14	*F1-3* (1)	....-A.............
FN15	*F1-3* (1)	....T.-..............
FN16	*F1-3* (1)	....A...........-..............
FN17	*F1-3* (1)	....A...........-..............
FN18	*F4* (10)	AGCTTGAGGAAGATAGTAGCTAAATACTTGAGTGAATGTGAATCGCTGGTCACATGAAACGTAGCTGGGTTGAACGCAAG-GGTGAGAACCAAAG
FN19	*F4* (6)	....-..............
FN20	*F4* (4)	....-..............
FN21	*F5* (6)	GGCTTGAGGAAGATAGTAGCTAAATACTTGAGTGAATGTGAATCGCTGGTCACATGAAACATAGCTGGATTGAACGCAAGTGGTGGGAACCAAAG
SN01	*S* (68)	AGCTTGAAGAAGATAGTAGCTATATACTAGAGTGAATGTGAATCGCTGGTCACATGAA-CATAGCTGGATTGAACGCAAG-TGTAGGAACCAATG
SN02	*S* (14)	....-.....................-..............
XN01	*X* (63)	AGCTTGAAGAAGATAGTAGCTAAACACTAGAGTGAATGTGAATCGCTGGTCACATGAA-CATAGCTTGGTTGAACGCAAG-CGCAGGAAACAAAG
XN02	*X* (23)	....-.....................-..............

Table A-1 *Continued*

Alleles	Lineage/freq.(2*n*)	Sequence alignment: Nucleotide position nos.286–378
		222222222222223333333333333333333333333333333333333333333333333333333333333333333333333333333333333333333333
		888899999999990000000000111111111122222222223333333333444444444455555555556666666666777777777
		678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678
FN01	*F1-3* (24)	CCA--GGCTTTGGGTCTCAATTCTTTGTCAGGACATGCCTTTGGAGAAGGCAGTTAAGCCGTC--TTA-----GAACGGATCG-TTCTACCCT
FN02	*F1-3* (14)	...--.....--...-----.....-.....
FN03	*F1-3* (10)	...--.....C.....--..T-----.....-.....
FN04	*F1-3* (9)	...--.....--...-----.....-.....
FN05	*F1-3* (8)	...--.....C.....--...-----.....-.....
FN06	*F1-3* (4)	...--.....C.....--...-----.....-.....
FN07	*F1-3* (3)	...--.....C.....--..T-----.....-.....
FN08	*F1-3* (2)	...--.....--...-----.....-.....
FN09	*F1-3* (2)	...--.....--...-----.....-.....
FN10	*F1-3* (2)	...--.....C.....--...-----.....-.....
FN11	*F1-3* (2)	...--.....C.....--...-----.....-.....
FN12	*F1-3* (1)	...--.....C.....A.....--...-----.....-.....
FN13	*F1-3* (1)	...--.....--...-----.....-.....
FN14	*F1-3* (1)	...--.....C.....--...-----.....-.....
FN15	*F1-3* (1)	...--.....C.....--..T-----.....-.....
FN16	*F1-3* (1)	...--.....--...-----.....-.....
FN17	*F1-3* (1)	...--.....--...-----.....-.....
FN18	*F4* (10)	CCA--GTCTTTGGGTCTCAATTCCTTGTCAGGACATGCCTTTGGAGAAGGCAGTTAAGCCGTC--TTA-----GAACGGATCG-TTCTACACT
FN19	*F4* (6)	.A.--.....G.....G.AG...AGCCG.CTTA..AT.-.....
FN20	*F4* (4)	...--.....G.AG...AGCCG.CTTA..AT.-.....
FN21	*F5* (6)	CCA--GGCTTTGGTTTCCAATTCCTTGTCAGGACGTGCCTTGGGAGAAGGCAGTTGAACCGTT--GTAA----GAACGGCTCA-TTCTACCCT
SN01	*S* (68)	TCTTTGACTTTGGTTTCATATTCCTTGTCAAAACGAG-----GGCAGTGTGTGGGAGCCCGTCTGTTCAA---GGACAGATCAATTCCACCCT
SN02	*S* (14)	.....G.....-----.....---
XN01	*X* (63)	CAA--GGCTTTGGTTTCGTATTCCTTGTCAGAATGTGGGCAT-AAGGAGGCAAGGCACCCGTCCATA-----GGGCGGTTCT-TCCTACCCT
XN02	*X* (23)	...--.....-.....-----.....

Appendix B

Delineation of estuarine systems and estimation of physical parameters

To understand the interaction between copepod distribution and estuarine environment, we reviewed major estuarine systems involved in this study and described their basic environmental features. A big challenge to this work is the inconsistency of estuary delineations, parameter definitions and their estimation methods across literatures. Numerical values of a parameter for an estuary often varied substantially due to this inconsistency problem. In this appendix, I specify the details of estuary delineation, parameter definition, and estimation methods adopted in this study.

Delineation of estuarine systems, salinity zones and estimation of sizes

From the literature review, I found that Bricker *et al*. (2007)'s project "Effects of nutrient enrichment in the nation's estuaries" provided the best framework for my study. I followed most of its delineation of estuaries, salinity zone structures, and cited their major results for most of the estuarine systems I listed in Table 4-3. There are two types of estuary involved in this study: river estuary and lagoon estuary. According to Bricker *et al*. (2007), a river estuary is delineated from its river mouth opened to the ocean or large transient water body (e.g. sounds, lagoons) upstream to the upmost freshwater under tidal effect. Therefore a river estuary typically has three salinity zones: tidal fresh (0~0.5 PSU), mixing (0.5~25 PSU), and seawater (> 25 PSU). Lagoon estuaries, such as Long Island Sound, Albemarle Sound, Pamlico Sound, were delineated based on geographic morphology. Its estuarine length was defined as the riparian distance along

the salinity gradient from the freshest to the saltiest point, and width as the distance perpendicular to the estuarine length. Delineations of salinity zone were obtained from Bricker *et al*. (2007) and/or other literature sources.

To estimate size (length, width, depth, area, and volume) of an estuary lack of literature data, I used bathymetric data for the estuary from Coastal Relief Model provided by National Geographic Data Center. Coastal Relief Model was developed from multiple hydrographic data sources and provides 3-arc-second gridded bathymetric data set of the US Coastal Zone (*http://www.ngdc.noaa.gov/mgg/coastal/coastal.html*). Gridded data retrieved from NGDC was manually selected for a delineated estuary with the guide of NOAA/NOS medium resolution coastline (1:70,000) (http://rimmer.ngdc.noaa.gov/mgg/coast/getcoast.html). Estuary depth was estimated as an average of all data points. To estimate area and volume, each data point was treated as a 3 x 3 arc-second x depth cell and cell areas and volumes were summed up. To estimate estuary length, representative points along the river length were first selected, then distances between adjacent points were calculated and summed up for total length. To estimate estuary width, representative points from both banks were selected to calculate widths along the river length, then averaged for the overall width of an estuary. All distances between two points bearing longitude and latitude were calculated based on the sphere of law with the earth radius of 6371 km.

Tidal range

Tidal range refers to the tidal height difference between mean high water level and mean low water level. Values for major estuaries were cited from Bricker *et al*. (2007)

and other literatures. The author estimated tidal ranges of Thames estuary and Savannah estuary using daily tide data recorded at NOAA monitoring station nos. 8461490 (New London, CT) and 8670870 (Fort Pulaski, GA) during the period of 30 days before sampling date.

Freshwater input and flushing time

Freshwater inputs (FWI) to estuaries often fluctuate seasonally and inter-annually. To reflect the levels related to study period, averages of 30 days' data prior to sampling date were calculated using USGS stream water records at the representative stations for the estuaries. Table A-1 listed all stations used for FWI estimation. Flushing time is the time required to replace the existing freshwater in the estuary at the rate of freshwater input (Dyer 1997). Values for major estuaries were cited from Bricker *et al*. (2007) and other literatures. I approximately calculated the flushing time of Thames estuary using water and salt budget method for two layer exchange estuaries (Dyer 1997). Required surface and bottom salinity data were obtained from the monitoring station LTR (Lower Thames River) of MYSound Project (Tedesco *et al*. 2003) for a full-year time from Sept.1, 2003 to Aug.31, 2004.

Temperature and temperature duration for Acartia *development*

Temperature range of an estuary was estimated using daily temperature (surface, or average of water column where applicable) recorded at monitoring stations available for each estuary for a full year or multiple years. Same data sets were applied to estimate

temperature duration (number of days with no less than 20°C). Missing data in considered time series were interpolated using cubic spline method (Emery & Thompson 1997; Forsythe *et al*. 1977)

Table B-1 Environmental monitoring stations used in estimation of tidal range, freshwater input, temperature range and duration for estuaries. USGS data was retrieved from http://waterdata.usgs.gov. NOS and NDBC data were retrieved from http://seaboard.ndbc.noaa.gov. Abbreviations: CBP – Chesapeake Bay Program, HRECOS: Hudson River Environmental Conditions Observing System, NDBC – NOAA's National data buoy center, NOS – NOAA's National Ocean Service, USGS – United States Geological Survey.

	Stations for freshwater input estimates	Stations for temperature estimates
Long Island sound		
Thames Estuary	USGS: 01122500, 01127000, 01127500, 01123000	MYSound: LTR[a]
Connecticut R. Estuary	USGS: 01184000	NOS: 8465705
Long Island Sound	USGS: 01122500, 01127000, 01127500, 01123000, 01184000, 50120550	MYSound: 44022, 44040,44039, 44060[b]
Hudson-Raritan System		
Hudson R. Estuary	USGS: 01358000	HRECOS: Castle Point Buoy, GW Bridge[c]; NOS: 8518750, 8519483;
Entire system	USGS: 01358000, 01403060, 01389500, 01378500	HRECOS: Castle Point Buoy, GW Bridge[c]; NOS: 8518750, 8519483, 8531680
Chesapeake Bay		
Mainstem	-	CBP: CB1.0, CB1.1, CB2.2, CB3.1-3.2, CB4.1-4.4, CB5.1-5.5, CB6.1-6.4, CB7.1-7.4, CB8.1[d]
Choptank R. Tributary	USGS: 01491000	CBP: EE2.1, ET5.0, ET5.1, ET5.2[d]
Albemarle-Pamlico System		
Pamlico/Pungo Estuary	USGS: 02084000	USGS: 02084472, 0208455155, 0208453300
Neuse Estuary	USGS: 02091814	USGS: 02092162, 0209262905, 0209265810
Albemarle Sound	USGS: 02080500, 0208111310, 02053200, 02047000, 02052000, 02053500, 0204382800	NDBC-DUCN7
Pamlico Sound	USGS: 02084000, 02091814	FERRYMON: Cedar-Ocracoke Ferry route[e]
Savannah R. Estuary	USGS: 02198500	NOS: 8670870

a. Data was retrieved from http://sounddata.uconn.edu
b. Data was retrieved from NDBC http://seaboard.ndbc.noaa.gov
c. Data was retrieved from http://hudson.dl.stevens-tech.edu/hrecos
d. Data was retrieved from http://www.chesapeakebay.net/data_waterquality.aspx
e. Ferry-based monitoring route of Ferrymon project, University of North Carolina (http://www.unc.edu/ims/paerllab/research/ferry.mon/images/data.

LIST OF REFERENCES

Ahel M, Barlow RG, Mantoura RFC (1996) Effect of salinity gradients on the distribution of phytoplankton pigments in a stratified estuary. *Marine Ecology Progress Series* **143**, 289–295.

Antunes A, Ramos MJ (2005) Discovery of a large number of previously unrecognized mitochondrial pseudogenes in fish genomes. *Genomics* **86**, 708–717.

Atkinson D (1994) Temperature and organism size: a biological law for ectotherms. *Advances in Ecological Research* **25**, 1–58.

Avise JC (2000) *Phylogeography: The History and Formation of Species.* Harvard University Press, Cambridge, Massachusetts.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution.* 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Avise JC, Ball RM (1990) Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Survey of Evolutionary Biology* **7**, 45–67.

Baker P, Austin JD, Bowen BW, Baker SM (2008) Range–wide population structure and history of the northern quahog (*Merceneria merceneria*) inferred from mitochondrial DNA sequence data. *Ices Journal of Marine Science* **65**, 155–163.

Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82**, 247–277.

Ballard JWO, Rand DM (2005) The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology Evolution and Systematics* **36**, 621–642.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16**, 113–148.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. *Monographs in Systematic Botany from the Missouri Botanical Garden* **53**, 289–303.

Bazin E, Glemin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570–572.

Bazinet AL, Neel MC, Shaw KL, Cummings MP (2008) The genealogical sorting index: Software and website for quantifying the exclusivity of lineages. *Http://www.genealogicalSorting.org*.

Beaugrand G (2004) The North Sea regime shift: Evidence, causes, mechanisms and consequences. *Progress in Oceanography* **60**, 245–262.

Beaugrand G, Ibanez F, Lindley JA, Reid PC (2002) Diversity of calanoid copepods in the North Atlantic and adjacent seas: Species associations and biogeography. *Marine Ecology Progress Series* **232**, 179–195.

Bekkevold D, Andre C, Dahlgren TG, Clausen LAW, Torstensen E, Mosegaard H, Carvalho GR, Christensen TB, Norlinder E, Ruzzante DE (2005) Environmental correlates of population differentiation in Atlantic herring. *Evolution* **59**, 2656–2668.

Bergmann C (1847) Über die Verhältnisse der Wärmeökonomie der Thiere zu ihrer Grösse. *Göttinger Studien* **3**, 595–708.

Bermingham E, Moritz C (1998) Comparative phylogeography: Concepts and applications. *Molecular Ecology* **7**, 367–369.

Bernardi G, Sordino P, Powers DA (1993) Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 9271–9274.

Bertness MD, Gaines SD (1993) Larval dispersal and local adaptation in acorn barnacles. *Evolution* **47**, 316–320.

Bierne N, Bonhomme F, David P (2003) Habitat preference and the marine speciation paradox. *Proceedings of the Royal Society B* **270**, 1399–1406.

Bilton DT, Paula J, Bishop JDD (2002) Dispersal, genetic differentiation and speciation in estuarine organisms. *Estuarine Coastal and Shelf Science* **55**, 937–952.

Blackburn TM, Gaston KJ (1994) Animal body size distributions: Patterns, mechanisms and implications. *Trends in Ecology and Evolution* **9**, 471–474.

Blanckenhorn WU (2000) The evolution of body size: What keeps organisms small? *Quarterly Review of Biology* **75**, 385–407.

Bonnet D, Frid C (2004) Seven copepod species considered as indicators of water mass influence and changes: Results from a Northumberland coastal station. *Ices Journal of Marine Science* **61**, 485–491.

Bradbury IR, Laurel B, Snelgrove PVR, Bentzen P, Campana SE (2008) Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. *Proceedings of the Royal Society B* **275**, 1803–1809.

Bradford-Grieve JM (1999) *ICES Identification Leaflets for Plankton. No. 181 Copepoda: Calanoida: Acartiidae: Acartia, Paracartia, Pteriacartia.* International Council for the Exploration of the Sea, Copenhagen, Denmark.

Bricker S, Longstaff B, Dennison W, Jones A, Boicourt K, Wicks C, Woemer J (2007) Effects of nutrient enrichment in the nation's estuaries: a decade of change, National estuarine eutrophication update. NOAA Coastal Ocean Program Decision Analysis Series No. 26. National Centers for Coastal Ocean Service, Silver Spring, Maryland.

Briggs JC (1974) *Marine Zoogeography.* McGraw-Hill, New York.

Briggs JC (1995) *Global Biogeography.* Elsevier, Amsterdam ; New York.

Briggs JC (2007) Marine longitudinal biodiversity: Causes and conservation. *Diversity and Distributions* **13**, 544–555.

Brownlee DC, Jacobs F (1987) Mesozooplankton and microzooplankton in the Chesapeake Bay. In: *Contaminant Problems and Management of Living Chesapeake Bay Resources* (eds. Majumdar SK, Hall LW, Austin CM), pp. 217–269. The Pennsylvania Academy of Science, Philadelphia, Pennsylvania.

Bucklin A, Frost BW, Bradford-Grieve J, Allen LD, Copley NJ (2003) Molecular systematic and phylogenetic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. *Marine Biology* **142**, 333–343.

Bucklin A, LaJeunesse TC, Curry E, Wallinga J, Garrison K (1996) Molecular diversity of the copepod, *Nannocalanus minor*: Genetic evidence of species and population structure in the North Atlantic Ocean. *Journal of Marine Research* **54**, 285–310.

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* **325**, 31–36.

Carlton JT, Geller JB (1993) Ecological roulette: the global transport of nonindigenous marine organisms. *Science* **261**, 78–82.

Castric V, Bonney F, Bernatchez L (2001) Landscape structure and hierarchical genetic diversity in the brook charr, *Salvelinus fontinalis*. *Evolution* **55**, 1016–1028.

Castro-Longoria E (2003) Egg production and hatching success of four *Acartia* species under different temperature and salinity regimes. *Journal of Crustacean Biology* **23**, 289–299.

Caudill CC, Bucklin A (2004) Molecular phylogeography and evolutionary history of the estuarine copepod, *Acartia tonsa*, on the Northwest Atlantic coast. *Hydrobiologia* **511**, 91–102.

Chen G, Hare MP (2008) Cryptic ecological diversification of a planktonic estuarine copepod, *Acartia tonsa*. *Molecular Ecology* **17**, 1451–1468.

Clark AG (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution* **7**, 111–122.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657–1659.

Colborn J, Crabtree RE, Shaklee JB, Pfeiler E, Bowen BW (2001) The evolutionary enigma of bonefishes (*Albula* spp.): Cryptic species and ancient separations in a globally distributed shorefish. *Evolution* **55**, 807–820.

Coleman AW, Vacquier VD (2002) Exploring the phylogenetic utility of ITS sequences for animals: A test case for abalone (*Haliotis*). *Journal of Molecular Evolution* **54**, 246–257.

Cordell JR, Morrison SM (1996) The invasive Asian copepod *Pseudodiaptomus inopinus* in Oregon, Washington, and British Columbia estuaries. *Estuaries* **19**, 629–638.

Coyne JA, Orr HA (2004) *Speciation.* Sinauer Associates, Sunderland, Massachusetts.

Cronin TM (1988) Evolution of marine climates in the U. S. Atantic coast during the past four million years. In: *The Past Three Million Years: Evolution of Climate Variability in the North Atantic Region* (eds. Shackleton NJ, West RG, Owens DQ), pp. 327–356. The Royal Society, London, United Kingdom.

Cuker BE, Watson NA (2002) Diel vertical migration of zooplankton in contrasting habitats of the Chesapeake Bay. *Estuaries* **25**, 296–307.

Cummings MP, Neel MC, Shaw KL (2008) A genealogical approach to quantifying lineage divergence. *Evolution* **62**, 2411–2422.

Dalebout ML, Ruzzante DE, Whitehead H, Oien NI (2006) Nuclear and mitochondrial markers reveal distinctiveness of a small population of bottlenose whales (*Hyperoodon ampullatus*) in the western North Atlantic. *Molecular Ecology* **15**, 3115–3129.

Darling KF, Kucera M, Wade CM (2007) Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5002–5007.

Darling KF, Wade CM, Stewart IA, Kroon D, Dingle R, Brown AJL (2000) Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature* **405**, 43–47.

David V, Sautour B, Chardy P (2007) Successful colonization of the calanoid copepod *Acartia tonsa* in the oligo-mesohaline area of the Gironde estuary (SW France): Natural or anthropogenic forcing? *Estuarine Coastal and Shelf Science* **71**, 429–442.

Dawson MN, Jacobs DK (2001) Molecular evidence for cryptic species of *Aurelia aurita* (Cnidaria, Scyphozoa). *Biological Bulletin* **200**, 92–96.

Day JW, Hall CAS, Kemp WM, Yanez-Arancibia A (1989) *Estuarine Ecology*. John Wiley and Sons, New York.

de Vargas C, Norris R, Zaninetti L, Gibb SW, Pawlowski J (1999) Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2864–2868.

de Vargas C, Renaud S, Hilbrecht H, Pawlowski J (2001) Pleistocene adaptive radiation in *Globorotalia truncatulinoides*: genetic, morphologic, and environmental evidence. *Paleobiology* **27**, 104–125.

Dodson SI, Grishanin AK, Gross K, Wyngaard GA (2003) Morphological analysis of some cryptic species in the *Acanthocyclops vernalis* species complex from North America. *Hydrobiologia* **500**, 131–143.

Dover G (1982) Molecular drive: A cohesive mode of species evolution. *Nature* **299**, 111–117.

Durbin EG, Durbin AG, Smayda TJ, Verity PG (1983) Food limitation of production by adult *Acartia tonsa* in Narragansett Bay, Rhode Island. *Limnology and Oceanography* **28**, 1199–1213.

Ederington MC, Mcmanus GB, Harvey HR (1995) Trophic transfer of fatty acids, sterols, and a triterpenoid alcohol between bacteria, a ciliate, and the copepod *Acartia tonsa*. *Limnology and Oceanography* **40**, 860–867.

Edmands S, Harrison JS (2003) Molecular and quantitative trait variation within and among populations of the intertidal copepod *Tigriopus californicus*. *Evolution* **57**, 2277–2285.

Endler JA (1977) *Geographic Variation, Speciation, and Clines.* Princeton University Press, Princeton, New Jersey.

Endler JA (1989) Conceptual and other problems in speciation. In: *Speciation and its Consequences* (eds. Otte D, Endler JA), pp. 625–648. Sinauer Associates, Sunderland, Massachusetts.

Engle VD, Summers JK (1999) Latitudinal gradients in benthic community composition in western Atlantic estuaries. *Journal of Biogeography* **26**, 1007–1023.

Enrique Carrillo BG, Miller CB, Wiebe PH (1974) Failure of interbreeding between Atlantic and Pacific populations of marine calanoid copepod *Acartia clausi* Giesbrecht. *Limnology and Oceanography* **19**, 452–458.

Epifanio CE, Garvine RW (2001) Larval transport on the Atlantic continental shelf of North America: A review. *Estuarine Coastal and Shelf Science* **52**, 51–77.

Ewens WJ (1979) *Mathematical Population Genetics.* Springer-Verlag, Berlin; New York.

Excoffier L, Larval G, Schneider S (2005) Arlequin version 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Fenchel T (2005) Cosmopolitan microbes and their 'cryptic' species. *Aquatic Microbial Ecology* **41**, 49–54.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294–299.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

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

Garmew TG, Hammond S, Mercantini A, Morgan J, Neunert C, Fornshell JA (1994) Morphological variability of geographically distinct populations of the estuarine copepod *Acartia tonsa*. *Hydrobiologia* **293**, 149–156.

Garvine RW (1975) Distribution of salinity and temperature in Connecticut River estuary. *Journal of Geophysical Research* **80**, 1176–1183.

Gaudy R, Verriopoulos G (2004) Spatial and seasonal variations in size, body volume and body proportion (prosome : urosome ratio) of the copepod *Acartia tonsa* in a semi-closed ecosystem (Berre Lagoon, western Mediterranean). *Hydrobiologia* **513**, 219–229.

Gauze GF (1934) *The Struggle for Existence.* The Williams & Wilkins company, Baltimore, Maryland.

Geider RJ (1987) Light and temperature-dependence of the carbon to chlorophyll *a* ratio in microalgae and cyanobacteria: Implications for physiology and growth of phytoplankton. *New Phytologist* **106**, 1–34.

Gerbi SA (1985) Evolution of ribosomal DNA. In: *Molecular Evolutioanry Genetics* (ed. MacIntyre RJ), pp. 419–517. Plenum, New York.

Gillespie MC (1971) Analysis and treatment of zooplankton of estuarine waters of Louisiana. In: *Cooperative Gulf of Mexico Estuarine Inventory and Study, Louisiana. Phase IV. Biology*, pp. 108–175. Louisiana Wildlife and Fisheries Commission, New Orleans, Louisiana.

Goetze E (2003) Cryptic speciation on the high seas: Global phylogenetics of the copepod family Eucalanidae. *Proceedings of the Royal Society of London B* **270**, 2321–2331.

Goetze E (2005) Global population genetic structure and biogeography of the oceanic copepods *Eucalanus hyalinus* and *E. spinifer*. *Evolution* **59**, 2378–2398.

Gomez A, Serra M, Carvalho GR, Lunt DH (2002) Speciation in ancient cryptic species complexes: Evidence from the molecular phylogeny of *Brachionus plicatilis* (Rotifera). *Evolution* **56**, 1431–1444.

Gomez S, Fleeger JW, Rocha-Olivares A, Foltz D (2004) Four new species of *Cletocamptus Schmankewitsch*, 1875, closely related to *Cletocamptus deitersi* (Richard, 1897) (Copepoda : Harpacticoida). *Journal of Natural History* **38**, 2669–2732.

Gonzalez JG (1974) Critical thermal maxima and upper lethal temperatures for calanoid copepods *Acartia tonsa* and *A. clausi*. *Marine Biology* **27**, 219–223.

Gosner KL (1971) *Guide to Identification of Marine and Estuarine Invertebrates: Cape Hatteras to the Bay of Fundy.* Wiley–Interscience, New York.

Grice GD, Marcus NH (1981) Dormant eggs of marine copepods. *Oceanography and Marine Biology Annual Review* **19**, 125–140.

Hardin G (1960) Competitive exclusion principle. *Science* **131**, 1292–1297.

Hare MP, Avise JC (1998) Population structure in the American oyster as inferred by nuclear gene genealogies. *Molecular Biology and Evolution* **15**, 119–128.

Hare MP, Palumbi SR (1999) The accuracy of heterozygous base calling from diploid sequence and resolution of haplotypes using allele-specific sequencing. *Molecular Ecology* **8**, 1750–1752.

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution* **4**, 6–11.

Hayden BP, Dolan R (1976) Coastal marine fauna and marine climates of Americas. *Journal of Biogeography* **3**, 71–81.

Hays GC (2003) A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations. *Hydrobiologia* **503**, 163–170.

Hebert PDN (1998) Variable environments and evolutionary diversification in inland waters. In: *Advances in Molecular Ecology* (ed. Carvalho GR), pp. 267–290. IOS Press, Amsterdam, Netherlands.

Hedgpeth JW (1957) Marine Biogeography. In: *Treatise on Marine Ecology and Paleoecology* (ed. Hedgpeth JW), pp. 359–382. Memoir 67 of the Geological Survey of America.

Heinle DR (1966) Production of a calanoid copepod, *Acartia tonsa*, in the Patuxent River estuary. *Chesapeake Science* **7**, 59–74.

Herke SW, Foltz DW (2002) Phylogeography of two squid (*Loligo pealei* and *L. plei*) in the Gulf of Mexico and northwestern Atlantic Ocean. *Marine Biology* **140**, 103–115.

Hershkovitz MA, Zimmer EA, Hahn WJ (1999) Ribosomal DNA sequences and angiosperm systematics. In: *Molecular Systematics and Plant Evolution* (eds. Hollingsworth PM, Bateman RM, Gornall RJ), pp. 268–326. Taylor and Francis, London, United Kingdom.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography or seeing genes in space and time. *Molecular Ecology* **10**, 537–549.

Hey J (2001) The mind of the species problem. *Trends in Ecology and Evolution* **16**, 326–329.

Hilbish TJ, Bayne BL, Day A (1994) Genetics of physiological differentiation within the marine mussel genus *Mytilus*. *Evolution* **48**, 267–286.

Hill RS (2004) *Genetic Diversity and Structure of Calanoid Copepods: Molecular Evolutionary Patterns in Coastal Estuaries (*Acartia tonsa*) and the open ocean (*Calanus *spp.).* PhD dissertation, University of New Hampshire.

Hillis DM, Dixon MT (1991) Ribosomal DNA: Molecular evolution and phylogenetic inference. *Quarterly Review of Biology* **66**, 411–453.

Hirst AG, Sheader M, Williams JA (1999) Annual pattern of calanoid copepod abundance, prosome length and minor role in pelagic carbon flux in the Solent, UK. *Marine Ecology Progress Series* **177**, 133–146.

Ho J, Perkins PS (1985) Symbionts of marine copepoda: An overview. *Bulletin of Marine Science* **37**, 586–598.

Hobbs CH (2004) Geological history of Chesapeake Bay, USA. *Quaternary Science Reviews* **23**, 641–661.

Hoffmeyer MS, Figueroa P (1997) Integumental structures in the oral field of *Eurytemora affinis* and *Acartia tonsa* (Copepoda, Calanoida) in relation to their trophic habits. *Crustaceana* **70**, 257–271.

Hough AR, Naylor E (1991) Field studies on retention of the planktonic copepod *Eurytemora affinis* in a mixed estuary. *Marine Ecology Progress Series* **76**, 115–122.

Hudson RR (1990) Gene genealogy and the coalescent process. *Oxford Survey of Evolutionary Biology* **7**, 1–44.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164.

Hutchinson GE (1957) Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology* **22**, 415–427.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* **53**, 1898–1914.

Jacobs J (1961) Laboratory cultivation of the marine copepod *Pseudodiaptomus coronatus* Williams. *Limnology and Oceanography* **6**, 443–446.

Jeffries HP (1962) Succession of two *Acartia* species in estuaries. *Limnology and Oceanography* **7**, 354–364.

Johnson WS, Allen DM (2005) *Zooplankton of the Atlantic and Gulf coasts : A Guide to their Identification and Ecology*. Johns Hopkins University Press, Baltimore, Maryland.

Jonsson PR, Tiselius P (1990) Feeding behavior, prey detection and capture efficiency of the copepod *Acartia tonsa* feeding on planktonic ciliates. *Marine Ecology Progress Series* **60**, 35–44.

Kalinowski ST (2004) Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics* **5**, 539–543.

Kendall MA, Aschan M (1993) Latitudinal Gradients in the structure of macrobenthic communities: A comparison of Arctic, temperate and tropical sites. *Journal of Experimental Marine Biology and Ecology* **172**, 157–169.

Kimmel DG, Roman MR (2004) Long-term trends in mesozooplankton abundance in Chesapeake Bay, USA: Influence of freshwater input. *Marine Ecology Progress Series* **267**, 71–83.

Kimmerer WJ, Mckinnon AD (1987) Zooplankton in a marine bay. 2. Vertical migration to maintain horizontal distributions. *Marine Ecology Progress Series* **41**, 53–60.

Kingman JFC (1982) The coalescent. *Stochastic Processes and their Applications* **13**, 235–248.

Kinlan BP, Gaines SD, Lester SE (2005) Propagule dispersal and the scales of marine community process. *Diversity and Distributions* **11**, 139–148.

Kleppel GS, Burkart CA (1995) Egg production and the nutritional environment of *Acartia tonsa*: The role of food quality in copepod nutrition. *Ices Journal of Marine Science* **52**, 297–304.

Kleppel GS, Burkart CA, Houchin L (1998) Nutrition and the regulation of egg production in the calanoid copepod *Acartia tonsa*. *Limnology and Oceanography* **43**, 1000–1007.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology* **11**, 2623–2635.

Knowlton N (1993) Sibling species in the sea. *Annual Review of Ecology and Systematics* **24**, 189–216.

Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London B* **265**, 2257–2263.

Knowlton N, Weigt LA, Solorzano LA, Mills DK, Bermingham E (1993) Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. *Science* **260**, 1629–1632.

Kobari T, Tadokoro K, Shiomoto A, Hashimoto S (2003) Geographical variations in prosome length and body weight of *Neocalanus* copepods in the North Pacific. *Journal of Oceanography* **59**, 3–10.

Koehn RK, Newell RIE, Immermann F (1980) Maintenance of an aminopeptidase allele frequency cline by natural selection. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 5385–5389.

Koufopanou V, Burt A, Taylor JW (1997) Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 5478–5482.

Kruse I, Reise K (2003) Reproductive isolation between intertidal and subtidal *Scoloplos armiger* (Polychaeta, Orbiniidae) indicates sibling species in the North Sea. *Marine Biology* **143**, 511–517.

Kuehl RO (2000) *Design of Experiments: Statistical Principles of Research Design and Analysis.* 2nd edn. Duxbury/Thomson Learning, Pacific Grove, California.

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park, Pennsylvania.

Lakkis S (1994) Coexistence and competition within *Acartia* (Copepoda, Calanoida) congeners from Lebanese coastal water: Niche overlap measurements. *Hydrobiologia* **293**, 481–490.

Lalli CM, Parsons TR (1993) *Biological Oceanography: An Introduction.* 1st edn. Pergamon Press, New York.

Lambeck K, Chappell J (2001) Sea level change through the last glacial cycle. *Science* **292**, 679–686.

Lance J (1964) The salinity tolerances of some estuarine planktonic crustaceans. *Biological Bulletin* **127**, 108–118.

Lavoie DM, Smith LD, Ruiz GM (1999) The potential for intracoastal transfer of non-indigenous species in the ballast water of ships. *Estuarine Coastal and Shelf Science* **48**, 551–564.

Lazoski C, Sole-Cava AM, Boury-Esnault N, Klautau M, Russo CAM (2001) Cryptic speciation in a high gene flow scenario in the oviparous marine sponge *Chondrosia reniformis*. *Marine Biology* **139**, 421–429.

Lee CE (1999) Independent invasions of fresh water: Comparison of sodium pumping ability among lineages of the copepod *Eurytemora affinis*. *American Zoologist* **39**, 17a.

Lee CE (2000) Global phylogeography of a cryptic copepod species complex and reproductive isolation between genetically proximate "populations". *Evolution* **54**, 2014–2027.

Lee CE, Frost BW (2002) Morphological stasis in the *Eurytemora affinis* species complex (Copepoda: Temoridae). *Hydrobiologia* **480**, 111–128.

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites genomic distribution, putative functions and mutational mechanisms: A review. *Molecular Ecology* **11**, 2453–2465.

Lin J, Xie L, Pietrafesa LJ, Ramus JS, Paerl HW (2007) Water quality gradients across Albemarle-Pamlico estuarine system: Seasonal variations and model applications. *Journal of Coastal Research* **23**, 213–229.

Llewellyn CA, Fishwick JR, Blackford JC (2005) Phytoplankton community assemblage in the English Channel: A comparison using chlorophyll *a* derived from HPLC-CHEMTAX and carbon derived from microscopy cell counts. *Journal of Plankton Research* **27**, 103–119.

Lloyd SS (2006) *Zooplankton Ecology in the Chesapeake Bay Estuarine Turbidity Maximum, with Emphasis on the Calanoid Copepod* Eurytemora affinis. PhD Dissertation, University of Maryland, College Park.

Marcus NH (1996) Ecological and evolutionary significance of resting eggs in marine copepods: Past, present, and future studies. *Hydrobiologia* **320**, 141–152.

Marko PB (2004) 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Molecular Ecology* **13**, 597–611.

Mauchline J (1998) The biology of calanoid copepods. *Advances in Marine Biology* **33**, 1–710.

Mayr E (1942) *Systematics and the Origin of Species.* Columbia University Press, New York.

McAlice BJ (1981) On the post-glacial history of *Acartia tonsa* (Copepoda, Calanoida) in the Gulf of Maine and the Gulf of St. Lawrence. *Marine Biology* **64**, 267–272.

McMillen-Jackson AL, Bert TM (2004) Mitochondrial DNA variation and population genetic structure of the blue crab *Callinectes sapidus* in the eastern United States. *Marine Biology* **145**, 769–777.

Moore WS (1995) Inferring phylogenies from mtDNA variation: Mitochondrial gene trees versus nuclear gene trees. *Evolution* **49**, 718–726.

Moritz C, Hillis DM (1996) Molecular systematics: Context and controversies. In: *Molecular Systematics* (eds. Hillis DM, Moritz C, Mable BK), pp. 1–16. Sinauer Associates, Sunderland, Massachusetts.

Muylaert K, Sabbe K, Vyverman W (2000) Spatial and temporal dynamics of phytoplankton communities in a freshwater tidal estuary (Schelde, Belgium). *Estuarine Coastal and Shelf Science* **50**, 673–687.

Naylor E (2006) Orientation and navigation in coastal and estuarine zooplankton. *Marine and Freshwater Behaviour and Physiology* **39**, 13–24.

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

Nguyen TTT, Murphy NP, Austin CM (2002) Amplification of multiple copies of mitochondrial Cytochrome *b* gene fragments in the Australian freshwater crayfish, *Cherax destructor* Clark (Parastacidae: Decapoda). *Animal Genetics* **33**, 304–308.

Norris RD (1999) Hydrographic and tectonic control of plankton distribution and evolution. In: *Reconstructing Ocean History: A Window into the Future* (eds. Abrantes F, Mix A), pp. 173–193. Kluwer/Plenum, New York.

Ortells R, Gomez A, Serra M (2003) Coexistence of cryptic rotifer species: ecological and genetic characterisation of *Brachionus plicatilis*. *Freshwater Biology* **48**, 2194–2202.

Paerl HW, Bales JD, Ausley LW, Buzzelli CP, Crowder LB, Eby LA, Fear JM, Go M, Peierls BL, Richardson TL, Ramus JS (2001) Ecosystem impacts of three sequential hurricanes (Dennis, Floyd, and Irene) on the United States' largest lagoonal estuary, Pamlico Sound, NC. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5655–5660.

Paffenhöfer GA (1991) Some characteristics of abundant subtropical copepods in estuarine, shelf and oceanic waters. *Bulletin of Plankton Society of Japan* **Special Volume**, 201–216.

Paffenhöfer GA, Stearns DE (1988) Why is *Acartia tonsa* (Copepoda, Calanoida) restricted to nearshore environments. *Marine Ecology Progress Series* **42**, 33–38.

Palumbi SR (1992) Marine speciation on a small planet. *Trends in Ecology and Evolution* **7**, 114–118.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* **25**, 547–572.

Palumbi SR, Cipriano F, Hare MP (2001) Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. *Evolution* **55**, 859–868.

Palumbi SR, Martin A, Romano S, McMillan WO, Stice L, Grabowski G (1996) *The Simple Fool's Guide to PCR.* Kewalo Marine Laboratory and Unviersity of Hawaii, Honolulu, Hawaii.

Parrish KK, Wilson DF (1978) Fecundity studies on *Acartia tonsa* (Copepoda: Calanoida) in standardized culture. *Marine Biology* **46**, 65–81.

Peijnenburg KTCA, Breeuwer JAJ, Pierrot-Bults AC, Menken SBJ (2004) Phylogeography of the planktonic chaetognath *Sagitta setosa* reveals isolation in European seas. *Evolution* **58**, 1472–1487.

Perry RI, Smith SJ (1994) Identifying habitat associations of marine fishes using survey data: An application to the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 589–602.

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**, 844–855.

Quinlan EL, Phlips EJ (2007) Phytoplankton assemblages across the marine to low-salinity transition zone in a blackwater dominated estuary. *Journal of Plankton Research* **29**, 401–416.

R Development Core Team (2008) R: A language and envrionment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raymond M, Rousset F (1995a) An exact test for population differentiation. *Evolution* **49**, 1280–1283.

Raymond M, Rousset F (1995b) Genepop (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–249.

Reaugh ML (2005) *The Effects of Freshwater Flow and Grazing on the Plankton Community Structure of Chesapeake Bay Tributaries.* PhD Dissertation, University of Maryland, College Park.

Reeb CA, Avise JC (1990) A genetic discontinuity in a continuously distributed species: Mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics* **124**, 397–406.

Rising JD, Avise JC (1993) Application of genealogical concordance principles to the taxonomy and evolutionary history of the sharp tailed sparrow (*Ammodramus caudacutus*). *Auk* **110**, 844–856.

Robinson-Rechavi M, Huchon D (2000) RRTree: Relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**, 296–297.

Robinson M, Gouy M, Gautier C, Mouchiroud D (1998) Sensitivity of the relative-rate test to taxonomic sampling. *Molecular Biology and Evolution* **15**, 1091–1098.

Rocha-Olivares A, Fleeger JW, Foltz DW (2001) Decoupling of molecular and morphological evolution in deep lineages of a meiobenthic harpacticoid copepod. *Molecular Biology and Evolution* **18**, 1088–1102.

Rocha–Olivares A, Fleeger JW, Foltz DW (2004) Differential tolerance among cryptic species: A potential cause of pollutant-related reductions in genetic diversity. *Environmental Toxicology and Chemistry* **23**, 2132–2137.

Roman MR (1984) Utilization of detritus by the copepod, *Acartia tonsa*. *Limnology and Oceanography* **29**, 949–959.

Rosel PE, Dizon AE, Heyning JE (1994) Genetic analysis of sympatric morphotypes of common dolphins (Genus *Delphinus*). *Marine Biology* **119**, 159–167.

Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3**, 380–390.

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

Roy K (2008) Evolution: Dynamics of body size evolution. *Science* **321**, 1451–1452.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497.

Ruiz GM, Fofonoff PW, Carlton JT, Wonham MJ, Hines AH (2000) Invasion of coastal marine communities in North America: Apparent patterns, processes, and biases. *Annual Review of Ecology and Systematics* **31**, 481–531.

Ruzzante DE, Taggart CT, Cook C, Goddard S (1996) Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland: Microsatellite DNA variation and antifreeze level. *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 634–645.

Ruzzante DE, Taggart CT, Cook D (1999) A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St. Lawrence. *Fisheries Research* **43**, 79–97.

Sabatini ME (1990) The developmental stages (Copepodid I to Copepodid VI) of *Acartia tonsa* Dana, 1849 (Copepoda, Calanoida). *Crustaceana* **59**, 53–61.

Saez AG, Probert I, Geisen M, Quinn P, Young JR, Medlin LK (2003) Pseudo-cryptic speciation in coccolithophores. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 7163–7168.

Saunders NC, Kessler LG, Avise JC (1986) Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, *Limulus polyphemus*. *Genetics* **112**, 613–627.

Scheltema RS (1975) Relationship of larval dispersal, gene flow and natural selection to geographic variation of benthic invertebrates in estuaries and along coastal regions. In: *Estuarine Research Volume 1. Chemistry, Biology, and the Estuary System* (ed. Cornin LE), pp. 372–391. Academic Press, New York.

Schizas NV, Chandler GT, Coull BC, Klosterhaus SL, Quattro JM (2001) Differential survival of three mitochondrial lineages of a marine benthic copepod exposed to a pesticide mixture. *Environmental Science & Technology* **35**, 535–538.

Schizas NV, Street GT, Coull BC, Chandler GT, Quattro JM (1999) Molecular population structure of the marine benthic copepod *Microarthridion littorale* along the southeastern and Gulf coasts of the USA. *Marine Biology* **135**, 399–405.

Schmidt PS, Rand DM (2001) Adaptive maintenance of genetic polymorphism in an intertidal barnacle: Habitat- and life-stage-specific survivorship of Mpi genotypes. *Evolution* **55**, 1336–1344.

Schmitz J, Piskurek O, Zischler H (2005) Forty million years of independent evolution: A mitochondrial gene and its corresponding nuclear pseudogene in primates. *Journal of Molecular Evolution* **61**, 1–11.

Schroth W, Jarms G, Streit B, Schierwater B (2002) Speciation and phylogeography in the cosmopolitan marine moon jelly, *Aurelia* sp.. *BMC Evolutionary Biology*, 1.

Schubart CD, Diesel R, Hedges SB (1998) Rapid evolution to terrestrial life in Jamaican crabs. *Nature* **393**, 363–365.

Schubel JR, Pritchard DW (1987) A brief physical description of the Chesapeake Bay. In: *Contaminant Problems and Management of Living Chesapeake Bay Resources* (eds. Majumdar SK, Hall LW, Austin HM), pp. 1–32. The Pennsylvania Academy of Science, Philadelphia, Pennsylvania.

Schulenburg JHGV, Englisch U, Wagele JW (1999) Evolution of ITS1 rDNA in the Digenea (Platyhelminthes: Trematoda): 3' end sequence conservation and its phylogenetic utility. *Journal of Molecular Evolution* **48**, 2–12.

Sellner KG (1987) Phytoplankton in the Chesapeake Bay: Role in carbon, oxygen and nutrient dynamics. In: *Contaminant Problems and Management of Living Chesapeake Bay Resources* (eds. Majumdar SK, Hall LW, Austin HM), pp. 134–157. The Pennsylvania Academic of Science, Philadelphia, Pennsylvania.

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49**, 369–381.

Simonsen KL, Churchill GA, Aquadro CF (1995) Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**, 413–429.

Sin Y, Wetzel RL, Anderson IC (2000) Seasonal variations of size-fractionated phytoplankton along the salinity gradient in the York River estuary, Virginia (USA). *Journal of Plankton Research* **22**, 1945–1960.

Šlapeta J, López–García P, Moreira D (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Molecular Biology and Evolution* **23**, 23–29.

Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **75**, 733–756.

Sponer R, Roy MS (2002) Phylogeographic analysis of the brooding brittle star *Amphipholis squamata* (Echinodermata) along the coast of New Zealand reveals high cryptic genetic variation and cryptic dispersal potential. *Evolution* **56**, 1954–1967.

Stearns SC, Koella JC (1986) The evolution of phenotypic plasticity in life-history traits: Predictions of reaction norms for age and size at maturity. *Evolution* **40**, 893–913.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**, 978–989.

Sullivan BK, Costello JH, Van Keuren D (2007) Seasonality of the copepods *Acartia hudsonica* and *Acartia tonsa* in Narragansett Bay, RI, USA during a period of climate change. *Estuarine Coastal and Shelf Science* **73**, 259–267.

Sullivan BK, McManus LT (1986) Factors controlling seasonal succession of the copepods *Acartia hudsonica* and *Acartia tonsa* in Narragansett Bay, Rhode Island: Temperature and resting egg production. *Marine Ecology Progress Series* **28**, 121–128.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Syrjala SE (1996) A statistical test for a difference between the spatial distributions of two populations. *Ecology* **77**, 75–80.

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

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599.

Templeton AR (2002) Out of Africa again and again. *Nature* **416**, 45–51.

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

Templeton AR (2008) Nested clade analysis: an extensively validated method for strong phylogeographic inference. *Molecular Ecology* **17**, 1877–1880.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence Data .3.Cladogram estimation. *Genetics* **132**, 619–633.

Tester PA, Turner JT (1991) Why is *Acartia tonsa* restricted to estuarine habitats. *Bulletin of Plankton Society of Japan* **Special Volume**, 603–611.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**, 4876–4882.

Thuesen EV, Numachi K, Nemoto T (1993) Genetic variation in the planktonic chaetognaths *Parasagitta elegans* and *Eukrohnia hamata*. *Marine Ecology Progress Series* **101**, 243–251.

Trujillo-Ortiz A, Burton RS, de la Rosa-Velez J, Correa-Sandoval F (1999) Interbreeding between two populations of *Acartia californiensis* (Copepoda: Calanoida): A laboratory study. *Journal of the Marine Biological Association of the United Kingdom* **79**, 945–948.

Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology and Evolution* **16**, 330–343.

Turner JT (1984) Zooplankton feeding ecology: Contents of fecal pellets of the copepods *Acartia tonsa* and *Labidocera aestiva* from continental shelf waters near the mouth of the Mississippi River. *Marine Ecology Pubblicazioni Della Stazione Zoologica Di Napoli I* **5**, 265–282.

van der Spoel S, Heyman RP (1983) *An Atlas of Comparative Zooplankton: Biological Patterns in the Oceans.* Springer-Verlag, Berlin; New York.

van der Spoel S, Pierrot-Bults AC (1979) *Zoogeography and Diversity in Plankton.* Halsted Press, New York.

Veldhuis MJW, Kraay GW (2004) Phytoplankton in the subtropical Atlantic Ocean: Towards a better assessment of biomass and composition. *Deep–Sea Research I* **51**, 507–530.

Walker ML, Chirhart SE, Moore AF, Honeycutt RL, Greenbaum IF (2006) Genealogical concordance and the specific status of *Peromyscus sejugis*. *Journal of Heredity* **97**, 340–345.

Wares JP (2002) Community genetics in the Northwestern Atlantic intertidal. *Molecular Ecology* **11**, 1131–1144.

Weir BS, Cockerham CC (1984) Estimating F statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Weissman P, Lonsdale DJ, Yen J (1993) The effect of peritrich ciliates on the production of *Acartia hudsonica* in Long Island Sound. *Limnology and Oceanography* **38**, 613–622.

Werner FE, Blanton JO, Lynch DR, Savidge DK (1993) A numerical study of the continental shelf circulation of the United States South Atlantic Bight during the Autumn of 1987. *Continental Shelf Research* **13**, 971–997.

White TJ, Burns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR protocols: A Guide to Methods and Applications* (eds. Innis MA, Gelfand DH, Shinsky JJ, White TJ), pp. 315–322. Academic Press, San Diego, California.

Whitfield J (2005) Biogeography: Is everything everywhere? *Science* **310**, 960–961.

Wilhelm R, Hilbish TJ (1998) Assessment of natural selection in a hybrid population of mussels: Evaluation of exogenous vs endogenous selection models. *Marine Biology* **131**, 505–514.

William J, Ballard O, Kreitman M (1995) Is mitochondrial DNA a strictly neutral marker. *Trends in Ecology and Evolution* **10**, 485–488.

Williams GC (1975) *Sex and Evolution.* Princeton University Press, Princeton, New Jersey.

Wooldridge T, Erasmus T (1980) Utilization of tidal currents by estuarine zooplankton. *Estuarine and Coastal Marine Science* **11**, 107–114.

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

Young I, Coleman AW (2004) The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a *Drosophila* example. *Molecular Phylogenetics and Evolution* **30**, 236–242.

Zane L, Ostellari L, Maccatrozzo L, Bargelloni L, Battaglia B, Patarnello T (1998) Molecular evidence for genetic subdivision of Antarctic krill (*Euphausia superba* Dana) populations. *Proceedings of the Royal Society of London B* **265**, 2387–2391.

Zane L, Ostellari L, Maccatrozzo L, Bargelloni L, Cuzin–Roudy J, Buchholz F, Patarnello T (2000) Genetic differentiation in a pelagic crustacean (*Meganyctiphanes norvegica*: Euphausiacea) from the North East Atlantic and the Mediterranean Sea. *Marine Biology* **136**, 191–199.

Zhang DX, Hewitt GM (1996) Nuclear integrations: Challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution* **11**, 247–251.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: Practice, problems and prospects. *Molecular Ecology* **12**, 563–584.

Zillioux EJ, Gonzalez JG (1972) Egg dormancy in a neritic calanoid copepod and its implications to overwintering in boreal waters. In: *The fifth European Marine Biology Symposium, Piccin Editore, Padova*, pp. 217–230.

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the alpha-chains of Hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 2158–2162.