MOLECULAR SYSTEMATICS AND BIOGEOGRAPHY OF THE HOLARCTIC SMELT FAMILY OSMERIDAE (PISCES).

by

KATRIINA LARISSA ILVES

B.Sc., The University of British Columbia, 2000
B.A., The University of British Columbia, 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA

December 2007

ABSTRACT

Biogeographers have long searched for common processes responsible for driving diversification in the Holarctic region. Although terrestrial flora and fauna have been well studied, much of the marine biogeographic work addresses patterns and processes occurring over a relatively recent timescale. A prerequisite to comparative biogeographic analysis requires well-resolved phylogenies of similarly distributed taxa that diverged over a similar timeframe. The overall aim of my Ph.D. thesis was to address fundamental questions in the systematics and biogeography of a family of Holarctic fish (Osmeridae) and place these results in a broad comparative biogeographic framework. With eight conflicting morphological hypotheses, the northern hemisphere smelts have long been the subjects of systematic disagreement. In addition to the uncertainty in the interrelationships within this family, the relationship of the Osmeridae to several other families remains unclear.

Using DNA sequence data from three mitochondrial and three nuclear genes from multiple individuals per species, I reconstructed the phylogenetic relationships among the 6 genera and 15 osmerid species. Phylogenetic reconstruction and divergence dating yielded a well-resolved phylogeny of the osmerid genera and revealed several interesting evolutionary patterns within the family: (1) *Hypomesus chishimaensis* and *H. nipponensis* individuals are not reciprocally monophyletic, suggesting that they are conspecific and *H. chishimaensis* is a recently evolved freshwater ecotype that invaded the Kuril Islands following the last glaciation, (2)

The trans-Pacific sister relationships in *Hypomesus* based on lateral line scale counts are not supported, implying that this phenotype evolved in parallel on each side of the North Pacific Ocean, (3) The Plecoglossidae are the Osmeridae sister group, (4) Over half of the characters from previous studies show evidence of parallel evolution; however, 27 traits reflect ancestral relationships, (5) Multiple divergences within the Osmeridae date to both the mid-Miocene cooling period and the Pliocene Bering Seaway opening, suggesting these events were important in the evolution of these fishes, and (6) Divergences in many marine taxa for which dated phylogenies are available are also correlated with these time periods. Future research should target additional Holarctic marine taxa for further comparative analysis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to thank my Ph.D. supervisor, Eric (Rick) Taylor. Rick's undergraduate Diversity and Evolution of Fishes course inspired my initial interest in fish systematics and biogeography, which was then reinforced through my work with him at the UBC Fish Museum. Through this work I also gained an enormous appreciation for the value of natural history collections and discussed possibilities for graduate work, which turned into the topic for my Ph.D. thesis. I am sincerely grateful for the opportunities I've had working in Rick's lab, and he deserves special recognition for his financial support and patience with my numerous crises of faith over the years. I am also thankful for the insight provided by the members of my supervisory committee, Mary Berbee, Wayne Maddison, Sarah (Sally) Otto, and Chris Stephens, who assisted me with methods and interpretation and kept me on track. In particular, Sally Otto has acted as a significant mentor to me since my undergraduate days working with her and Jana Vamosi (formerly Heilbuth). Sally and Jana first exposed me to the research process and Sally has continued to inspire me with her enthusiasm, support and ability to view problems from varied perspectives. My labmates Jen Gow, Ramona de Graaf, Jen McLean, Mike Stamford, Allan Costello, Les Harris, Jon Mee, Patrick Tamkee, Patricia Woodruff and Sara Northrup have been an amazing source of intellectual and emotional support throughout this process. In particular, Jen Gow and Ramona de Graaf have put up with me for numerous years and have even dared to take me along on their field excursions. I would especially like to thank Jen Gow for being a mentor and a

great friend. Mike Stamford and Les Harris were also a great help in the field and I appreciate their assistance and company in what were often trying circumstances. I also thank Don McPhail for numerous insightful discussions. Don's knowledge of fish and his ability to communicate his enthusiasm are truly inspiring. I acknowledge Brad Davis for his assistance with the logistic model for the LAGRANGE analysis in Chapter 5 and also for his friendship. Many thanks also to Marika Younker, who showed me the initial ropes of PCR and provided cytochrome *b* sequences for many samples, Jason Weir for many interesting conversations about phylogenetic and biogeographic inference, and Rick Ree for assistance with and discussion about LAGRANGE. My thesis research would not have been possible without the generous sample donations from many people across the northern hemisphere: R. Gross, D. Hay, K. Hieb, J. Jurvelius, E. Lowery, K. Maslenikov, J. McPhail, S. Mori, A. Morton, P. Moyle, L. Plough, T. Quinn, K. Takata, P. Tamkee, J. Taylor, M. Thompson, B. Urbain, and P. Westley. Katherine Maslenikov at the University of Washington Fish Collection was particularly helpful. Fieldwork collecting longfin smelt from Harrison and Pitt Lakes was conducted with the assistance of J. Hume and S. MacLellan from the Department of Fisheries and Oceans, G. Scholten, P. Woodruff, and G. Wilson from the BC Ministry of Environment, and Mike Stamford. J. Taylor and L. Harris also assisted me in the Fraser River. This fieldwork would not have been possible without the assistance of these numerous people and I am very grateful to everyone who volunteered his/her time. The staff at the UBC NAPS sequencing facility deserve huge thanks for all their work, especially Marian van der Geest who was incredibly helpful with troubleshooting and general discussion. The

data analysis for this thesis would not have been possible without the facilities at the UBC Zoology Computing Unit (ZCU). I am grateful to Alistair Blatchford and Sanja LeBlanc from the ZCU for their assistance and patience. I am also grateful to Allison Barnes, the Zoology department's graduate secretary, for all her work and support. I would also like to extend a sincere thank you to Lynn Norman from the first year biology program for her teaching guidance. In addition to the financial support from Rick Taylor, the MacLean Fraser Scholarship and University Graduate Fellowship allowed me to pursue my thesis research. Finally, I would not have made it through this process without the support and encouragement from my family and friends. I first thank my parents, Maie and Riivo Ilves and my uncle Aser Niiholm, for their unending support in all its forms, and my brother and friend Eerik for always being there. My friends Colleen Fulton, Matt Goodmurphy, Jane Cipollone, and Ryan Stephenson have been an amazing support network. Thank you to everyone.

For my parents, Maie and Riivo Ilves, for all their support and encouragement

CO-AUTHORSHIP STATEMENT

Chapters 2 and 3 were co-authored in published form with Dr. Eric B. Taylor. For both of these chapters Dr. Taylor and I co-developed the idea and approach. I collected all of the molecular data, performed the analyses and wrote the manuscripts. Dr. Taylor contributed significantly to the interpretation of the results and editing of the manuscripts. Likewise, the initial ideas for Chapters 4 and 5 were co-developed by Dr. Taylor and myself. I collected all of the molecular data, performed the analyses and wrote the content for these chapters, which are currently in preparation for submission.

CHAPTER 1: GENERAL INTRODUCTION

"Should we begin by studying a whole genus in common before going on to study the special properties of the different species? Or should we study the particular species one at a time? For, as things are, this has not been determined."

-Aristotle (Parts of Animals, Book 1, Chapter 1, 639b, 5, tr. T. Irwin and G. Fine)

1.1 INTRODUCTION

1.1.1 Brief introduction to systematics and taxonomy

The fundamental goal of systematic biology is to understand the relationships among organisms through time. This field can be traced at least to Aristotle (384-322 BC), and while the primary objective has not changed significantly, the philosophy of how these relationships are determined has undergone many shifts. Early beginnings resting on ideas of immutable species have now been almost completely replaced by an evolutionary perspective first put forth by Jean-Baptiste Lamarck and later refined by Charles Darwin and Alfred Russell Wallace (but see Brower 2000). Although an evolutionary approach to systematics (phylogenetics) is now almost universally accepted, methodological debates regarding how relationships among organisms are best determined have formed a large part the literature since the latter half of the 20th century.

Willi Hennig's (1950) publication of *Grundzüge einer Theorie der Phylogenetischen Systematik* and the 1966 English translation *Phylogenetic*

Systematics (Hennig 1966) synthesized ideas and outlined the principles of how phylogenetic inference should be conducted. He argued that clades should be defined on the basis of synapomorphies (shared derived traits), and that taxonomic designations should only be applied to monophyletic groups of taxa, with rankings scaled relative to time (Hennig 1966). The computation of phylogenies began in the 1950s. Sokal and Sneath's (1963) work *Numerical Taxonomy* was the first major introduction of a phenetic (overall similarity) approach to inferring phylogenies from morphological data. Shortly following, a parsimony method was described by Camin and Sokal (1965). Since that time, molecular techniques for analyzing protein composition, and more recently DNA sequences, spurred the development of many other phylogenetic methods, including distance (Fitch and Margoliash 1967; Jukes and Cantor 1969), parsimony (Eck and Dayhoff 1966; Kluge and Farris 1969; Farris 1970), maximum likelihood (Edwards and Cavalli-Sforza 1964; Neyman 1971; Felsenstein 1981), and Bayesian approaches (reviews in Larget and Simon 1999; Huelsenbeck et al. 2001). A detailed history of phylogenetic inference is found in Felsenstein (2004), which is the source of much of the above (abbreviated) information.

Taxonomy and systematics are separate, yet increasingly related, branches of biology. While taxonomy has been defined as the theory and practice of identifying, classifying and naming organisms (e.g., Simpson 1961; Mayr 1969; both cited in Mayr 1982), recent work has aimed to disentangle taxonomy [classification] and nomenclature [naming] (e.g., de Queiroz 2006). The currently implemented hierarchical system of taxonomy, with binomial species names, originated with the

Swedish botanist, anatomist, and zoologist Carolus Linnaeus' *Systema Naturae* (Linnaeus 1758), although a number of earlier biologists, including Andrea Cesalpino and John Ray, also developed classification systems based on measures of relatedness among organisms (Mayr 1982). As the theory of evolution had not yet been developed at Linnaeus' time, and he held the prevailing view that life is static, his classification system was primarily based on the physical similarities of different organisms (Mayr 1982). With the works of Darwin and Wallace came a greater understanding that organisms are related by a process of descent with modification, and this knowledge has been increasingly incorporated into taxonomic practice such that classifications are intended to reflect the current hypothesis of evolutionary relationships among taxa.

The Linnaean system of biological classification remains the standard used by present day systematists and taxonomists. There have, however, been numerous critiques (e.g., de Queiroz and Gauthier 1990; Hibbett and Donoghue 1998; Ereshefsky 2001) and increasing calls for either improvement (Hennig 1966; Mayr 1982; Avise and Johns 1999; Avise and Mitchell 2007) or complete replacement by alternatives (e.g., de Queiroz and Gauthier 1992; Ereshefsky 2001). Although the Linnaean system has been criticized for a number of reasons, including philosophical objections to the distinction between species and higher taxa (Ereshefsky 1991), two common pragmatic objections to its retention are that it is (1) unstable and, (2) inconsistently applied across taxonomic groups. The first objection refers to the fact that as our understanding of evolutionary relationships changes with improvements in our ability to infer phylogenies, it is often found that, for

example, species once thought to belong to the same genus are actually more closely related to species in other genera. Thus, if the classification is to reflect the phylogenetic relationships of the species, they require a name change; hence, the system is unstable. The second major criticism stems from the practice of systematists working on different groups of organisms. There is no objective standard for how to determine, for example, a genus, family, or order; therefore, it is not possible to make meaningful comparisons of diversity across different taxonomic groups.

These objections have recently spurred a large body of literature either defending (with improvements) or advocating the complete replacement of the Linnaean system of classification. The latter camp rejects the idea of ranked names, and the most prominent suggested replacement is the PhyloCode. The PhyloCode (Cantino and de Queiroz 2006), which stems from ideas put forth by de Queiroz and Gauthier (1990, 1992, 1994), following Hennig (1966), that taxonomy should reflect our understanding of the evolutionary relationships among taxa, is a system of phylogenetic nomenclature with the goal of applying uninomial, unranked names to monophyletic clades above the species level. Some of the benefits of this approach are said to include: (1) increased stability, as situations that occur in the Linnaean system where a monophyletic group's designation change (e.g. from a family to a subfamily) also requires a name change (e.g., –idae to –inae in zoology) are avoided because the name of a monophyletic group remains constant, (2) monophyletic groups that fall in between major categories of the Linnaean system can be named without excessive addition of ranks (e.g., subgenus, section, subsection, etc.), (3)

complete knowledge of the phylogeny surrounding a well-supported monophyletic group is unnecessary to attach a name to this group and, (4) it is more consistent with the evolutionary view currently adopted by most systematists (Hibbett and Donoghue 1998; Bryant and Cantino 2002).

The PhyloCode and phylogenetic nomenclature in general are receiving support in the systematics literature and have already been implemented to some extent for some taxa (e.g., Engstrom et al. 2004; Hillis and Wilcox 2005); however, there has also been vigorous opposition (e.g., Benton 2000; Monsch 2006; Rieppel 2006; Dubois 2007). Many of the criticisms expressed by these authors have been addressed in responses by Bryant and Cantino (2002) and Hillis (2007), but it is clear that debate between Linnaean and phylogenetic nomenclatures remains active.

Although the PhyloCode aims to solve the first pragmatic objection to the Linnaean system listed above, that it is unstable, it does not address the second, which is the impossibility of meaningfully comparing diversity across different taxonomic groups because Linnaean taxonomy is not standardized. Although Hennig (1966) discussed the possibility that taxonomic ranks be determined by geological age, Avise and Johns (1999) and more recently Avise and Mitchell (2007) propose 'temporal-banding' and 'timeclip' schemes, respectively, with the goal of standardizing taxonomic designations above the species level with geologic time periods, while retaining the Linnaean hierarchy. Briefly, the taxonomic procedure for temporal-banding involves assigning taxonomic rank based on a clade's date of origination (Avise and Johns 1999). For example, monophyletic groups dating to the

Pliocene would be designated a sub-genus, while those dating to the mid-early Miocene would be considered a genus (Avise and Johns 1999). The timeclip suggestion, on the other hand, involves attaching a geologic time designation to a clade named under the Linnaean system of classification (Avise and Mitchell 2007). Both of these alternatives require dated phylogenies, which Avise and Mitchell (2007) note are becoming standard practice in molecular phylogenetic studies. Although Avise and Johns' (1999) temporal-banding proposal has been widely cited in the literature (59 citations, ISI Web of Science database), only a few studies have made an attempt to implement the procedure (e.g., Biju and Bossuyt 2003; Schneider et al. 2001; Pitra et al. 2004). The more recent timeclip alternative has not yet been discussed in the literature, although Avise and Mitchell (2007) suggest that it could be used under both the Linnaean and unranked systems of nomenclature, making it likely that future additions to the PhyloCode, or other approaches to phylogenetic nomenclature, will consider the possibility of including an indication of time.

1.1.2 Brief introduction to historical biogeography

Modern historical biogeography had its beginnings in the early 20th century and was dominated by the goal of determining a taxon's 'centre of origin'. Underlying this goal was a belief in a static earth over which organisms dispersed from their area of origination and then subsequently diversified (Lomolino et al. 2006). Several authors (e.g., Adams 1902, 1909; Matthew 1915; Cain 1944, all cited in Lomolino et al. 2006) outlined specific criteria by which such areas could be

identified, such as location of greatest number of species, location of maximum size of individuals, and location of least dependence on restricted habitat, to name a few (Cain 1944, cited in Lomolino et al. 2006). Serious objections to such approaches began to develop in the 1950s and 1960s with Leon Croizat's (e.g., 1958, 1964) development of pan-biogeography and continued into the 1970s and 1980s when vicariance (or cladistic) biogeography dominated the discipline. Croizat's approach differed significantly from that of earlier 'centre of origin' biogeographers, as he was focused on identifying biogeographic patterns, mainly patterns of disjunction, across numerous taxonomic groups. Although Croizat took his views to an extreme, arguing that all patterns of trans-oceanic disjunction could be explained by earlier landbridge connections between areas, thereby completely discounting dispersal as a mechanism that could explain distributions, his method of examining biogeographic patterns across different groups of organisms was highly influential to the development of the vicariance biogeography era that followed (Lomolino et al. 2006).

Vicariance biogeography arose out of several key advancements: Hennig's (1966) work detailing phylogenetic systematic practice, the acceptance of the mechanism of plate tectonics to explain continental drift, and Croizat's comparative biogeography. In particular, Brundin's (1966) classic analysis of chironomid midge biogeography in the southern hemisphere, where he applied Hennig's (1966) cladistic methods to areas of distribution, motivated a new generation of biogeographers to develop analytical methods for determining general area cladograms by integrating information contained in individual taxon phylogenies

(e.g., Platnick and Nelson 1978; Rosen 1978; Nelson and Platnick 1981; Nelson and Ladiges 1991, cited in Lomolino et al. 2006; reviewed in Humphries and Parenti 1999 and Crisci et al. 2003). This group strongly opposed the 'story-telling' centre of origin approaches from earlier in the century, and following Croizat, rejected the role of dispersal in historical biogeographic analysis as this process was considered idiosyncratic and specific to individual taxa, and therefore not informative for comparative biogeography (e.g., Nelson 1979).

Historical biogeography is currently undergoing yet another paradigm shift. With the increased ability to estimate divergence times from phylogenetic trees in combination with recently developed analytical approaches to explicitly model biogeographic history through time, comparative biogeography appears to be moving away from a focus of examining patterns of area cladograms and towards an approach where timing is (almost) everything. Although there are some vicariance biogeographers who deny the relevance of time in the comparison of area cladograms (e.g., Humphries and Parenti 1999), it is generally recognized that topological congruence among area cladograms alone is not evidence that a group of taxa share a similar vicariant history ('pseudo-congruence'); however, synchronous timing of such divergences would suggest a similar process was responsible for generating the patterns of divergence and distribution (Cunningham and Collins 1994; Donoghue and Moore 2003). Further, there is increasing evidence that dispersal has not only been an important process in shaping organismal distributions (e.g., McDowall 2002; de Queiroz 2005; Sanmartín et al. 2007) but also that dispersals are not random, idiosyncratic events unique to

particular taxa, as dispersal opportunities for numerous groups may arise at the same time given changes in environmental or geological conditions (Lomolino et al. 2006).

Searching for centres of origin remains a goal of many biogeographic studies, either explicitly (e.g., Briggs 2000, 2003, 2004; Smith et al. 2005; Baker et al. 2006) or implicitly (Lomolino et al. 2006), although such approaches are not without their critics (e.g., Ebach 1999; Humphries and Parenti 1999). Recent developments in historical biogeography include event-based methods that allow explicit modeling of a taxon's biogeography (e.g., Ronquist 1997; Ree et al. 2005). Such DIVA (dispersal-vicariance) methods incorporate processes of dispersal, vicariance, and extinction, now widely accepted to affect organismal distributions. The original parsimony-based approach, which optimized ancestral areas on a phylogenetic tree based on the minimum number of dispersal and extinction events (DIVA; Ronquist 1997) have been joined by a maximum likelihood reconstruction method that incorporates divergence times and includes a geologic and climatic model of area connections through a taxon's evolutionary history (LAGRANGE; Ree et al. 2005). A Bayesian approach along the lines of LAGRANGE is also in preparation (I. Sanmartín, 2007 International Biogeography Society meeting). These biogeographic modeling approaches have been widely implemented (e.g., Sanmartín et al. 2001, 2007; Burbrink and Lawson 2007; Ilves and Taylor 2007) and with increasing availability of dated phylogenies, will likely become standard biogeographic practice. Further, as genetic data for similarly distributed organisms continue to accrue, coalescent modeling to identify synchronous timing of

divergences is another approach likely to receive considerable future attention (e.g., Hickerson and Cunningham 2005; Hickerson et al. 2006a,b).

1.1.3 Northern hemisphere smelts: a systematic and biogeographic case study

The Osmeridae, as currently defined, contains six genera (*Allosmerus*, *Hypomesus*, *Osmerus*, *Mallotus*, *Spirinchus*, and *Thaleichthys*) and 15 species (Eschmeyer 2006), which are distributed throughout cool temperate coastal marine and freshwaters of the North Pacific and North Atlantic Oceans, with two species also extending into the Arctic region (McAllister 1963; Froese and Pauly 2006). They are small (< 30 cm) and have an elongate, silvery body, like many other pelagic fishes, such as herring. Most species are anadromous (spawn in freshwater and mature in the marine environment), but two are strictly marine (*A. elongatus* and *S. starksi*), one strictly freshwater (*H. olidus*), and others have populations that appear to have secondarily become freshwater residents (*H. nipponensis*, *O. eperlanus*, *O. mordax*, and *S. thaleichhtys*). They have a planktivorous diet and are an important forage fish for larger fishes and marine mammals. In general, there is relatively little known about most of the Osmeridae species, although recent work has provided insight into the ecology and/or genetics of several species: rainbow smelt, *Osmerus mordax* (Taylor and Bentzen 1993; Saint-Laurent et al. 2003; Curry et al. 2004; Lecomte and Dodson 2004); eulachon, *Thaleichthys pacificus* (McLean et al. 1999; McLean and Taylor 2001); wakasagi, *Hypomesus nipponensis* (Kudo

and Mizuguchi 2000; Torao 2000; Arai et al. 2006); and delta smelt, *H. transpacificus* (Trenham et al. 1998; Swanson et al. 2000; Feyrer et al. 2007).

Several species have important subsistence fisheries, including *Osmerus mordax* in eastern North America and *Spirinchus starksi* in the Pacific Northwest. The eulachon, *Thaleichthys pacificus*, has been of great traditional importance to many First Nations in British Columbia; however, recent population declines have closed fisheries and threatened the livelihoods and traditions of many communities. The delta smelt, *Hypomesus transpacificus*, has the most restricted distribution, found only in the San Joaquin – Sacramento estuary, and has been impacted by both the introduction of a Japanese congener *H. nipponensis* (Moyle et al. 1992; Moyle 2002) and general environmental degradation (Feyrer et al. 2007).

Although relatively little is known about the ecology of most species, their systematic relationships have been the subject of intense scrutiny by ichthyologists for much of the last century. A detailed discussion of Osmeridae systematics is presented in Chapter 4, but as a brief introduction, there are currently eight published morphological phylogenies (Johnson and Patterson 1996, and references therein). These hypotheses conflict to the point where not a single clade is supported by all studies.

In addition to interest in the systematics of the Osmeridae, their Holarctic distribution, with apparent trans-oceanic and Pacific-Atlantic sister relationships, has also led many authors to speculate on their biogeographic history (McAllister 1963; Klyukanov 1975; Taylor and Dodson 1994). McAllister (1963) proposed a compression mechanism by which a more widely spread ancestor had its range

compressed during periods of climatic cooling in the Pliocene and Pleistocene, splitting the population on opposite sides of the ocean basin. The opening of the Bering Seaway in the Pliocene/Pleistocene has also been implicated as important for allowing osmerids to disperse between the Pacific and Atlantic Oceans via the Arctic (McAllister 1963; Klyukanov 1975; Taylor and Dodson 1994). These events have also been suggested as significant drivers of evolutionary change in a number of other temperate northern hemisphere taxa (e.g., Vermeij 1991, 2005; Stearly 1992; Collins et al. 1996; Väinölä 2003). Aspects of osmerid and comparative Holarctic biogeography are discussed in detail in Chapters 3 and 5.

More specifically, using phylogenetic analysis of mitochondrial (mtDNA) and nuclear (nDNA) gene sequences, this thesis addresses several issues in the systematics and biogeography of the northern hemisphere smelts and aims to place the results in a comparative biogeographic context of other North Pacific and Holarctic taxa. The following gives a brief summary of the major objectives of each of the subsequent thesis chapters.

Chapter 2 - Are *Hypomesus chishimaensis* and *H. nipponensis* (Pisces: Osmeridae) distinct species? A molecular assessment using comparative sequence data from five genes.

In Chapter 2, I address the question of whether a newly identified freshwater species of *Hypomesus*, *H. chishimaensis,* from the Kuril Islands of Japan, is distinct from its anadromous congener *H. nipponensis*. This is accomplished

through sequence comparison of five genes (two mtDNA, three nDNA) from individuals identified as each of the two putative species. The results are discussed in relation to the geologic history of the islands and conjoined with recent morphological data to assess whether this species designation should be considered valid. Due to the lack of morphological, ecological, behavioural, and genetic data for distinguishing the two putative species, and the complex geologic history of the Kuril Islands, I conclude that *H. chishimaensis* should not be designated at the species level and should instead be considered a synonym of *H. nipponensis*.

Chapter 3 - Evolutionary and biogeographic patterns within the smelt genus *Hypomesus* (Pisces: Osmeridae) in the North Pacific Ocean.

In Chapter 3, I reconstruct the phylogenetic relationships of *Hypomesus*, the most species-rich genus of northern hemisphere smelts, through analysis of sequence data from five genes (two mtDNA, three nDNA). The phylogenetic results are then used to examine the biogeographic history of the genus in the North Pacific through the latter half of the Cenozoic using two methods of biogeographic reconstruction: (1) parsimony step-matrix, and (2) maximum likelihood modeling. This chapter addresses hypotheses of area of origin, parallel evolution of phenotypes, and possible causes of trans-Pacific distributions. The biogeographic results are discussed in the context of patterns found in other similarly distributed marine taxa.

Chapter 4 – Molecular systematics of the Osmeridae

In chapter 4, I reconstruct the phylogenetic relationships of the Osmeridae from analysis of sequence data from six genes (three mtDNA, three nDNA). The results are statistically compared to the previous eight hypotheses of Osmeridae interrelationships and the characters used in the reconstruction of several of these earlier studies are discussed in relation to the molecular phylogeny. The resulting molecular phylogeny significantly differs from all previous hypotheses of Osmeridae interrelationships and suggests that parallel evolution of morphological traits is a common occurrence in these fishes.

Chapter 5 – Holarctic biogeography of the Osmeridae

In Chapter 5, I use the Osmeridae phylogeny resulting from Chapter 4 to estimate divergence dates for each of the nodes. The biogeographic history of the group is then reconstructed using two methods: (1) analysis of a parsimony step-matrix, and (2) maximum likelihood modeling. The results are compared to previous hypotheses of Osmeridae biogeography and to data available for other Holarctic marine taxa, and are discussed with regard to the impact of fossil data on biogeographic interpretations. Direct comparative data on the Cenozoic timescale of Osmeridae evolution are lacking; however, the divergence date and biogeographic analyses suggest the development of cool water conditions in the Paleocene/Eocene, the mid-Miocene cooling

period, and the opening of the Bering seaway in the late Miocene have been significant events in the diversification of extant lineages.

1.2 REFERENCES

Adams, C. C. 1902. Southeastern United States as a center of geographical distribution of fauna and flora. Biol. Bull. 3:115-131.

———. 1909. An ecological survey of Isle Royale, Lake Superior. Report of Board of Geological Survey for 1908, Lansing, Michigan.

Arai, T., J. Yang, and N. Miyazaki. 2006. Migration flexibility between freshwater and marine habitats of the pond smelt *Hypomesus nipponensis*. J. Fish Biol. 68:1388-1398.

Avise, J. C., and G. C. Johns. 1999. Proposal for a standardized temporal scheme of biological classification for extant species. Proc. Natl. Acad. Sci. 96:7358-7363.

Avise, J. C., and D. Mitchell. 2007. Time to standardize taxonomies. Syst. Biol. 56:130-133.

Baker, A. J., S. L. Pereira, O. P. Haddrath, and K.-A. Edge. 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. Proc. R. Soc. B. 273:11-17.

Benton, M. J. 2000. Stems, nodes, crown clades, and rank-free lists: is Linnaeus dead? Biol. Rev. 75:633-648.

Biju, S. D., and F. Bossuyt. 2003. New frog family from India reveals an ancient biogeographical link with the Seychelles. Nature 425:711-714.

Briggs, J. C. 2000. Centrifugal speciation and centres of origin. J. Biogeogr. 27:1183-1188.

———. 2003. Marine centres of origin as evolutionary engines. J. Biogeogr. 30:1-18.

———. 2004. Older species: a rejuvenation on coral reefs? J. Biogeogr. 31:525-530.

Brower, A. V. Z. 2000. Evolution is not a necessary assumption of cladistics. Cladistics 16:143-154.

Brundin, L.1966. Transantarctic relationships and their significance, as evidenced by chironomid midges with a monograph of the subfamilies Podonominae and Aphroteniinae and the austral Heptagyiae, Kungl. Svenska Vetens. 11:1-472.

Bryant, H. N., and P. D. Cantino. 2002. A review of criticisms of phylogenetic nomenclature: is taxonomic freedom the fundamental issue? Biol. Rev. 77:39-55.

Burbrink, F. T., and R. Lawson. 2007. How and when did Old World ratsnakes disperse into the New World? Mol. Phylogenet. Evol. 43:173-189.

Cain, S. A. 1944. Foundations of plant geography. Harper and Brothers, New York.

Camin, J. H., and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. Evolution 19:311-326.

Cantino, P. D., and K. de Queiroz. 2006. *PhyloCode*. http://ohiou.edu/phylocode

Collins, T. M., K. Frazer, A. R. Palmer, G. J. Vermeij, and W. M. Brown. 1996. Evolutionary history of northern hemisphere *Nucella* (Gastropoda, Muricidae): molecular, morphological, ecological, and paleontological evidence. Evolution 50:2287-2304.

Crisci, J. V., L. Katinas, and P. Posadas. 2003. Historical biogeography. Harvard University Press, Cambridge, MA.

Croizat, L. 1958. Panbiogeography. Vol. I, IIa, IIb. Published by the author, Caracas, Venezuela.

———. 1964. Space, time, form: the biological synthesis. Published by the author, Caracas, Venezuela.

Cunningham, C. W., and T. M. Collins. 1994. Developing model systems for molecular biogeography: vicariance and interchange in marine invertebrates, Pp. 405-433 in B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle, eds. Molecular ecology and evolution: approaches and applications. Birkhauser Verlad, Basel, Switzerland.

Curry, R. A., S. L. Currie, L. Bernatchez, and R. Saint-Laurent. 2003. The rainbow smelt, *Osmerus mordax*, complex of Lake Utopia: threatened or misunderstood? Env. Biol. Fish. 69: 153-166.

de Queiroz, A. 2005. The resurrection of oceanic dispersal in historical biogeography. Trends Ecol. Evol. 20:68-73.

de Queiroz, K. 2006. The PhyloCode and the distinction between taxonomy and nomenclature. Syst. Biol. 55:160-162.

de Queiroz, K., and J. Gauthier. 1990. Phylogeny as a central principle in taxonomy: phylogenetic definitions of taxon names. Syst. Zool. 39:307-322.

———. 1992. Phylogenetic taxonomy. Annu. Rev. Ecol. Syst. 23:449-480.

———. 1994. Toward a phylogenetic system of biological nomenclature. Trends Ecol. Evol. 9:27-31.

Donoghue, M. J., and B. R. Moore. 2003. Toward an integrative historical biogeography. Integr. Comp. Biol. 43:261-270.

Dubois, A. 2007. Naming taxa from cladograms: a cautionary tale. Mol. Phylogenet. Evol. 42:317-330.

Ebach, M. C. 1999. Paralogy and the centre of origin concept. Cladistics 15:387-391.

Eck, R. V., and M. O. Dayhoff. 1966. *Atlas of protein sequence and structure 1966*. National Biomedical Research Foundation, Silver Spring, Maryland.

Edwards, A. W. F., and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. Pp. 67-76 in V.H. Heywood and J. McNeill, eds. *Phenetic and phylogenetic classification*. Systematics Association Publ. No. 6, London.

Engstrom, T. N., H. B. Shaffer, and W. P. McCord. 2004. Multiple data sets, high homoplasy, and the phylogeny of softshell turtles (Testudines: Trionychidae). Syst. Biol. 53:693-710.

Ereshefsky, M. 1991. Species, higher taxa, and the units of evolution. Philos. Sci. 58:84-101.

———. 2001. *The poverty of the Linnaean hierarchy*. Cambridge University Press, Cambridge.

Eschmeyer, W. N. 2006. Catalogue of fishes, online version. California Academy of Sciences, San Francisco. Available from http//www.calacademy.org/research/ichthyology/catalog

Farris, J. S. 1970. Methods for computing Wagner trees. Syst. Zool. 19:83-92.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368-376.

———. 2004. Inferring phylogenies. Sinauer Associates Inc., Sunderland, MA.

Feyrer, F., M. L. Nobriga, and T. R. Sommer. 2007. Multidecadal trends for three declining fish species: habitat patterns and mechanisms in the San Francisco Estuary, California, USA. Can. J. Fish. Aquat. Sci. 64:723-734.

Fitch, W. M., and E. Margoliash. 1967. Construction of phylogenetic trees. Science 155:279-284.

Froese, R. and D. Pauly. (Eds.). 2006. FishBase. World Wide Web electronic publication. www.fishbase.org

Hennig, W. 1950. Grundzüge einer theorie der phylogenetischen systematik. Deutscher Zentralverlag, Berlin.

———. 1966. Phylogenetic systematics. translated by D.D. Davis and R. Zangerl. University of Illinois Press, Urbana.

Hibbett, D. S., and M. J. Donoghue. 1998. Integrating phylogenetic analysis and

classification in fungi. Mycologia 90:347-356.

Hickerson, M. J., and C. W. Cunningham. 2005. Contrasting quaternary histories in an ecologically divergent sister pair of low-dispersing intertidal fish (*Xiphister*) revealed by multilocus DNA analysis. Evolution 59:344-360.

Hickerson, M. J., G. Dolman, and C. Moritz. 2006a. Comparative phylogeographic summary statistics for testing simultaneous vicariance. Mol. Ecol. 15:209-223.

Hickerson, M. J., E. A. Stahl, and H. A. Lessios. 2006b. Test for simultaneous divergence using approximate Bayesian computation. Evolution 60:2435-2453.

Hillis, D. M. 2007. Constraints in naming parts of the Tree of Life. Mol. Phylogenet. Evol. 42:331-338.

Hillis, D. M., and T. P. Wilcox. 2005. Phylogeny of the New World true frogs (*Rana*). Mol. Phylogenet. Evol. 34:299-314.

Huelsenbeck, J. P., F. Ronquist, R. Nielsen, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310-2314.

Humphries, C. J., and L. R. Parenti. 1999. Cladistic biogeography, 2nd edn. Oxford University Press, Oxford.

Ilves, K. L., and E. B. Taylor. 2007. Evolutionary and biogeographic patterns within the smelt genus *Hypomesus* (Pisces: Osmeridae) in the North Pacific Ocean. *J. Biogeogr.*, in press.

Irwin, T., and G. Fine (tr.). 1996. Parts of Animals, Pp. 104-114 in Aristotle: introductory readings. Hackett Publishing Company, Inc. Indianapolis, Indiana.

Johnson, G. D., and C. Patterson. 1996. Relationships of lower Euteleostean fishes, Pp. 251-332 in M. Stiassny, L. J. Parenti, L. R. Johnson, and G. David, eds. Interrelationships of fishes. Academic Press, Toronto.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21-132 in M. N. Munro, ed. Mammalian protein metabolism, Vol. III. Academic Press, New York.

Kluge, A. G,. and J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18:1-32.

Klyukanov, V. A. 1975. Taxonomy and evolutionary relations between the smelts of the genera *Osmerus* and *Hypomesus* and their dispersion. Zool. J. 54:590-596.

Kudo, T. , and K. Mizuguchi. 2000. Growth of large and small forms of pond smelt *Hypomesus nipponensis* in Lake Kasumigaura, Japan. Fish. Sci. 66:432-441.

Larget, B. and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750-759.

Lecomte, F., and J. J. Dodson. 2003. Role of early life-history constraints and resource polymorphism in the segregation of sympatric populations of an estuarine fish. Evol Ecol. Res. 6:631-658.

Linnaeus, C. 1758. *Systema naturae*. 10th ed. Stockholm.

Lomolino, M. V., B. R. Riddle, and J. H. Brown Biogeography, 3rd edn. Sinauer Associates Inc., Sunderland, MA.

Mayr, E. 1969. Principles of systematic zoology. McGraw-Hill, New York.

———. 1982. The growth of biological thought. Belknap Press of Harvard University Press, Cambridge, MA.

Matthew, W. D. 1915. Climate and evolution. Ann. New York Acad. Sci. 24:171-318.

McAllister, D. E. 1963. A revision of the smelt family, Osmeridae. Bull. Nat. Mus. Can. 191:1-53.

McDowall, R. M. 2002. Accumulating evidence for a dispersal biogeography of southern cool temperate freshwater fishes. J. Biogeogr. 29:207-219.

McLean, J. E., D. E. Hay, and E. B. Taylor. 1999. Marine population structure in an anadromous fish: life-history influences patterns of mitochondrial DNA variation in the eulachon, *Thaleichthys pacificus*. Mol. Ecol. 8:S143-S158.

McLean, J. E., and E. B. Taylor. 2001. Resolution of population structure in a species with high gene flow: microsatellite variation in the eulachon (Osmeridae: *Thaleichthys pacificus*). Mar. Biol. 139:411-420.

Monsch, K. A. 2006. The PhyloCode, or alternative nomenclature: why it is not beneficial to paleontology, either. Acta Paleontologica Polonica 51: 521-524.

Moyle, P. B. 2002. Inland fishes of California. 2nd ed. University of California Press, Berkeley, CA.

Moyle, P. B., B. Herbold, D. E. Stevens, and L. W. Miller. 1992. Life history and status of the delta smelt in the Sacramento-San Joaquin Estuary, California. Trans. Am. Fish. Soc. 121:67-77.

Nelson, G. 1979. From Candolle to Croizat: comments on the history of biogeography. J. Hist. Biol. 11:269-305.

Nelson, G., and P. Y. Ladiges. 1991. Three-area statements: standard assumptions for biogeographic analysis. Syst. Zool. 40:470-485.

Nelson, G., and N. Platnick. 1981. Systematics and biogeography: cladistics and vicariance. Columbia University Press, New York.

Neyman, J. 1971. Molecular studies of evolution: a source of novel statistical problems. Pp. 1-27 in S. S. Gupta and J. Yackel, eds. Statistical decision theory and related topics. Academic Press, New York.

Pitra, C., J. Fickel, E. Meijaard, and P. C. Groves. 2004. Evolution and phylogeny of old world deer. Mol. Phylogenet. Evol. 33:880-895.

Platnick, N. I. and G. Nelson. 1978. A method of analysis for historical biogeography. Syst. Zool. 27:1–16.

Ree, R. H., B. R. Moore, C. O. Webb, and M. J. Donoghue. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59:2299-2311.

Rieppel, O. 2006. The PhyloCode: a critical discussion of its theoretical foundation. Cladistics 22:186-197.

Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Syst. Biol. 46:195-203.

Rosen, D. E. 1978. Vicariant patterns and historical explanations in biogeography. Syst. Zool. 27:159 –188.

Sanmartín, I., H. Enghoff, and F. Ronquist. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol. J. Linn. Soc. 73:345-390.

Sanmartín, I., L. Wanntorp, and R. C. Winkworth. 2007. West Wind Drift revisited: testing for directional dispersal in the southern hemisphere using event-based tree fitting. J. Biogeogr. 34: 398-416.

Saint-Laurent, R., M. Legault, and L. Bernatchez. 2003. Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). Mol. Ecol. 12:315-330.

Schneider, H., F. C. Canavez, I. Sampaio, M. A. M. Moreira, C. H. Tagliaro, and H. N. Seuánez. 2001. Can molecular data place each neotropical monkey in its own branch? Chromosoma 109:515-523.

Simpson, G. G. 1961. Principles of animal taxonomy. Columbia University Press, New York.

Smith, S. A., P. R. Stephens, and J. J. Wiens. 2005. Replicate patterns of species richness, historical biogeography, and phylogeny in holarctic treefrogs. Evolution 59:2433-2450.

Sokal. R. R., and P. H. A. Sneath. 1963. Numerical taxonomy. W. H. Freeman, San Francisco.

Stearly, R. F. 1992. Historical ecology of Salmoninae, with special reference of *Oncorhynchus*, Pp. 622-658 in R. L. Mayden , ed. Systematics, historical ecology and North American freshwater fishes. Stanford University Press, Stanford, CA.

Swanson, C., T. Reid, P. S. Young, and J. J. Cech, Jr. 2000. Comparative environmental tolerances of the threatened delta smelt (*Hypomesus transpacificus)* and introduced wakasagi *(H. nipponensis*) in an altered California estuary. Oecologia 123:384-390.

Taylor E. B., and P. Bentzen.1993. Evidence for multiple origins and sympatric divergence of trophic ecotypes of smelt (*Osmerus*) in northeastern North America. Evolution 47:813–832.

Taylor, E. B., and J. J. Dodson. 1994. A molecular analysis of relationships and biogeography within a species complex of Holarctic fish (genus *Osmerus*). Mol. Ecol. 3:235–248.

Torao, M. 2000. Early developmental phase of wakasagi, *Hypomesus nipponensis*, in Lake Abashiri. Fish. Sci. 66:605-607.

Trenham, P. C., H. B. Shaffer, and P. B. Moyle. 1998. Biochemical identification and assessment of population subdivision in morphologically similar native and invading smelt species (*Hypomesus*) in the Sacramento–San Joaquin estuary, California. Trans. Am. Fish. Soc. 127:417-424.

Väinölä, R. 2003. Repeated trans-Arctic invasions in littoral bivalves: molecular zoogeography of the *Macoma balthica* complex. Mar. Biol. 143:935-946.

Vermeij, G. J. 1991. Anatomy of an invasion: the trans-Arctic interchange. Paleobiology 17:281-307.

———. 2005. From Europe to America: Pliocene to Recent trans-Atlantic expansion of cold-water North Atlantic molluscs. Proc. R. Soc. Lond. Ser. B. Biol. Sci. 272:2545-2550.

CHAPTER 2: ARE *HYPOMESUS CHISHIMAENSIS* AND *H. NIPPONENSIS* (PISCES: OSMERIDAE) DISTINCT SPECIES? A MOLECULAR ASSESSMENT USING COMPARATIVE SEQUENCE DATA FROM FIVE GENES[1]

2.1 INTRODUCTION

2.1.1 Ecotypes and cryptic species

It is particularly challenging to delineate species boundaries for organisms that differ in their life-history characteristics. Difficulties in determining the extent of morphological, ecological and/or genetic differentiation necessary and sufficient to merit the distinction of a Latin binomial are common to all organisms (e.g. Sites and Marshall 2003; Dayrat 2005). The study of taxa that display more than one life-history characteristic presents an additional problem: an alternate life-history found in similar environments in separate geographic areas, i.e. an ecotype (Turesson 1922; Keeley et al. 2005), allows for the possibility that it may have had more than one evolutionary origin (e.g. Mayr 1963), and such ecotypes, if recognized taxonomically, would be polyphyletic.

The phenomenon of multiple ecotypes within species is well documented in many north temperate fishes (Behnke 1972; Bernatchez and Wilson 1998; Taylor 1999). Although molecular analyses have provided great insight into the number of

---

[1] A version of this chapter has been published as Ilves, K. L., and E. B. Taylor. 2007. Are *Hypomesus chishimaensis* and *H. nipponensis* (Pisces: Osmeridae) distinct species? A molecular assessment using comparative sequence data from five genes. *Copeia* 2007:180-185.

origins of ecotypes in many groups (e.g. Bernatchez and Dodson 1990; Taylor and Bentzen 1993; Douglas et al. 2005), there is ongoing debate as to whether life-history variants of even well-studied fishes should be separately classified (e.g. Behnke 1972; Taylor 1999). For those groups where relatively few data are available, it is even more likely that current taxonomic designations incorrectly or incompletely classify diversity (Wheeler et al. 2004; Wilson 2003).

2.1.2 Osmeridae systematics background

The small, elongate, silvery fishes of the Osmeridae family display diverse life-history characteristics. There are marine, anadromous and freshwater forms of these northern hemisphere smelts, which are generally found in near-shore marine and coastal freshwaters throughout the Holarctic (McAllister 1963). Because of their disjunct distributions and general morphological similarity, the relationships among them have been a source of systematic confusion and taxonomic uncertainty since at least the time of Linnaeus. A full understanding of the evolutionary history of this group requires investigation into a number of nested biological questions: interpreting within species divergences that have led to ecologically and phenotypically differentiated forms (e.g. 'dwarf' and 'normal' rainbow smelt *Osmerus mordax*: Lanteigne and McAllister 1983; Taylor and Bentzen 1993); resolving evolutionary relationships within and among the genera; studying the role of geography in shaping their evolution; and addressing problems concerning which higher taxa belong within the family.

A starting point for such an analysis is to focus on recent developments in our understanding of the most speciose genus, *Hypomesus*. In comparison to *Osmerus*, which has been the subject of several morphological, genetic, and biogeographic studies (McAllister 1963; Klyukanov 1975; Luey et al. 1982; Taylor and Dodson 1994; Curry et al. 2004), most systematic studies of *Hypomesus* have been confined to morphological assessments with a relatively narrow regional scope (Chereshnev et al. 2001; Shed'ko 2001; Sidorov and Pichugin 2004). A 1997 revision by Saruwatari et al. described a new species, *H. chishimaensis*, from the Kuril Islands of Japan, bringing the total number of species in the genus to six. Of these six species, three are found in the western Pacific (*H. chishimaensis, H. nipponensis* and *H. japonicus*) and two in the eastern Pacific (*H. pretiosus* and *H. transpacificus*). The eastern range limit of the sixth species, *H. olidus*, is reached at the Pacific and Arctic drainages of North America; however, its southwestern limit is unresolved as being either Korea (McAllister 1963) or the northern Sea of Japan (Chereshnev et al. 2001).

Initially collected from freshwaters on the southernmost Kuril Islands, Kunashir and Iturup, the new species *H. chishimaensis* was distinguished from *H. nipponensis* primarily on the basis of the presence of small teeth in the middle of the posterior portion of the glossohyal, heavier pigmentation of the body and larger eye diameter (Saruwatari et al. 1997). A subsequent morphological study at four Kuril Island lakes, however, failed to detect any significant phenotypic differences between the two putative species, leading the authors to conclude that *H. chishimaensis* is best regarded as an ecological variant of *H. nipponensis* (Sidorov

and Pichugin 2004). This conclusion was also echoed by Chereshnev et al. (2001) in Peter the Great Bay, who found that morphological characteristics of *H. nipponensis* overlap with those thought to distinguish *H. chishimaensis*.

Because of these conflicting morphological assessments, it is necessary to characterize the species genetically relative to the other members of *Hypomesus* in order to gain further understanding of the status of *H. chishimaensis*. There have been many instances of so-called 'cryptic species' identified through molecular, ecological and behavioural analyses (Mayr 1948; Martin and Bermingham 2000; Arnegard et al. 2005), showing that explicit phenotypic differences are not always evident between true biological species. Being indistinguishable morphologically, therefore, does not preclude genetic differentiation between *H. chishimaensis* and *H. nipponensis*, as long as they have been reproductively isolated for enough time for differences to accumulate. Because the two southernmost Kuril Islands, Kunashir and Iturup, from which *H. chishimaensis* was identified, have relatively recent geologic origins, rising above sea level no earlier than the Pliocene (Pietsch et al. 2001) the possibility for differences to arise in these newly colonized habitats may be restricted.

Here I present an analysis of a recently described lacustrine smelt *Hypomesus chishimaensis* (Osmeridae) and its anadromous congener *H. nipponensis*, where I used mitochondrial (mtDNA) and nuclear (nDNA) sequence data to test the hypothesis that *H. chishimaensis* and *H. nipponensis* are separate species. Following the flow-chart proposed by Wiens and Penkrot (2002) for assessing species boundaries with molecular phylogenetic data, species designation

of *H. chishimaensis* would be supported if the individuals assigned to this species based on morphology form a well-supported monophyletic group separate from the individuals of *H. nipponensis*.

2.2 MATERIALS AND METHODS

2.2.1 Taxon sampling and molecular analysis

Morphologically defined samples representing *Hypomesus nipponensis*, *H. chishimaensis*, *H. olidus* and the outgroup *Mallotus villosus* were analyzed for two mitochondrial (cytochrome *b* [cyt*b*] and16S rRNA [16S]) and three nuclear (internal transcribed spacer 2 [ITS2], S7 ribosomal protein, intron 1 [S71], and recombination-activating gene 1 [RAG1]) markers. Thirteen samples of *H. chishimaensis* were from the two southernmost main-chain Kuril Islands, Kunashir (3) and Iturup (4), and one of the Habomai group islands, Zelionyi (3), of Japan, and Sakhalin Island (3), Russia. Six *H. nipponensis* samples came from Hokkaido, Japan, and four *H. olidus* individuals from Kamchatka, Russia (Fig. 1). The outgroup *M. villosus* sample was obtained from Tribune Channel, British Columbia, Canada. This species is the taxon most closely related to *Hypomesus* (K. Ilves, unpubl. data). *Hypomesus olidus*, the most closely related species to *H. chishimaensis* and *H. nipponensis* (K. Ilves, unpubl. data), was included in the study to ensure the markers used are appropriate for distinguishing well-recognized species.

DNA was extracted from ethanol-preserved or frozen tissue using either the Gentra PUREGENE® DNA Purification Kit or the Qiagen DNeasy® Tissue Kit following the manufacturer's instructions. Amplification reactions for all markers contained 50 -300 ng of genomic DNA template. Primers used for the amplification of cyt*b* (cyt*b*2, GluDG), 16S (16Sar, 16Sbr), ITS2 (5.8sr, 28s), S71 (S7RPEX1F, S7RPEX2R) and RAG1 (RAG1F, RAG1R) were obtained from Kocher et al. (1989), Waters et al. (2002), Presa et al. (2002), Chow and Hazama (1998), and Quenouille et al. (2004), respectively. Cyt*b* was amplified in 50 μl reactions containing final concentrations of 800 μM of dNTPs, 800 nM of each primer, 1 unit of Invitrogen™ *Taq* DNA polymerase, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, and 2.5 mM $MgCl_2$ under the following conditions: 95 C for 3 min, 52 C for 1 min and 1 min at 72 C, followed by 5 cycles of 1 min each of 94 C, 52 C and 72 C, and 30 30-sec cycles of 92 C, 55 C and 72 C. A final extension step was performed at 72 C for 5 min. 16S was amplified using a protocol modified from Waters and Cambray (1997): 5 cycles of 94 C (30 sec), 61 C (1 min), 72 C (2 min) followed by 35 cycles of 30 sec each at the same temperatures. The procedure for ITS2 was modified from Presa et al. (2002): 95 C for 5 min, 5 cycles of 95 C (1 min 30 sec), 60 C (2 min), 72 C (3 min) followed by 35 cycles of 30 sec each at the same temperatures. For some samples a 55 C annealing temperature was used. Procedures for amplifying S71 and RAG1 were unmodified from Chow and Hazama (1998) and Quenouille et al. (2004), respectively. PCR products were purified using the Qiagen QIAquick® PCR Purification Kit and sequenced at the NAPS Unit at the University of British Columbia

on an ABI Prism 377 automated sequencer. All sequences have been deposited in GenBank (see Appendix 1 for accession numbers).

2.2.2 Phylogenetic analysis

Sequences were aligned using ClustalX (Thompson et al. 1997) or manually with MacClade v4.06 (Maddison and Maddison 2003) and edited with Se-Al v2.0a11 (Rambaut 1996) or MacClade. The alignments for cyt*b*, 16S and RAG1 were unambiguous. For extra confidence in the 16S rRNA alignment, a secondary structure model presented by Waters et al. (2002) was followed, showing that the few indels occurred within loops. There were many indels in the alignments of ITS2 and S71 making positional homology uncertain in several locations; therefore, 80 and 21 characters were excluded from the two alignments, respectively. PAUP* v.4.0b10 (Swofford 2002) was used to calculate pair-wise distances and perform maximum likelihood (ML) analyses and MrBayes v3.1.1 (Huelsenbeck and Ronquist 2001) was used for the Bayesian estimates of phylogeny.

Separate analyses were conducted by marker as well as by combined mitochondrial, combined nuclear, and all markers combined. Modeltest v3.6 (Posada and Crandall 1998) was run for all data partitions (individual and combined) to select a model of sequence evolution for use in the ML and Bayesian analyses. Posada and Buckley (2004) suggest that the hierarchical Likelihood Ratio Tests (hLRT) implemented in previous versions of Modeltest are inferior to the Akaike Information Criterion (AIC) methods of model testing. We therefore implemented the

model chosen by AIC in our analyses. For the combined analyses this model corresponded to TVM + I (mtDNA), GTR + Γ (nDNA) and GTR + I (all markers).

Bayesian analysis was conducted on all sequences for each marker separately and for the combined datasets with a reduced number of individuals (see below) using the general model chosen by the AIC method in Modeltest, allowing MrBayes to calculate the exact parameter values. Two parallel analyses were run for $5x10^6$ generations with four MCMC chains, a sample frequency of 100 and a burnin of 5,000. For ML analyses on the combined data sets, heuristic searches were conducted with five random replicates of stepwise taxon addition. Confidence in groupings was assessed using bootstrap pseudo-replicates (1,000 for all data combined and 300 for the mtDNA and nDNA partitions), retaining only groupings that appeared with a frequency of at least 50%. We considered Bayesian posterior probabilities of at least 95% (Huelsenbeck and Ronquist 2001) and bootstrap values of at least 70% (Hillis and Bull 1993) to indicate well-supported nodes.

2.3 RESULTS

2.3.1 Sequences

The individual marker Bayesian analyses with all sequences showed that all *H. olidus* individuals either had identical sequences or clustered in 99--100% of the trees, therefore a single individual was used for subsequent combined ML and Bayesian analyses. Similarly, because preliminary Bayesian analysis of each marker with all sequences resulted in a single unstructured clade of *H.*

chishimaensis and *H. nipponensis* individuals in each case and little variation was found among the *H. chishimaensis* individuals (with one variant for cyt*b* and ITS2, two for 16S and RAG1 and four for S71), six were included for the final combined analyses. These six individuals had sequences for all five gene regions and encompassed all of the haplotype diversity apart from one 16S variant. The other seven individuals had sequences for at least one mitochondrial and two nuclear loci. Since they shared haplotypes at these three loci with the other *H. chishimaensis* individuals, obtaining sequences for the remaining two would not have provided any additional information and was therefore deemed unnecessary. All six individuals of *H. nipponensis* were included for the final analyses, with two missing a RAG1 sequence. These final analyses included 15 sequences (14 for RAG1) ranging from 286--425 bp (cyt*b*), 529--548 bp (16S), 339--441 bp (ITS2), 338--678 bp (S71), and 1254--1431 bp (RAG1). Combined mitochondrial, nuclear, and all data analyses contained 973, 2,590, and 3,563 characters, respectively, including indels.

Uncorrected pair-wise differences between the samples show that the markers used in this study have varying degrees of divergence, with greater differences within cyt*b*, ITS2 and S71 than the more slowly evolving 16S and RAG1. Furthermore, for each marker at least one individual assigned to *H. chishimaensis* has an identical sequence to an individual assigned to *H. nipponensis*.

2.3.2 Phylogenetic analyses

The ML and Bayesian analyses of all markers yielded a phylogeny where all *H. chishimaensis* and *H. nipponensis* fell into a single, unstructured monophyletic

group (Fig. 2). The individual gene Bayesian analyses produced the same topology. The only slight variations on this topology were with the ML analysis of mtDNA and the ML and Bayesian analyses of nDNA, which added some poorly supported structure to the *H. chishimaensis*/*H. nipponensis* grouping. In the mtDNA ML phylogeny, one individual assigned to *H. nipponensis* fell outside the rest of the group with a bootstrap support of 69%, while in the nDNA analyses two individuals assigned to *H. chishimaensis* grouped outside the other sequences with bootstrap support and Bayesian posterior probabilities of 54 and 51%, respectively (data not shown). These values fall below accepted standards for supported clades (Hillis and Bull 1993; Huelsenbeck and Ronquist 2001), and it is clear that even with some structure in the group, there is no differentiation among the individuals named *H. chishimaensis* and those named *H. nipponensis* (Fig. 2).

2.4 DISCUSSION

Analyses of five gene regions, both individually and in combination, show no genetic divergence between the newly described smelt *Hypomesus chishimaensis* and its congener *H. nipponensis*. These results together with Wiens and Penkrot's (2002) framework for delimiting species, and the apparent lack of morphological divergence (Chereshnev et al. 2001; Sidorov and Pichugin 2004), strongly suggest that these two named species are, in fact, conspecific.

2.4.1 Relevance of geologic history of Kuril Islands

The Kuril Islands, reaching 1200 kilometres from Hokkaido, Japan to Russia's Kamchatka peninsula (Fig. 1), are volcanic in origin and have a fascinating geological and biological history (Kimura and Tamaki 1985; Bulgakov 1996; Pietsch et al. 2001). The maximum timeframe for the evolution of the populations assigned to *H. chishimaensis* is determined by the origin of the islands on which they are found and events that would allow isolation in freshwater habitats. Although the first of the islands appeared as early as the Upper Cretaceous, the southernmost main-chain islands Kunashir and Iturup, where many of the *H. chishimaensis* samples were taken (Fig. 1), likely arose sometime in the Pliocene and Pleistocene, respectively (Pietsch et al. 2001). After the formation of the southernmost main-chain islands major sea-level regressions in the late Pleistocene connected the terrestrial habitats of the Kuril archipelago, Hokkaido, and Kamchatka, when Hokkaido and Sakhalin were attached to mainland Asia (Briggs 1974; Pietsch et al. 2001). The most recent regression occurred between 10,000 and 30,000 years ago (Pietsch et al. 2001). Assuming *H. nipponensis* invaded the freshwaters of some of these islands prior to this last sea-level regression, the period of isolation resulting from this drop in sea-level may have been too short for detectable morphological and molecular divergence to develop (Behnke 1972).

2.4.2 Taxonomic status of *Hypomesus chishimaensis*

Although there has been relatively little evolutionary time for divergence to occur between the populations assigned to *H. chishimaensis* and *H. nipponensis*,

this explanation for why we see no genetic (or morphological) differences is incomplete. Differentiation into separate ecotypes on comparable timescales is a fairly widespread phenomenon in north temperate fish faunas (Bernatchez and Wilson 1998; Taylor 1999), with examples from smelts (Taylor and Bentzen 1993), sticklebacks (Taylor and McPhail 1999), salmonids (Bernatchez and Dodson 1990) and lampreys (Salewski 2003). In each of these examples there is a considerable body of evidence for quantifiable morphological, ecological, behavioural and/or genetic differences, all of which are lacking for *H. chishimaensis*. A further complication, relevant to many of these other cases, is the possibility that the freshwater populations reflect independent evolution of each lake-resident population (Sidorov and Pichugin 2004). The problem of how to deal with a similar ecological form that has evolved in separate geographic locations has caused taxonomic difficulties in many groups. In Osmeridae in particular, genetic data showed that assigning a single taxonomic name (*Osmerus spectrum*; Lanteigne and McAllister 1983) to a 'dwarf' version of the rainbow smelt *O. mordax* did not accurately reflect its multiple-origin evolutionary history (Taylor and Bentzen 1993).

The assignment of a Linnaean binomial to a taxon has significant theoretical and practical implications in terms of cataloging, understanding and protecting biodiversity (Mayr 1963; Agapow et al. 2004; Isaac et al. 2004); therefore, the process of assigning and retracting such designations is not a trivial matter. Given the lack of morphological and genetic divergence between the populations assigned to *Hypomesus chishimaensis* and *H. nipponensis*, I feel that Chereshnev et al. (2001) and Sidorov and Pichugin (2004) were correct in concluding that *H.*

chishimaensis is a lake-resident life-history type of *H. nipponensis* that does not merit recognition at the species level. Further, assigning the freshwater populations as a subspecies of *H. nipponensis* is not justified because there are no morphological, genetic, or ecological traits that distinguish them, and further, each lake population may have had an independent origin.

Although I have argued that the freshwater populations are not a separate species from the anadromous *H. nipponensis*, there are many interesting questions yet to be answered. There remains a clear need for basic ecological research on these lacustrine populations. In particular, investigations to distinguish between single and multiple origin hypotheses would further clarify their evolutionary history. Moreover, microsatellite analysis between the freshwater and anadromous populations can provide an estimate of the amount of gene flow between the life-history types, which could further contribute to understanding the process of ecological specialization (Schluter 1996; Rundle and Nosil 2005).

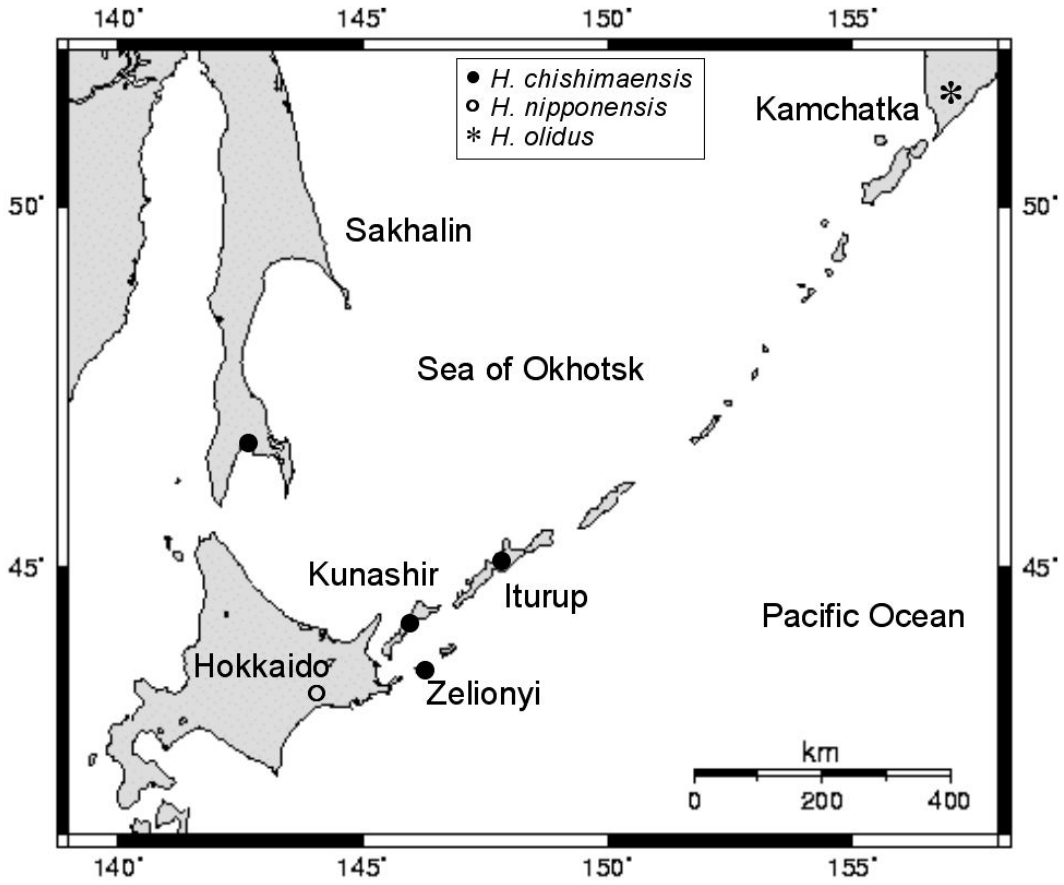


Figure 2.1 Sample locations of *Hypomesus chishimaensis* (•), *H. nipponensis* (○), and *H. olidus* (*****).

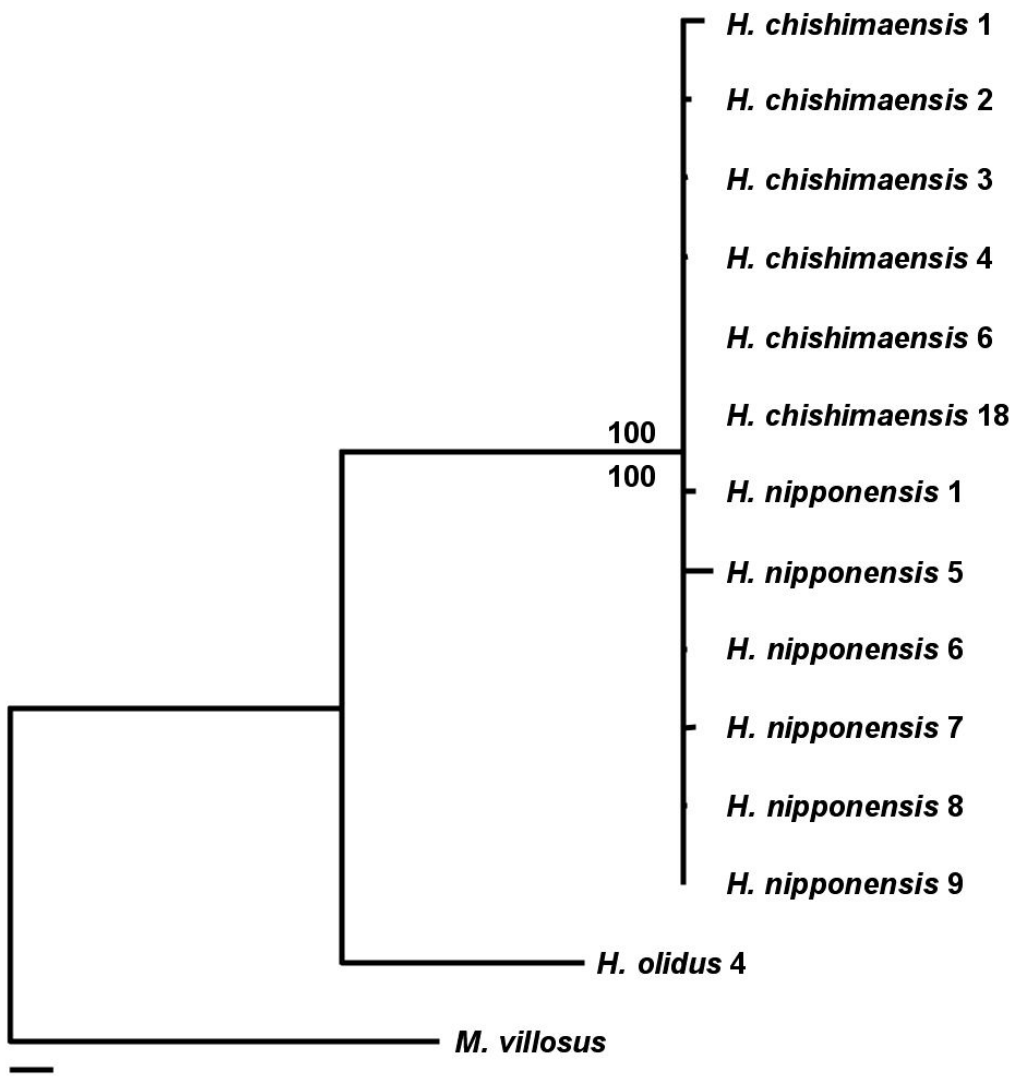


Figure. 2.2 Combined mitochondrial (cyt*b* and 16S) and nuclear (ITS2, S71 and RAG1) phylogeny of *Hypomesus chishimaensis*, *H. nipponensis* and *H. olidus*, with outgroup *Mallotus villosus* based on a GTR + I model of sequence evolution. Numbers following taxon names identify the individual fish used for the final analyses (see Results). The number above the *H. chishimaensis*/*H. nipponensis* node represents the ML bootstrap value (n=1,000), while that below the node represents the Bayesian posterior probability (%; 45,000 trees). Scale bar represents 10 changes.

2.5 REFERENCES

Agapow, P.-M., O. R. P. Bininda-Emonds, K. A. Crandall, J. L. Gittleman, G. M. Mace, J. C. Marshall, and A. Purvis. 2004. The impact of species concept on biodiversity studies. Quart. Rev. Biol. 79:161-179.

Arnegard, M. E., S. M. Bogdanowicz, and C. D. Hopkins. 2005. Multiple cases of striking genetic similarity between alternate electric fish signal morphs in sympatry. Evolution 59:324-343.

Behnke, R. J. 1972. The systematics of salmonid fishes in recently glaciated lakes. J. Fish. Res. Board Can. 29:639-671.

Bernatchez, L., and J. J. Dodson. 1990. Allopatric origin of sympatric populations of lake whitefish (*Coregonus clupeaformis*) revealed by mitochondrial DNA restriction analysis. Evolution 44:1263-1271.

Bernatchez, L., and C. C. Wilson. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. Mol. Ecol. 7:431-452.

Briggs, J. C. 1974. Marine Zoogeography. McGraw-Hill. Inc, Toronto.

Bulgakov, R. 1996. Reconstruction of Quaternary history of Southern Kuril Islands. J. Coastal Res. 12:930-939.

Chereshnev, I., A. A. V. Shestakov, and S. V. Frolov. 2001. On the systematics of species of the genus *Hypomesus* (Osmeridae) of Peter the Great bay, Sea of Japan. Russ. J. Mar. Biol. 27:296-302.

Chow, S., and K. Hazama. 1998. Universal PCR primers for S7 ribosomal protein introns in fish. Mol. Ecol. 7: 1247-1263.

Curry, R. A., S. L. Currie, L. Bernatchez, and R. Saint-Laurent. 2003. The rainbow smelt, *Osmerus mordax*, complex of Lake Utopia: threatened or misunderstood? Env. Biol. Fish. 69: 153-166.

Dayrat, B. 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85:407-415.

Douglas, M. R., P. C. Brunner, and M. E. Douglas. 2005. Evolutionary homoplasy among species flocks of central alpine *Coregonus* (Teleostei: Salmoniformes). Copeia 2005:347-358.

Hillis, D. M., and J. J. Bull. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42:182-192.

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

Isaac, N. J., B. J. Mallet, and G. M. Mace. 2004. Taxonomic inflation: its influence on macroecology and conservation. Trends Ecol. Evol. 19:464-469.

Keeley, E. R., E. A. Parkinson, and E. B. Taylor. 2005. Ecotypic differentiation of native rainbow trout (*Oncorhynchus mykiss*) populations from British Columbia. Can. J. Fish. Aquat. Sci. 62:1523-1539.

Kimura, G., and K. Tamaki. 1985. Tectonic framework of the Kuril Arc since its initiation, Pp. 641-676 in N. Nasu, K. Kobayashi, S. Uyeda, I. Kushiro, and H. Kagami, eds. Formation of active ocean margins. Terra Scientific Publishing Co., Tokyo.

Klyukanov, V. A. 1975. The systematic position of the Osmeridae in the order Salmoniformes. J. Ichthyol. 15: 1-17.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86: 6196-6200.

Lanteigne, J., and D. E. McAllister. 1983. The pygmy smelt, *Osmerus spectrum* Cope, 1870, a forgotten sibling species of eastern North American fish. Syllogeus 45:1-32.

Luey, J. E., C. C. Krueger and D. R. Schreiner. 1982. Genetic relationships among smelt, genus *Osmerus*. Copeia 1982:725-728.

Maddison, D. R., and W. P. Maddison. 2003. MacClade 4.06: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA.

Martin, A. P., and E. Bermingham. 2000. Regional endemism and cryptic species revealed by molecular and morphological analysis of a widespread species of Neotropical catfish. Proc. R. Soc. Lond. B. 267:1135-1141.

Mayr, E. 1948. The bearing of the new systematics on genetical problems the nature of species. Adv. Genet. 2:205-237.

Mayr, E. 1963. Animal Species and Evolution. Belknap Press, Cambridge, MA.

McAllister, D. E. 1963. A revision of the smelt family, Osmeridae. Bull. Nat. Mus. Can. 191:1-53.

Pietsch, T. W., K. Amaoka, D. E. Stevenson, E. L. MacDonald, B. K. Urbain, and J. A. López. 2001. Freshwater fishes of the Kuril Islands and adjacent regions. Species Diversity 6:133-164.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53:793-808.

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

Presa, P., B. G. Pardo, P. Martinez, and L. Bernatchez. 2002. Phylogeographic congruence between mtDNA and rDNA ITS markers in Brown Trout. Mol. Biol. Evol. 19:2161-2175.

Quenouille, B., E. Bermingham, and S. Planes. 2004. Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. Mol. Phyl. Evol. 31: 66-88.

Rambaut, A. 1996. Se-Al: a manual sequence alignment editor, v.2.0a11. Available from http://tree.bio.ed.ac.uk/software/seal/.

Rundle, H. D., and P. Nosil. 2005. Ecological speciation. Ecol. Lett. 8:336-352.

Salewski, V. 2003. Satellite species in lampreys: a worldwide trend for ecological speciation in sympatry? J. Fish Biol. 63:267-279.

Saruwatari, T., J. A. López, and T. W. Pietsch. 1997. A revision of the osmerid genus *Hypomesus* Gill (Teleostei: Salmoniformes), with the description of a new species from the Southern Kuril Islands. Species Diversity 2:59-82.

Schluter, D. 1996. Ecological speciation in postglacial fishes. Proc. Roy. Soc. Lond. B. 351:807-814.

Shed'ko, S. V. 2001. On species composition of smelts (Osmeridae) in waters of Primor'e. J. Ichthyol. 41:164-167.

Sidorov, L. K., and Y. Pichugin. 2004. Morphological traits of lacustrine forms of smelts of the genus *Hypomesus* (Salmoniformes) from the Southern Kurils. J. Ichthyol. 44:433-443.

Sites, J. W., Jr., and J. C. Marshall. 2003. Delimiting species: a Renaissance issue in systematic biology. Trends Ecol. Evol. 18:462-470.

Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (* and other methods), Version 4.0b10. Sinauer Associates, Sunderland, MA.

Taylor, E. B. 1999. Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation. Rev. Fish Biol. Fish. 9:299-324.

Taylor, E. B., and P. Bentzen. 1993. Evidence for multiple origins and sympatric divergence of trophic ecotypes of smelt (*Osmerus*) in northeastern North America. Evolution 47:813-832.

Taylor, E. B., and J. J. Dodson. 1994. A molecular analysis of relationships and biogeography within a species complex of Holarctic fish (genus *Osmerus*). Mol. Ecol. 3:235–248.

Taylor, E. B., and J. D. McPhail. 1999. History of an adaptive radiation in species pairs of threespine sticklebacks (*Gasterosteus aculeatus*): insights from mitochondrial DNA. Biol. J. Linn. Soc. 66:271-291.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX-Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 25:4876-4882.

Turesson, G. 1922. The species and the variety as ecological units. Heriditas 3:100-113.

Waters, J. M., and J. A. Cambray. 1997. Intraspecific phylogeography of the Cape galaxias from South Africa: evidence from mitochondrial DNA sequences. J. Fish Biol. 50:1329-1338.

Waters, J. M., T. Saruwatari, T. Kobayashi, I. Oohara, R. M. McDowall, and G. P. Wallis. 2002. Phylogenetic placement of retropinnid fishes: data set incongruence can be reduced by using asymmetric character state transformation costs. Syst. Biol. 51:432-449.

Wheeler, Q. D., P. H. Raven, and E. O. Wilson. 2004. Taxonomy: impediment or expedient. Science 303:285.

Wiens, J. J., and T. A. Penkrot. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51:69-91.

Wilson, E. O. 2003. The encyclopedia of life. Trends Ecol. Evol. 18:77-80.

CHAPTER 3: EVOLUTIONARY AND BIOGEOGRAPHIC PATTERNS WITHIN THE SMELT GENUS *HYPOMESUS* (PISCES: OSMERIDAE) IN THE NORTH PACIFIC OCEAN[2]

3.1 INTRODUCTION

As an important evolutionary centre for marine diversity (Briggs 2003), the temperate and boreal North Pacific Ocean offers an ideal setting for conducting comparative biogeographic analysis (Amano and Vermeij 2003). In contrast to the focus on diversity 'hotspots' in the southern oceans (e.g. Briggs 2000, 2004; Santini and Winterbottom 2002) and biogeographic studies on North Atlantic taxa (e.g. Wares and Cunningham 2001; Addison and Hart 2005; Vermeij 2005), there has been relatively little recent work exploring patterns of diversity across the entire North Pacific (although see Amano et al. 1993 Collins et al. 1996; Stepien et al. 2000; Hyde and Vetter 2007).

3.1.1 Trans-Pacific distributions

A long-recognized distributional pattern in the North Pacific, however, is the existence of amphi-or trans-Pacific taxa (e.g. Andriashev 1939; Golikov and Tzvetkova 1972; Briggs 1974), where a taxon is found in the eastern and western

[2] A version of this chapter has been accepted for publication. Ilves, K. L., and E. B. Taylor. 2007. Evolutionary and biogeographic patterns within the smelt genus *Hypomesus* (Pisces: Osmeridae) in the North Pacific Ocean. J. Biogeogr., in press.

Pacific, but is absent from the northern region. This distribution has been documented in a variety of taxonomic groups, including fishes (e.g. Andriashev 1939), crustaceans (e.g. Schweitzer 2001), molluscs (e.g. Amano and Vermeij 1998, 2003), polychaetes (Uschakov 1971), and mammals (Deméré et al. 2003).

Explanations for these disjunct distributions have generally postulated that changes in climate compressed the range of a more widely distributed ancestor during periods of global cooling, thereby facilitating differentiation and/or allopatric speciation on opposite sides of the Pacific (Andriashev 1939; McAllister 1963; Briggs 1974; Amano et al. 1993). Additional hypotheses, such as changes in oxygen distribution (White 1987), long-distance dispersal (Rosenblatt and Waples 1986) and reinterpretations of the earth's geologic history (McCarthy 2003, 2005) have also been suggested to explain amphi-Pacific distributions.

3.1.2 Centres of origin

In addition to questions about the timing and possible causes of amphi-Pacific distributions, identifying the area of origin of amphi- or pan-Pacific taxa within the North Pacific has also been of interest. For many organisms, it is assumed that the area that currently houses the highest diversity is the area of origin (e.g. Briggs 1974), although this assumption often is not tested within a phylogenetic framework, a prerequisite to understanding a group's biogeography. Furthermore, many analyses of groups that have amphi- and/or pan-Pacific distributions have focused on different questions and have therefore had a narrow or incomplete taxonomic and/or geographic focus, or have not explicitly linked phylogeny and geography,

thereby making it difficult to make basin-level inferences of origin from the results. Exceptions to this general lack of data are the large body of work on intertidal molluscs (e.g. Amano et al. 1993; Titova 1994; Collins et al. 1996; Reid et al. 1996) and recent studies on Pacific perciform fishes (Bernardi and Bucciarelli 1999; Stepien et al. 2000; Kai et al. 2003; Hyde and Vetter 2007)

3.1.3 Study system

Here I conduct a molecular phylogenetic analysis of the pan-Pacific smelt genus *Hypomesus* Gill (1862) [Pisces: Osmeridae] to further our understanding of evolution within this group. I also explore general questions regarding amphi- and pan-Pacific taxa and important areas of origin in the North Pacific Ocean. With six currently recognized species (Eschmeyer 2006) *Hypomesus* is the most species rich genus of the northern hemisphere smelts, Osmeridae, a family of small, silvery lower euteleost fishes found in near-shore marine and coastal freshwaters throughout the Holarctic (McAllister 1963). Of these six species, three are found in the western Pacific [*H. chishimaensis* Saruwatari et al. 1997*, H. nipponensis* McAllister (1963) and *H. japonicus* Brevoort (1856)], two in the eastern Pacific [*H. pretiosus* Girard (1854) and *H. transpacificus* McAllister (1963)], and one [*H. olidus* Pallas (1814)] has a northern Pacific and Arctic distribution (Fig. 1).

Hypomesus has a complicated taxonomic history. Original sub-species classifications *H. pretiosus pretiosus, H. p. japonicus*, *H. transpacificus transpacificus*, and *H. t. nipponensis*, with *H. olidus* at full species rank, imply sister relationships between species on opposite sides of the North Pacific Ocean

(McAllister 1963). If these classifications reflect ancestry, they require at least two replicate divergences across the ocean basin. Subsequent revisions led to the elevation of all taxa to full species status (Klyukanov 1970; Saruwatari et al. 1997). Saruwatari et al. (1997) also reinforced the pre-eminence of lateral line scale-counts in ascertaining relationships, as suggested by McAllister (1963), by introducing the idea of species-groups based on these counts. A 'high' scale-count group contained *H. japonicus* and *H. pretiosus*, *H. nipponensis* and *H. transpacificus* were classified in a 'low' scale-count group and *H. olidus* was placed in its own group. These relationships reflected the original sub-species designations. Saruwatari et al. (1997) also identified a new species from the freshwaters of the Kuril Islands, *H. chishimaensis*, which was placed in the 'nipponensis' grouping. Recent morphological (Sidorov and Pichugin 2004) and genetic (Ilves and Taylor 2007) work failed to detect any differences between the populations assigned to this new species and those assigned to *H. nipponensis*; therefore, in this study we consider *H. chishimaensis* to be a synonym of *H. nipponensis*.

3.1.4 Justification of methods

Inferring ancestral states is a rapidly developing field in phylogenetic analysis, with parsimony (Maddison and Maddison 2003, 2006), maximum likelihood (ML, Schluter et al. 1997; Ree et al. 2005; Maddison and Maddison 2006) and Bayesian (Bollback 2006) methods available. Methods in which geographic area is mapped onto a phylogeny and optimized on nodes, are frequently used (e.g., Smith et al. 2005; Baker et al. 2006; Jones et al. 2006) but have also been criticized for a

number of reasons, particularly because distributional area is not inherited in the same manner as Mendelian characters (Ree et al. 2005). Further, these particular approaches, which assume a dispersalist or centre of origin mechanism have been much criticized in the vicariance biogeography literature (e.g. Croizat et al. 1974; Nelson and Platnick 1981; Ebach 1999; Humphries and Parenti 1999).

The decision of whether to follow a strict dispersalist or a vicariance biogeography approach for a particular study can be made in several ways, individually or in combination:

1. An ideological attachment to one idea over the other would lead to the choice and exclusion of particular methods.

2. There may be *a priori* information that makes one scenario biologically and/or geologically more likely for the group in question (e.g. known timing of divergences correlates to fragmentation of continents [vicariance]; volcanic origin of islands [dispersal]).

3. *A posteriori* information may suggest one mechanism was more important for generating the distributions of the group in question (e.g. a phylogeny shows one area occupied only by most recently diverged species [dispersal]).

Early hypotheses of Osmeridae evolution (McAllister 1963) developed from a centre of origin perspective, and we therefore think it is appropriate to test those hypotheses from that framework (c.f. point 1 above). Furthermore, I agree with McDowall (2002), de Queiroz (2005), Cowie and Holland (2006), and others, that in addition to vicariance, dispersal is and has been an important mechanism in

generating distributions and that a widespread ancestor is unlikely in many cases, such as when a taxon is composed of geographically restricted taxa (Bremer 1992). As this is the situation with *Hypomesus*, I do not feel that a widespread ancestor for this genus should necessarily be the null hypothesis (c.f. point 2 above). Finally, cladistic biogeographic approaches, which aim to discern bifurcating relationships between areas from taxon cladograms through a number of methods (e.g. Nelson and Platnick 1981; Zandee and Roos 1987; Page 1988; Nelson and Ladiges 1991, 1996; Humphries and Parenti 1999, and references therein), do not capture the complexity of areas that undergo repeated and/or differential fragmentation through time (Donoghue and Moore 2003). I do not, however, start from the assumption that vicariance was unimportant in the history of *Hypomesus*, and therefore also attempt to model the biogeographic history of the genus in a maximum likelihood framework, correlate the timing of divergences within the genus to known cooling events in the Cenozoic history of the North Pacific Ocean, and compare these divergence times to those of other similarly distributed organisms.

To address the question of area of origin, I use two basic approaches: a dispersalist, centre of origin approach, and a newly developed ML modeling method that incorporates both dispersal and vicariance (Ree et al. 2005). The first parsimony-based approach, implemented in Mesquite v1.12 (Maddison and Maddison 2006), yields several testable predictions (Hypotheses A-D; Fig. 2):

Hypothesis A: The proposal by McAllister (1963), subsequently supported by Saruwatari et al. (1997), requires two replicate divergences

across the North Pacific to generate the scale-count relationships between putative amphi-Pacific sister-species. In a parsimony framework, the expected phylogeny, provided this phenotype reflects ancestry, shows sister-relationships between species with similar scale-counts (Fig. 2A).

Hypothesis B: What we consider an *a priori* most parsimonious explanation (Fig. 2B), consistent with the idea of range compression due to climate cooling, begins with a northern ancestor followed by a single split across the North Pacific and differentiation on each side.

Hypothesis C: Eastern Pacific origin of *Hypomesus*, followed by dispersal to northern and western areas (Fig. 2C).

Hypothesis D: Western Pacific origin of *Hypomesus*, followed by dispersal to northern and eastern areas (Fig. 2D).

The ML approach in LAGRANGE, on the other hand, incorporates phylogenetic and palaeogeographic information about the timing and extent of connections between different areas and calculates fractional likelihoods for different biogeographic scenarios at each node of the phylogenetic tree (Ree et al. 2005). This approach may yield both dispersalist and vicariance interpretations, which can then be compared to patterns seen in other similarly distributed taxa.

3.1.5 Aims

From a phylogenetic analysis of five gene regions (two mitochondrial, three nuclear) with multiple individuals per species, I aim to (1) clarify the evolutionary relationships within a genus of systematically problematic fishes, (2) determine whether the sub-species designations based on scale-counts reflect ancestry, or instead whether the shared phenotype is a result of parallel evolution, (3) use several methods to test different biogeographic hypotheses of *Hypomesus* evolution to determine the likely area of origin of this genus, and (4) put these results in a comparative framework to contribute towards an understanding of biogeographic relationships in the North Pacific.

3.2 MATERIALS AND METHODS

3.2.1 Taxon sampling

Samples used (Table 3.1) represent the five *Hypomesus* species and the outgroup *Mallotus villosus* Müller (1776) [a species within the Osmeridae, outside *Hypomesus*, K. Ilves unpubl. data based on phylogenetic analysis of five gene regions]. Samples for "*H. chishimaensis*" were obtained from the two southernmost main chain Kuril Islands (Kunashir and Iturup), one of the Habomai group islands (Zeliyoni) of Japan, and Sakhalin Island, Russia. *H. nipponensis* and *H. japonicus* samples came from two locations on Hokkaido, Japan, and *H. olidus* samples originated from Kamchatka, Russia and Chignik, Alaska (Fig. 3.1). The two eastern

Pacific species were represented by *H. pretiosus* from Sumner Strait, Alaska, USA, two locations in British Columbia, Canada (Wreck Beach, Vancouver, and Bamfield, west coast of Vancouver Island), and from the San Juan Islands, Washington, USA and by *H. transpacificus* from the Sacramento River, California, USA. The two *Mallotus villosus* outgroup individuals were from locations in British Columbia, Canada: the central coast and Trevor Channel on the west coast of Vancouver Island.

GenBank Accession numbers for all sequences used in this study are provided in Appendix 2. Phylogenetic trees from analysis of the mitochondrial, nuclear, and all data combined have been submitted to TreeBase (S1807 [study], M3307-M3309 [mtDNA, nDNA, allDNA, respectively]).

3.2.2 DNA sequence data

Sequence data for two mitochondrial and three nuclear gene regions were obtained for multiple individuals of the five species within the genus (Table 3.1). These regions included mitochondrial (mtDNA) protein-coding cytochrome *b* [cyt*b*; Kocher et al. 1989] and large ribosomal subunit [16S; Waters et al. 2002] and nuclear (nDNA) internal transcribed spacer [ITS2; Presa et al. 2002], first intron of the ribosomal S7 protein [S71; Chow and Hazama 1998] and the protein-coding recombination activating gene [RAG1; Quenouille et al. 2004]. Use of these markers in a previous study on *Hypomesus* showed them to have varying degrees of divergence, making their combined use fruitful for phylogenetic studies (Ilves and

Taylor 2007). DNA extraction, PCR and sequencing protocols were performed as described in Ilves and Taylor (2007).

3.2.3 Phylogenetic analyses

Sequences were aligned using ClustalX (Thompson et al. 1997) or manually with MacClade v4.06 (Maddison and Maddison 2003) and edited with Se-Al v2.0a11 (Rambaut 1996) or MacClade v4.06 (Maddison and Maddison 2003). The alignments for cyt*b*, 16S and RAG1 were unambiguous. For extra confidence in the 16S rRNA alignment, a secondary structure model presented by Waters et al. (2002) was followed, which showed that the few indels occurred within loop regions. Indels in the ITS2 and S71 alignments made positional homology uncertain in several locations; therefore, 104 and 13 characters were excluded from the two alignments, respectively. PAUP* v4.0b10 (Swofford 2002) was used to calculate pairwise distances, perform maximum likelihood (ML), parsimony, and neighbour-joining (NJ) analyses and conduct Shimodaira-Hasegawa tests (SH tests) for comparing tree topologies. MrBayes v3.1.1 (Huelsenbeck and Ronquist 2001) was used for Bayesian estimates of phylogeny.

Phylogenetic analyses were conducted by individual locus as well as by combined mitochondrial, combined nuclear, and all data combined. Modeltest v3.6 (Posada and Crandall 1998) was run for all data partitions (individual and combined) to select a model of sequence evolution. The model chosen by the Akaike Information Criterion (AIC) method of model testing was implemented for the

analyses (Posada and Buckley 2004), which corresponded to TVM + I (mtDNA), GTR + Γ (nDNA) and GTR + I (all data).

Due to the large amount of sequence data (Table 3.1), NJ and Bayesian analyses were run for all sequences for all of the data partitions to identify individuals of each species that represent a large range of the intraspecific variation to be included for more rigorous ML and Bayesian analyses. This first step of analysis included 1000 NJ bootstrap replicates of all sequence for each gene separately, and Bayesian analyses of all sequences for all genes combined ($5x10^5$ generations, burnin 500, 4500 sampled trees). Apart from unresolved relationships within *H. transpacificus* for the 16S region and *H. pretiosus* for the ITS2 region, all species were monophyletic for each data partition. To represent intraspecific variation, therefore, the three most divergent individuals within each species clade for cyt*b*, 16S, ITS2, and S71 were chosen for inclusion in more thorough analyses (see Results). Only two *H. olidus* individuals were included as there was very little intraspecific variation for these genes. RAG1 sequences were not available for all individuals, so to avoid a situation of missing data, in several cases a sequence from a conspecific was substituted. Similarly, for the final analyses of ITS2, a single substitution was made for *H. olidus*. These substitutions are indicated in Appendix 2.

Bayesian analysis was conducted on all sequences for each gene region separately, for the combined datasets with all individuals, and for the combined datasets with a reduced number of individuals (see below) using the general model chosen by the AIC method in Modeltest, allowing MrBayes to calculate the exact parameter values. Two parallel analyses were run for 10^6 (single locus) or $2x10^6$

(combined datasets) generations with four MCMC chains, a sample frequency of 100 and a burnin of 3000. As the stationarity of likelihoods was always observed by 2000 generations, the sampled trees were found well after stabilization.

For ML analyses on the combined datasets, heuristic searches were conducted with six random replicates of stepwise taxon addition. Confidence in groupings was assessed using 1000 bootstrap pseudo-replicates, retaining groupings that appeared with a frequency of at least 50%. To confirm concordance with the Bayesian analyses, ML analyses by locus were also conducted (100 bootstrap pseudo-replicates).

Parsimony analyses were also conducted on all reduced data sets: by gene region (branch and bound searches), and combined datasets (heuristic searches, simple stepwise addition), and were assessed by 1000 bootstrap pseudo-replicates.

Finally, to assess whether or not the species-groups based on lateral line scale counts reflect ancestry within *Hypomesus*, the high and low scale-count groups were mapped onto the resulting phylogenies.

3.2.4 Evaluating biogeographic scenarios

I used two basic approaches to statistically evaluate biogeographic hypotheses for the evolution of *Hypomesus*: dispersalist, which optimizes areas onto nodes like a character (Maddison and Maddison 2003, 2006) and a maximum likelihood method for inferring ancestral areas (Ree et al. 2005). The first approach involved parsimony reconstruction of ancestral areas. First, using the SH test

(Shimodaira and Hasegawa 1999) implemented in PAUP* v4.0b10 (Swofford 2002), I compared the expected topologies under several biogeographic scenarios (Fig. 3.2) with topologies generated from the molecular analyses. This test compares the likelihoods of different tree topologies for a particular dataset and can assess whether one topology is significantly better than another (Felsenstein 2004). The parsimony reconstruction method in Mesquite v1.12 (Maddison and Maddison 2006) with a step-matrix describing transitions between areas was also used to infer the area of origin for *Hypomesus* with *H. olidus* both restricted to the northern Pacific and polymorphic for the three areas. The step-matrix assumed coastal dispersal, so a transition between either the west or east Pacific to the northern Pacific was one step, while dispersal from west to east or vice-versa was two steps.

For the second basic approach to evaluate different biogeographic scenarios I implemented the ML method in the LAGRANGE program described in Ree et al. (2005). This method improves on the character mapping approaches described above in several ways. Most significantly, LAGRANGE allows different modes of inheritance so that daughter species do not necessarily inherit identical geographic ranges, multiple character states are permitted, and palaeogeographic information about when dispersal between areas was possible is integrated into the analysis (Ree et al. 2005). To conduct this analysis the following parameters are required: an ultrametric phylogenetic tree (tree that assumes a molecular clock where all tips are contemporaneous) with branch lengths, a set of areas assigned to each taxon, divergence time of root node, times and probabilities when dispersal between areas was possible, and probabilities of lineage dispersal (from) and extinction (within)

areas (Ree et al. 2005). By integrating all this information, LAGRANGE calculates fractional likelihoods of ancestral areas for each node of the tree.

My implementation of this analysis was as follows. I chose a single individual to represent each species (see Appendix 2) and, based on all of the sequence data combined, produced an ultrametric tree using the Bayesian analysis program Beast v.1.4.1 (Drummond and Rambaut 2003) under a lognormal relaxed molecular clock (Drummond et al. 2006) and a Yule pure birth model of speciation. The results of three runs of 5×10^7 generations were compiled. Resulting divergence times for each node were calibrated based on a divergence time of 15 million years ago (mya), the origin of the Sea of Japan (Itoh *et al*. 1997), for *H. japonicus* and the two eastern Pacific species, under the assumption that the ancestor of *H. japonicus* evolved in this basin (Table 2; see Discussion). *H. japonicus* and *H. nipponensis* were assigned a western Pacific distribution and *H. pretiosus* and *H. transpacificus* an eastern Pacific distribution. Simulations were run with *H. olidus* restricted to the North Pacific and Arctic and as polymorphic for all regions of the Pacific and Arctic. The outgroup *M. villosus* was excluded from this analysis because it is not the sister group to *Hypomesus*. The dispersal connections between areas were parameterized based on documented land bridge connections and cooling events of the Cenozoic (Fig. 3.3). Connections between areas have associated probabilities of dispersal success through time, which is different from the probability of dispersal from one area to another, as discussed below. A connection between the Arctic and North Pacific oceans was allowed with a probability of 1.0 between 7 – 2 mya, as there was no Bering land bridge during this period prior to the onset of the Pleistocene

glacial cycles (Marincovich and Gladenkov 1999). Dispersal between the northern Pacific and the western and eastern Pacific was deemed possible during relatively warm periods based on Zachos et al. (2001): the probability of dispersal success was 1.0 between 25 – 15 mya, corresponding to the Miocene climatic optimum, and then decreased to 0.5 at 15 mya where it remained until 5 mya, with a final linear decrease to 0 between 5 – 2 mya [Pliocene cooling to the beginning of the Pleistocene glacial period]. Allowing connections during the early Pliocene warm period (5 – 3 mya; Ravelo et al. 2004) with a linear decrease to zero between 3 – 2 mya did not change the results (data not shown). Two scenarios of dispersal between the western and eastern North Pacific were tested, one allowing minimal dispersal and one assuming that all dispersal between the east and west had to go through the north Pacific roughly along the arc formed by the Aleutian and Kuril island chain, a scenario I believe more plausible for near-shore anadromous fishes. When dispersal across the Pacific was permitted, the periods corresponded to the same warm periods as described for northern – (western/eastern) connections, although with half the probability, and with a decreasing probability from 0.5 – 0 between 15 – 5 mya (Fig. 3). Following Ree et al. (2005), several combinations of dispersal (λ_D) and extinction (λ_E) probabilities were run ($\lambda_D = 0.09$, $\lambda_E = 0.01$; $\lambda_D = 0.009$, $\lambda_E = 0.001$; $\lambda_D = 0.05$, $\lambda_E = 0.05$; $\lambda_D = 0.005$, $\lambda_E = 0.005$; $\lambda_D = 0.01$, $\lambda_E = 0.09$; $\lambda_D = 0.001$, $\lambda_E = 0.009$). For each λ_D, λ_E combination I ran 10^5 iterations. This was repeated five times and for each combination the result with the highest likelihood was compiled for interpretation.

3.2.5 Comparisons with other North Pacific taxa

As a supplement to broad comparisons with published literature on pan- and trans-Pacific taxa, for a more direct approach we identified two groups of fishes, the surfperches (Embiotocidae) and thornyhead rockfishes (*Sebastolobus*), with amphi-Pacific relationships for which cyt*b* sequence data were available. Based on cyt*b* and 16S sequences, in the Embiotocidae, one clade within the family showed a sister-relationship between two Japanese genera (the monotypic *Ditrema* and *Neoditrema*) and a clade of genera from the eastern Pacific (with the monotypic *Hypsurus* as most basal; Bernardi and Bucciarelli, 1999). In *Sebastolobus*, Stepien et al. (2000) found the western Pacific *S. macrochir* to be basal to the two eastern Pacific species (*S. alascanus* and *S. altivelis*) using mtDNA control region sequence divergence. Cyt*b* sequences were available for all three surfperch species (GenBank accession numbers AF159340, AF159341, AF159335; Bernardi and Bucciarelli, 1999), and two *Sebastolobus* species, *S. macrochir* (GenBank accession number AB096136; Kai et al. unpubl.) and *S. alascanus* (GenBank accession number AF031497; Rocha-Olivares et al. 1999). To compare cyt*b* sequence divergence among amphi-Pacific taxa in these two groups with that of *Hypomesus*, uncorrected pairwise differences were calculated in PAUP* v4.0b10 (Swofford 2002).

3.3 RESULTS

3.3.1 Sequences

The three individuals chosen per species for the final phylogenetic analyses were representative of the intra-specific diversity for each locus in all cases, including *H. transpacificus* and *H. pretiosus* where initial analyses were unable to determine the relationships among individuals of these species for 16S and ITS2, respectively (see Methods). These final analyses included 15 sequences (14 for RAG1) ranging from 286-425 bp (cyt*b*), 531-548 bp (16S), 287-441 bp (ITS2), 432-706 bp (S71), and 916-1431 bp (RAG1). Including indels, the combined mtDNA, nDNA and all data partitions included for analysis (117 characters excluded from the nDNA and all data partitions due to questionable alignment) contained 973, 2615, and 3588 characters, respectively.

3.3.2 *Hypomesus* phylogeny

The combined mtDNA and nDNA topologies produced through ML and Bayesian analyses (Fig. 3.4) initially appeared to differ in the placement of *H. japonicus*. The mtDNA tree indicated a close relationship between *H. japonicus* and the *H. nipponensis*-*H. olidus* grouping, whereas the nDNA data resulted in a sister relationship between *H. japonicus* and the two eastern Pacific species (Fig. 3.4). To determine whether or not these differences represented a true conflict between the mitochondrial and nuclear data, SH tests were performed in each data set,

comparing the topology generated separately under nDNA and mtDNA data to a constraint tree that had the alternative topology (i.e. a tree constrained to the nDNA-based topology was tested with the mtDNA data, and vice-versa). With the mtDNA dataset, there was no significant difference between the nDNA-based constraint tree and that generated with the mtDNA data (-2355.5 vs. -2358.1, $P = 0.21$); however, with the nuclear dataset, the nDNA tree was significantly better than the tree constrained to the mtDNA-based topology (-5669.5 vs. -5685.9, $P < 0.02$). The likelihoods of the topologies generated with nDNA and mtDNA are not significantly different from one another when tested against the mtDNA dataset. By contrast, the nDNA-generate topology is significantly better than the mtDNA-generated topology when tested against the nDNA data, which suggests that there is no true conflict between the datasets.

Combining all data yielded a tree with the same topology as the nDNA dataset, with very high support for a grouping of *H. japonicus* with the eastern Pacific species, and increased support for a sister-relationship between *H. nipponensis* and *H. olidus* relative to the mtDNA and nDNA partitions (Fig. 3.5). After mapping the high (*) and low (+) scale-count groups onto the phylogeny, it was clear that the trans-Pacific species with similar scale-counts were not monophyletic (Figs. 3.4 and 3.5).

Analyses by individual gene partition produced several interesting results. First, it appears that 16S evolves too slowly to provide a good estimate of phylogeny for the timeframe of evolution of *Hypomesus* as *H. pretiosus* and *H. transpacificus* individuals grouped as a polytomy and support levels for other relationships were

also relatively low in both ML and Bayesian analyses (data not shown). Furthermore, support for a monophyletic group of *H. japonicus*, *H. nipponensis*, and *H. olidus* was evident in only the two most rapidly evolving gene regions, cyt*b* and the Bayesian analysis of ITS2 (though very poorly supported, posterior probability < 0.7). Analyses of S71 and RAG1 individually were generally concordant with the results of all data combined though they supported a topology with *H. nipponensis* as basal to the rest of the species (data not shown). This topology also appeared in the set of trees generated from the Bayesian analysis of all the data combined, though at very low frequency, and was strongly supported by parsimony analysis (92% [nDNA], 83% [all DNA], 1000 bootstrap pseudo-replicates; data not shown, allDNA tree available in TreeBase, M3309). Because of this uncertainty, I consider the placement of *H. olidus* as unresolved (see below).

3.3.3 Biogeographic scenarios

SH tests comparing the molecular phylogeny (Fig. 3.5) to the expected topologies under each *a priori* biogeographic hypothesis (Fig. 3.2) showed the molecular phylogeny is significantly better than McAllister's (1963) hypothesis of scale-count relationships (Fig. 3.2A), the *a priori* northern Pacific ancestor hypothesis (Fig. 3.2B) and the scenario of an eastern Pacific origin (Fig. 3.2C; $P << 0.001$ in all cases). For the eastern Pacific hypothesis, I also tested a constraint tree with *H. pretiosus* in the basal position as well as trees with *H. olidus* in between the two eastern Pacific taxa, with the same result. Comparison of the molecular topology (Fig. 3.4) with *H. nipponensis* and *H. olidus* as sister taxa to a tree with *H.*

nipponensis basal to the rest of the species (Fig. 3.2D) showed no significant difference (-8096.7 vs. -8101.0, $P = 0.2$); therefore, the position of *H. olidus* in the phylogeny of *Hypomesus* as either sister or derived relative to *H. nipponensis* remains uncertain. Taken together, the results of the SH tests suggest a western Pacific origin for *Hypomesus*.

Results from LAGRANGE (Ree et al. 2005) suggested a geographically more widespread ancestor (Fig. 3.6), although this analysis also supported a role for the western Pacific as a centre of evolution for the group (see Discussion). Relatively high values of dispersal and extinction ($\lambda_D = 0.09$, $\lambda_E = 0.01$) yielded significantly more likely scenarios (likelihood ratio test, $P < 0.05$, df = 4); therefore, the other parameter combinations are not discussed further. The simulations allowing some dispersal across the Pacific produced similar results, but with a western Pacific and Arctic ancestor at node 1 and a western/northern Pacific and Arctic ancestor at node 2. Results from simulations where *H. olidus* was polymorphic for all areas were similar to those where *H. olidus* was restricted to the North Pacific and Arctic, except that the most likely scenario for the ancestor at node 2 (*H. olidus – H. nipponensis* divergence) was an ancestor present in all areas and the ancestor at node 3 (*H. japonicus – H. pretiosus*/*H. transpacificus* divergence) had a distribution in the west, north and east Pacific (data not shown). Although these are the most likely scenarios, for all nodes apart from node 4 there were others that fell within two log-likelihoods of the most likely scenario. I attribute the multitude of scenarios to the small number of taxa in the phylogeny and relatively high amount of homoplasy (R.

Ree, pers. comm. 2006). Further analysis in the context of an Osmeridae and/or Osmeroidea phylogeny may help further clarify *Hypomesus* biogeography.

3.3.4 Embiotocidae and *Sebastolobus* cyt*b* divergence

For a direct comparison with other trans-Pacific taxa, we identified two similarly distributed groups of fishes for which cyt*b* sequence data were available. Uncorrected pairwise cyt*b* sequence divergence in the Embiotocidae between the two basal western Pacific *Ditrema* and *Neoditrema* and the eastern Pacific *Hypselurus* were approximately 14.7%, whereas in *Sebastolobus*, the eastern (*S. alascanus*) and western (*S. macrochir*) Pacific species differed by around 6%. This compares to uncorrected cyt*b* sequence divergence between the western Pacific *Hypomesus japonicus* and the eastern Pacific *H. pretiosus* and *H. transpacificus* of approximately 10.5%.

3.4 DISCUSSION

3.4.1 *Hypomesus* phylogeny

The well-supported phylogeny from combined ML and Bayesian analyses of all five gene regions (Fig. 3.5) provides several key insights into *Hypomesus* evolutionary patterns. First, the apparent conflict in the placement of *H. japonicus* between the mt and nDNA phylogenies (Fig. 3.4) emphasizes the need for using multiple loci, both mitochondrial and nuclear, for inferring phylogeny. The sequence

divergence between *H. japonicus*, *H. nipponensis* and *H. olidus* is large (10 – 14%, uncorrected cyt*b*) and speciation seems to have happened on a relatively narrow timescale relative to the divergence of *H. pretiosus* and *H. transpacificus* (~2.5%, uncorrected cyt*b*; Figs. 3.4 and 3.5), which suggests that the only markers that show even a poorly supported relationship between these three species (cyt*b* and ITS2) may have reached a saturation point that obscures the phylogenetic signal. I interpret the differences between the mtDNA and nDNA phylogenies as a false conflict and it is clear that if only the mtDNA gene regions had been included, a very different picture of the relationships within this genus would have emerged. Furthermore, analyses with all of the data (Fig. 3.5) led to increased support values for all relationships relative to analyses by individual gene region (data not shown) and the mtDNA and nDNA partitions (Fig. 3.4), indicating that uncertainties within partitions are sometimes resolved when all information is combined.

One uncertainty that was not resolved from combining all data was the placement of *H. olidus* within the *Hypomesus* phylogeny. Although a sister-relationship between *H. olidus* and *H. nipponensis* is apparently strongly supported (Fig. 3.5), an alternative topology with *H. nipponensis* as basal (Fig. 3.2D) could not be rejected statistically (see Results). Further, parsimony analysis supports a sister relationship between *H. nipponensis* and the rest of the *Hypomesus* species (83% bootstrap support, data not shown). This conflict is interesting as *H. olidus* has the widest distribution of all the *Hypomesus* species, with populations in both North Pacific and Arctic drainages (Fig. 3.1; McAllister 1963). For the similarly distributed cisco (*Coregonus autumnalis*), populations from drainages in the Arctic Ocean differ

from those further south in the Bering Sea in a number of morphological traits, to the extent that they are now considered separate species [*C. autumnalis* and *C. laurettae*] (McPhail 1966; Eschmeyer 2006). These two cisco species are thought to have diverged in allopatry from a common ancestor distributed throughout the Arctic and Bering regions, whose range was divided by the formation of the Bering land bridge at the onset of the Pleistocene glaciations (McPhail 1966). A similar north – south split has been suggested for the Dolly Varden char complex (*Salvelinus malma* [Behnke 1980; Phillips et al. 1999]). Although the two populations of *H. olidus* that we sampled occurred on opposite sides of the Pacific (Fig. 3.1), we did not have information from populations in intervening areas (e.g. Arctic Ocean, Aleutian Archipelago) and, consequently, the uncertainty may result from a lack of data for this area. Additional sampling of Arctic populations may help clarify the position of *H. olidus* within the *Hypomesus* phylogeny and allow more direct comparisons with similarly distributed North Pacific taxa.

Regardless of the exact placement of *H. olidus*, it is clear from mapping the lateral line scale-counts onto the phylogenies that species with similar counts are not monophyletic (Figs. 3.4 and 3.5). Because the species-groups based on these counts are not supported by the phylogeny of *Hypomesus*, the data suggest that scale-count phenotypes (“high” and “low” count forms) have evolved in parallel on each side of the North Pacific Ocean. Parallel evolution of phenotypes is a well-documented phenomenon in many taxa (e.g., Losos et al. 1998; Wiens et al. 2006), and particularly in north temperate fishes (e.g., Behnke 1972; Taylor 1999). The difference in this instance is that lateral line scale-counts are presumed neutral,

whereas parallel evolution is usually associated with traits that evolve due to similar selective environments in different geographic regions (Simpson 1953). Although less likely, parallel evolution in neutral DNA regions has been demonstrated as theoretically possible (Orr 2005); however, the prevalence of this phenomenon in neutral traits in nature is unknown. Lateral line scale-counts are used as identifying characters in taxonomic keys for a number of different fishes (e.g., Scott and Crossman 1998). Although this trait may be useful for distinguishing species in the field, its utility as a systematic character can only be known through comparative phylogenetic analyses of different fishes. Recent work on rockfish (*Sebastes*) evolution has also shown parallel evolution of many traits (Hyde and Vetter 2007). Mapping morphological characters onto well-resolved phylogenies across multiple groups of fish would be informative in identifying both traits that are prone to evolve in parallel as well as those that are good indicators of evolutionary relationships.

Original classifications based on sub-species (McAllister 1963) and scale-count groups (Klyukanov 1970; Saruwatari et al. 1997) that required two trans-Pacific relationships in *Hypomesus* (Fig. 3.2A) are not supported by the molecular phylogeny. There is, however, at least one trans-Pacific disjunction in the genus: that between *H. japonicus* and the two eastern Pacific species, *H. pretiosus* and *H. transpacificus* (Fig. 3.5; discussed further below). My finding that the two eastern Pacific species are sister taxa is in agreement with allozyme data (Stanley et al. 1995; Trenham et al. 1998) that showed *H. pretiosus* and *H. transpacificus* are more similar to each other than are *H. transpacificus* and *H. nipponensis*, which were formerly sub-species of *H. nipponensis* (McAllister, 1963). My data also indicate a

relatively recent divergence of the eastern Pacific species (Figs. 3.4 and 3.5; discussed further in *Hypomesus* biogeography section).

3.4.2 Biogeography of *Hypomesus*

The paucity of fossil osmerids and the disagreement about extant osmerid relationships makes the formation of a plausible biogeographic scenario for these fishes particularly challenging. There are no fossil *Hypomesus* and the oldest fossil osmerid dates to the Palaeocene and is postulated to be related to the Japanese ayu *Plecoglossus altivelis* (Wilson and Williams 1991); however, the relationship of *P. altivelis* to other osmerids remains controversial (Johnson and Patterson 1996 and references therein), preventing even rough calibration of an osmerid-specific molecular clock from these data. Although inconclusive, two reasonable rates of cyt*b* evolution are available: 0.5 – 0.9% per million years (pmy) for ectothermic vertebrates (Martin and Palumbi 1993) and 1.56% pmy for thornyhead rockfishes (Stepien et al. 2000). To assess which of these rates may be most appropriate for *Hypomesus*, I calibrated a molecular clock based on the divergence of *H. japonicus* from *H. pretiosus* and *H. transpacificus* (Table 3.2) and the origin of the Sea of Japan in the Miocene around 15 mya (Itoh et al. 1997) under the assumption that the ancestor of *H. japonicus* evolved in this basin, an assumption that may be refuted or supported based on future fossil osmerid discoveries and/or further biogeographic analyses. This calibration yielded a rate of 1.77% pmy. As this rate is close to the 1.56% pmy recently proposed for rockfishes, we applied the range of these approximations to the corrected cyt*b* sequence divergence within *Hypomesus*.

The cyt*b* and Bayesian estimates suggest that speciation events in *Hypomesus* are roughly associated with climatic changes in the Cenozoic (Table 3.2).

The results from LAGRANGE (Fig. 3.6) suggest a more complicated biogeographic history than the western Pacific origin obtained from the dispersalist approach (Fig. 3.2, SH tests in Results). In general, they suggest a scenario in which pan-Pacific or generally more widely distributed ancestors repeatedly underwent speciation following range fragmentation (Fig. 3.6). With *Hypomesus olidus* restricted to the northern Pacific and Arctic, the divergence between *H. olidus* and *H. nipponensis* is congruent with a hypothesis where a widespread ancestor had its range fragmented due to cooling temperatures in the mid-Miocene, where *H. olidus* inherited the northern part of the range and then dispersed into the Arctic while *H. nipponensis* evolved in the western Pacific (node 2, Fig. 3.6). The trans-Pacific divergence between *H. japonicus* and *H. pretiosus*/*H. transpacificus* occurred through a split within an ancestor that already had a disjunct distribution across the North Pacific (node 3, Fig. 3.6). This distribution may also have been inherited from the ancestor at node 1; however, it is possible that the lineage between nodes 1 and 3 underwent a range expansion to the northern Pacific that was divided before the divergence at node 3 (Fig. 3.6). The most recent speciation event (*H. pretiosus – H. transpacificus*) occurred in the eastern Pacific during the Pliocene or early Pleistocene (Table 3.2). *H. pretiosus* is found along the western coast of North America from Alaska to California and *H. transpacificus* is endemic to the San Francisco estuary region of the Sacramento – San Joaquin River basin (Moyle 2002; Fig. 3.1). These distributions, coupled with the estimated time (Table 3.2) and

location of divergence (Fig. 3.6), suggest that a plausible speciation scenario involves a vicariant split of a widespread eastern Pacific ancestor when a population was isolated in a freshwater basin in western California. The extensive inland seas of the late Miocene had receded by the Pliocene (Oakeshott 1978; Norris and Webb 1990), which predates the estimated divergence time; however, the most recent glacial period is associated with drops in sea-level and it is possible that an ancestral population of these fishes was isolated in the series of Pleistocene lakes in the southern San Joaquin Valley (Norris and Webb 1990), thereby leading to differentiation of the marine and estuarine species. Comparing divergence times of other apparent marine-derived species endemic to the Central Valley region of the Sacramento – San Joaquin basin, such as the Kern brook lamprey (*Entosphenus hubbsi*; Moyle 2002), would clearly be of interest, although to my knowledge, no such estimates are available.

Mapping the dispersalist model from Fig. 3.2 onto the *Hypomesus* phylogeny used for the LAGRANGE analysis shows this approach yields a similar interpretation, although it supports a northern and western ancestor at node 1 and a widespread Pacific ancestor at node 3 (Fig. 3.6). Neither the LAGRANGE nor dispersalist models of *Hypomesus* support the expectations depicted in Fig. 3.2; however, if *H. olidus* is given a widespread Pacific and Arctic distribution, the dispersalist model supports a western Pacific origin for the genus (data not shown). Further, although the most likely reconstructions for the root node in the likelihood framework is an ancestor with a western and eastern Pacific distribution, a western

Pacific ancestor falls within two log-likelihoods; therefore, a western Pacific ancestor for *Hypomesus* cannot be rejected.

3.4.3 Comparative biogeography of the North Pacific

Comparative biogeographic analysis invokes concordance in patterns of distribution across different taxa to support the role of large-scale geologic events in shaping their evolutionary patterns (Lomolino et al. 2006). In assessing concordance two relevant questions are: what is the timing of the divergences, and what is the area of origin?

To address the first question on timing, Cenozoic cooling periods beginning in the mid-Eocene are thought to have led to diversification across a variety of taxonomic groups, with a large literature on fishes, molluscs and crustaceans in particular. A well-preserved fossil record has allowed relatively accurate dating of species radiations in a number of molluscan (e.g. MacNeil 1965; Titova 1994; Oleinik 2001; Amano and Vermeij 2003) and crustacean (e.g. Schweitzer 2001) taxa. Relatively early radiations in the Eocene and Oligocene epochs have been suggested for some gastropod (e.g. Titova 1994) and decapod (e.g. Schweitzer 2001) groups; however, much of the literature on North Pacific evolution has centered on the importance of the Miocene and to some extent, the Pliocene, in generating diversity (Andriashev 1939; Briggs 1974). Cooling beginning in the mid-Miocene is thought to have been a particularly important evolutionary pump in a number of North Pacific taxa, including, whelks (Collins et al. 1996), periwinkles (Reid et al. 1996), decapods (Schweitzer 2001), kelps (Estes and Steinberg 1988),

and fishes such as poachers (Laroche 1986), salmonids (Stearly 1992), and rockfishes (Kai et al. 2003). For fishes, early hypotheses for many groups emphasized the importance of climate and associated sea-level changes in the Plio- and Pleistocene epochs (e.g. Andriashev 1939; Tarp 1952; Lindberg 1953, cited in Briggs 1974; Neave 1958), and more recently the possible role of tectonic events in promoting diversification, particularly in the eastern North Pacific, has been highlighted for the Pacific salmon genus *Oncorhynchus* (Montgomery 2000).

Of particular interest in the current study was the phenomenon of trans-Pacific distributions. Although the original trans-Pacific sister relationships based on lateral line scale-counts were not supported by our phylogenetic analysis, a trans-Pacific relationship between the western *H. japonicus* and the two eastern species *H. pretiosus* and *H. transpacificus* was identified (Fig. 3.5). Comparison of uncorrected cyt*b* distances between this *Hypomesus* disjunction (10.5%) and that in surfperches (14.7%) and thornyhead rockfishes (6%) shows that, if rates of molecular evolution are similar across these groups, the trans-Pacific divergences occurred during different time intervals. Our divergence estimates suggest a mid-Miocene divergence for the *Hypomesus* disjunction. Bernardi and Bucciarelli (1999) and Stepien et al. (2000) both estimated the divergence between their respective western and eastern taxa in the early Pliocene at approximately 5 mya, which clearly assumes different cyt*b* evolutionary rates (Stepien et al. (2000) estimated the divergence based on control region sequence divergence). With the estimated rates of cyt*b* evolution, the trans-Pacific divergences in these three groups of fishes are apparently much older than the Pleistocene glaciations that have been thought

responsible for generating many trans-Pacific relationships (Bernardi and Bucciarelli 1999). The relative paucity of phylogenetic work on other pan- and trans-Pacific taxa that includes all species prevents additional comparisons, although Kai et al. (2003) and Hyde and Vetter (2007) showed that early hypotheses of several trans-Pacific sister relationships in the rockfish genus *Sebastes* based on morphological analyses (Matsubara 1943) are also unsupported by molecular analyses. Many studies of organisms with such distributions have contributed to Pacific – Atlantic or trans-equatorial relationships (e.g., Harrison and Crespi 1999; Møller and Gravlund 2003; Burridge 2002; Waters et al. 2002; Väinölä 2003; Grant et al. 2005) or have addressed questions focused on a more local scale (e.g. Johns and Avise 1998). Additional molecular data for trans-Pacific sister taxa will allow future tests for simultaneous vicariance (e.g. Hickerson et al. 2006) to assess whether Cenozoic climate changes resulted in the current distributions. Further, I am sensitive to the much-discussed problems with molecular clocks (reviewed in Arbogast et al. 2002), possible problems of cyt*b* saturation, and uncertainty regarding the calibration point. Analysis within a larger phylogenetic framework including the Osmeridae and Osmeroidea, where fossil calibration points are available, will allow additional confirmation, or refutation, of the biogeographic interpretations.

Incomplete geographic and/or taxonomic sampling also limits comparisons of ancestral area of origin. My biogeographic analysis of *Hypomesus* suggests either a western Pacific or a pan-Pacific ancestor for the genus, depending on the reconstruction approach. Although often not tested in a phylogenetic framework and in many cases still controversial, the western Pacific has been suggested an

important area of origin for a number of fishes, including, *Oncorhynchus* (Neave 1958), *Salvelinus malma* (Oleinik et al. 2005), *Gasterosteus aculeatus* (Haglund et al. 1992; Ortí et al. 1994; Higuchi and Goto 1996), and *Sebastes* (Barsukov 1981; Hyde and Vetter 2007), and many invertebrates (e.g. MacNeil 1965; Titova 1994; Oleinik 2001). Furthermore, Gladenkov (1994) infers that from the Oligocene through the Neogene trans-Pacific migrations of molluscan fauna appear to be more frequent from west to east than vice-versa. Although the eastern Pacific has been supported as an area of origin for a number of other taxa (e.g. Embiotocidae [Tarp 1952; Bernardi and Bucciarelli 1999], Scorpaenidae [Briggs 1974], *Nucella* [Amano et al. 1993], *Littorina* [Reid et al. 1996]), the relatively complex geologic history of the western Pacific, with basins such as the Sea of Japan and possibly the Sea of Okhotsk undergoing periodic isolation (Lindberg 1953, cited in Briggs 1974; Zenkevitch 1963), suggests that formation of such barriers has been an important mechanism for generating diversity in the North Pacific, including that in *Hypomesus*, under an allopatric model of speciation.

3.4.4 Conclusions

In this study I present a phylogenetic reconstruction of *Hypomesus*, a genus within a systematically problematic group of fishes, which showed that species-group designations based on lateral line scale-counts do not reflect ancestry of the species. The placement of the widely distributed *H. olidus* is not yet resolved, though I expect that including samples of this species from Arctic drainages will help clarify its placement in the phylogeny and allow direct comparison with other North

Pacific/Arctic species. My biogeographic analyses suggest an important role for the western Pacific and maximum likelihood reconstructions of ancestral ranges support the idea that Cenozoic climatic changes have been important drivers of diversification. Precise dating of divergences in *Hypomesus* is currently not feasible due to a lack of fossil information for these fishes, although best estimates strongly suggest they considerably pre-date the climate fluctuations of the Plio- and Pleistocene epochs historically thought to be responsible for generating trans-Pacific relationships in many fishes. The hypothesized time periods of *Hypomesus* diversification are implicated in divergences in a number of other similarly distributed North Pacific taxa, although the extent of comparison is limited due to a lack of phylogenetic data in many groups. Not surprisingly, it is often the western Pacific species that are missing, likely due to the difficulty of obtaining samples from many remote areas in this vast region. Along with Uschakov (1971) and Amano and Vermeij (2003) we view the North Pacific as an ideal arena for comparative biogeographic work and encourage trans-Pacific collaboration whenever possible. Accumulation of phylogenetic and biogeographic data on pan-Pacific taxa will facilitate comparisons of both timing and area of divergences within the North Pacific Ocean.

Table 3.1 Number of sequences for the five gene regions for *Hypomesus* species and the outgroup *Mallotus villosus*. The range number of base pairs for each gene (excluding indels) is indicated in parentheses.

	Gene region (bp)				
	Cyt*b* (285-425)	16S (495-547)	ITS2 (247-441)	S71 (338-732)	RAG1 (742-1440)
Species					
H. japonicus	6	6	6	6	5
H. nipponensis (*H. chishimaensis*)	20	25	25	24	14
H. olidus	8	5	4	8	3
H. pretiosus	5	5	5	5	2
H. transpacificus	7	7	7	7	5
M. villosus	2	2	2	2	2

Table 3.2 Corrected (TIM + Γ model) cyt*b* divergences with timeframe and associated climatic events (Kennett, 1982; Tsuchi, 1997; Zachos et al. 2001) corresponding to *Hypomesus* divergences based on two rates of cyt*b* evolution: 1.77% (calibrated based on *H. japonicus-H. pretiosus*/*H. transpacificus* split [see text], and 1.56% per my (Stepien et al. 2000), and Bayesian estimation under a log-normal relaxed molecular clock model, based on all sequence data combined, implemented in Beast v.1.4.1(Drummond and Rambaut, 2003, http://evolve.zoo.ox.ac.uk/beast). Numbers in brackets indicate 95% confidence intervals for Beast estimations.

Node	Corrected mean cyt*b* sequence divergence (%)	Cyt*b* divergence estimate (mya)	Beast divergence estimate (mya)	Geologic time period	Climatic events
1	23.6 (*H. olidus - H. japonicus*)	13.3-15.1	22.9 (14.5-34.4)	early - mid-Miocene	mid-Miocene climatic optimum
2	22.2 *(H. nipponensis - H. olidus)*	12.5-14.2	16.1 (9.3-23.9)	mid-Miocene	mid-Miocene climatic optimum followed by cooling with onset of Antarctic glaciation and formation of Sea of Japan
3	26.5 *(H. japonicus - H. pretiosus/H. transpacificus)*	15.0-17.0	15 (8.9-22.3)		
4	2.8 (*H. pretiosus - H. transpacificus*)	1.6-1.8	3.9 (1.1-8.2)	mid-Pliocene – early Pliocene	early Pliocene warm period followed by beginning of Pleistocene glacial cycles

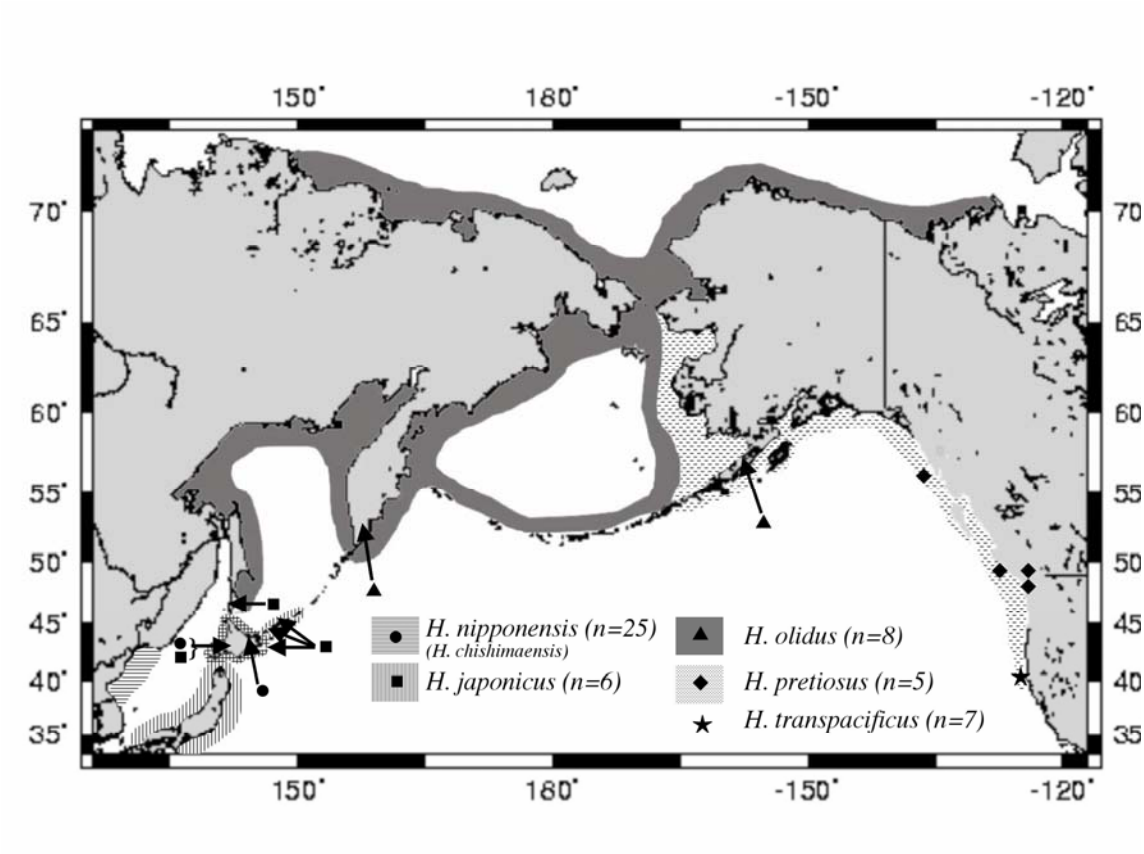


Figure 3.1 Approximate distribution map of the five species of *Hypomesus* in the North Pacific Ocean with sample numbers and locations. Distribution information from McAllister (1963), Saruwatari et al. (1997), and FishBase (Froese and Pauly 2005).

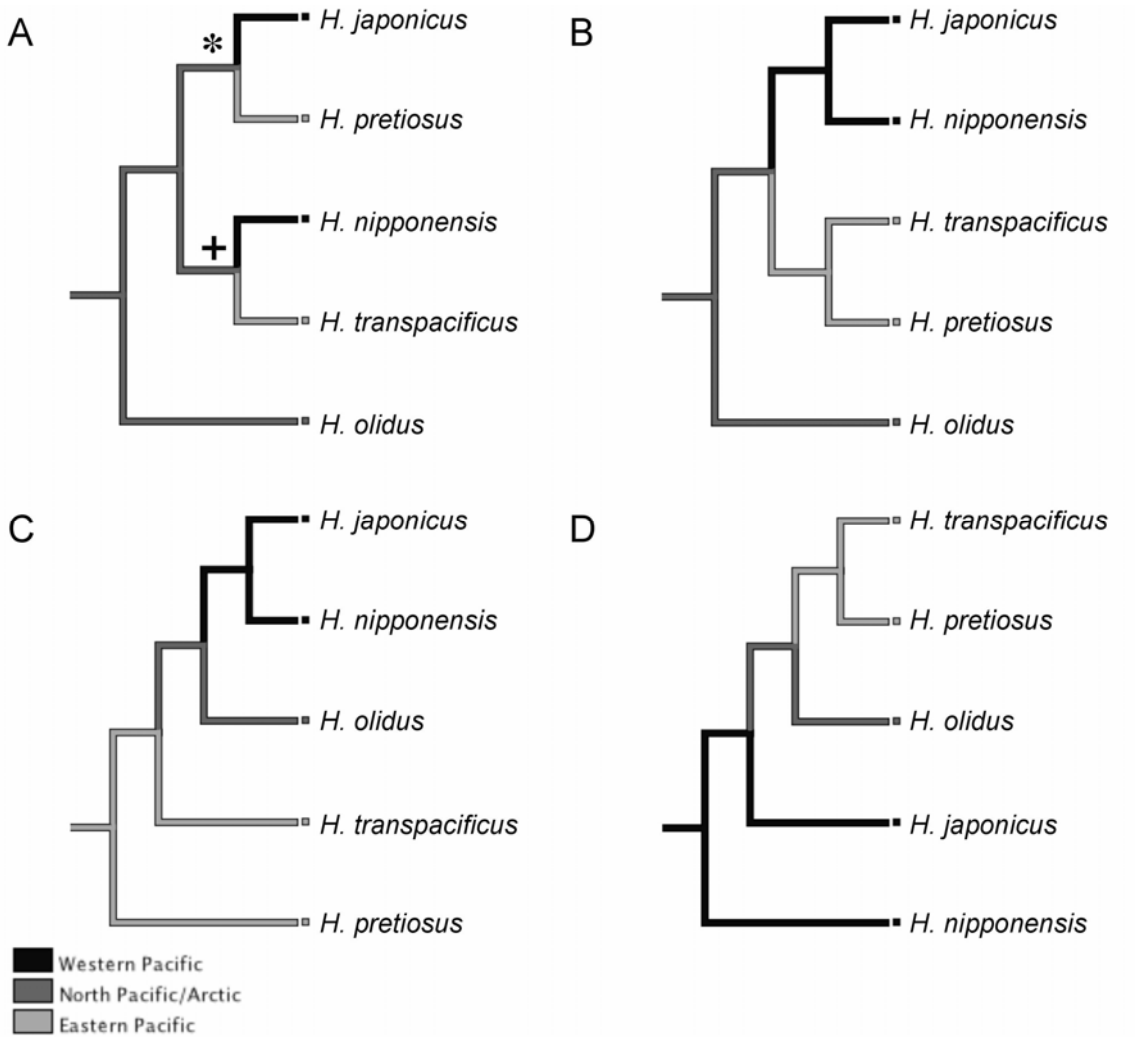


Figure 3.2 Parsimony-based expected phylogenies of *Hypomesus* with the geographic ranges of the species. (A) McAllister (1963), three species groups with two replicate divergences across the North Pacific required. * and + represent high and low scale count groups, respectively (B) *a priori* hypothesis with single split across the North Pacific (C) eastern Pacific origin (D) western Pacific origin. *H. olidus* is assigned a northern Pacific distribution. Characters are ordered such that two steps are required between the west and east Pacific, and one step between the north and west or east Pacific. Interchanging the terminal taxa from the same geographic range in (B-D) would result in the same interpretation. See text for further discussion.

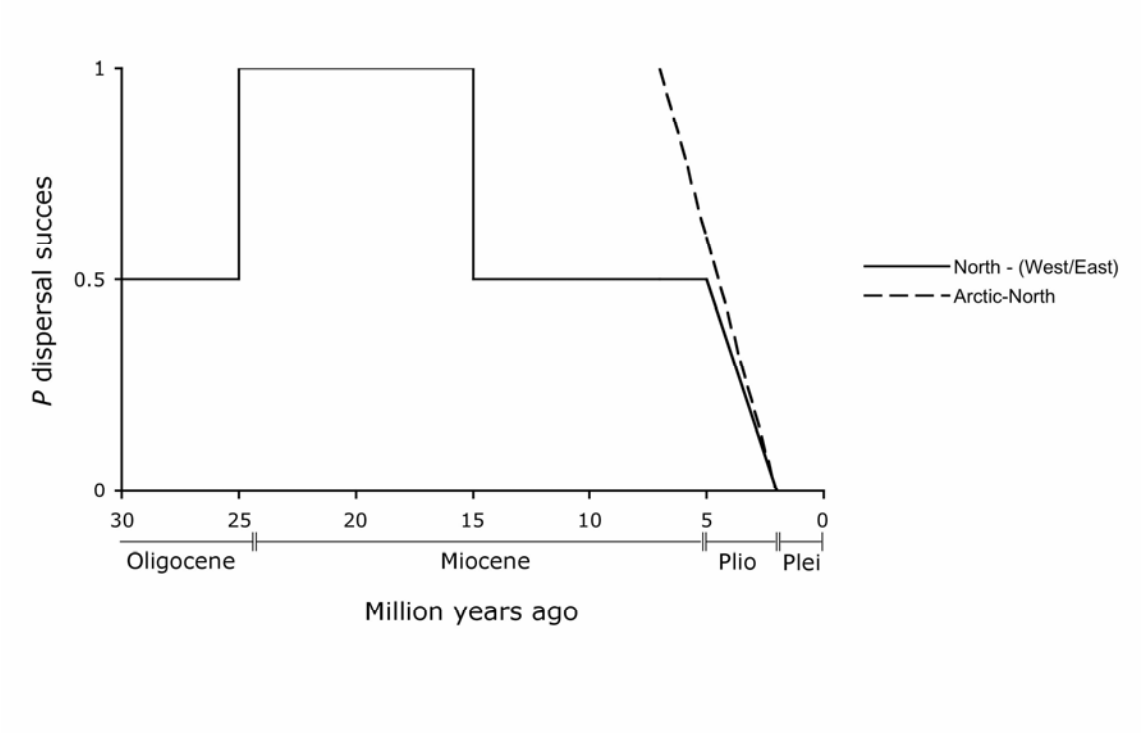


Figure 3.3 Probability of dispersal success (connection) between the four areas defined for the LAGRANGE analysis of ancestral area: Arctic Ocean, north Pacific, west Pacific and east Pacific oceans. The corresponding probability of dispersal success between Arctic–north Pacific and north Pacific–(west/east) Pacific through the mid-late Cenozoic is plotted. The west–east Pacific connection is described in text. Increases and decreases in probability of dispersal success between regions correspond to periods of global warming and cooling, respectively, based on Zachos et al. (2001).

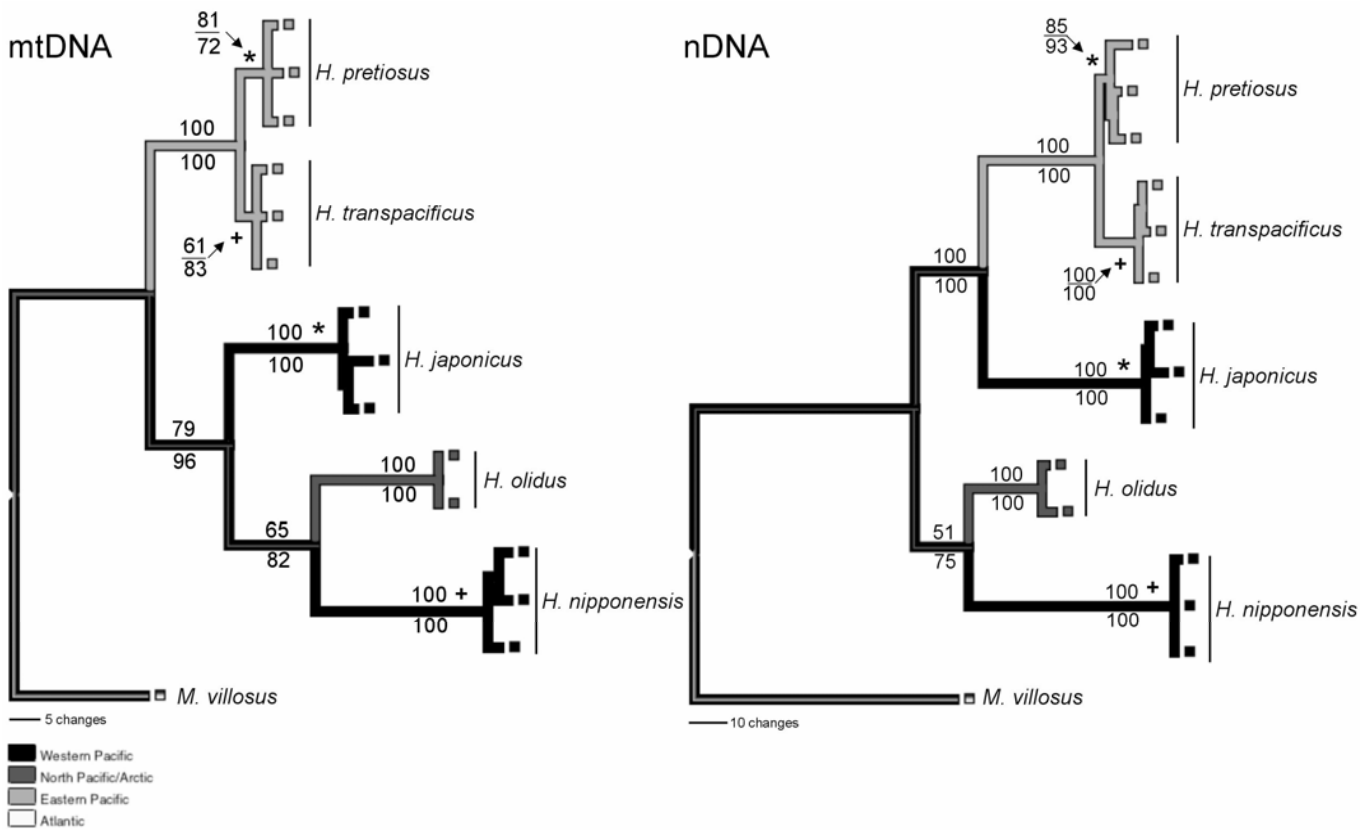


Figure 3.4 Combined mitochondrial (mtDNA) [cyt*b* and 16S] and nuclear (nDNA) [ITS2, S71 and RAG1] phylogenies of *Hypomesus*. * and + represent species with high and low lateral line scale counts, respectively, and WPac, NPac, and EPac correspond to geographic distributions in the western, northern and eastern Pacific, respectively. Numbers above nodes represent bootstrap support values from 1000 pseudo-replicates in a maximum likelihood analysis and those below nodes are posterior probabilities from a consensus of 17000 post-burnin trees generated through Bayesian analysis (2×10^6 generations). *Mallotus villosus* is the outgroup. Geographic areas are mapped onto the phylogenies following the model described in Figure 3.2.

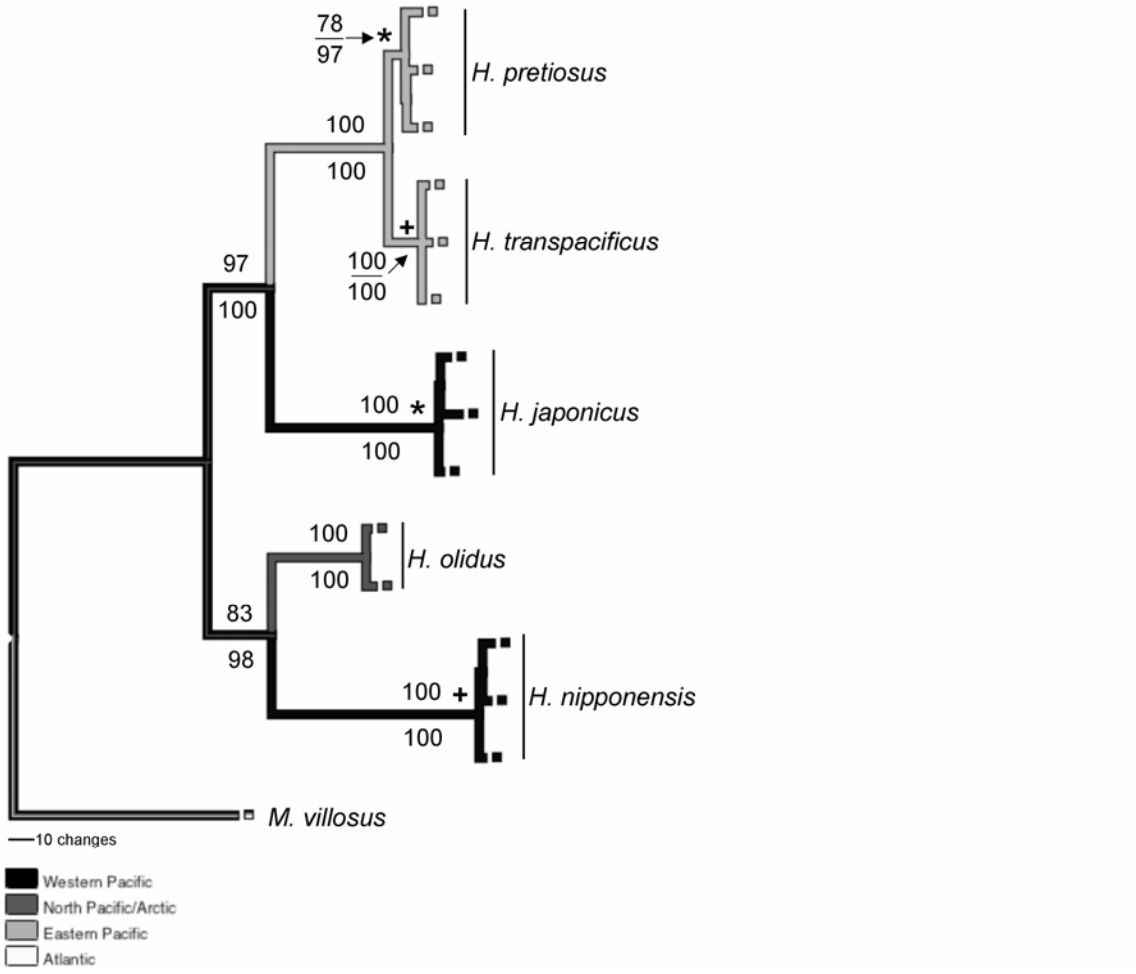


Figure 3.5 *Hypomesus* phylogeny from combined mtDNA [cyt*b* and 16S] and nDNA datasets [ITS2, S71 and RAG1]. * and + represent species with high and low lateral line scale counts, respectively, and WPac, NPac, and EPac correspond to geographic distributions in the western, northern and eastern Pacific, respectively. Numbers above nodes represent bootstrap support values from 1000 pseudo-replicates in a maximum likelihood analysis and those below nodes are posterior probabilities from a consensus of 17000 post-burnin trees generated through Bayesian analysis (2×10^6 generations). *Mallotus villosus* is the outgroup. Geographic areas are mapped onto the phylogeny following the model described in Figure 3.2.

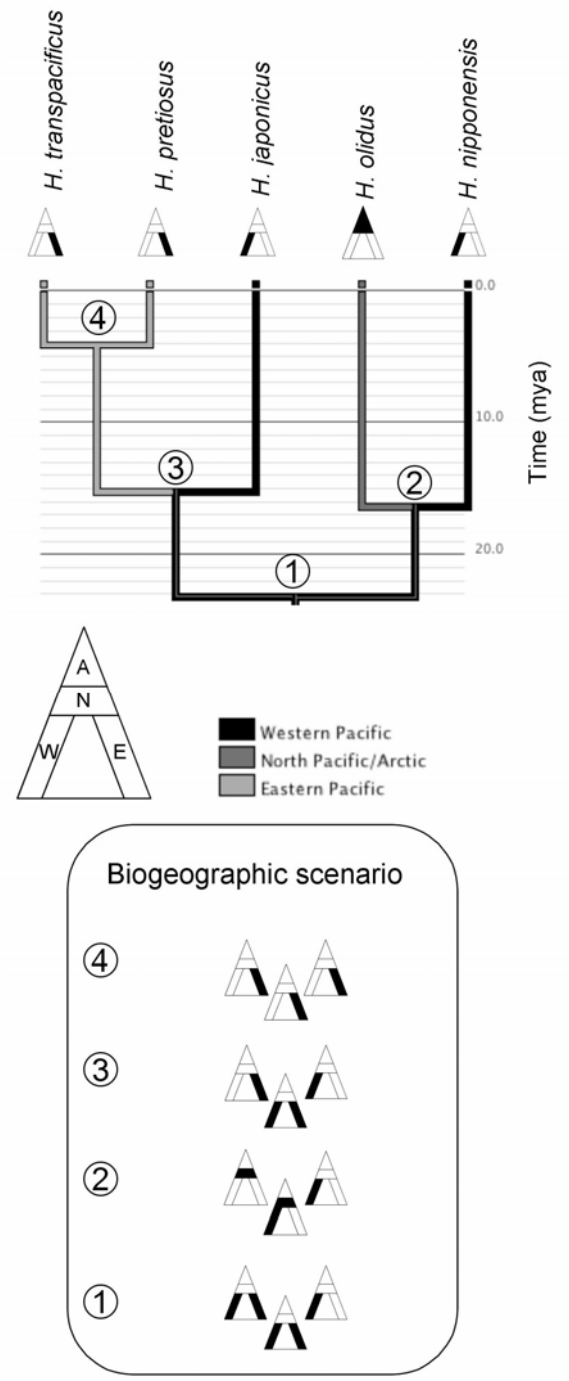


Figure 3.6 LAGRANGE maximum likelihood estimates of ancestral range and inheritance scenarios based on a model with a root node age of 22.9 mya, no west – east Pacific dispersal, and relatively high rates of dispersal and extinction ((λ_D = 0.09, λ_E = 0.01). Reconstructions with highest likelihoods shown for each node. Each node diagram depicts the ancestral geographic range and the range inherited by each daughter lineage. Geographic areas are mapped onto the phylogeny following the model described in Figure 3.2.

3.5 REFERENCES

Addison, J. A., and M. W. Hart. 2005. Colonization, dispersal, and hybridization influence phylogeography of North Atlantic sea urchins (*Stronglyocentrotus droebachiensis*). Evolution 59:532-543.

Amano, K. and G. J. Vermeij. 1998. Origin and biogeographic history of *Ceratostoma* (Gastropoda: Muricidae). Venus (Tokyo*)* 57**:**209-223.

———. 2003. Evolutionary adaptation and geographic spread of Cenozoic buccinid genus *Lirabuccinum* in the North Pacific. J. Paleontol. 77:863-872.

Amano, K., G. J. Vermeij, and K. Narita. 1993. Early evolution and distribution of the gastropod genus *Nucella*, with special reference to Miocene species from Japan. Trans. Proc. Paleontol. Soc. Jpn, New Ser. 171:237-248.

Andriashev, A. P. 1939. On the amphipacific (Japan-Oregonian) distribution of sea-fauna in the northern part of the Pacific Ocean. Zool. Zh. 18(2):1-16. [In Russian with English Summary]

Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002 Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annu. Rev. Ecol. Syst. 33:707-740.

Baker, A. J., S. L. Pereira, O. P. Haddrath, and K. A. Edge. 2006. Multiple gene evidence for expansion of extant penguins out of Antarctica due to global cooling. Proc. R. Soc. Lond. Ser. B. Biol. Sci. 273:11-17.

Barsukov, V. V. 1981. A brief review of the subfamily Sebastinae. *J. Ichthyol.* 21:1-26.

Behnke, R. J. 1972. The systematics of salmonid fishes in recently glaciated lakes. J. Fish. Res. Board Can. 29:639-671.

———. 1980. A systematic review of the genus *Salvelinus*, Pp. 441-481 in E. K. Balon, ed. Charrs: salmonid fishes of the genus *Salvelinus*. W. Junk, The Hague.

Bernardi, G. and G. Bucciarelli. 1999. Molecular phylogeny and speciation of the surfperches (Embiotocidae, Perciformes). Mol. Phylogenet. Evol. 13:77-81.

Bollback, J. 2006. SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7:88.

Bremer, K. 1992. Ancestral areas: a cladistic reinterpretation of the center of origin concept. Syst. Biol. 41:436-445.

Brevoort, J. C. 1856. Notes on some figures of Japanese fish. Pp. 253-256 In *Narrative of Commodore M.C. Perry's expedition to Japan. Vol. II.* U.S. Senate Ex. Doc. No. 79, 33rd Congress, 2nd Session. Beverley Tucker, Washington, D.C.

Briggs, J. C. 1974. Marine Zoogeography. McGraw-Hill Inc., Toronto.

———. 2000. Centrifugal speciation and centres of origin. J. Biogeogr. 27:1183-1188.

———. 2003. Marine centres of origin as evolutionary engines. J. Biogeogr. 30:1-18.

———. 2004. Older species: a rejuvenation on coral reefs? J. Biogeogr. 31:525-530.

Burridge, C. P. 2002. Antitropicality of Pacific fishes: molecular insights. Env. Biol. Fish. 65:151-164.

Chow, S., and K. Hazama. 1998. Universal PCR primers for S7 ribosomal protein introns in fish. Mol. Ecol. 7: 1247-1263.

Collins, T. M., K. Frazer, A. R. Palmer, G. J. Vermeij, and W. M. Brown. 1996. Evolutionary history of northern hemisphere *Nucella* (Gastropoda, Muricidae): molecular, morphological, ecological, and paleontological evidence. Evolution 50:2287-2304.

Cowie, R. H., and B.S. Holland. 2006. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. J. Biogeogr. 33:193-198.

Croizat, L., G. Nelson, and D. E. Rosen. 1974. Centers of origin and related concepts. Syst. Zool. 23:267-287.

Deméré, T. A., A. Berta, and P. J. Adam. 2003. Pinnipedimorph evolutionary biogeography. Bull. Am. Mus. Nat. Hist. 279:32-76.

de Queiroz, A. 2005. The resurrection of oceanic dispersal in historical biogeography. Trends Ecol. Evol. 20:68-73.

Donoghue, M.J. and Moore, B.R. (2003). Toward an integrative historical biogeography. Integr. Comp. Biol. 43:261-270.

Drummond, A. J., and A. Rambaut. 2003. BEAST v1.4, Available from http://beast.bio.ed.ac.uk/

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLOS 4:e88.

Ebach, M. C. 1999. Paralogy and the centre of origin concept. Cladistics 15:387-391.

Eschmeyer, W. N. 2006. Catalogue of fishes, online version. California Academy of Sciences, San Francisco. Available from http//www.calacademy.org/research/ichthyology/catalog

Estes, J. A. and P. D. Steinberg. 1988. Predation, herbivory, and kelp evolution. Paleobiol. 14:19-36.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Inc., Sunderland, MA.

Froese, R. and D. Pauly. (Eds.). 2006. FishBase. World Wide Web electronic publication. www.fishbase.org, version

Gill, T. N. 1862. On the subfamily Argentininae. Proc. Natl. Acad. Sci. USA 14:14-15.

Girard, C. F. 1854. Observations upon a collection of fishes made on the Pacific coast of the United States, by Lieut. W.P. Trowbridge, U.S.A. for the museum of the Smithsonian Institution. Proc. Natl. Acad. Sci. USA 7:142-156.

Gladenkov, Y. B. 1994. Cenozoic paleogeography and climatic change in the North Pacific Ocean. Palaeogeogr., Palaeoclimatol., Palaeoecol. 108:311-318.

Golikov, A. N. and N. L. Tzvetkova. 1972. The ecological principle of evolutionary reconstruction as illustrated by marine animals. Mar. Biol. 14:1-9.

Grant, W. S., R. W. Leslie, and B. W. Bowen. 2005. Molecular genetic assessment of bipolarity in anchovy genus *Engraulis*. J. Fish Biol. 67:1242-1265.

Haglund, T. R., D. G. Buth, and R. Lawson. 1992. Allozyme variation and phylogenetic reconstruction of Asian, North American, and European populations of the threespine stickleback, *Gasterosteus aculeatus*. Copeia 1992:432-443.

Harrison, M. K. and B. J. Crespi. 1999. Phylogenetics of *Cancer* crabs (Crustacea: Decapoda: Brachyura). Mol. Phylogenet. Evol. 12:186-199.

Hickerson, M. J., G. Dolman, and C. Moritz. 2006. Comparative phylogeographic summary statistics for testing simultaneous vicariance. Mol. Ecol. 15:209-223.

Higuchi, M. and A. Goto. 1996. Genetic evidence supporting the existence of two distinct species in the genus *Gasterosteus* around Japan. Environ. Biol. Fish. 47:1-16.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754-755.

Humphries, C. J., and L. R. Parenti. 1999. Cladistic biogeography, 2nd edn. Oxford University Press, Oxford.

Hyde, J. R., and R. D. Vetter. 2007. The origin, evolution, and diversification of rockfishes of the genus *Sebastes* (Cuvier). Mol. Phylogenet. Evol. 44:790-811.

Ilves, K. L., and E. B. Taylor. 2007. Are *Hypomesus chishimaensis* and *H. nipponensis* (Pisces: Osmeridae) distinct species? A molecular assessment using comparative sequence data from five genes. *Copeia* 2007:180-185.

Itoh, Y., N. Nakajima, and A. Takemura. 1997. Neogene deformation of the back-arc shelf of Southwest Japan and its impact on the palaeoenvironments of the Japan Sea. Tectonophysics 281:71-82.

Johns, G. C. and J. C. Avise. 1998. Tests for ancient species flocks based on molecular phylogenetic appraisals of *Sebastes* rockfishes and other marine fishes. Evolution 52:1135-1146.

Johnson, G. D., and C. Patterson. 1996. Relationships of lower Euteleostean fishes, Pp. 251-332 in M. Stiassny, L. J. Parenti, L. R. Johnson, and G. David, eds. Interrelationships of fishes. Academic Press, Toronto.

Jones, W. J., Y. J. Won, P. A. Y. Maas, P. J. Smith, R. A. Lutz, and R. C. Vrijenhoek. 2006. Evolution of habitat use by deep-sea mussels. Mar. Biol. 148:841-851.

Kai, Y., K. Nakayama, and T. Nakobo. 2003. Molecular phylogenetic perspective on speciation in the genus *Sebastes* (Scorpaenidae) from the northwest Pacific and the position of *Sebastes* within the subfamily Sebastinae. Ichthyol. Res. 50:239-244.

Kennett, J. P. 1982. Marine geology. Prentice-Hall, Englewood Cliffs, NJ.

Klyukanov, V. A. 1970. Morphological basis of the classification of smelts of the genus *Hypomesus*. Zool. Zh. 49:1534-1542.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca,

and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86: 6196-6200.

Laroche, W. A. 1986. A preliminary morphological investigation of the Agonidae: towards reconstruction of agonid phylogeny and biogeographic history of neritic/littoral cold marine fishes. M. Sc. Thesis, Humboldt State University, Arcata.

Lindberg, G. V. 1953. Principles of the distribution of fishes and the geological history of the far-eastern seas. Akademia Nauk SSR, Ikhthiolog. Komm., Moscow-Leningrad. [In Russian].

Lomolino, M.V., B. R. Riddle, and J. H. Brown. 2006. Biogeography. 3rd ed. Sinauer Associates, Sunderland, MA.

Losos, J. B., T. R. Jackman, A. Larson, K. de Queiroz, and L. Rodríguez-Schettino. 1998. Historical contingency and determinism in replicated adaptive radiations of island lizards. Science 279:2115-2118.

MacNeil, F. S. 1965. Evolution and distribution of the genus *Mya*, and Tertiary migrations of Mollusca. Prof. Pap. U.S. Geol. Surv. 483G:1-51.

Maddison, D. R., and W. P. Maddison. 2003. MacClade 4.06: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA.

Maddison, W. P., and D. R. Maddison. 2006. Mesquite: a modular system for phylogenetic analysis, Version 1.12. http://mesquiteproject.org

Marincovich, L., Jr., and A. Y. Gladenkov. 1999. Evidence for an early opening of the Bering Strait. Nature 397:149-151.

Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate , generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA 90:4087-4091.

Matsubara, K. 1943. Studies on the scorpaenoid fishes of Japan. Sigenkagaku Kenkyusyo (Research Institute for Natural Resources), Tokyo, Japan.

McAllister, D. E. 1963. A revision of the smelt family, Osmeridae. Bull. Nat. Mus. Can. 191:1-53.

McCarthy, D. 2003. The trans-Pacific zipper effect: disjunct sister taxa and matching geological outlines that link the Pacific margins. J. Biogeogr. 30:1545-1561.

———. 2005. Biogeographical and geological evidence for a smaller, completely enclosed Pacific Basin in the Late Cretaceous. J. Biogeogr. 32:2161-2177.

McDowall, R. M. 2002. Accumulating evidence for a dispersal biogeography of southern cool temperate freshwater fishes. J. Biogeogr. 29:207-219.

McPhail, J. D. 1966. The *Coregonus autumnalis* complex in Alaska and northwestern Canada. J. Fish. Res. Board Canada. 23:141-148.

Møller, P. R. and P. Gravlund. 2003. Phylogeny of the eelpout genus *Lycodes* (Pisces, Zoarcidae) as inferred from mitochondrial cytochrome b and 12S rDNA. Mol. Phylogenet. Evol. 26:369-388.

Montgomery, D. R. 2000. Coevolution of the Pacific salmon and Pacific Rim topology. Geology 28:1107-1110.

Moyle, P. B. 2002. Inland fishes of California. 2nd ed. University of California Press, Berkeley, CA.

Müller, O. F. 1776. *Zoologiae Danicae prodromus, seu animalium Daniae et Norvegiae indigenarum characteres, nomina, et synonyma imprimis popularium.* Havniae. Zool. Danicae Prodromus.

Neave, F. 1958. The origin and speciation of *Oncorhynchus*. Trans. R. Soc. Can. Third Ser. 52:25-39.

Nelson, G., and P. Y. Ladiges. 1991. Three-area statements: standard assumptions for biogeographic analysis. Syst. Zool. 40:470-485.

———. 1996. Paralogy in cladistic biogeography and analysis of paralogy-free subtrees. *Am. Mus. Novit.* 3167:1-58.

Nelson, G., and N. Platnick. 1981. Systematics and biogeography: cladistics and vicariance. Columbia University Press, New York.

Norris, R. M. and R. W. Webb. 1990. Geology of California. 2nd ed. John Wiley & Sons, Inc., Toronto.

Oakeshott, G. B. 1978. California's changing landscapes. 2nd ed. McGraw-Hill Book Company, Toronto.

Oleinik, A. E. 2001. Eocene gastropods of western Kamchatka – implications for high-latitude North Pacific biostratigraphy and biogeography. Palaeogeogr., Palaeoclimatol., Palaeoecol. 166:121-140.

Oleinik, A. G., L. A. Skurikhina, Vl. A. Brykov, P. A. Crane, and J. K. Wenburg. 2005. Differentiation of Dolly Varden char *Salvelinus malma* from Asia and North America inferred from PCR-RFLP analysis of mitochondrial DNA. Russ. J. Genet. 41:501-508.

Orr, H. A. 2005. The probability of parallel evolution. Evolution 59:216-220.

Ortí, G., M. A. Bell, T. E. Reimchen, and A. Meyer. 1994. Global survey of mitochondrial DNA sequences in the threespine stickleback: evidence for recent migrations. Evolution 48:608-622.

Page, R. D. M. 1988. Quantitative cladistic biogeography: constructing and comparing area cladograms. Syst. Zool. 37:254-270.

Pallas, P. S. (1814) *Zoographia Rosso-Asiatica, sistens omnium animalium in extenso Imperio Rossico et adjacentibus maribus observatorum recensionem, domicilia, mores et descriptiones anatomen atque icones plurimorum*, 3 vols. In Officina Caes. Academie Scientiarum Impress, St. Petersburg.

Phillips, R. B., L. I. Gudex, K. M. Westrich, and A. L. DeCicco. 1999. Combined phylogenetic analysis of ribosomal ITS1 sequences and new chromosome data supports three subgroups of Dolly varden char (*Salvelinus malma*). Can. J. Fish. Aquat. Sci. 56:1504-1511.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53:793-808.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

Presa, P., B. G. Pardo, P. Martinez, and L. Bernatchez. 2002. Phylogeographic congruence between mtDNA and rDNA ITS markers in Brown Trout. Mol. Biol. Evol. 19:2161-2175.

Quenouille, B., E. Bermingham, and S. Planes. 2004. Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. Mol. Phylogenet. Evol. 31:66-88.

Rambaut, A. 1996. Se-Al: a manual sequence alignment editor, v.2.0a11. Available from http://tree.bio.ed.ac.uk/software/seal/.

Ravelo, A. C., D. H. Andreasen, M. Lyle, A. O. Lyle, and M. W. Wara. 2004. Regional climate shifts caused by gradual global cooling in the Pliocene epoch. Nature 429:263-267.

Ree, R. H., B. R. Moore, C. O. Webb, and M. J. Donoghue. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59:2299-2311.

Reid, D. G., E. Rumbak, and R. H. Thomas. 1996. DNA, morphology and fossils: phylogeny and evolutionary rates of the gastropod genus *Littorina*. Philos. Trans. R. Soc. Lond. Ser. B. Biol. Sci. 351:877-895.

Rocha-Olivares, A., R. H. Rosenblatt, and R. D. Vetter. 1999. Molecular evolution, systematics, and zoogeography of the rockfish subgenus *Sebastomus* (*Sebastes*, Scorpaenidae) based on mitochondrial cytochrome *b* and control region sequences. Mol. Phylogenet. Evol. 11:441-458.

Rosenblatt, R. H., and R. S. Waples. 1986. A genetic comparison of allopatric populations of shore fish species from the eastern and central Pacific Ocean: dispersal or vicariance? Copeia. 1986: 275-284.

Santini, F., and R. Winterbottom. 2002. Historical biogeography of Indo-western Pacific coral reef biota: is the Indonesian region a centre of origin? J. Biogeogr. 29:189-205.

Saruwatari, T., J. A. López, and T. W. Pietsch. 1997. A revision of the osmerid genus *Hypomesus* Gill (Teleostei: Salmoniformes), with the description of a new species from the Southern Kuril Islands. Species Divers. 2:59-82.

Schluter, D., T. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699-1711.

Schweitzer, C. E. 2001. Paleobiogeography of Cretaceous and Tertiary decapod crustaceans of the North Pacific Ocean. J. Paleontol. 75:808-826.

Scott, W. B., and E. J. Crossman. 1998. Freshwater fishes of Canada. Galt House Publications, Oakville, ON.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114-1116.

Sidorov, L. K., and Y. Pichugin. 2004. Morphological traits of lacustrine forms of smelts of the genus *Hypomesus* (Salmoniformes) from the Southern Kurils. J. Ichthyol. 44:433-443.

Simpson, G. G. 1953. The major features of evolution. Columbia Univ. Press, New York.

Smith, S. A., P. R. Stephens, and J. J. Wiens. 2005. Replicate patterns of species richness, historical biogeography, and phylogeny in holarctic treefrogs. Evolution 59:2433-2450.

Stanley, S. E., P. B. Moyle, and H. B. Shaffer. 1995. Allozyme analysis of delta smelt, *Hypomesus transpacificus* and longfin smelt, *Spirinchus thaleichthys* in the Sacramento-San Joaquin estuary, California. Copeia 1995:390-396.

Stearly, R. F. 1992. Historical ecology of Salmoninae, with special reference of *Oncorhynchus*, Pp. 622-658 in R. L. Mayden , ed. Systematics, historical ecology and North American freshwater fishes. Stanford University Press, Stanford, CA.

Stepien, C. A., A. K. Dillon, and A. K. Patterson. 2000. Population genetics, phylogeography, and systematics of the thornyhead rockfishes (*Sebastolobus*) along the deep continental slopes of the North Pacific Ocean. Can. J. Fish. Aquat. Sci. 57:1701-1717.

Swofford, D. L. 2002. PAUP*:phylogenetic analysis using parsimony. Ver. 4.0b10. Sinauer Associates, Sunderland, MA.

Tarp, F. H. 1952. A revision of the family Embiotocidae (the surfperches). Calif. Dep. Fish Game Fish Bull. 88:1-99.

Taylor, E. B. 1999. Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation. Rev. Fish Biol. Fish. 9:299-324.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX-Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.

Titova, L. V. 1994. Cenozoic history of Turritelloidea and Buccinoidea (Mollusca: Gastropoda) in the North Pacific. Palaeogeogr., Palaeoclimatol., and Palaeoecol. 108:319-334.

Trenham, P. C., H. B. Shaffer, and P. B. Moyle. 1998. Biochemical identification and

assessment of population subdivision in morphologically similar native and invading smelt species (*Hypomesus*) in the Sacramento–San Joaquin estuary, California. Trans. Am. Fish. Soc. 127:417-424.

Tsuchi, R. 1997. Marine climatic responses to Neogene tectonics of the Pacific Ocean seaways. Tectonophysics 281:113-124.

Uschakov, P. V. 1971. Amphipacific distribution of polychaetes. J. Fish. Res. Board Canada 28:1403-1406.

Väinölä, R. 2003. Repeated trans-Arctic invasions in littoral bivalves: molecular zoogeography of the *Macoma balthica* complex. Mar. Biol. 143:935-946.

Vermeij, G. J. 2005. From Europe to America: Pliocene to Recent trans-Atlantic expansion of cold-water North Atlantic molluscs. Proc. R. Soc. Lond. Ser. B. Biol. Sci. 272:2545-2550.

Wares, J. P., and C. W. Cunningham. 2001. Phylogeography and historical ecology of the North Atlantic intertidal. Evolution 55:2455-2469.

Waters, J. M., T. Saruwatari, T. Kobayashi, I. Oohara, R. M. McDowall, and G. P. Wallis. 2002. Phylogenetic placement of retropinnid fishes: data set incongruence can be reduced by using asymmetric character state transformation costs. Syst. Biol. 51:432-449.

White, B. N. 1987. Oceanic anoxic events and allopatric speciation in the deep sea. Biol. Oceanogr. 5:243-259.

Wiens, J. J., M. C. Brandley, and T. W. Reeder. 2006. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. Evolution 60:123-141.

Wilson, M.V. H., and R. R. G. Williams. 1991. New Paleocene genus and species of smelt (Teleostei: Osmeridae) from freshwater deposits of the Paskapoo Formation, Alberta, Canada, and comments on osmerid phylogeny. J. Vertebr. Paleontol. 11:434-451.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups. 2001. Trends, rhythms, and aberrations in global climate 65 ma to present. Science 292:686-693.

Zandee, M., and M. Roos. 1987. Component compatibility in historical biogeography. Cladistics 3:305-332.

Zenkevitch, L. A.1963. Biology of the seas of the U.S.S.R. George Allen & Unwin, Ltd., London (translated from Russian by S. Botcharskaya).

CHAPTER 4: MOLECULAR SYSTEMATICS OF THE OSMERIDAE[3]

4.1 INTRODUCTION

The northern hemisphere smelts (Osmeridae) are small (<30 cm), elongate, silvery fishes that display diverse life-history characteristics with marine, anadromous and freshwater forms, often within the same species. Osmerids are planktivores found in nearshore marine and coastal freshwaters and are important forage fishes for a number of fishes and marine mammals. The family has a Holarctic distribution, meaning they are found in cool-temperate and Arctic waters throughout the Northern Hemisphere; however, all species, with the exception of the capelin (*Mallotus villosus*) and Arctic rainbow smelt (*Osmerus dentex*), have more limited distributions, generally along single coastlines of the North Pacific and North Atlantic Oceans. The species composition and phylogenetic relationships among the taxa variously attributed to the Osmeridae have been much debated in the fish systematics literature (e.g., McAllister 1963, 1966; Begle 1991; Wilson and Williams 1991; Johnson and Patterson 1996).

In this chapter I conducted a molecular phylogenetic analysis of all Osmeridae species and evaluated the utility of the morphological characters used in previous systematic analyses of the family.

---

[3] A version of this chapter will be submitted for publication. K. L. Ilves, and E. B. Taylor. Molecular systematics of the Osmeridae.

4.1.1 History of osmerid systematics

Linnaeus (1758) originally classified the smelts within the salmonid genus *Salmo*, and although Cuvier (1817) eventually introduced a separate osmerid genus, *Osmerus*, the grouping of the smelts was still retained within the family Salmonidae. The osmerids continued to receive systematic and taxonomic attention throughout the early 20th century (Hubbs 1925; Kendall 1927; Chapman 1941); however, the most recent comprehensive revision of the Osmeridae to date is that of McAllister (1963). McAllister's (1963) systematic revision of the entire family was based on meristic and morphometric characteristics and resulted in a change in the number of recognized species from 14-16 to 10, and naming of a new species and subspecies (McAllister 1963).

However monumental, McAllister's (1963) study did not quell the debate about the interrelationships of the osmerids, as the uncertainty surrounding the phylogenetic placement and biogeography of this group of fishes continues to the present day. Although the systematic relationships among the teleost fishes have been the focus of more research effort than any other vertebrate group (Parenti 1986; Begle 1991), the difficulties encountered by evolutionary biologists studying the smelts have led Johnson and Patterson (1996) to lament that the "osmerids are unique in the disparity of opinion on their interrelationships".

4.1.2 Relationships within the Osmeroidea

Although alternatively placed in the order Osmeriformes (Begle 1991; Johnson and Patterson 1996; Helfman et al. 1997) or Salmoniformes (Nelson 1984), there appears to be general agreement that the northern hemisphere smelts fall into the superfamily Osmeroidea. Below the level of superfamily there exists much disagreement about how the various putative families and genera are related. The eight published phylogenies of the osmeroids are based on morphological characteristics (Fig. 4.1). The phylogenies disagree about the relationships among the species and genera, and how the monotypic Plecoglossidae, the Japanese Ayu *Plecoglossus altivelis*, and Salangidae, a northwestern Pacific temperate and subtropical family of icefishes, are related to the Osmeridae (Johnson and Patterson 1996). Uncertainty also extends to the relationships among the Osmeridae and the southern hemisphere smelts and galaxiids, families Retropinnidae and Galaxiidae, respectively. Although the Plecoglossidae and Salangidae are listed as separate families (Eschmeyer 2006), a classification that also has historical support (Chapman 1941; McAllister 1963; Klyukanov 1975), other morphological analyses have either nested both families within the Osmeridae (Howes and Sanford 1987; Johnson and Patterson 1996) or found the closest affinities of the Salangidae to be with the southern hemisphere group, with *Plecoglossus* relating most closely to members of the Osmeridae (Greenwood *et al*. 1966; Roberts 1984; Begle 1991).

Recent molecular analyses have increased our understanding of the interrelationships among these families. Based on mitochondrial DNA sequences Waters et al. (2002) determined that the Osmeridae (northern hemisphere) and

Retropinnidae (southern hemisphere) together are sister to the southern hemisphere Galaxiidae. Subsequent phylogenetic analyses (López et al. 2004; Fu et al. 2005) have confirmed that the two southern hemisphere families are not sister taxa, but found a more complex relationship among the other families. Waters et al. (2002) found the Salangidae to be nested within the Osmeridae (based on limited taxonomic sampling of both families), and the Plecoglossidae to be their sister group. Also using mitochondrial sequence data, *Galaxias* (Galaxiidae) as an outgroup, and including a more thorough taxonomic sampling of the Salangidae and Osmeridae, Fu et al. (2005) showed a poorly supported sister relationship between the Osmeridae and the Salangidae, which were in turn sister to the Plecoglossidae. In a broader lower euteleostean context, further analyses of mitochondrial and nuclear sequences showed strongly supported sister relationships between the Salangidae and Plecoglossidae, and these two families with the Osmeridae (López at al. 2004). López et al. (2004) also found surprising evidence that the Stomiiform fishes, not the Galaxiidae, are sister to the (Retropinnidae, Plecoglossidae, Salangidae, Osmeridae) clade.

4.1.3 Relationships within polytypic osmerid genera

Of the six osmerid genera, *Spirinchus*, *Osmerus*, and *Hypomesus* contain more than one species. While *Spirinchus* has received relatively little systematic attention, the relationships within *Osmerus* and *Hypomesus* have been extensively debated in the fish systematics literature.

Osmerus in particular has been the focus of considerable taxonomic effort, both morphological and molecular. This genus has a Holarctic distribution, found in near-shore marine habitats and coastal freshwaters of the Pacific, Atlantic and Arctic Ocean drainages. Earlier studies of the relationships among *Osmerus* species was complicated by unresolved taxonomic issues. In his 1963 revision of the family, McAllister recognized two forms: the European *O. eperlanus eperlanus* and *O. e. mordax* from the Arctic, Pacific and western Atlantic. Subsequent revisions (Klyukanov 1969, cited in Scott and Crossman 1973; McAllister et al. 1980) divided the genus into *O. eperlanus* (Europe) and *O. mordax mordax* (western Atlantic) and *O. m. dentex* (Pacific and Arctic). These latter taxonomic designations imply that the western Atlantic and Pacific and Arctic forms are more closely related to each other than either is to the European *Osmerus*.

Luey et al. (1982) performed the first molecular analysis of the relationships within *Osmerus*, using electrophoretic analysis of 13 allozyme loci. They found, in contrast to the taxonomy of Klyukanov (1969), Scott and Crossman (1963) and McAllister et al. (1980), that the two Atlantic forms are genetically each other's closest relatives. More recent study of mitochondrial cytochrome *b* sequence divergence and restriction fragment length polymorphisms (RFLPs) (Taylor and Dodson 1994) conflicted with the allozyme data of Luey et al. (1982) by finding a sister relationship between the western Atlantic and Pacific and Arctic forms, as suggested by earlier morphological analyses (Klyukanov 1969, cited in Scott and Crossman 1973; McAllister et al. 1980). Although the most recent molecular work supports the relationships found in these latter studies, full consensus on the

taxonomy of the genus has not yet been reached, as some authors retain the North Pacific and western North Atlantic populations as subspecific forms (*O. mordax dentex* and *O. m. mordax*, respectively) (Haldorson and Craig 1984; Eschmeyer 2006), while others consider the two North American marine forms to belong to separate species (*O. dentex* and *O. mordax*) (Nellbring 1989; Taylor and Dodson 1994). For this study, I consider the three forms as separate species.

As discussed in Chapters 2 (Ilves and Taylor 2007a) and 3 (Ilves and Taylor 2007b), a general consensus on the relationships among the species of *Hypomesus* has also been elusive. Representatives of this genus have a pan-Pacific distribution, ranging from Korea to the Arctic and down the western coast of North America to northern California (McAllister 1963). Like other osmerids, members of *Hypomesus* are found in marine, brackish and freshwaters (Hart and Clemens 1973). Systematic analyses based on morphological characteristics (McAllister 1963; Klyukanov 1970; Saruwatari et al. 1997) have led to changes in the taxonomy of this genus, with McAllister's original sub-species designations elevated to full species status by Klyukanov (1970) and Saruwatari et al. (1997) [detailed in Chapter 3; Ilves and Taylor 2007b]. Although there are currently six recognized species within the genus, recent morphological (Sidorov and Pichugin 2004) and molecular analyses (Chapter 2; Ilves and Taylor 2007a) found that a newly identified species from the Kuril Islands of Japan, *H. chishimaensis* (Saruwatari et al. 1997), is indistinguishable from *H. nipponensis*; therefore, in this study I consider the two species as conspecific under the *H. nipponensis* species designation.

4.1.4 Molecular systematics of the Osmeridae

Each of the eight previous morphological analyses of the species within the Osmeridae yielded a different hypothesis about their interrelationships (Fig. 4.1). Conflict has resulted not only from disagreements about character coding, but also, and perhaps more significantly, from the choice of which characters to include (Waters et al. 2002). Homoplasious characters that evolved independently in separate Osmeroidea lineages caused well-documented problems (Johnson and Patterson 1996, and references therein; Waters et al. 2002). With this pattern of character evolution, it is not surprising that determining the relationships within this group has proved challenging.

The aim of the current study was to conduct a phylogenetic analysis of the Osmeridae based on multiple independent molecular loci with the hope that this approach would lead to improved phylogenetic resolution. To this end, I sequenced segments of three mitochondrial (cytochrome *b* [*cytb*], 16S rRNA [16S], 12S rRNA [12S]) and three nuclear (internal transcribed spacer 2 [ITS2], S7 ribosomal protein intron 1 [S71], recombination activating gene 1 [RAG1]) genes from multiple individuals of each species within the Osmeridae. Because of the debate surrounding the phylogenetic position of the Plecoglossidae, Salangidae, Retropinnidae and Galaxiidae, I initially analysed *cytb*, 16S, 12S, and RAG1 sequences to determine an appropriate outgroup. After the preliminary analysis, the aims of this study have been to (1) conduct a molecular phylogenetic analysis of the Osmeridae (2) compare the resulting phylogeny to previous hypothesis of Osmeridae interrelationships and (3) examine the evolution of characters used in

previous studies using the molecular phylogeny and assess their utility for determining the systematic relationships within the Osmeridae.

4. 2 MATERIALS AND METHODS

4.2.1 Taxon sampling

A list of all samples and their geographic locations used in the phylogenetic analyses appears in Table 4.1. To take into account potential problems with lineage sorting (Maddison 1997), 4-25 individuals per species were sequenced for at least a subset of the six genes used in this study (Table 4.2). The samples used in this study came from at least two geographic locations per species, with the following exceptions: (1) *Hypomesus transpacificus*, which is endemic to the Sacramento – San Joaquin River, CA, (2) *H. japonicus*, for which samples were only available from a single location in Hokkaido, Japan, (3) *Spirinchus lanceolatus*, found only along eastern Hokkaido, and (3) the outgroup *Plecoglossus altivelis*, for which samples were only available from Lake Biwa, Japan. Although most species have samples from multiple locations, full range coverage was not available, and it is unknown how much of the intraspecific diversity they represent.

4.2.2 DNA sequence data

DNA extraction, PCR, and sequencing protocols were performed as described in Chapter 2 (Ilves and Taylor 2007a), with the following additions and

modifications. The mitochondrial 12S gene was amplified with the primers 12SF and 12SR (Table 4.3) in 50 μl reactions containing 50-300 ng of genomic DNA and final concentrations of 800 μM of dNTPs, 800 nM of each primer, 1.25 units of New England Biolabs (NEB) *Taq* DNA polymerase, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, and 2.5 mM $MgCl_2$ under the following conditions: 95 C for 5 min, followed by 30 cycles of 1 min each of 95 C, 50 C and 72 C, and a final extension at 72 C for 5 min. Amplification of the nuclear RAG1 gene using the RAG1F, RAG9R combination (Table 4.3) proved difficult for *Osmerus dentex*, *O. eperlanus*, and *Plecoglossus altivelis*, so additional primer pairs from López et al. (2004) were used for these species. *O. dentex* was amplified with primers RAG1F1 and RAG1R1 (Table 4.3) in 25 μl reactions with final concentrations of 800 μM of dNTPs, 800 nm of each primer, 1 unit of NEB *Taq* DNA polymerase, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, and 2.5 mM of $MgCl_2$ under the following conditions: 94 C for 3 min, followed by 35 cycles of 45 s at 95 C, 45 s at 52 C and 1 min 15 s at 72 C, and a final extension at 72 C for 5 min. *P. altivelis* and *O. eperlanus* were amplified with the primers RAG1F1 and RAG1R2 (Table 4.3) with the same reaction conditions described for *O. dentex*. The PCR profile for these two species was the same as that for *O. dentex*, except a 53 C annealing temperature was used for *P. altivelis*. The mitochondrial 16S and nuclear ITS2 regions were amplified with the primers and conditions described in Chapter 2 (Ilves and Taylor 2007a), except that a 55 C annealing temperature was used for most samples.

4.2.3 Phylogenetic analyses

Sequence alignment and general phylogenetic methods

Sequences were aligned using ClustalX (Thompson et al. 1997) or manually with MacClade v4.06 (Maddison and Maddison 2003) and edited with Se-Al v2.0a11 (Rambaut 1996) or MacClade v4.06 (Maddison and Maddison 2003). Alignments of protein-coding genes cyt*b* and RAG1 and the mitochondrial 12S rRNA gene were unambiguous. By contrast, the presence of indels and repeats in the 16S rRNA, ITS2 and S71 regions made it difficult to determine positional homology in many regions. As a result, such ambiguously aligned characters were excluded from subsequent analyses (see Results for details).

Neighbour-joining (NJ), parsimony, and maximum likelihood (ML) phylogenetic analyses were all implemented in PAUP* v.4.0b10 (Swofford 2002). Bayesian analysis was conducted using MrBayes v.3.1.2 (Huelsenbeck and Ronquist 2001). Modeltest v3.6 (Posada and Crandall 1998) was used to select a model of sequence evolution for each data partition (individual and combined) to use for NJ, Bayesian and ML analyses. In each analysis, the model chosen by the Akaike Information Criterion (AIC) method was implemented (Posada and Buckley 2004). The models corresponding to each partition were as follows: TrN+I+G (cyt*b* and cyt*b*+16S), TrNef+I (16S), TVMef+I+G (12S), TVM+I+G (ITS2 and all mtDNA), TrN+G (S71 and RAG1), GTR+I+G (all nDNA and allDNA).

Support for monophyletic groups was assessed using 1000 bootstrap pseudo-replicates (Felsenstein 1985) for NJ, parsimony and ML analyses, and by posterior probabilities for Bayesian analyses. Heuristic searches for parsimony and

ML analyses were conducted with 10 random replicates of stepwise taxon addition, except for ML bootstrap analyses where 5 such replicates were used. NJ and Bayesian analyses were conducted on all datasets and parsimony bootstrap analysis was used for all reduced datasets. ML heuristic searches were conducted for all reduced data partitions, and ML bootstrapping was performed for the combined mtDNA, nDNA and allDNA (mtDNA + nDNA) partitions using a single individual per species (see following sections for more detail). Bayesian analyses on all datasets were run in two parallel analyses for 2 $x10^6$ generations, trees sampled every 100 generations, with a burnin of 2000. A graphical plot of likelihood over generations indicated that stabilization of likelihoods apparently occurred by 2000 generations, meaning that all trees sampled are from well after this stabilization (100x number of generations required for stabilization).

Justification of the outgroup

Recent molecular phylogenetic work has shown that the Plecoglossidae and Salangidae are sister to the Osmeridae (López et al. 2004), but due to continued debate regarding the relationships among the Osmeridae, Salangidae and Plecoglossidae (e.g. Johnson and Patterson 1996, and references therein; López et al. 2004; Fu et al. 2005) preliminary analyses were conducted to determine an appropriate outgroup. Not all species of the Osmeridae were considered in López et al.'s (2004) study, but my preliminary analyses supported choosing *Plecoglossus altivelis* as a suitable outgroup. NJ, parsimony and Bayesian analyses of *cytb*, 16S, 12S and RAG1 from *Galaxias fasciatus* (southern hemisphere smelt, family

Galaxiidae), *Retropinna retropinna* (southern hemisphere smelt, family Retropinnidae), *Salangichthys microdon* (western Pacific icefish, family Salangidae), *P. altivelis* (Japanese ayu, family Plecoglossidae), and one individual from each of the Osmeridae proper were conducted. Sequences for *G. fasciatus* and *S. microdon* were obtained from GenBank (Table 4.4), while those for *R. retropinna*, *P. altivelis*, and all species of the Osmeridae were generated for the current study using available tissue samples (Table 4.1), apart from the 12S sequence for *R. retropinna* (Table 4.4). Analysis of the nuclear RAG1 gene with *G. fasciatus* as the outgroup showed that *P. altivelis* was a member of the sister group to the Osmeridae, likely as sister to the Salangidae (Fig. 4.2B). Similarly, analysis of all mtDNA (cyt*b*, 16S, 12S) and nDNA (ITS2, S71, RAG1) gene regions sequenced in this study with *R. retropinna* as the outgroup also confirmed that *P. altivelis* was not embedded within the Osmeridae (Fig. 4.3). As *P. altivelis* falls outside the Osmeridae, is one of the most closely related taxa to this group, and tissue samples for this species were available, it was chosen as the outgroup for the current phylogenetic study.

Sequence selection

Due to the large amount of sequence data (Table 4.2), NJ and Bayesian analyses were run for all sequences for each gene to identify individuals of each species that represent a range of the intraspecific variation to be included for more rigorous ML and parsimony analyses. This first step of analysis included 1000 NJ bootstrap pseudo-replicates and Bayesian analyses (2×10^6 generations, burnin 2000, 18001 sampled trees) of all sequences for each gene separately (Appendices

4-9). A reduced dataset, consisting of two to three individuals per species, was then designed to include a range of intraspecific variation based on branching patterns within each species for each gene. In most cases the same individuals were used across genes, although in a few cases a sequence from a conspecific was substituted in order to capture intraspecific variation and sequence quality. These substitutions are listed in Appendix 3. To allow thorough ML bootstrap analysis of the combined mtDNA, nDNA and allDNA datasets, a further reduction was made where a single individual was chosen to represent each species. NJ, Bayesian and parsimony analyses were also performed for these further reduced datasets to compare topologies resulting from different reconstruction methods.

Partition homogeneity tests

Partition homogeneity tests [incongruence length difference tests; (Farris et al. 1994; Swofford 2002)] were performed on the single individual per species datasets to determine whether it was appropriate to combine the data into a single concatenated matrix for the mtDNA, nDNA and allDNA partitions. These tests were implemented in PAUP* v.4.0b10 (Swofford 2002) with 1000 replicates using a heuristic search with 10 replicates of random sequence addition and TBR branch swapping.

Topology tests

Shimodaira-Hasegawa (SH) tests (Shimodaira and Hasegawa 1999) were performed in cases where placements of particular taxa conflicted and also to compare the molecular phylogeny of the Osmeridae generated in this chapter to previous hypotheses. The SH test compares the likelihoods of alternative topologies for a given dataset and was used to evaluate whether some topologies were significantly better than others (Felsenstein 2004). This test was implemented in PAUP* v.4.0b10 (Swofford 2002) using the likelihood model selected by the AIC method in Modeltest v3.6 (Posada and Crandall 1998) for the corresponding dataset, with 10000 RELL bootstrap replicates.

4.2.4 Character mapping

To compare the results of the molecular phylogenetic analysis with previous morphological studies of Osmeridae systematics, the characters used by McAllister (1963, 1966), Begle (1991), Wilson and Williams (1991) and Johnson and Patterson (1996) were mapped onto the molecular phylogeny by entering the characters and their states into a character matrix using MacClade (Maddison and Maddison 2003). Multistate characters were coded as unordered with *Plecoglossus altivelis* as the outgroup. Begle (1991) and Johnson and Patterson (1996) included additional taxa not used in the current study, therefore, their matrices were scanned for characters that showed variation among the Osmeridae and Plecoglossidae species sequenced for the molecular systematic analysis. The most comprehensive morphological assessment was conducted by Johnson and Patterson (1996), which in conjunction

with Patterson and Johnson (1997) included a vigorous critique of Begle's (1991) work. Johnson and Patterson (1996) also noted differences among their analysis and those of McAllister (1963, 1966), Howes and Sanford (1987), and Wilson and Williams (1991) in how some characters are coded. In places of disagreement between studies, I included Johnson and Patterson's (1996) and Patterson and Johnson's (1997) character state designations as these studies include the most thorough assessments of the Osmeridae to date. Further, because Johnson and Patterson's (1996) phylogeny included the Salangidae and Plecoglossidae nested within the Osmeridae, I separately reviewed the characters used in their 1996 and 1997 (Patterson and Johnson) studies with a phylogeny that includes the Salangidae. Appendix 16 lists the characters and state names. Appendices 17 and 18 contain the data matrices for the entire character dataset and the dataset from Johnson and Patterson (1996) and Patterson and Johnson (1997), respectively.

4.3 RESULTS

4.3.1 Sequences

The number and length of sequences for each gene and species are listed in Table 4.2. Due to ambiguous sequence alignments in the 16S, ITS2, and S71 matrices, 28, 242, and 65 characters were excluded from the respective datasets with all sequences. The regions of the excluded characters remained the same in the reduced datasets, however, the number of excluded sites changed slightly due

to the presence of empty positions as a result of the deletion of particular sequences. Because 12S sequences were not available for all individuals used in the reduced analyses based on 2-3 individuals per species, combined mtDNA and allDNA partitions included only *cytb* and 16S sequences; however, the further reduced partitions with only a single individual representing each species included 12S. Including indels, the reduced mtDNA (*cytb* + 16S), nDNA and mtDNA + nDNA datasets contained 959, 2597, and 3556 characters, respectively, while the mtDNA (*cytb* + 16S + 12S), nDNA, and allDNA datasets based on a single individual contained 1352, 2597, and 3949 characters, respectively.

4.3.2 Species monophyly

Analyses based on all individuals and a reduced number of individuals showed that all species were reciprocally monophyletic for each gene in almost all cases (Appendices 4-15). There was evidence for incomplete lineage sorting between *Hypomesus pretiosus* and *H. transpacificus* for all genes apart from S71 (Appendices 10-15), and some other species were not reciprocally monophyletic for 16S (*Thaleichthys pacificus*; Appendix 11), 12S (*T. pacificus*; Appendix 12), ITS2 (*Osmerus dentex*, *O. mordax*, *Spirinchus lanceolatus*; Appendix 13), and S71 (*S. starski*, *S. thaleichthys*; Appendix 14). Phylogenies from the complete (Appendices 4-9) and reduced datasets (Appendices 10-15) were the same apart from a few changes at relatively poorly supported nodes. As a result, it is unlikely that the choice of sequences for the reduced datasets had a substantial effect on the main results and conclusions from the analyses.

4.3.3 Individual gene analyses

Analysis of the individual genes generally resulted in poorly resolved phylogenies (Appendices 10-15). Of the six genes, topologies from the nuclear S71 (Appendix 14) and RAG1 (Appendix 15) loci most closely matched those of the combined dataset analyses (Figs. 4.4-4.6). Although the intergeneric relationships were not fully resolved, the monophyly of the three polytypic genera, *Hypomesus*, *Osmerus* and *Spirinchus*, was supported by several genes: *cytb* (*Hypomesus*, *Osmerus*; Appendix 10), 16S (*Osmerus*; Appendix 11), 12S (*Hypomesus*; Appendix 12), and S71 (Appendix 14) and RAG1 (Appendix 15) for all three genera. Analyses of the individual genes also showed apparent conflicts within these polytypic genera, and in the placement of *Mallotus villosus* (discussed further in following sections). Analysis of ITS2 yielded relationships not seen in any of the other analyses, individual or combined, although with very low levels of support (Appendix 13).

4.3.4 Combined dataset analyses

Partition homogeneity tests showed no significant conflict among the mtDNA ($P = 0.54$) or nDNA ($P = 0.41$) genes, but did suggest conflict between the mtDNA and nDNA datasets ($P = 0.01$). Because there was little phylogenetic resolution from the mtDNA dataset and the nDNA and allDNA data resulted in the same topology apart from poorly supported relationships in *Osmerus* (discussed below),

this suggests a significance level at $P = 0.05$ is too stringent for these data. Thus, I considered it acceptable to combine all of the data for analysis.

The phylogenetic trees generated by analysis of the combined mtDNA (*cytb*, 16S, 12S), nDNA and allDNA datasets based on a single individual per species are shown in Figures 4.4-4.6, respectively. Analysis of the single individual/species mtDNA dataset did not produce a fully resolved phylogeny, but showed some support for the monophyly of *Hypomesus* and *Osmerus* (Fig. 4.4). By contrast, a well-supported phylogeny of the Osmeridae resulted from analysis of the nDNA (Fig. 4.5) and allDNA (Fig. 4.6) datasets. A summary tree showing the hypothesis of osmerid relationships that emerged from this study is shown in Fig. 4.7.

The different methods of phylogenetic reconstruction for the nDNA and allDNA datasets yielded the same intergeneric relationships, apart from the interesting exception of *Mallotus villosus*. For both of these partitions, *M. villosus* appears as sister to the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*)) [hereby designated as OTAS] clade using NJ, ML and Bayesian methods, but as sister to all genera with parsimony analysis (Figs. 4.5 and 4.6). All phylogenetic reconstruction methods of the nDNA and all DNA datasets show *Allosmerus* and *Spirinchus* as the most derived sister taxa, which are sister to *Thaleichthys*. *Osmerus* is shown as sister to this latter clade and *Hypomesus* appears as sister to the rest of the genera (Figs. 4.5 and 4.6). Support values for these relationships were generally high across reconstruction methods with Bayesian posterior probabilities of 0.92-1.0 and ML bootstrap values 84-100% for the nDNA dataset (Fig. 4.5) and 0.97-1.0 and 88-100%, respectively, for the allDNA dataset (Fig. 4.6).

4.3.5 Placement of *Mallotus villosus*

As mentioned in sections 4.3.3 and 4.3.4, the placement of *Mallotus villosus* as either the sister to the OTAS clade or as sister to all of the genera is unclear. To test whether one of these topologies is significantly better than the other, SH tests were performed using the nDNA, all DNA, and allDNA datasets with *P. altivelis* and *Retropinna retropinna* as the outgroup, using the topologies from the single individual per species analyses. For the nDNA and allDNA datasets with *Plecoglossus altivelis* as the outgroup, the topology with *M. villosus* sister to the OTAS clade had a higher likelihood, whereas the alternative topology had a higher likelihood with *R. retropinna* as the outgroup; however, the likelihoods were not significantly different from each other (Table 4.5).

4.3.6 Relationships within *Osmerus*

Relationships among the three *Osmerus* species were poorly resolved. At least one tree from the *cytb*, 16S, S71, mtDNA, and allDNA datasets showed *O. dentex* and *O. eperlanus* as sister species, but *O. dentex* and *O. mordax* appeared as sister species in cyt*b* and nDNA analyses and *O. mordax* and *O. eperlanus* were sister taxa in trees from RAG1 analyses (Figs. 4.4-4.6; Appendices 10-15). All relationships were poorly supported, apart from the *O. dentex-O. eperlanus* relationship with ML analysis of the allDNA dataset, which received a moderate 79% bootstrap support (Fig. 4.6).

SH tests performed on the mtDNA, nDNA and allDNA partitions with two outgroups: *Plecoglossus altivelis* and reduced datasets with only *Mallotus villosus* and *Osmerus*. Table 4.6 shows that the *O. dentex* – *O. eperlanus* sister relationship has the highest likelihood in most (but not all) comparisons; however, none of the topologies are significantly different.

4.3.7 Relationships within *Spirinchus*

Similar to the situation in *Osmerus*, there is apparent conflict among genes and combined data partitions regarding the relationships within *Spirinchus*. The *cytb*, 16S, RAG1, and mtDNA datasets support a sister relationship between *S. lanceolatus* and *S. thaleichthys*, while the S71, nDNA and allDNA partitions suggest a *S. starksi* – *S. thaleichthys* relationship. The *S. lanceolatus* – *S. thaleichthys* relationship is poorly supported in all analyses and datasets, whereas the *S. starksi* – *S. thaleichthys* relationship is moderately to highly supported in the S71, nDNA and allDNA partitions (Appendix 14, Figs. 4.5 and 4.6, respectively). Even with the apparent high level of support, SH tests between the two topologies under the mtDNA, nDNA and allDNA datasets, with either the entire phylogeny with *Plecoglossus altivelis* as the outgroup or a reduced matrix with only *Allosmerus* and *Spirinchus*, the only significant difference was with the mtDNA dataset with *P. altivelis* as the outgroup. In this instance, a tree with *S. lanceolatus* and *S. thaleichthys* as sister taxa was significantly better than the alternative tree with *S. thaleichthys* and *S. starksi* as sister species (Table 4.7). Because the topology of

this tree was generally unresolved (Fig. 4.4), this significant difference should be interpreted with caution.

4.3.8 Relationships within *Hypomesus*

As discussed in Chapter 3, the placement of *Hypomesus olidus*, as either sister to *H. nipponensis* or to the (*H. japonicus* (*H. pretiosus*, *H. transpacificus*)) clade, was unclear. This uncertainty also appears to extend to the family level phylogeny, where different partitions showed support for both of these arrangements. The 12S (Appendix 12) , ITS2 (Appendix 13), and combined mtDNA (Fig. 4.4) datasets showed low levels of support for a *H. olidus* – *H. nipponensis* sister relationship, while there was mixed support for the alternative topology for the S71, RAG1, nDNA and allDNA partitions (Appendices 14 and 15; Figs. 4.5 and 4.6, respectively). The relationship between these species was unresolved for *cytb* (Appendix 10) and 16S (Appendix 11) and analysis of ITS2 showed a poorly supported sister relationship between *H. olidus* and *H. nipponensis*, although *Hypomesus* was not monophyletic with this locus (Appendix 13).

SH tests were performed using the mtDNA, nDNA and allDNA datasets to test whether the likelihoods of the two alternative topologies for *Hypomesus* are significantly different. Table 4.8 shows that the topology with *H. olidus* sister to (*H. japonicus* (*H. pretiosus*, *H. transpacificus*)) has a higher likelihood under the nDNA and allDNA partitions, while the *H. olidus* – *H. nipponensis* sister relationship has a higher likelihood with the mtDNA dataset; however, these likelihoods are not significantly different from one another.

4.3.9 Comparison of molecular phylogeny and previous hypotheses

A main objective of this chapter was to compare a molecular phylogeny of the Osmeridae with eight previous morphological hypotheses of their interrelationships. To this end, I compared the molecular topology generated from Bayesian and ML analysis of the allDNA dataset (Fig. 4.6) to the topologies shown in Figure 4.1. Because four of the phylogenies in Figure 4.1 did not include *Plecoglossus altivelis*, I removed this taxon from the molecular phylogeny for those comparisons. In the phylogeny of Johnson and Patterson (1996), the Salangidae were embedded within the Osmeridae, so in order to test the molecular phylogeny against their topology, I used a dataset consisting of *cytb*, 16S, 12S, and RAG1, for which gene sequences from *Salangichthys microdon*, a species within the Salangidae, were available from GenBank. The molecular phylogeny I tested placed *S. microdon* and *P. altivelis* as sister species.

Results of the SH tests showed that my molecule-based phylogeny had a significantly higher likelihood than all eight previous hypotheses of Osmeridae interrelationships (Table 4.9). Excluding *Mallotus villosus*, the placement of which was uncertain, yielded the same result (data not shown).

4.3.10 Character mapping

An initial matrix consisting of 155 characters was generated from the data of McAllister (1963, 1966), Begle (1991), Wilson and Williams (1991), Johnson and

Patterson (1996), and Patterson and Johnson (1997). After scanning this matrix for duplicate characters and incorporating the character deletions and coding changes suggested by Johnson and Patterson (1996) and Patterson and Johnson (1997), a final matrix of 121 characters was mapped onto the molecular phylogeny. A trace of all characters with unambiguous changes onto the molecular tree was 185 steps (Fig. 4.8). Tree lengths from mapping these same characters onto the topologies from the earlier studies were: 158 [McAllister 1963, 1966; *Plecoglossus* excluded]; 221 [Begle 1991]; 182 [Wilson and Williams 1991]; 182 [Johnson and Patterson (1996) and Patterson and Johnson (1997) – Salangidae data not available for most taxa]. If *Plecoglossus* is included at the base of the McAllister (1963) and (1966) trees their lengths increase to 187 and 189, respectively. With only data from Johnson and Patterson (1996) and Patterson and Johnson (1997), which included the Salangidae, their phylogeny was 111 steps, while the molecular tree with *Plecoglossus* and the Salangidae as sister taxa was 120 steps.

Although not shown on the phylogeny, *Plecoglossus altivelis* differed from the species of the Osmeridae by 14 characters. Fig. 4.8 showed one shared character state uniting the (*Mallotus*, (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, Spirinchus)))) clade, 21 character states defining the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) clade, and five character states shared by *Thaleichthys*, *Allosmerus,* and *Spirinchus*. Of the 121 characters, none exclusively defined the *Allosmerus* – *Spirinchus* sister relationship, although they share the 'no cucumber odor' state of character 64 along with *P. altivelis* and have a head length of 4.7 or greater than standard length (character 100), which is also shared by at least one species of

Osmerus (state of this character is missing for *P. altivelis*). *Osmerus* apparently lacks unique character states in this dataset. The numerous homoplasies in the morphological characters studied are indicated by the triangular markings on Fig. 4.8.

4.4 DISCUSSION

While there have been several recent molecular studies that have included members of the Osmeridae (e.g., Waters et al. 2002; López et al. 2004; Fu et al. 2005), my study is the first to include all members of the family in a multi-locus phylogenetic analysis.

4.4.1 Relationships of the Osmeridae and related families

Phylogenetic analysis of nDNA RAG1 sequences from the Osmeridae, Plecoglossidae and a single representative of the Salangidae, using *Galaxias fasciatus* as the outgroup, supported the results of López et al. (2004) that the Plecoglossidae and Salangidae fall outside the Osmeridae proper (Fig 4.2B), while analysis of combined mtDNA genes cyt*b*, 12S, and 16S nested the two former families within the Osmeridae with very low support (Fig. 4.2A). A sister relationship between the Plecoglossidae and Salangidae, as suggested by López et al. (2004) was also generally supported by these analyses; however, parsimony analysis of the mtDNA and RAG1 genes indicated a poorly supported sister relationship between

the Salangidae and Osmeridae (data not shown). Additional nuclear sequence data for the salangids may further clarify their position relative to *Plecoglossus* and the osmerids.

4.4.2 Molecular systematics of the Osmeridae

Intergeneric relationships

Although analyses of the combined mitochondrial sequences produced generally unresolved phylogenies (Fig. 4.4), all analyses of the combined nDNA and all DNA datasets produced a highly supported phylogeny of the Osmeridae (Figs. 4.5 and 4.6, respectively). As noted in the Results, the placement of *Mallotus villosus* differs between the phylogenies resulting from parsimony, where it appears as sister to all other genera, and the other reconstruction methods, where it is sister to the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*)) [OTAS] clade (Figs. 4.5 and 4.6). A possible explanation for the difference in topology may be long branch attraction, where the high substitution rate along the *M. villosus* lineage is interpreted by the parsimony reconstruction method as indicative of a closer relationship between this taxon and the outgroup *Plecoglossus altivelis* (Felsenstein 2004). Although this possibility cannot be rejected outright, the evidence from these datasets points more strongly towards a phylogeny where *Mallotus* is sister to the OTAS clade based on high Bayesian and ML bootstrap support values for the nDNA and allDNA combined datasets (Figs. 4.5 and 4.6, respectively). The summary topology in Fig. 4.7 shows the hypothesis of Osmeridae interrelationships that I will

use for subsequent analyses of character mapping (section 4.4.3) and biogeographic reconstruction (Chapter 5).

Relationships within polytypic genera

While the intergeneric relationships within the Osmeridae are generally highly supported, the same cannot be said of the relationships within the three polytypic genera *Spirinchus*, *Osmerus* and *Hypomesus*. Different phylogenetic reconstruction methods and different loci yield different relationships among the species of all of these genera, and SH tests under a variety of conditions failed to show significant differences in almost every case (Tables 4.5-4.8, respectively). Conflicting hypotheses of the species relationships within these genera are not unique to my molecular analysis.

Spirinchus

Of the three polytypic genera, *Spirinchus* has received the least systematic attention. McAllister (1963, 1966) designated S*. thaleichthys* and *S. lanceolatus* as sister species, a relationship that also appeared in most phylogenies of the current study, albeit with generally low levels of support (Appendices 10, 12, 15; Figs. 4.4, 4.6). The alternative *S. thaleichthys – S. starksi* relationship, on the other hand, is highly supported in Bayesian analysis of the nDNA and allDNA partitions and moderately supported by ML analysis of these datasets (Figs. 4.5 and 4.6, respectively). Bayesian analysis of all sequence data excluding S71, also strongly

suggests *S. thaleichthys* and *S. starksi* are sister species (0.98 posterior probability based on Bayesian analysis, data not shown).

Osmerus

In contrast to *Spirinchus*, *Osmerus* has been the focus of considerable taxonomic effort, both morphological and molecular. A sister relationship between *O. dentex* and *O. mordax* was supported by the morphological analyses of McAllister (1963) and Klyukanov (1969) and subsequently by Taylor and Dodson's (1994) mitochondrial sequence and RFLP analyses. By contrast, Luey et al.'s (1982) allozyme analysis suggested *O. mordax – O. eperlanus* are sister species.

Unfortunately, the current molecular study of mitochondrial and nuclear genes sequences does little to resolve the conflicts in our understanding of the relationships among the three species of *Osmerus*. While an *O. dentex – O. eperlanus* relationship appears most frequently from analysis of the various data partitions in this study, there is also evidence for both alternative relationships and SH tests did not find any significant differences between the phylogenies (Table 4.6). Subsequent analysis of character evolution is not dependent on the relationships among these species as the characters are at the genus level (Section 4.4.3).

Hypomesus

While the discrepancies among the relationships of the genus *Hypomesus* are not as extensive as those for *Spirinchus* and *Osmerus*, there is some question

as to whether *H. olidus* and *H. nipponensis* are sister taxa or whether *H. nipponensis* is sister to the rest of the *Hypomesus* species. Although SH tests failed to find a significant difference between the topologies with *H. olidus* as sister to *H. nipponensis* and the alternative with *H. nipponensis* as sister to the rest of the species (Table 4.7), the former arrangement is poorly supported when it appears, while the latter is highly supported by Bayesian (1.0) and ML (89%) analysis of the nDNA dataset (Fig. 4.5) and by Bayesian analysis (0.99) of the allDNA partition (Fig. 4.6). Therefore, I consider the topology with *H. nipponensis* as sister to the rest of *Hypomesus* to be the best hypothesis of the relationships within this genus based on the current dataset.

Understanding the discrepancies

Unlike the situation with the placement of *Mallotus*, long branch attraction does not appear to play a role in the uncertainties within *Spirinchus*, *Osmerus* and *Hypomesus*, as parsimony analysis did not yield different results from those of the other reconstruction methods. Mitochondrial and nuclear phylogenies did not conflict so there was no evidence for an explanation based on an hypothesis of historical hybridization between species. Alternatively, a possible explanation for the discrepancies is that the divergences within these genera may have occurred in a relatively short span of time compared to the mutation rates of the sequenced genes, thereby obscuring the relationships among the species. Increased geographic sampling of widely distributed *Spirinchus*, *Osmerus* and *Hypomesus* species may help capture more of the genetic variation in these species; however,

sequencing of additional loci will be necessary to overcome difficulties in determining branching relationships among species with short internodes.

4.4.3 Review of characters used in previous systematic studies

The molecular phylogeny from this study is incongruent with all previous hypotheses of relationships among Osmeridae genera. Thus, a review of the morphological characters used in the most comprehensive examination of these fishes to date, Johnson and Patterson (1996) and Patterson (1997), with comments on studies by McAllister (1963, 1966), Begle (1991), Howes and Sanford (1987), and Wilson and Williams (1991) may identify possible reasons for the disagreements. Johnson and Patterson's (1996) hypothesis (Fig. 4.9) conflicts with the molecular phylogeny in three respects: (1) They place *Hypomesus* as the most basal taxon as opposed to *Plecoglossus* and the Salangidae, (2) In their analysis, the Salangidae are sister to *Mallotus* as opposed to sister to *Plecoglossus*, and (3) They consider *Allosmerus* and *Thaleichthys* as sister taxa, as opposed to an *Allosmerus-Spirinchus* relationship from the molecular analysis (Figs. 4.5 and 4.6).

(1) *Hypomesus* as most basal taxon

Mapping the 52 characters from Johnson and Patterson (1996) and Patterson and Johnson (1997) that vary among the taxa included in this study onto their phylogeny did not reveal any uncontradicted synapomorphies that separated *Hypomesus* from the other taxa and only one character [median keels of laminar

bone present on distal parts of last few neural and haemal spines] unambiguously separated *Hypomesus* and *Plecoglossus* from the other taxa (character 23; data not shown because state autapomorphic for both *Hypomesus* and *Plecoglossus*). This contradicted the phylogenies in Fig. 19 of Johnson and Patterson (1996) where, depending on the optimization used (favouring either reversals or forward changes), a topology with *Hypomesus* sister to *Plecoglossus* and the rest of the osmerids is depicted as sharing two or four such synapomorphies, while the separation of the osmerids from *Hypomesus* and *Plecoglossus* is defined by two or three uncontradicted character states. *Hypomesus* does have a unique state for character 12 (endopterygoid teeth concentrated along dorsal margin of bone, with a patch of teeth posteriorly; Fig. 4.9); however other osmeroid fishes outside the Osmeridae, Plecoglossidae and Salangidae also share this trait. A review of their character matrix (Appendix 1, Johnson and Patterson 1996) revealed that for almost all suggested uncontradicted synapomorphies, there is homoplasy among the taxa. For example, character 84 of their matrix (character 27, Fig. 4.9) concerning the articulation point of the fourth pectoral radial, which is indicated as an uncontradicted synapomorphy separating *Hypomesus* from the rest under both optimizations, is coded as '0' for *Hypomesus, Spirinchus* and the Salangidae, and '1' for all other taxa. A similar situation applies to characters 13 (their 25) and 19 (their 43; *Hypomesus* + others; character state shared by *Hypomesus* and Salangidae or *Mallotus*, respectively; Fig. 4.9) and 29 (*Hypomesus*, *Plecoglossus* + others; character state shared by *Hypomesus*, *Plecoglossus* and *Osmerus*; Fig. 4.9). The Salangidae are coded as missing data for characters 4 (their 9) and 19 (their 43;

Hypomesus + others) and 8 (their 16; *Hypomesus*, *Plecoglossus* + others). Thus, from Johnson and Patterson's (1996) dataset there appears to be limited evidence, at least on the basis of synapomorphies, for the placement of *Hypomesus* as basal to all other taxa. The placement of *Plecoglossus* and the Salangidae outside the Osmeridae clade has been supported by previous molecular studies (López et al. 2004; Fu et al. 2005) as well as by the preliminary analyses conducted in this chapter to assess the suitability of using *Plecoglossus* as the outgroup of the Osmeridae (Fig. 4.2).

(2) Placement of *Mallotus* and Salangidae

Molecular analyses placed the Salangidae outside the Osmeridae as defined in this chapter (Fig. 4.2; López et al. 2004; Fu et al. 2005). Johnson and Patterson (1996), on the other hand, considered the Salangidae a sister taxon to *Mallotus*. The phylogeny in Figure 4.9 shows five apparent synapomorphies uniting these taxa. For two of these traits (enlarged scales on anal fin base of mature males [34], and skeletal ontogeny retarded relative to sexual maturity [39]; Fig. 4.9), the *Mallotus*/salangid character state is shared by other osmeroid taxa (Appendix 1 in Johnson and Patterson 1996), indicating these traits are homoplasious in closely related fishes. The long unossified ethmoid endoskeleton [character 2], distally multifid fourth pectoral radial [character 28], and an adipose cartilage that is a transversely arched, fenestrate plate [character 30] appear to be synapomorphies for the *Mallotus*/Salangidae clade (Fig. 4.9); however, these character states are not all unique among fishes as a similar modification of the fourth pectoral radial is seen

in the alepocephaloid *Platytroctes* (Sazanov 1986, cited in Johnson and Patterson 1996), and adipose cartilage is found in other teleosts (Matsuoka and Iwai 1983, cited in Johnson and Patterson 1996). Further, Johnson and Patterson (1996) indicate that a transversely arched adipose cartilage is not found in two salangid genera (*Salanx* and *Salangichthys*), which makes it unclear how the coding of states for this family was determined for other characters.

Johnson and Patterson (1996) also discussed their confidence in the different parts of their Osmeridae phylogeny. They were confident that *Allosmerus*, *Thaleichthys* and *Spirinchus* were derived taxa, and *Hypomesus* was a basal osmerid, conclusions also supported by molecular data (Figs. 4.5 and 4.6; Appendices 13-15). Johnson and Patterson (1996) were less certain, however, about the arrangement of *Osmerus*, *Mallotus* and the Salangidae because exchanging the positions of *Osmerus* and (*Mallotus* + Salangidae) only increased the tree length by a small amount. Figure 4.9 shows that four synapomorphies from Johnson and Patterson's (1996) data define the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) clade. Including characters from other studies further increases the number of shared character states to 21 and exchanging the positions of *Osmerus* and *Mallotus* on this phylogeny increases the tree length from 185 to 203 steps (data not shown). Thus, while there may be limited support for the monophyly of the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) clade from Johnson and Patterson's (1996) dataset, this grouping is strongly supported by the molecular analyses from this chapter and numerous shared morphological characters from McAllister (1963) and Wilson and Williams (1991) [Fig. 4.8].

No unambiguous synapomorphies from the characters examined support *Plecoglossus* and the Salangidae as sister to the Osmeridae (data not shown); however, when the Salangidae are placed as sister to the Osmeridae, one character state supports the osmerids + salangids clade (character 46; fusion of 5th epibranchial to 4th to form circular foramen for efferent artery) and another defines the Osmeridae (character 45; metapterygoid with prominent diagonal lateral shelf) [data not shown]. An Osmeridae – Salangidae sister relationship was suggested by the mtDNA analyses of Fu et al. (2005); however, analyses of the nDNA RAG1 gene for lower euteleosts by López et al. (2004) supports the conclusions from this chapter (Fig. 4.2) that *Plecoglossus* and the Salangidae together are sister to the Osmeridae. No characters support an *Plecoglossus* – Salangidae sister relationship. *Plecoglossus* and the Salangidae may have diverged morphologically to such an extent that they no longer share any defining characters. An alternative explanation is that the short internodes among the Osmeridae, Plecoglossidae and Salangidae are obscuring the branching pattern among these families. Sequencing of additional loci may help further clarify these relationships.

(3) Relationships among *Allosmerus*, *Spirinchus* and *Thaleichthys*

Johnson and Patterson (1996) found a sister relationship between *Spirinchus* and *Thaleichthys* supported by two (globose otic bulla [8], fontanelles in cartilaginous roof of otic region remaining closed during ontogeny [10]) or three (+ pterosphenoid without ventral process or flange from anterior half of ventral margin [4]) traits depending on the optimization (Fig. 4.9; their Fig. 19). The first two

character states appear to be synapomorphies within the osmeroids, although *Osmerus* and members of the Galaxiidae also share the last trait. By contrast, the molecular phylogeny strongly supports a sister-relationship between *Allosmerus* and *Spirinchus* (Figs. 4.5 and 4.6; Appendices 10 and 15), an arrangement not unambiguously supported by any character states (Figs. 4.8 and 4.9). *Allosmerus* and *Spirinchus* do however share two homoplastic traits: lack of cucumber odour (64, Fig. 4.8; shared with *Plecoglossus*) and head length greater than 4.7 in standard length (100, Fig. 4.8; shared with *Osmerus mordax* and *O. dentex*, data missing for *Plecoglossus*). As stated by Johnson and Patterson (1996), the cucumber odour is only detectable in fresh specimens and may have been overlooked in some taxa. The odour may also be one of the more subjective characters assessed for these fishes. Information about relative head length was unavailable for *Plecoglossus*; however, relatively small head sizes are also shared by species of *Osmerus*. Because of the strong support for an *Allosmerus*-*Spirinchus* sister relationship based on molecular analyses, I conclude that the globose otic bulla and fontanelles in the cartilaginous roof of the otic region remaining closed during ontogeny either evolved independently in *Spirinchus* and *Thaleichthys* or subsequently changed in the lineage leading to *Allosmerus*. Further morphological examination of *Allosmerus* and *Spirinchus* may identify synapomorphic traits.

4.4.4 General comments on morphological analyses of the Osmeridae

As indicated by the triangular markings on Fig. 4.8, numerous characters used in previous systematic studies of the Osmeridae are homoplasious. This will

not come as a surprise to anyone who has studied osmeroid fishes, as repeated losses and gains of particular character states have been noted by many previous authors (Johnson and Patterson 1996, and references therein; Waters et al. 2000). Waters et al. (2002) reduced incongruence between molecular and morphological analysis of the southern hemisphere smelts by differentially weighting gains and losses. They also suggested that some studies might have been inadvertently biased by individual systematists who sought characters to define anticipated relationships among particular taxa. On a more positive note, even though homoplasy has complicated the inference of osmerid relationships from morphology, the combined data from several studies showed that 27 characters unambiguously supported the intergeneric relationships inferred from molecular analysis and a further 27 characters were autapomorphic for particular genera (Fig. 4.8).

As an aside, in this study the difficulty of distinguishing among osmerid species was shown by apparent misidentifications in several species. Based on mitochondrial and nuclear sequence data, four individuals collected from Hokkaido, off Ohtanoshike near Kushiro, Japan identified as *Hypomesus japonicus* were genetically *H. nipponensis*, one individual from Puget Sound, WA identified as *Spirinchus thaleichthys* was genetically *S. starksi*, and two individuals from the Bering Sea and Sakhalin, Russia, respectively, identified as *Osmerus mordax*, were genetically *O. dentex*. The latter case, however, most likely resulted from taxonomic confusion within *Osmerus*, as some authors considered the North Pacific and western Atlantic forms to fall under the same name, *O. mordax*. I concluded these samples were likely misidentifications and not a result of incomplete lineage sorting

or hybridization because there is little intraspecific variation for these genes, and they shared haplotypes with their respective congeners for both mitochondrial and nuclear genes.

4.4.5 Taxonomic considerations

As discussed in section 4.1.4, some authors nested both the Plecoglossidae and Salangidae within the Osmeridae. Further, some authors who considered one or both the former 'families' as sister to the Osmeridae (i.e., recognize all taxa as reciprocally monophyletic), have argued for their inclusion in the family Osmeridae (e.g., Fu et al. 2005). This situation highlights one of the much discussed failings of the current system of biological classification, that it is arbitrary above the species level. Avise and Johns (1999), developing a suggestion of Hennig (1966), proposed a temporal banding scheme where higher level taxonomic designations refer to lineages that diverged in a standardized timespan. For example, a genus designation could refer to a monophyletic group of taxa that arose in the Pliocene, while a family name might represent a group of organisms that share a common ancestor in the Eocene (Avise and Johns 1999). Thus, in order to decide upon appropriate taxonomic rankings, this approach requires a dated phylogeny. As divergence time data for the Osmeridae, Plecoglossidae and Salangidae are not currently available, I will refrain from making any strong conclusions on this matter at the present time. Osmeridae taxonomy will be discussed in Chapter 6 in conjunction with the results from Chapter 5.

4.4.6 Conclusions

Molecular phylogenetic analysis of mitochondrial and nuclear gene sequences of all species of the Osmeridae yielded a generally well-resolved phylogeny of the genera. This hypothesis conflicts with all previous morphological hypotheses, likely due to homoplasious traits used in the construction of the earlier phylogenies. Additional sequencing of nuclear genes may help better determine the position of the problematic *Mallotus villosus* and uncertainties in the relationships among the species of the polytypic *Spirinchus*, *Osmerus* and *Hypomesus*. Relationships among the Plecoglossidae, Salangidae and Osmeridae are not yet clear, and neither is their taxonomy. Nuclear gene sequences, particularly from salangid species, will not only aid in resolving their systematic relationships but will also provide additional data for estimating the timeframe of their divergence, which can then be used to assess the appropriate taxonomic designations for these taxa.

Table 4.1 Identification, geographic locality and source of all samples used in phylogenetic analysis of the Osmeridae.

Species	Sample ID	Geographic location	Sample Source	Catalogue # (if applicable)
Allosmerus elongatus	AE4L	BC, Canada		
	AEBMS1	Imperial Eagle Channel, BC, Canada	K. Ilves, E. Taylor	
	AEOR1,2	Oregon, USA	K. Maslenikov	UW 112170, UW 112171
Hypomesus japonicus	HJ5-HJ10	Hokkaido, off Ohtanoshike, near Kushiro, Japan	K. Takata	
H. nipponensis (*H. chishimaensis*)	HC1-HC4	Iturup Island, Kuybyshevskoe Lake, Japan	K. Maslenikov	UW 043710
	HC5-HC7	Kunashir Island, Lake Serebryanoye, Japan	K. Maslenikov	UW 041862
	HC10-HC12	Zelionyi Island, stream of Lake Srednoye, Japan	K. Maslenikov	UW 041869
	HC18-HC20	Sakhalin Island, Shlyuzovka River, Russia	K. Maslenikov	UW 046336
	HN1-HN3; HN5-HN9	Hokkaido, Harutori River, Japan	S. Mori	
	HN11-HN14	Hokkaido, off Ohtanoshike, near Kushiro, Japan	K. Takata	
H.olidus	HO1-HO4	Kamchatka, Yavinskoye Lake, Russia	B. Urbain	UW 043724
	HOAK1-HOAK4	AK, Chignik Lake, USA	E. Lowery, P. Westley, T. Quinn	
H. pretiosus	HP1	AK Sumner Strait, USA	D. Hay, M. Thompson	
	HP3B	BC, Wreck Beach, Vancouver, Canada	P. Tamkee	
	HP4,5,7			
	HPBMS2	Imperial Eagle Channel, BC, Canada	K. Ilves, E. Taylor	
	HPJB1,2	Puget Sound, San Juan Islands, WA, USA	L. Plough	

Table 4.1 con't. Identification, geographic locality and source of all samples used in phylogenetic analysis of the Osmeridae.

H. transpacificus	HT1-HT8	Sacramento River, CA, USA	P. Moyle	
Mallotus villosus	MV1-MV3	BC, Central Coast, Canada	D. Hay, M. Thompson	
	MVTC2	BC, Trevor Channel, Canada	A. Morton	
	MVAT1	Atlantic Ocean, Canada		
Osmerus dentex	ODAK27, 28, 31, 35, 43	AK, USA		
	ODBER1	AK Bering Sea, AK, USA	K. Maslenikov	UW 112173
	ODSAK1	Sakhalin Island, Russia	K. Maslenikov	UW 044763
O. eperlanus	OE1, 14; OERT	Lake Ijsselmeer, Netherlands		
	OELK1	Lake Kolovesi, Finland	J. Jurvelius	
	OELP1, 2	Lake Paasivesi, Finland	J. Jurvelius	
	OEPJ1	Lake Peipsi, Estonia	R. Gross	
O. mordax	OM13, 19	ME, USA	E. Taylor	
	OMLH3, 18	QC, Lac Heney, Canada	E. Taylor	
	OMML8	ON, Meech Lake, Canada	E. Taylor	
	OMRT	ON, Canada	E. Taylor	
Spirinchus lanceolatus	SL1-SL5	Japan		
S. starksi	SSVI1, 2	Vancouver Island	D. Hay	
	SSCA1, 2	CA	P. Moyle	
	SSPUG1 (STPUG1)	WA, Puget Sound, USA	K. Maslenikov	UW 048781
S. thaleichthys	STAN1, 2	BC, Canada	J. Taylor	
	STFR1	BC, Fraser River, Canada	J. Taylor	
	STHL1	BC, Harrison Lake, Canada	J.D. McPhail	
	STLW1	WA, Lake Washington, USA	J.D. McPhail	
Thaleichthys pacificus	TPBR2	AK, Bering Sea, USA		
	TPBMS2	BC, Bamfield, Canada		
	TPCOWA	WA, Cowlitz River, USA		
	TPPUG1, 2	WA, Puget Sound, USA	K. Maslenikov	UW 048013, UW 048014
Plecoglossus altivelis	PLEC3, 4	Honshu, Lake Biwa, Japan	B. Urbain	UW 012704

Table 4.2 Number of sequences analyzed for each species for the six gene regions. * indicates that at least one sequence was obtained GenBank (12S sequences), or form an earlier study (Osmerus cyt*b*; Taylor and Dodson 1994).

	mtDNA genes			nDNA genes			
Species	**Cyt*b*** (286-438)	**16S** (498-550)	**12S** (374-397)	**ITS2** (247-441)	**S71** (297-732)	**RAG1** (725-1455)	
Allosmerus elongatus	4	2	2	4	2	2	
Hypomesus japonicus	6	6	2	6	6	6	
H. nipponensis (H. chishimaensis)	20	25	2	25	24	14	
H. olidus	8	5	3*	4	8	3	
H. pretiosus	5	8	2	5	5	2	
H. transpacificus	8	7	2	7	7	5	
Mallotus villosus	5	4	2	4	3	2	
Osmerus dentex	8*	4	2	5	1	1	
O. eperlanus	7*	3	2	5	4	2	
O. mordax	6*	3	2	5	3	2	
Spirinchus lanceolatus	5	4	2*	4	4	3	
S. starksi	4	4	1	3	4	1	
S. thaleichthys	5	4	3*	3	3	2	
Thaleichthys pacificus	5	4	2*	4	4	2	
Plecoglossus altivelis	2	2	3*	1	2	1	
Retropinna retropinna	2	2	2	2	2	2	
TOTAL	**100**	**87**	**34**	**87**	**82**	**50**	**440**

Table 4.3 Primers used to amplify the six gene regions sequenced in this chapter.

Gene	Primer	Sequence (5'-3')	Source
Cyt*b*	cytb2	CCC TCA GAA TGA TAT TTG TCC TCA	Kocher et al. (1989)
	GluDG	TGA CTT GAA RAA CCA YCG TTG	
16S	16Sar	CGC CTG TTT ATC AAA AAC AT	Waters et al. (2002)
	16Sbr	CCG GTC TGA ACT CAG ATC ACG T	
12S	12SF	AAA AAG CTT CAA ACT GGG ATT AGA TAC CCC ACT	Kocher et al. (1989)
	12SR	TGA CTG CAG AGG GTG ACG GGC GGT GTG T	
ITS2	5.8sr	CTA CGC CTG TCT GAG TGT C	Presa et al. (2002)
	28s	ATA TGC TTA AAT TCA GCG GG	
S71	S7RPEX1F	TGG CCT CTT CCT TGG CCG TC	Chow and Hazama (1998)
	S7RPEX2R	AAC TCG TCT GGC TTT TCG CC	
RAG1	RAG1F	AGC TGT AGT CAG TAY CAC AAR ATG	Quenouille et al. (2004)
	RAG9R	GTG TAG AGC CAG TGR TGY TT	
	RAG1F1	CTG AGC TGC AGT CAG TAC CAT AAG ATG T	López et al. (2004)
	RAG1R1	CTG AGT CCT TGT GAG CTT CCA TRA AYT T	
	RAG1R2	TGA GCC TCC ATG AAC TTC TGA AGR TAY TT	
	RAG1R3	GTC TTG TGS AGG TAG TTG GT	

Table 4.4 GenBank accession numbers for Galaxiidae, Salangidae and Retropinnidae samples used in outgroup determination.

species	**cyt*b***	**16S**	**12S**	**RAG1**
Galaxias fasciatus	AF267350	AF112333	AY430265	AY430218
Salangichthys microdon	AF454838	AY443566	AY430267	AY380539
Retropinna retropinna	n/a	n/a	NC004598	n/a

Table 4.5 Results of Shimodaira-Hasegawa topology tests (Shimodaira and Hasegawa 1999) for the placement of *Mallotus villosus*. + indicates highest likelihood score.

Comparison	-ln likelihood	*P* value
nDNA		
sister to OTAS	7484.8+	**0.16**
sister to all	7487.9	
all DNA		
sister to OTAS	11825.8+	**0.13**
sister to all	11829.4	
all DNA with *Retropinna*		
sister to OTAS	12516.7	**0.61**
sister to all	12516.0+	

Table 4.6 Results of Shimodaira-Hasegawa topology tests (Shimodaira and Hasegawa 1999) for relationships within *Osmerus*. + indicates highest likelihood score.

Comparison	-ln likelihood	*P* value
mtDNA		
Plecoglossus outgroup		
O. dentex - O. eperlanus	4158.5+	**0.18**
O. dentex - O. mordax	4161.9	
O. dentex - O. eperlanus	4158.5+	**0.12**
O. eperlanus - O. mordax	4162.4	
O. dentex - O. mordax	4161.9+	**0.3**
O. eperlanus - O. mordax	4162.4	
Mallotus outgroup		
O. dentex - O. eperlanus	2546.4+	**0.33**
O. dentex - O. mordax	2546.7	
O. dentex - O. eperlanus	2546.4+	**0.35**
O. eperlanus - O. mordax	2546.7	
O. dentex - O. mordax	2546.7	**0.44**
O. eperlanus - O. mordax	2546.7+	
nDNA		
Plecoglossus outgroup		
O. dentex - O. eperlanus	7485.2	**0.39**
O. dentex - O. mordax	7484.7+	
O. dentex - O. eperlanus	7485.2	**0.39**
O. eperlanus - O. mordax	7484.9+	
O. dentex - O. mordax	7484.8+	**0.38**
O. eperlanus - O. mordax	7484.9	
Mallotus outgroup		
O. dentex - O. eperlanus	4279.6	**0.28**
O. dentex - O. mordax	4278.9+	
O. dentex - O. eperlanus	4279.6	**0.49**
O. eperlanus - O. mordax	4279.6	
O. dentex - O. mordax	4278.9+	**0.28**
O. eperlanus - O. mordax	4279.6	
all DNA		
Plecoglossus outgroup		
O. dentex - O. eperlanus	11825.8+	**0.17**
O. dentex - O. mordax	11830.5	
O. dentex - O. eperlanus	11825.8+	**0.11**
O. eperlanus - O. mordax	11831.4	
O. dentex - O. mordax	11830.5+	**0.33**
O. eperlanus - O. mordax	11831.4	
Mallotus outgroup		
O. dentex - O. eperlanus	6870.9+	**0.36**
O. dentex - O. mordax	6871.1	
O. dentex - O. eperlanus	6870.9+	**0.36**
O. eperlanus - O. mordax	6871.1	
O. dentex - O. mordax	6871.1	**0.41**
O. eperlanus - O. mordax	6871.1	

Table 4.7 Results of Shimodaira-Hasegawa topology tests (Shimodaira and Hasegawa 1999) for relationships within *Spirinchus*. + indicates highest likelihood score, * indicates significant difference ($p<0.05$).

Comparison	-ln likelihood	*P* value
mtDNA		
Plecoglossus outgroup		
S. lanceolatus - *S. thaleichthys*	4160.6+	**0.03***
S. thaleichthys - *S. starksi*	4172.5	
Allosmerus outgroup		
S. lanceolatus - *S. thaleichthys*	2420.5+	**0.29**
S. thaleichthys - *S. starksi*	2421.1	
nDNA		
Plecoglossus outgroup		
S. lanceolatus - *S. thaleichthys*	7491.3	**0.17**
S. thaleichthys - *S. starksi*	7484.8+	
Allosmerus outgroup		
S. lanceolatus - *S. thaleichthys*	3900.6+	**0.40**
S. thaleichthys - *S. starksi*	3901.4	
all DNA		
Plecoglossus outgroup		
S. lanceolatus - *S. thaleichthys*	11828.5	**0.32**
S. thaleichthys - *S. starksi*	11825.8+	
Allosmerus outgroup		
S. lanceolatus - *S. thaleichthys*	6358.4+	**0.49**
S. thaleichthys - *S. starksi*	6358.5	

Table 4.8 Results of Shimodaira-Hasegawa topology tests (Shimodaira and Hasegawa 1999) for relationships within *Hypomesus*. + indicates highest likelihood score.

Comparison	-ln likelihood	*P* value
mtDNA		
H. nipponensis - H. olidus	4172.9+	**0.18**
H. olidus - H. japonicus(H. pretiosus, H. transpacificus)	4175.7	
nDNA		
H. nipponensis - H. olidus	7492.9	**0.12**
H. olidus - H. japonicus(H. pretiosus, H. transpacificus)	7485.5+	
all DNA		
H. nipponensis - H. olidus	11828.3	**0.32**
H. olidus - H. japonicus(H. pretiosus, H. transpacificus)	11825.8+	

Table 4.9 Results of Shimodaira-Hasegawa tests (Shimodaira and Hasegawa 1999) comparing the molecular topology (Fig. 4.12) with previous morphology-based hypotheses (Fig. 4.1). + indicates highest likelihood score, * indicates significant difference ($p<0.05$).

Comparison	-ln likelihood	*P* value
without *Plecoglossus*		
all DNA phylogeny	10482.1+	
Chapman (1941)	10518.5	0.005*
McAllister (1963)	10519.1	0.004*
McAllister (1966)	10502.5	0.024*
Klyukanov (1977)	10518.5	0.005*
with *Plecoglossus*		
all DNA phylogeny	11825.8+	
Wilson and Williams (1991)	11859.9	0.027*
Begle (1991)	11896.0	0.0004*
with Salangidae		
cyt*b* , 16S, 12S, RAG1 phylogeny	8681.5+	
Howes and Sanford (1987)	8709.8	0.015*
Johnson and Patterson (1996)	8699.1	0.045*

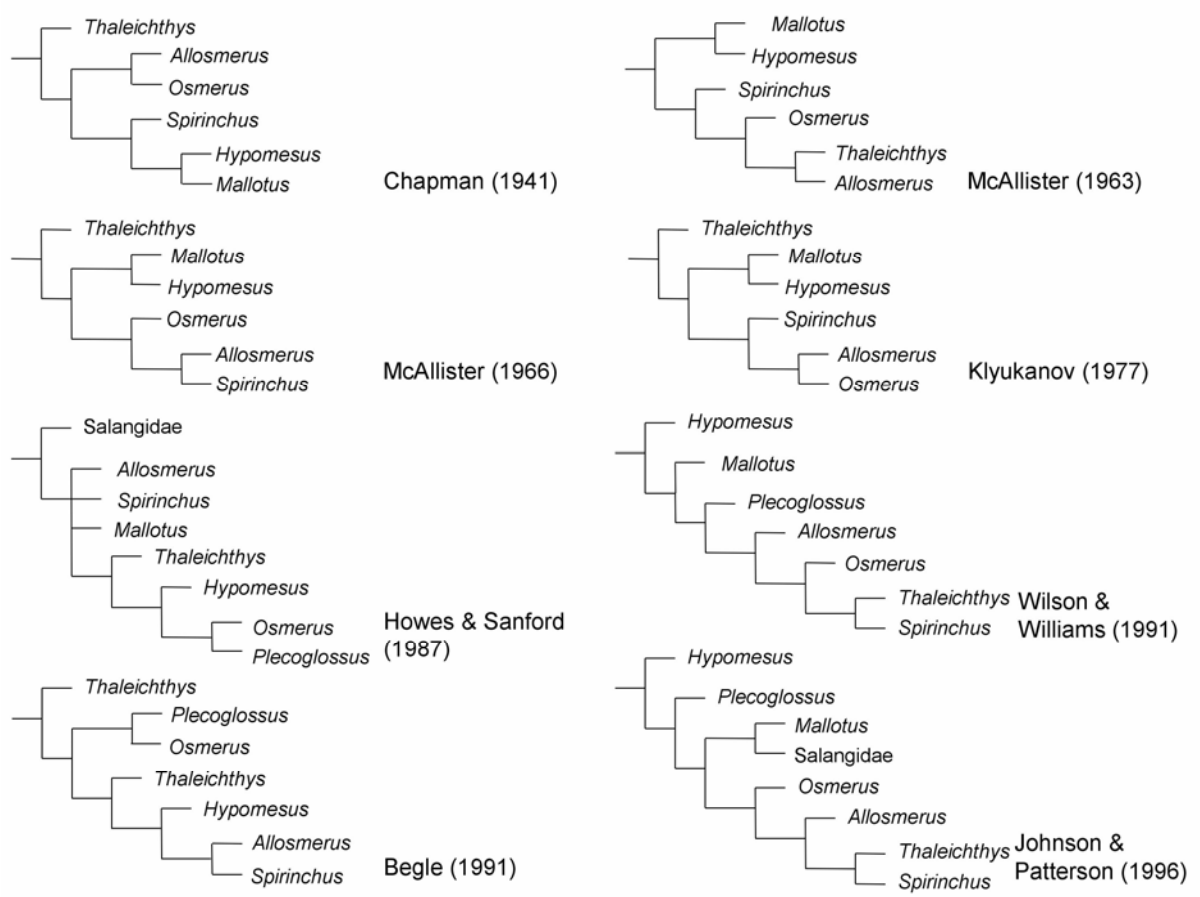


Figure 4.1 Eight morphology-based hypotheses of systematic relationships among Osmeridae genera. A. Chapman (1941) B. McAllister (1963) C. McAllister (1966) D. Klyukanov (1977) E. Howes and Sanford (1987) F. Wilson and Williams (1991) G. Begle (1991) H. Johnson and Patterson (1996). Figure adapted from Johnson and Patterson (1996).

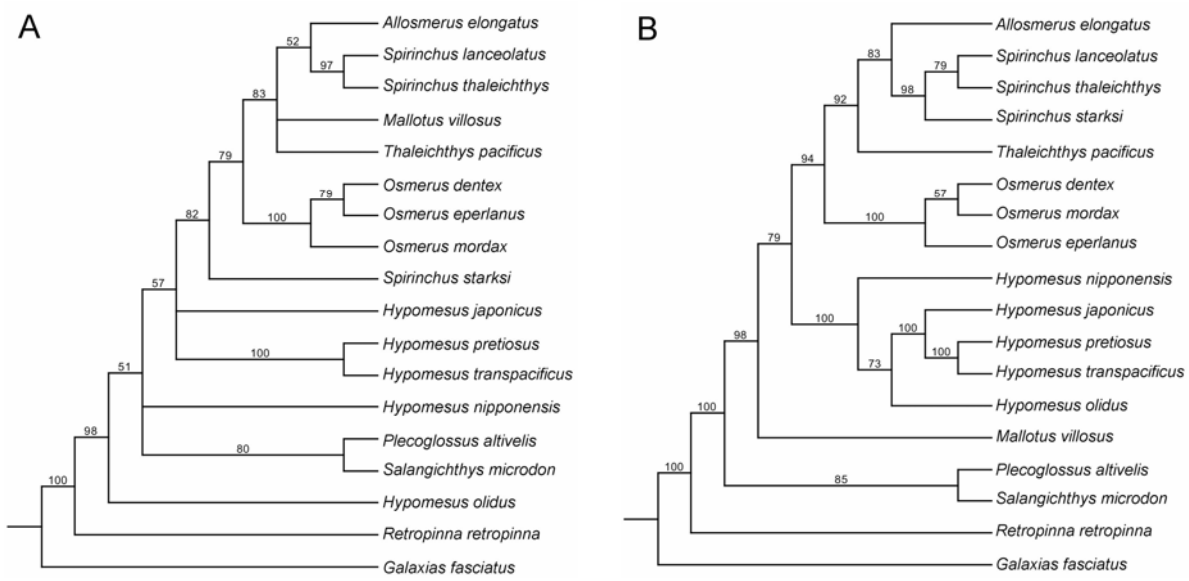


Figure 4.2 Relationships between Galaxiidae, Retropinnidae, Salangidae, Plecoglossidae and Osmeridae based on Bayesian reconstruction of (A) mtDNA (cytb, 16S, 12S) sequences, and (B) nDNA (RAG1) sequences. Numbers above nodes represent Bayesian posterior probability values from a majority-rule consensus of 18000 trees.

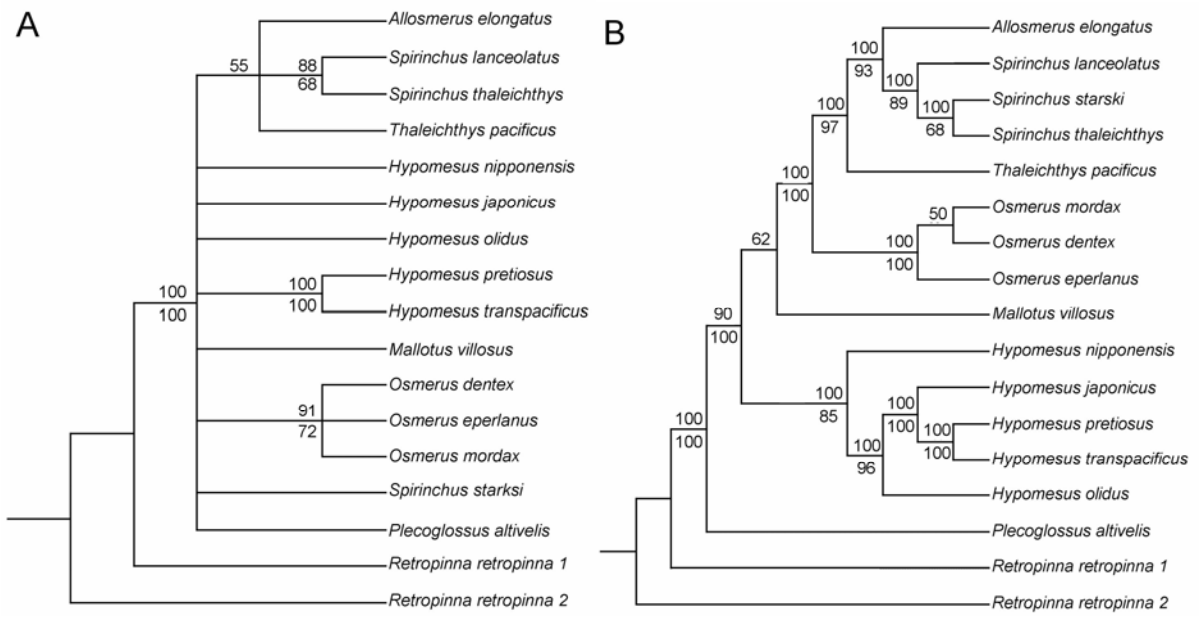


Figure 4.3 Relationships between Plecoglossidae and Osmeridae with outgroup *Retropinna retropinna* based on Bayesian and parsimony reconstruction of (A) mtDNA (cytb, 16S,12S) sequences, and (B) nDNA (ITS2, S71, RAG1) sequences. Numbers above nodes represent Bayesian posterior probability values from a majority-rule consensus of 18000 trees. Numbers below nodes represent support values based on 1000 bootstrap pseudo-replicates. Parsimony analysis of mtDNA data places *P. altivelis* outside the Osmeridae with 92% bootstrap support and with nDNA data places *M. villosus* as sister to all other Osmeridae taxa with 82% bootstrap support (data not shown).

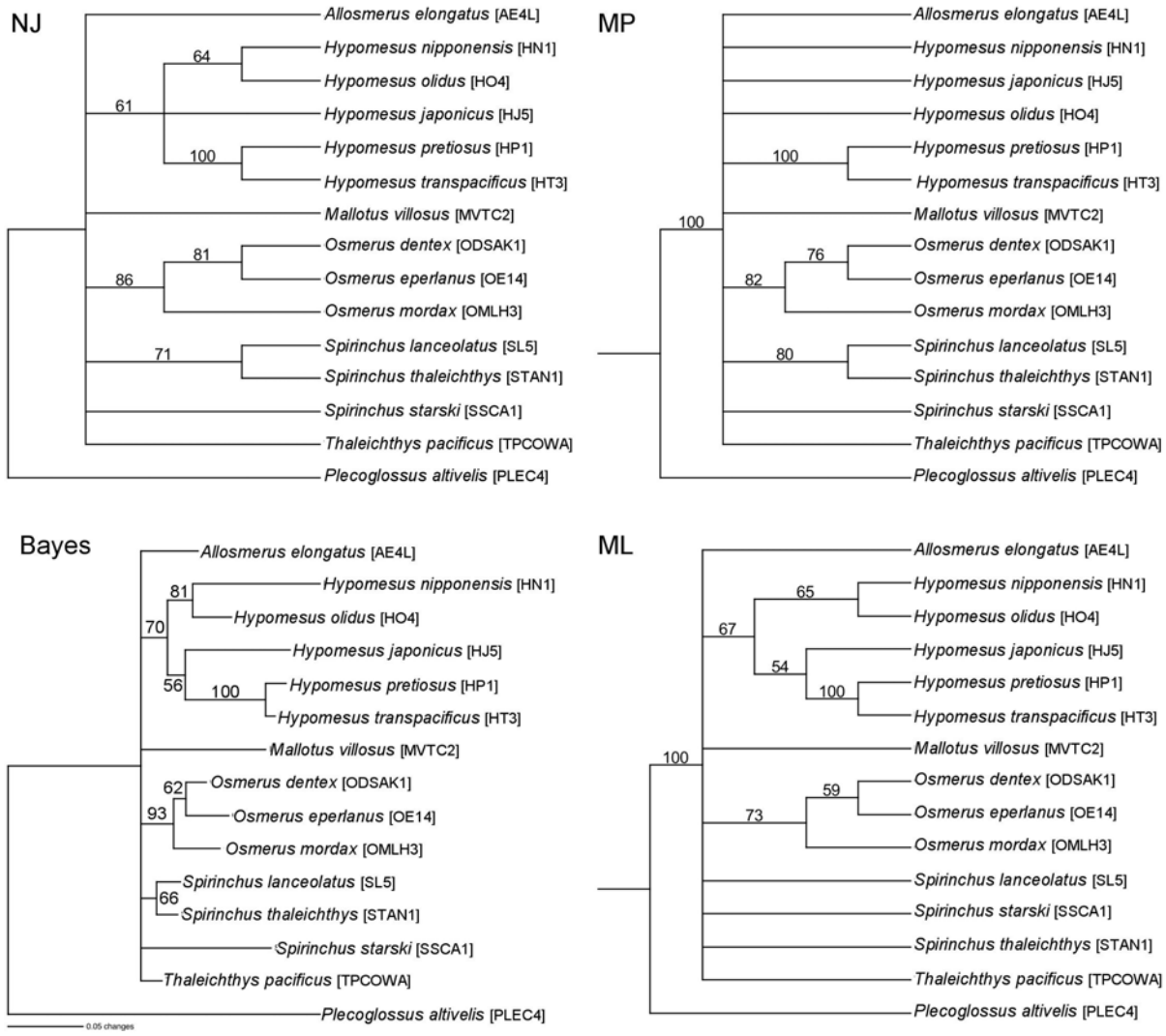


Figure 4.4 Combined mtDNA (cyt*b*, 16S, 12S) Osmeridae phylogenies based on a single individual/species resulting from neighbour-joining (NJ), parsimony (MP), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ, MP, ML) or posterior probabilities from a consensus of 18000 trees (Bayes).

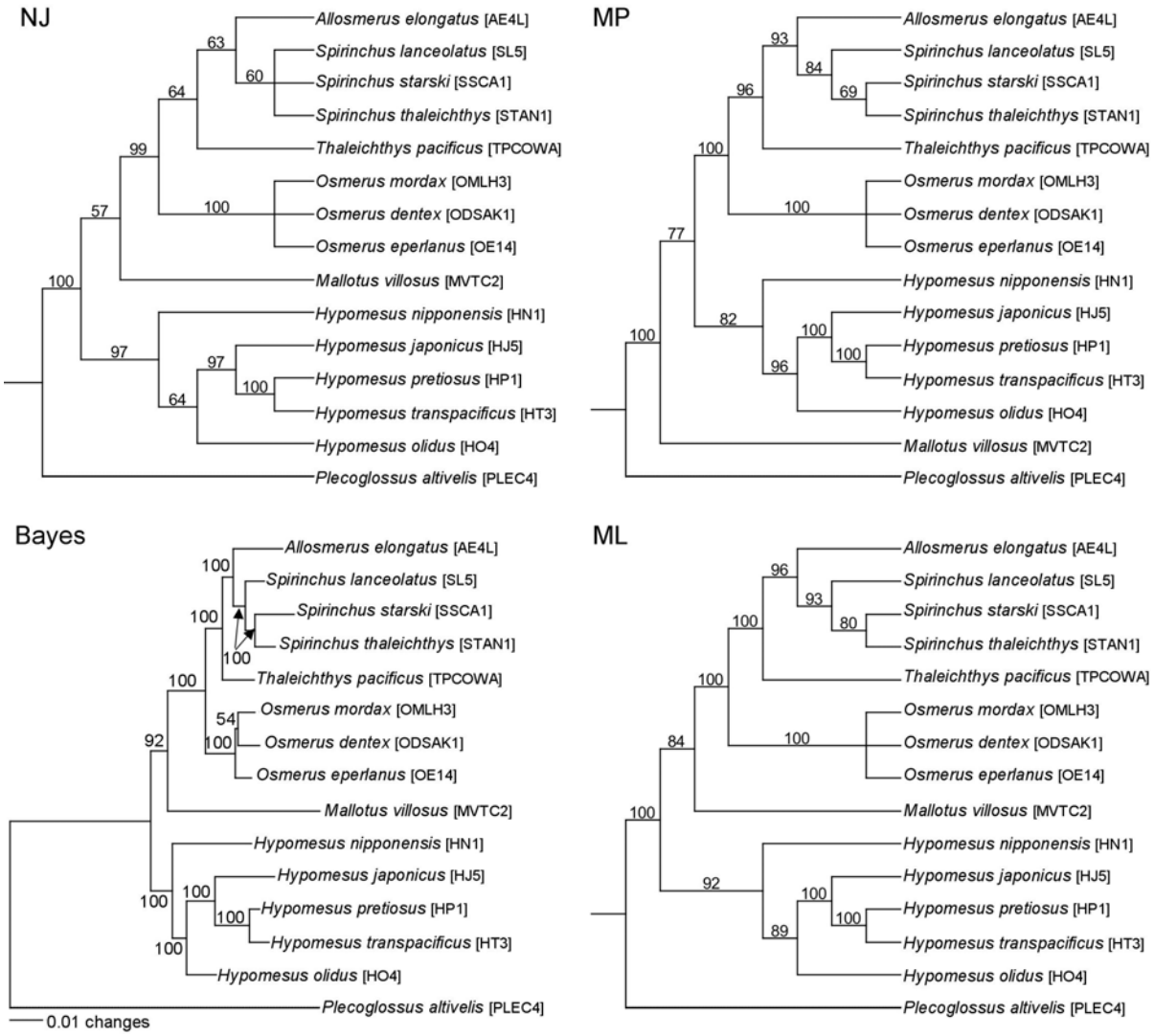


Figure 4.5 Combined nDNA (ITS2, S71, RAG1) Osmeridae phylogenies based on a single individual/species resulting from neighbour-joining (NJ), parsimony (MP), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ, MP, ML) or posterior probabilities from a consensus of 18000 trees (Bayes).

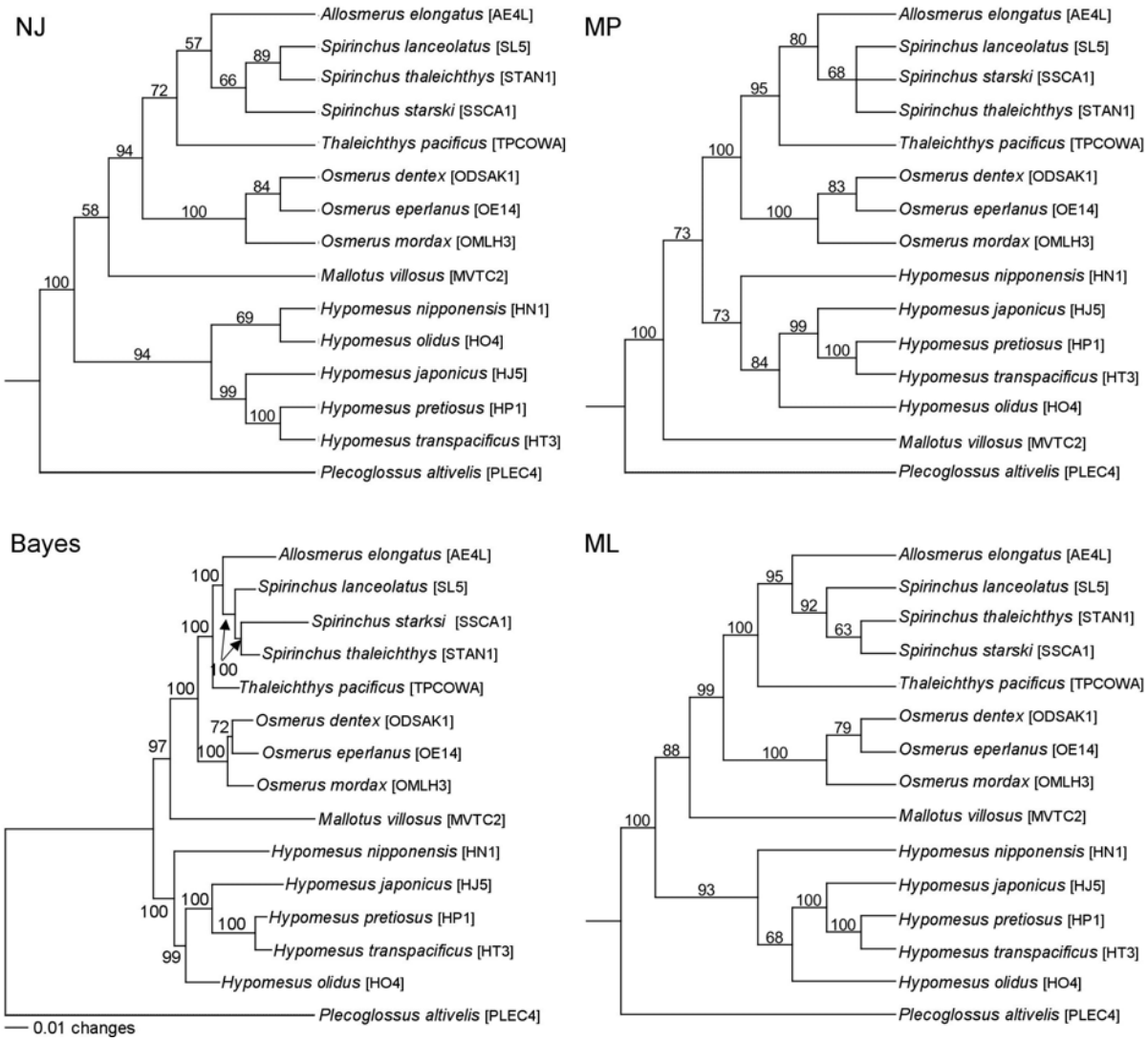


Figure 4.6 Combined mtDNA (cyt*b*, 16S, 12S) and nDNA (ITS2, S71, RAG1) Osmeridae phylogenies resulting from neighbour-joining (NJ), parsimony (MP), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ, MP, ML) or posterior probabilities from a consensus of 18000 trees (Bayes).

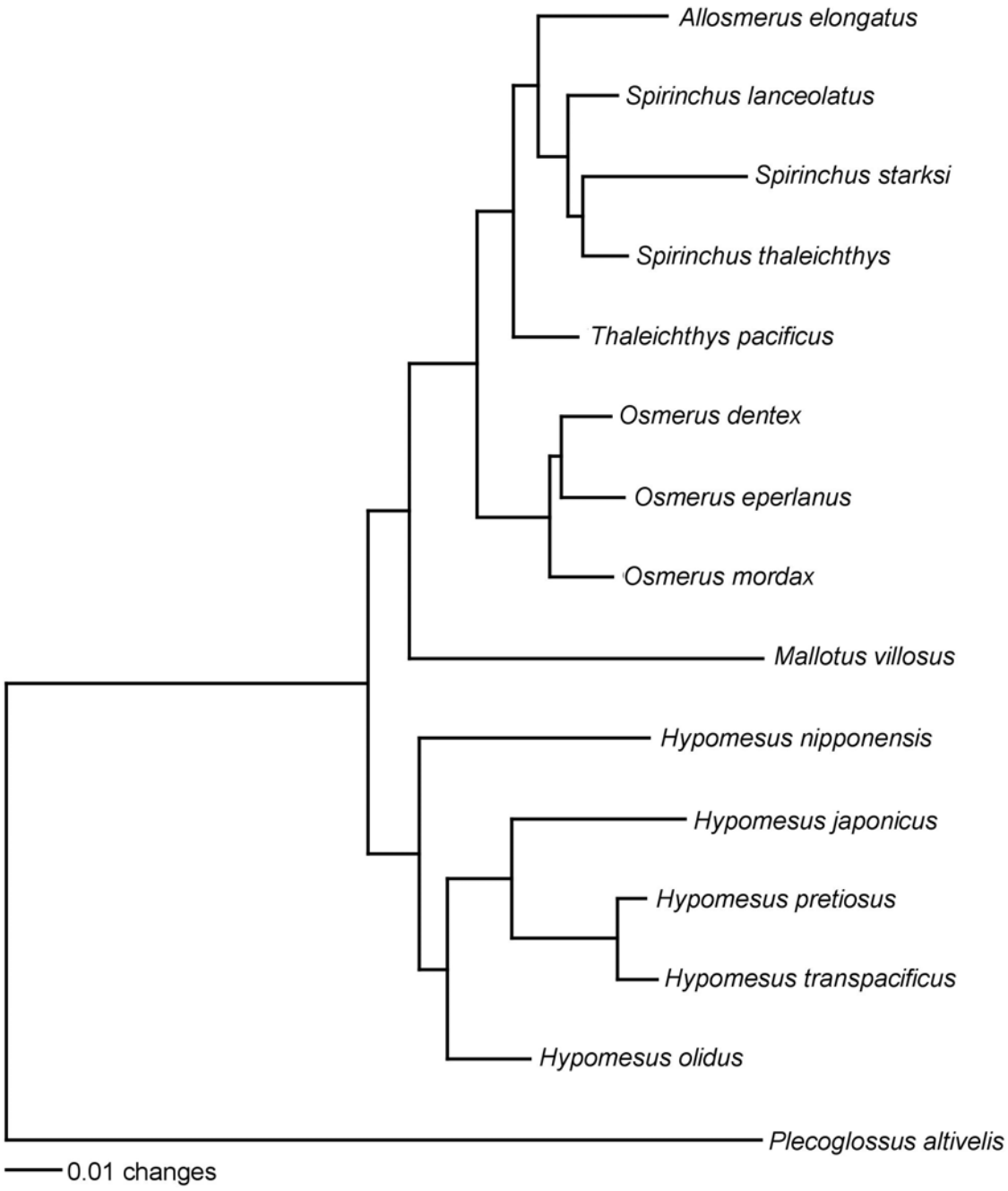


Figure 4.7 Osmeridae summary phylogeny based on Bayesian and ML analysis of cyt*b*, 16S, 12S, ITS2, S71, and RAG1 sequences. The nDNA topology is identical aside from the relationships within *Osmerus*. Branch lengths from Bayesian estimation.

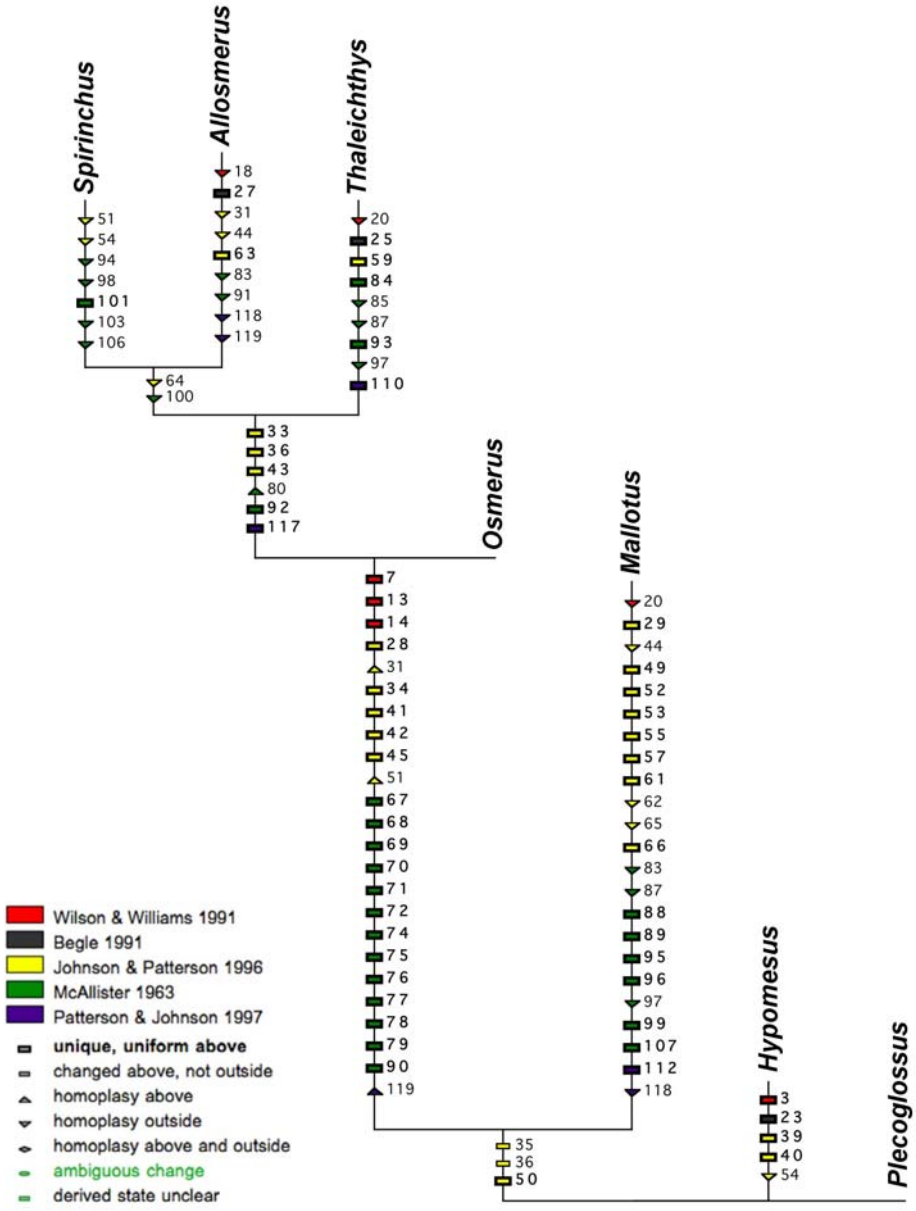


Figure 4.8 Molecular phylogeny of Osmeridae genera with characters from earlier morphological examinations mapped. Branch lengths are proportional to unambiguous changes. Numbers refer to characters listed in Appendices 16 and 17. Characters treated are unordered.

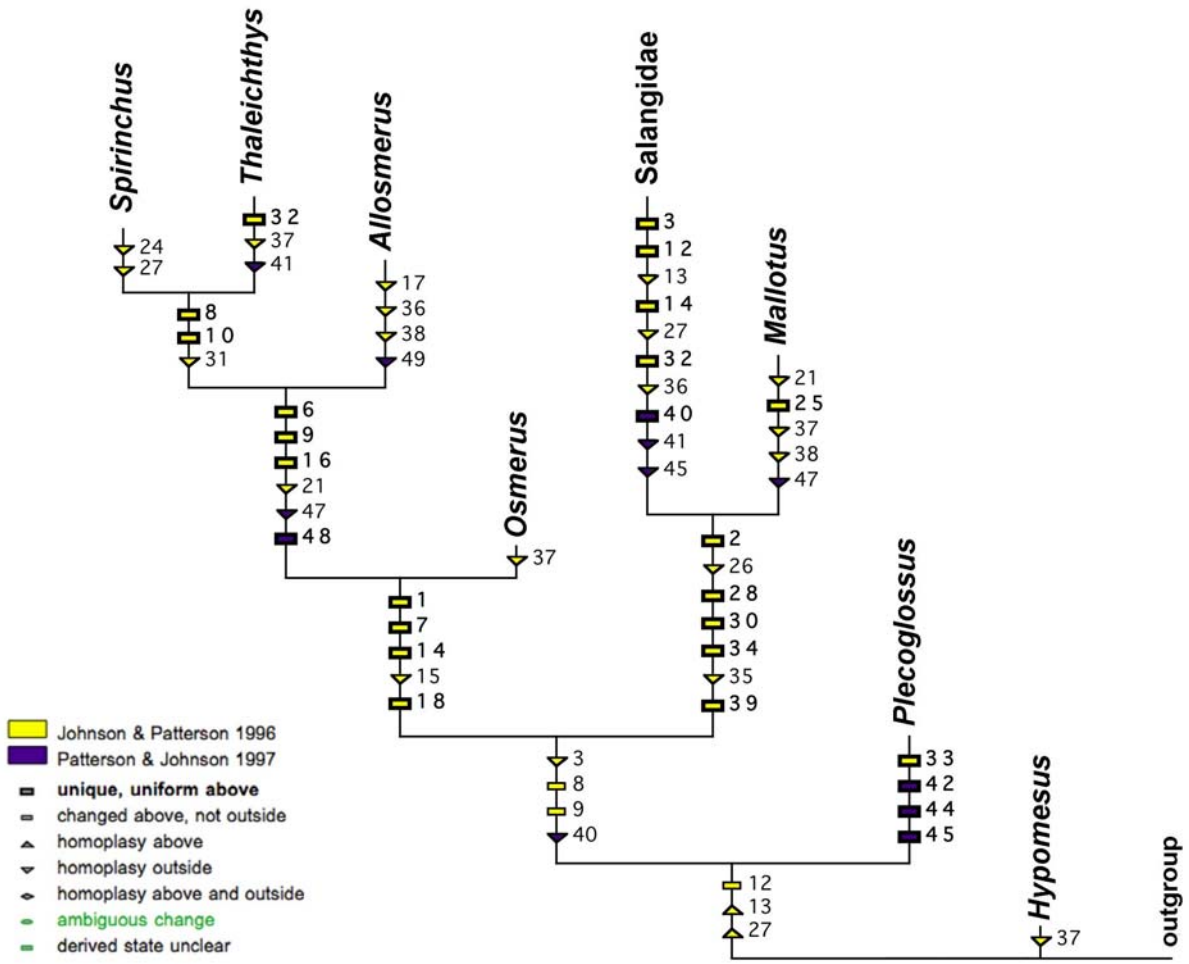


Figure 4.9 Hypothesis of Osmeridae, Salangidae and Plecoglossidae interrelationships from Johnson and Patterson (1996), with morphological characters from Johnson and Patterson (1996) and Patterson and Johnson (1997) mapped. Branch lengths are proportional to unambiguous changes. Numbers refer to characters listed in Appendices 16 and 18. Outgroup from Johnson and Patterson (1996) with all states coded as 0. Characters treated are unordered.

4.5 REFERENCES

Avise, J. C., and G. C. Johns. 1999. Proposal for a standardized temporal scheme of biological classification for extant species. Proc. Natl. Acad. Sci. 96:7358-7363.

Begle, D. P. 1991. Relationships of osmeroid fishes and the use of reductive characters in phylogenetic analysis. Syst. Zool. 40:33-53.

Chapman, W. M. 1941. The osteology and relationships of the osmerid fishes. J. Morph. 69:279-301.

Chow, S., and K. Hazama. 1998. Universal PCR primers for S7 ribosomal protein introns in fish. Mol. Ecol. 7:1247-1263.

Cuvier, G. 1817. Le regne animal. Deterville, Paris.

Eschmeyer, W. N. 2006. Catalogue of fishes, online version. California Academy of Sciences, San Francisco. Available from http//www.calacademy.org/research/ichthyology/catalog

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1994. Testing significance of congruence. Cladistics 10:315-319.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

———. 2004. Inferring phylogenies. Sinauer Associates, Inc., Sunderland, MA.

Fu, C., J. Luo, J. Wu, J. A. López, Y. Zhong, G. Lei, and J. Chen. 2005. Phylogenetic relationships of salangid fishes (Osmeridae, Salanginae) with comments on phylogenetic placement of the salangids based on mitochondrial DNA sequences. Mol. Phylogenet. Evol. 35:76-84.

Greenwood, P. H., D. E. Rosen, S. H. Weitzman, and G. S. Myers. 1966. Phyletic studies of teleostean fishes, with a provisional classification of living forms. Bull. Am. Mus. Nat. Hist. 131: 339-456.

Haldorson, L., and P. Craig. 1984. Life history and ecology of a Pacific-Arctic population of rainbow smelt in coastal waters of the Beaufort Sea. *Trans. Am. Fish. Soc*. 113:33-38.

Hart, J. L., and W. A. Clemens. 1973. Pacific fishes of Canada. Fish. Res. Board Canada, Ottawa.

Helfman, G. S., B. B. Collette, and D. E. Facey. 1997. The diversity of fishes. Blackwell Science, Toronto.

Hennig, W. 1966. Phylogenetic systematics. Translated by D. D. Davis and R. Zangerl. University of Illinois Press, Urbana.

Howes, G. J., and C. P. J. Sanford. 1987. The phylogenetic position of the Plecoglossidae (Teleostei, Salmoniformes), with comments on the Osmeridae and Osmeriodei. Proc. Congr. Eur. Ichthyol., 5th, Stockholm, pp. 17-30.

Hubbs, C. L. 1925. A revision of the osmerid fishes of the North Pacific. Proc. Biol. Soc. Wash. 38:49-56.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754-755.

Ilves, K. L., and E. B. Taylor. 2007a. Are *Hypomesus chishimaensis* and *H. nipponensis* (Pisces: Osmeridae) distinct species? A molecular assessment using comparative sequence data from five genes. *Copeia* 2007:180-185.

———. 2007b. Evolutionary and biogeographic patterns within the smelt genus *Hypomesus* (Pisces: Osmeridae) in the North Pacific Ocean. *J. Biogeogr.*, in press.

Johnson, G. D., and C. Patterson. 1996. Relationships of lower Euteleostean fishes, Pp. 251-332 in M. Stiassny, L. J. Parenti, L. R. Johnson, and G. David, eds. Interrelationships of fishes. Academic Press, Toronto.

Kendall, W. C. 1927. The smelts. U.S. Bur. Fish. Bull. 42: 217-375.

Klyukanov, V. A. 1969. Morphological bases of classification of smelts of the genus *Osmerus* (Osmeridae). Zool. J. 48 (1): 99-109.

Klyukanov, V. A. 1970. Morphological basis of the classification of smelts of the genus *Hypomesus*. Zool. J. 49: 1534-1542.

Klyukanov, V. A. 1975. The systematic position of the Osmeridae in the order Salmoniformes. J. Ichthyol. 15: 1-17.

Kocher, T. D. W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86: 6196-6200.

Linnaeus, C. 1758. *Systema naturae*. 10th ed. Stockholm.

López, J. A., W. Chen, and G. Ortí. 2004. Esociform phylogeny. Copeia 2004:449-464.

Luey, J. E., C. C. Krueger and D. R. Schreiner. 1982. Genetic relationships among smelt, genus *Osmerus*. Copeia 1982:725-728.

Maddison, D. R., and W. P. Maddison. 2003. MacClade 4.06: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA.

Maddison, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523-536.

Matsuoka, M., and T. Iwai. 1983. Adipose fin cartilage found in some teleostean fishes. Jpn. J. Ichthyol. 30:37-46.

McAllister, D. E. 1963. A revision of the smelt family, Osmeridae. Bull. Nat. Mus. Can. 191:1-53.

———. 1966. Numerical taxonomy and the smelt family, Osmeridae. Can. Field. Nat. 80:227-238.

McAllister, D. E., B. Parker, and K. E. Couvillion. 1980. *Osmerus mordax* (Mitchill), rainbow smelt. Pp.124- in D. S. Lee, C. R. Gilbert, C. H. Hocutt, R. E. Jenkins, D. E. McAllister, and J. R. Stauffer, eds. Atlas of North American freshwater fishes. North Carolina State Museum of Natural History, Raleigh, NC.

Nellbring, S. 1989. The ecology of smelts (genus *Osmerus*): a literature review. Nordic J. Freshw. Res. 65:116-145.

Nelson, J. S. 1984. Fishes of the world. 2nd ed. John Wiley and Sons, Toronto.

Parenti, L. R. 1986. The phylogenetic significance of bone types in euteleost fishes. Zool. J. Linn. Soc. 87: 37-51

Patterson, C., and G. D. Johnson. 1997. The data, the matrix, and the message: comments on Begle's "Relationships of osmeroid fishes". Syst. Biol. 46:358-365.

Posada, D., and T. R. Buckley. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53:793-808.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

Presa, P. B. G. Pardo, P. Martinez, and L. Bernatchez. 2002. Phylogeographic congruence between mtDNA and rDNA ITS markers in Brown Trout. Mol. Biol. Evol. 19:2161-2175.

Quenouille, B. E. Bermingham, and S. Planes. 2004. Molecular systematics of the damselfishes (Teleostei: Pomacentridae): Bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. Mol. Phyl. Evol. 31:66-88.

Rambaut, A. 1996. Se-Al: a manual sequence alignment editor, v.2.0a11. Available from http://tree.bio.ed.ac.uk/software/seal/.

Roberts, T. R. 1984. Skeletal anatomy and classification of the neotenic asian salmoniform superfamily Salangoidea (icefishes or noodlefishes). Proc. Cal. Acad. Sci. **43**(13): 179-220.

Saruwatari, T. J. A. López, and T. W. Pietsch. 1997. A revision of the osmerid genus *Hypomesus* Gill (Teleostei: Salmoniformes), with the description of a new species from the Southern Kuril Islands. Species Diversity 2:59-82.

Sazanov, Y. I. 1986. Morphology and classification of the fishes of the family Platytroctidae (Salmoniformes, Alepocephaloidei). Tr. P. P. Shirshov Inst. Oceanol. 121a:51-96. [ın Russian].

Scott, W. B., and E.J. Crossman. 1973. *Freshwater fishes of Canada*. Fish. Res. Board. Canada Bull. 184.

Shimodaira, H. and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114-1116.

Sidorov, L. K., and Y. Pichugin. 2004. Morphological traits of lacustrine forms of smelts of the genus *Hypomesus* (Salmoniformes) from the Southern Kurils. J. Ichthyol. 44:433-443.

Swofford, D. L. 2002. PAUP*:phylogenetic analysis using parsimony. Ver. 4.0b10. Sinauer Associates, Sunderland, MA.

Taylor, E.B. and J.J. Dodson. 1994. A molecular analysis of relationships and biogeography within a species complex of Holarctic fish (genus *Osmerus*). *Molec. Ecol*. **3**: 235-248.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX-Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.

Waters, J. M., J. A. López, and G. P. Wallis. 2000. Molecular phylogenetics and biogeography of galaxiid fishes (Osteichthyes: Galaxiidae): dispersal, vicariance, and the position of *Lepidogalaxias salamandroides*. Syst. Biol. 49:777-795.

Waters, J. M., T. Saruwatari, T. Kobayashi, I. Oohara, R. M. McDowall, and G. P. Wallis. 2002. Phylogenetic placement of retropinnid fishes: data set incongruence can be reduced by using asymmetric character state transformation costs. Syst. Biol. 51:432-449.

Wilson, M.V. H., and R. R. G. Williams. 1991. New Paleocene genus and species of smelt (Teleostei: Osmeridae) from freshwater deposits of the Paskapoo Formation, Alberta, Canada, and comments on osmerid phylogeny. J. Vertebr. Paleontol. 11:434-451.

CHAPTER 5: HOLARCTIC BIOGEOGRAPHY OF THE OSMERIDAE THROUGH THE CENOZOIC[4]

5.1 INTRODUCTION

5.1.1 Holarctic biogeography

Biogeographers have long searched for common processes responsible for driving diversification in the Holarctic region. Terrestrial (non-marine) flora and fauna have been particularly well-studied, with a vast literature on patterns of diversity for taxa ranging from plants (e.g. Wolfe 1975; Schultheis and Donoghue 2004; Lang et al. 2007), invertebrates (e.g. Hainey and Taylor 2003; Audzijonyte and Väinölä 2006; Hendrixson and Bond 2007), land vertebrates (e.g. Bowen et al. 2002; Smith et al. 2005; Burbrink and Lawson 2007), and freshwater fishes (e.g. Van Houdt et al. 2003, 2005; Stamford and Taylor 2004).

Marine organisms have also been extensively studied (e.g. Vermeij 1991; Ortí et al.1994; Brunner et al. 2001; Kontula and Väinölä 2003); however, much of this work addresses patterns and processes occurring over a relatively recent timescale. The disparity in diversity between the North Pacific and North Atlantic Oceans has received considerable interest. Biased dispersal from the North Pacific Ocean to the North Atlantic Ocean following the opening of the Bering Seaway in the late Miocene (Marincovich and Gladenkov 1999) has been documented for numerous taxa (e.g. Vermeij 1991; Rawson and Hilbish 1995; Collin 2003; Väinölä 2003), perhaps as a

[4] A version of this chapter will be submitted for publication. K. L Ilves, and E. B. Taylor. Holarctic biogeography of the Osmeridae through the Cenozoic.

consequence of increased ecological opportunity for North Pacific invaders owing to higher extinction rates in the North Atlantic basin (Briggs 1970; Vermeij 1989, 1991; Palumbi and Kessing 1991). By contrast, relatively little work has focused on marine Holarctic biogeographic patterns predating the trans-Arctic interchange (but see Grant and Leslie 2001, Arnason et al. 2006).

5.1.2 Osmeridae Holarctic biogeography

The six genera of the Osmeridae are found in cool temperate and Arctic coastal marine and freshwaters of the North Pacific, North Atlantic and Arctic Ocean basins. With fossils related to the closely related ayu, *Plecoglossus altivelis*, dating to the Paleocene (Wilson and Williams 1991), and osmerid fossils from the Oligocene (Gaudant 1985, cited in Wilson and Williams 1991) and Miocene (Uyeno and Sakamoto 1999), the Osmeridae lineage is considerably older than the opening of the Bering Seaway in the Miocene.

Not surprisingly, debates surrounding the phylogenetic relationships among osmerid taxa have also led to competing biogeographic hypotheses. To explain several apparent trans-Pacific sister relationships in *Hypomesus* and *Spirinchus*, McAllister (1963) proposed a 'compression' hypothesis whereby a widespread ancestor throughout the North Pacific underwent range compression during periods of climatic cooling, losing the northernmost portion of the range, thereby isolating populations on each side of the North Pacific basin. To account for the two trans-Pacific disjunctions he inferred for *Hypomesus*, McAllister (1963) postulated that this process must have occurred on at least two occasions. Molecular phylogenetic

analysis of the Osmeridae did not support the *Hypomesus* trans-Pacific relationships hypothesized by McAllister (1963) [Chapters 3 and 4], but did provide evidence for one trans-Pacific disjunction in *Hypomesus* and one in *Spirinchus*. If global climatic cooling had initiated these divergences, then they should coincide with known periods of cooling.

Of the Osmeridae, the Holarctic genus *Osmerus* has received the most biogeographic attention. *Osmerus* has a complicated taxonomic history (Chapter 4) but is here considered to contain three species: *O. dentex* (North Pacific and Arctic), *O. eperlanus* (eastern North Atlantic), and *O. mordax* (western North Atlantic). McAllister (1963) hypothesized a North Pacific origin of the genus, followed by northward and westward dispersal through the Arctic to Europe, leading to the divergence of the European smelt in the eastern Atlantic. The two North American forms were thought to have diverged more recently from an ancestor that dispersed from the eastern Pacific to the western Atlantic during the Pleistocene (McAllister 1963). Although Klyukanov (1975, cited in Taylor and Dodson 1994) also argued for a North Pacific centre of origin for *Osmerus*, according to his model, the first dispersal event was eastwards across the Canadian Arctic, with a subsequent divergence across the North Atlantic. A more recent mitochondrial DNA based phylogenetic analysis aimed to test the phylogenetic predictions of these hypotheses. Taylor and Dodson (1994) showed a sister relationship between the two North American species, which was dated to the late Pliocene and conformed to the hypothesis put forth by McAllister (1963). Further phylogenetic analysis of mitochondrial and nuclear sequences did not resolve the relationships among

Osmerus species (Chapter 4); however, through estimation of divergence times from DNA data, the hypothesized role of the Bering Seaway in the evolution of *Osmerus* can be tested.

5.1.3 Aims

The availability of a (mostly) well-resolved molecular phylogeny of the Osmeridae (Fig. 4.7; Chapter 4) allows further exploration of osmerid biogeography through the Cenozoic. This chapter aims to (1) estimate divergence times from nuclear DNA sequence data, (2) reconstruct the biogeographic history of the family using a dispersalist method based on a parsimony step-matrix and an ML modeling approach that includes both dispersal and vicariance interpretations, (3) compare these results to previous hypotheses of osmerid biogeography, and (4) discuss the results in the context of Holarctic marine biogeography, with a particular focus on the type of comparative data needed for a greater understanding of pre-Bering Seaway diversification in this region.

5.2 MATERIALS AND METHODS

5.2.1 Estimating dates of divergence

To estimate the date of divergence for each node in the osmerid phylogeny, I used a maximum likelihood (ML) approach implemented in r8s v.1.71 (Sanderson 2003) and a Bayesian approach using Beast v.1.4.4 (Drummond and Rambaut

2006). Both methods allowed the inclusion of calibration points, although the Bayesian method offered more flexibility in the distribution of the calibration point, for instance, a uniform, normal, and exponential distribution, among others could be implemented, whereas the ML approach only permitted fixed dates or hard minima and maxima. Both methods allowed for a relaxed molecular clock, which permitted the rates of evolution to vary across the tree, a clear benefit in cases where a phylogeny did not conform to a molecular-clock model. While the Bayesian approach used sequence alignments and could include data partitions, where, for example, different genes could be assigned different models of evolution, the ML method did not use sequence data in the calculation of divergence dates as these data were reduced to the information contained in the inferred branch lengths. While this leads to a potential loss in information, the ML method relied on fewer assumptions and fewer parameters were estimated in the calculation of divergence dates. There were two additional benefits of the ML approach. First was the ability to cross-validate different values of the 'smoothing' parameter, which represented the extent to which the rate of evolution varies across branches, to determine which value was most appropriate for the phylogeny under investigation. Second, it was possible to assess the validity of each of the assigned calibration points using the fossil-based model cross-validation procedure, which involved removing each calibration point in turn and recalculating the date at the calibrated node to check whether this estimate was consistent with the original constraint. This method could thus be used to evaluate whether the fossils used for calibration may have been incorrectly assigned. As each of these methods had both benefits and drawbacks, I

compared the results of each when estimating divergence dates within the Osmeridae. For both estimation techniques, only the nDNA data from Chapter 4 were included because there was little phylogenetic signal from the mtDNA dataset. Divergence date estimation and biogeographic inference using the allDNA dataset yielded similar results (i.e., divergence dates at all nodes always fell within 3 million years of the nDNA estimate) and did not affect the biogeographic interpretation (data not shown).

ML estimation using r8s

Branch lengths for the topology in Fig. 5.2 (the ML and Bayesian phylogenetic reconstruction from Chapter 4) were estimated from the nDNA dataset (ITS2, S71, RAG1) from a heuristic search with 10 random additions of taxa under the GTR+I+G model chosen by the AIC method in Modeltest (Posada and Crandall 1998), implemented in PAUP* v.4.0b10 (Swofford 2002). Three fossil osmerids from the literature were used to constrain nodes on the tree. The oldest osmerid fossil identified, *Speirsaenigma lindoei*, dated to the Paleocene and was thought to be most closely related to *Plecoglossus altivelis* (Wilson and Williams 1991); therefore, the root node of the Osmeridae + Plecoglossidae phylogeny was initially constrained to a minimum age of 55 million years ago (mya) and a maximum age of 65 mya. Early Oligocene fossils, thought to be basal osmerids related to either *Hypomesus* or *Mallotus* (Gaudant 1985, cited in Wilson and Williams 1991), were used to place a minimum age of 30 mya on the most recent common ancestor of the Osmeridae proper. A final fossil attributed to *Spirinchus* (*S. akagii*) from the mid-Miocene [12-16

mya], most closely resembling *S. lanceolatus* (Uyeno and Sakamoto 1999), was used to place a minimum age bound of 12 mya on the most recent common ancestor of *Spirinchus*.

To determine the most appropriate value for the smoothing parameter, cross-validation analysis (Sanderson 2003), under the penalized likelihood model (Sanderson 2002) with a logarithmic penalty function and tn (truncated Newton) algorithm, was conducted for smoothing values between 1 and 10^5, in increments of 0.5. This analysis yielded consistently decreasing cross-validation scores for these values, indicating that the phylogeny seems to fit the model of a molecular clock (Sanderson 2004). As such, to estimate divergence dates, the Langley-Fitch model (Langley and Fitch 1974), which assumes a molecular clock, was implemented with the tn algorithm and a gradient-check analysis. Analyses with the penalized likelihood model and tn algorithm with various smoothing values yielded similar results to those of the Langley-Fitch model (data not shown). Fossil constraint cross-validation required at least one fixed node in the phylogeny, so for this analysis the root node was constrained to either 55 or 65 mya.

Bayesian estimation using Beast

An input file generated by Beauti v.1.4.4 (Rambaut and Drummond 2007) was constructed from an alignment of the concatenated nDNA data, with a GTR+I+G substitution and site heterogeneity model, with four gamma categories, a lognormal relaxed clock (Drummond et al. 2006), and a Yule pure-birth tree prior, which assumes a constant speciation rate per lineage. The three fossil constraints were

implemented as follows: (1) root node – uniform prior between 55 and 65 mya, (2) Osmeridae most recent common ancestor – exponential prior with a minimum value (offset) of 30 mya, and a mean of 1.35, which was chosen so that 95% of the probability is contained between 30 mya and a 'soft' maximum bound (Yang and Rannala 2006; Ho 2007) of 34 mya [beginning of Oligocene]. (3) *Spirinchus* most recent common ancestor – exponential prior with a minimum of 12 mya, and mean of 1.35 [soft maximum bound of 16 mya, beginning of mid-Miocene]. The analysis was initiated with a tree conforming to the Bayesian nDNA topology from Fig. 4.11 (apart from an *O. dentex* – *O. eperlanus* sister relationship, which was supported by the allDNA analysis [Fig. 4.6]) with node dates that obeyed the above constraints and all operators referring to tree topology removed, thereby fixing the topology for the divergence date estimation analysis. Beast v.1.4.4 (Drummond and Rambaut 2007) was run for 5×10^7 MCMC generations, logging data every 1000 generations. Resulting data were analyzed using Tracer v.1.3 (Rambaut and Drummond 2004).

5.2.2 Reconstructing Osmeridae biogeographic history

Similar to the approach in Chapter 3 (Ilves and Taylor 2007), I used two methods to examine the biogeographic history of the Osmeridae: a dispersalist parsimony approach using a step-matrix implemented in MacClade v.4.06 (Maddison and Maddison 2003) and Mesquite (Maddison and Maddison 2006) and a ML approach in LAGRANGE (Ree et al. 2005) that uses an ultrametric tree and a model of connections between the different areas through time to calculate the most

likely ancestral areas at each node, from which a biogeographic scenario can be reconstructed.

Parsimony reconstruction

For the parsimony approach, I examined the evolution of geographic area on the nDNA Bayesian and ML tree (same as the allDNA except for the *Osmerus* dentex – *O. eperlanus* sister relationship) as well as a tree that included the three fossil taxa. Following the discussion in the MacClade v.4.06 manual (pp. 75-77; Maddison and Maddison 2003) about allowing polymorphic ancestral states, each species was coded with one of the following character states: 0 (western Pacific), 1 (eastern Pacific), 2 (northern Pacific), 3 (Arctic), 4 (western Atlantic), 5 (eastern Atlantic), 6 (northern Pacific and Arctic), 7 (western, eastern, northern Pacific and Arctic), or 8 (all areas). A step-matrix was constructed with the number of steps between each state shown in Table 5.1. For this model it was assumed that dispersal was only permitted through coastal areas, for example, a dispersal from the western Pacific to the eastern Pacific, and vice-versa, required passage through the northern Pacific (2 steps). For species found in more than one geographic area, transitions were assigned half a step, generally following Smith et al. (2005). For example, a transition from the western Pacific (state 0) to the northern Pacific and Arctic (state 6) required 1.5 steps (one step from the western to northern Pacific, and half a step to the Arctic).

Maximum likelihood reconstruction

For the ML analysis using LAGRANGE (Ree et al. 2005), an ultrametric tree was based on the divergence time estimates for each node from the r8s and Beast analyses for each of two trees: the nDNA (all DNA) tree topology and the same topology but with *Mallotus villosus* as sister to the rest of the species in the family. *Plecoglossus altivelis* was used as the outgroup and was appropriate to include in this analysis as it appears to be the sister taxon, with the similarly distributed Salangidae, to the Osmeridae (Chapter 4, Fig. 4.2).

As in Chapter 3 (Ilves and Taylor 2007), connections between areas were modeled on landbridge connections and climate changes in the Cenozoic. All climatic interpretations were from Zachos et al. (2001) unless otherwise specified. The probability of connection between the northern Pacific and western and eastern Pacific regions was implemented as follows: 1.0 prior to 50 mya, linear decrease from 1.0 to 0.5 between 50 and 34 mya [cooling following mid-Eocene optimum], 0.5 between 34 and 25 mya [relatively cool Oligocene period], linear increase from 0.5 to 1.0 between 25 and 15 mya [warming at end of Oligocene to mid-Miocene optimum], decrease from 1.0 to 0.5 at 15 mya [mid-Miocene cooling period], remaining at 0.5 until 5 mya, increasing from 0.5 to 1.0 at 5 mya [early Pliocene warm period; Ravelo et al. 2004], remaining at 1.0 between 5 and 3 mya, with a final linear decrease from 1.0 to 0.0 between 3 and 2 mya [beginning of Pleistocene glaciations]. A connection between the Arctic and North Pacific oceans was allowed with a probability of 1.0 between 7 – 2 mya, as there was no Bering land bridge during this period prior to the onset of the Pleistocene glacial cycles (Marincovich and Gladenkov 1999). Finally, dispersal between the Arctic and eastern and

western Atlantic regions was permitted with a logistic increase from 0.0 to 1.0 between 34 and 15 mya [opening of connection between the Arctic and Atlantic basins in the Oligocene]. The model from 15 mya to the present is the same as that for the northern Pacific – western and eastern Pacific regions.

Following Ree et al. (2005), several combinations of dispersal (λ_D) and extinction (λ_E) probabilities were run for each of the four trees ($\lambda_D = 0.09$, $\lambda_E = 0.01$; $\lambda_D = 0.009$, $\lambda_E = 0.001$; $\lambda_D = 0.05$, $\lambda_E = 0.05$; $\lambda_D = 0.005$, $\lambda_E = 0.005$; $\lambda_D = 0.01$, $\lambda_E = 0.09$; $\lambda_D = 0.001$, $\lambda_E = 0.009$). For each λ_D, λ_E combination I ran 10^5 iterations.

5.3 RESULTS

5.3.1 Divergence date estimates

Divergence dates for each node in the two Osmeridae phylogenies as estimated using ML and Bayesian approaches are listed in Table 5.2. In general, the estimates are largely congruent across the two methods, although the Bayesian analysis yielded slightly older ages for all nodes apart from the root (Table 5.2). Estimates for the phylogeny with *Mallotus villosus* sister to the rest of the genera were very similar to the alternative topology examined (Table 5.2).

The gradient-check analysis option in r8s (Sanderson 2003) showed that all of the constraints were 'active' but were not violated. Fossil cross-validation, on the other hand, revealed that the imposed fossil constraints might be questionable. Fixing the root node to either 55 or 65 mya yielded an estimated age of 6.95 mya for

Spirinchus when the minimum 12 mya constraint was removed and the age of the Osmeridae was estimated at 22.6 mya without the minimum age constraint of 30 mya. When either the *Spirinchus* or Osmeridae node was fixed at 12 or 30 mya, respectively, the estimate for the root node exceeded 100 mya.

With age estimates available for the Osmeridae phylogeny it was possible to estimate a molecular clock rate for each gene (Table 5.4). These were calculated using the pairwise corrected sequence data, calculating a rate per million years (pmy) for each node apart from the root, and then averaged across the phylogeny. The estimated rate of cyt*b* evolution (1.06% pmy) is more comparable to that calculated for ectothermic vertebrates (0.5-0.9% pmy; Martin and Palumbi 1993) than to the rates calculated for *Hypomesus* alone (1.77% pmy) in Chapter 3 (Ilves and Taylor 2007) and for rockfishes (1.56% pmy; Stepien et al. 2000). As with any molecular clock rate, these estimates should be treated with caution (Arbogast et al. 2002).

5.3.2 Biogeographic history reconstruction

Parsimony step-matrix

Mapping geographic area onto the phylogeny of extant Osmeridae species using the step-matrix in Table 5.1 yielded equivocal ancestral areas at several nodes. The root node and the putative ancestors of the Osmeridae and *Hypomesus* had either a western Pacific or northern Pacific distribution, while the ancestors of the (*Mallotus*, *Osmerus*, *Thaleichthys*, *Allosmerus*, *Spirinchus*) and (*Osmerus*, *Thaleichthys*, *Allosmerus*, *Spirinchus*) clades had equivocal reconstructions for

different regions of the Pacific and Arctic (Fig. 5.3A). Addition of fossil taxa in their inferred phylogenetic positions also did not result in unique most parsimonious reconstructions at these nodes but tended to support either northern Pacific or northern Pacific and Arctic ancestors (Fig. 5.3B).

Maximum likelihood biogeographic modeling

Of the various combinations of dispersal and extinction rate parameters run, the $\lambda_D = 0.09$, $\lambda_E = 0.01$ values yielded the biogeographic scenario with the highest likelihood for all trees, therefore, these results are presented. Analyses using the parameter combination $\lambda_D = 0.01$, $\lambda_E = 0.09$, with a relatively high extinction rate, failed (N/A result for all nodes), which suggests the extinction rate was likely too high. For most nodes, more than one scenario fell within two log-likelihoods of the most likely scenario, with a range from 1 (nodes 6, 9-11, and 14; Fig. 5.2) to >90 (node 7; Fig. 5.2). The most likely reconstructions for each node are listed in Table 5.3. Biogeographic reconstruction using the phylogeny with the Holarctic *Mallotus villosus* as sister to the rest of the osmerid genera resulted in the same ancestral distribution scenarios (data not shown).

5.4 DISCUSSION

5.4.1 Osmeridae divergence dates and Cenozoic climatic and geologic events

Many estimated divergence dates within the Osmeridae coincide with important events in the development of cold temperate biotas in the northern hemisphere. The origin of the Osmeridae in the early Oligocene (Table 5.2; node 2, Fig. 5.2) is consistent with evidence in a number of taxa that a cold temperate fauna was well established by this time (Briggs 1970; Kennett 1982; Oleinik 2001). Several subsequent divergences (Fig. 5.2) within *Hypomesus* (node 5), between *Allosmerus* and *Spirinchus* (node 12) and within *Spirinchus* (node 13) occurred during the mid-Miocene [16-12 mya], concurrent with global cooling (Kennett 1982; Zachos et al. 2001). Further, the timeframe of differentiation within Holarctic-distributed *Osmerus* [~5-7 mya] coincides with estimates of the first opening of the Bering Strait [4.8-7.4 mya] (Marincovich and Gladenkov 1999).

Estimated divergence dates in *Hypomesus* are largely concordant with those found from analysis of *Hypomesus* and *Mallotus* alone (cf. Tables 5.3 and 3.2, Chapter 3; Ilves and Taylor 2007), although estimates using the family level fossil-calibrated phylogeny yielded slightly older estimates for most nodes (0.3 – 3.2 my older; cf. Tables 5.3 and 3.2). The *H. japonicus – H. pretiosus*/*H. transpacificus* node calibrated at 15 mya in Chapter 3 (Ilves and Taylor 2007) was estimated to be younger with the current data, at 12.7 mya. While not an exact match, this estimate does not violate the assumption that the origin of the Sea of Japan (~15 mya; Itoh et al. 1997) was an important factor in this trans-Pacific divergence as the new

estimate post-dates this event. This timeframe also coincided with the mid-Miocene cooling period, therefore, with current data it is not possible to determine which (if either) event spurred the trans-Pacific disjunction seen in *Hypomesus*. As discussed in Chapter 3 (3.4.3), diversification in numerous marine taxa dated to this time period. Repeated isolation of the Sea of Japan from the Pacific Ocean and Sea of Okhotsk has been implicated as a significant driver of diversity in the western North Pacific (Lindberg 1953, cited in Briggs 1974), leading to the evolution of the endemic sea urchin *Hemicentrotus pulcherrimus* (Lee 2003) and to a speciation event in the gastropod genus *Tegula* (Hellberg 1998). Additional comparative data on the degree of endemism within the Sea of Japan in conjunction with divergence date estimates for taxa inhabiting this basin and surrounding areas may further clarify the role of this region in the diversification of North Pacific organisms.

The second trans-Pacific relationship within the Osmeridae, between the western Pacific *Spirinchus lanceolatus* and eastern Pacific *S. starksi/S. thaleichthys* clade (12.5 mya), appeared to be coincident with that seen in *Hypomesus* (12.7 mya), suggesting the same factor may have been responsible for initiating both divergences; however, further statistical comparisons with other trans-Pacific taxa are needed to assess the validity of this claim. *S. lanceolatus* is not currently found in the Sea of Japan, being restricted to the Pacific coast of Hokkaido, Japan, and no evidence suggests it had ever been located in this basin. This perhaps favours the hypothesis of the importance of cooling temperatures in the mid-Miocene in generating diversity in the North Pacific Ocean. The concordance in timing between the trans-Pacific splits in *Hypomesus* and *Sprinchus* following the mid-Miocene

cooling period was consistent with McAllister's (1963) compression hypothesis that postulated vicariant splitting of a widespread ancestor into populations on each side of the North Pacific Ocean.

Results of the fossil cross-validation analysis (Sanderson 2003) that I conducted suggest there may be problems with the age constraints placed on the root node. Estimated divergence dates for *Spirinchus* and the Osmeridae with the minimum age constraints removed [12 mya and 30 mya, respectively] resulted in much younger estimated ages for these taxa [6.95 and 22.6 mya, respectively], and the estimated age of the root was over 100 mya with removal of the 55 – 65 mya age constraint. Extending the maximum age to 250 mya using the ML divergence date estimation approach resulted in a root age of over 230 mya and also yielded much older ages for several other nodes (data not shown). The reasons for this discrepancy are unclear, but may result from the long branch connecting the outgroup *Plecoglossus altivelis* to the ingroup taxa. By contrast, rerunning the Bayesian divergence date analysis with a uniform prior over 55-150 mya for the root resulted in an older estimate for the root age (72.8 mya) but did not affect the other nodes by more than 1 my (data not shown). These new estimated ages did not affect the ML biogeographic analysis (data not shown). Because the Bayesian method uses sequence data directly and resulted in similar age estimates for all nodes apart from the root when the root age constraint was less stringent, I used the results from this analysis for the biogeographic model.

5.4.2 Osmeridae biogeographic reconstruction

The two methods of biogeographic reconstruction employed (dispersalist parsimony step-matrix and ML modeling that includes both dispersal and vicariance) yielded strikingly different interpretations of Osmeridae biogeography. Although both methods resulted in equivocal constructions for most nodes apart from the more recent divergences in the eastern Pacific, the ML approach suggested generally more widespread ancestors near the base of the phylogeny (Table 5.3, Fig. 5.3). The parsimony approach supported a Pacific or Arctic origin for the family (node 2, Fig. 5.3), whereas ML reconstruction suggested an ancestor distributed throughout the Pacific and Atlantic Oceans but missing from the Arctic region.

In principle, the ML analysis as implemented in LAGRANGE (Ree et al. 2005) is preferable to the more simplistic parsimony-based approach because it incorporates an explicit biogeographic model through time; however, the parsimony model yielded a more believable biogeographic reconstruction for the extant taxa (Fig. 5.3A). For instance, the LAGRANGE (Ree et al. 2005) results for the ancestors at the root and Osmeridae nodes (nodes 1 and 2, respectively, Table 5.3) highlighted a shortcoming of this method in its current incarnation. Given the biogeographic model implemented, which precluded dispersal between the Atlantic and Arctic until the early Oligocene (34 mya) because these basins were only connected after that time, it was surprising that the two basal nodes were widely distributed in the Pacific and Atlantic basins. The result of an ancestor in the Atlantic Ocean highlights a limitation of LAGRANGE in that it was not able to incorporate the possibility that areas had not existed for the entire timeframe under consideration,

but rather, appeared at a particular time. For the model of extant Osmeridae taxa, the Atlantic Ocean should effectively not exist until the Oligocene. This could be solved by the addition of a parameter describing the probability of a particular ancestral state through time, which most simply could be set to 0 or 1 for the time periods prior to and after, respectively, an area's appearance. This would be useful not only in the context of modeling oceanic connections, as in the case of the Osmeridae and other Holarctic organisms, but also for biogeographic modeling of island taxa, as many volcanic islands have known dates of emergence and clearly could not be part of an ancestral area prior to those dates.

Parsimony reconstruction, on the other hand, suggested a western or northern Pacific origin for the family. Only three species are currently found in the Atlantic Ocean basin: *Osmerus mordax* (western Atlantic), *O. eperlanus* (eastern Atlantic) and *Mallotus villosus* (circumpolar). According to both methods of estimation (Table 5.2), the divergences within *Osmerus* occurred in the late Miocene, after the opening of the Bering Strait, which would have allowed dispersal among the Pacific, Arctic and Atlantic Oceans. *M. villosus* could also have originated in the Pacific Ocean and colonized the other Holarctic regions when this connection was available.

In order for the Osmeridae ancestor to have had the distribution attributed to it by the ML analysis, one of two scenarios must have held in order to be consistent with geological knowledge that the Arctic and Atlantic Oceans were not connected until the Oligocene. First, the biogeographic model developed in this chapter assumed that all transitions between the Pacific and Atlantic basins were through a

northern connection across the Arctic region. It is, however, conceivable that dispersal between the oceans occurred via a southerly route along the southern edge of what is now Central America, prior to its connection with South America in the Pliocene (~3.5 mya). This seems unlikely as all known osmerids (extant and extinct) have lived in relatively cool environments. Temperatures at lower latitudes have been tropical throughout the estimated timeframe of Osmeridae evolution (Kennett 1982) and should have acted as a barrier against southward dispersal; however, the western North Pacific salangids, close relatives of the Osmeridae (Chapter 4; Fu et al. 2005), are represented in subtropical waters, which suggests that tolerance of a wider temperature range may have been present in ancestral osmerids. Dispersal from the Atlantic to the Pacific through the Panama Seaway in the mid – late Miocene was suggested as an explanation for Pacific – Atlantic relationships in the hake genus *Merluccius* (Grant and Leslie 2001). Evidence of fossil osmerids from more southerly regions, such as the Pacific and Atlantic sides of Central America, would provide the strongest support for this hypothesis.

A second possible scenario is that the Osmeridae ancestor had a wide-ranging distribution around the Holarctic prior to the Bering landbridge connection between eastern Asia and western North America in the mid-Cretaceous (~100 mya). The estimated divergence dates and known fossils do not support such an old origin for the family; however, it is possible that osmerid fossil evidence dating to this time period will show that the family has a much older origin than current data suggest. The interpretation of a Pacific Ocean origin for the family begins to unravel when known fossil evidence is taken into consideration. Although all extant Atlantic

species appear to have evolved relatively recently, as Wilson and Williams (1991) suggested, fossils of *Enoplopthalmus* from Europe dating to the Oligocene, when the Bering landbridge existed, prevented dispersal between the Pacific and Arctic basins, bringing a Pacific origin for the family into question. Inclusion of fossil taxa into the parsimony reconstruction (Fig. 5.3B) suggested a northern Pacific or northern Pacific and Arctic distribution for the Osmeridae; however, this model does not account for the timing of divergences, and thus also fails to accurately represent Osmeridae biogeography. In order to account for the distribution of extant and extinct osmerid taxa while remaining consistent with geologic evidence, the Osmeridae lineage must predate the connection of eastern Asia and western North America in the Cretaceous. Such a scenario thus suggests that Pacific and Arctic/Atlantic lineages within the family have been diverging considerably longer than suggested by extant species. Viewed from this perspective, results from the ML analysis may inadvertently be correct, even though on the timeframe of the model they are unsupported.

The two biogeographic reconstruction methods generally agreed with respect to divergences within *Hypomesus* and the (*Thaleichthys*,(*Allosmerus*,*Spirinchus*)) clade. The western and northern Pacific regions were postulated to have been important areas in *Hypomesus* evolution (Table 5.3, Fig. 5.3), and the scenarios are concordant with those estimated with the limited *Hypomesus* and *Mallotus* analysis in Chapter 3 (Ilves and Taylor 2007). By contrast, the eastern Pacific appears to be a centre of origin for the most recent divergences in the family (Table 5.3, Fig. 5.3).

5.4.3 Previous interpretations and comparisons across Holarctic marine taxa

Biogeographic analysis of the Osmeridae suggests three significant climatic and geologic events in the Cenozoic shaped the evolution of extant lineages in this family: (1) development of cold water conditions beginning in the Paleocene/Eocene, (2) further climatic cooling in the mid-Miocene, and (3) opening of the Bering Seaway in the late Miocene. If these large-scale changes were significant vicariant events for the Osmeridae, then they should also have affected other Holarctic marine taxa.

Much of the data on Holarctic-distributed taxa comes from the terrestrial literature; however, the development of cool temperate conditions has been implicated in the diversification of a number of marine taxa, including bivalves (MacNeil 1965), gastropods (Golikov and Tzvetkova 1972; Oleinik 2001), crabs (Schweitzer 2001), and fishes (Casier 1966, cited in Briggs 1970). Most studies were of hard-bodied invertebrates with well-preserved fossil records.

There has been extensive discussion in both the terrestrial botanical and zoological literature regarding patterns of diversification across the Holarctic through the Cenozoic. A recent review (Sanmartín et al. 2001) of 57 Holarctic terrestrial animal taxa, which employed an event-based biogeographic analysis similar to the ML model in LAGRANGE (Ree et al. 2005), examined distribution patterns through the Cenozoic and found concordance across these many taxa for the importance of landbridge connections between North America and Asia and North America and Europe in facilitating dispersal and diversification. To my knowledge, no similar comparisons are available for marine taxa on a similar timescale; however, a clear

case for biogeographic comparison to the Osmeridae is the relatively closely related salmon family Salmonidae, particularly the genera *Oncorhynchus* (Pacific salmon and trouts), *Salmo* (Atlantic salmon and trouts) and *Salvelinus* (chars). *Oncorhynchus* and *Salvelinus* have a North Pacific distribution, while *Salmo* species are restricted to North Atlantic drainages. Of the many studies of relationships among these genera (e.g., Neave 1958; Devlin 1993; Stearley and Smith 1993; Phillips and Oakley 1997; Oakley and Phillips 1999; Crespi and Fulton 2004), few have focused on the timing of divergences (e.g., Neave 1958; Devlin 1993; McPhail 1997). From nuclear growth hormone sequences, Devlin (1993) inferred a widely accepted sister relationship between *Oncorhynchus* and *Salmo* (e.g., Neave 1958; Stearley and Smith 1993; Murata et al. 1996), which was dated at ~20 mya. This divergence postdates the estimated Osmeridae date of origin (~30 mya; Table 5.2); however, it also falls into the time period when the Atlantic and Arctic Oceans were connected but the Pacific and Arctic basins were not. Without a trans-Arctic route, this again raises questions regarding the ancestral distribution and the process of lineage splitting. Recent nDNA evidence suggests *Salvelinus*, not *Salmo*, is the sister group to *Oncorhynchus* (Oakley and Phillips 1999; Crespi and Fulton 2004); however, neither of these studies estimate divergence dates or discuss biogeographic implications.

The family Gasterosteidae (sticklebacks) also has a Holarctic distribution, with species found in coastal marine and freshwaters throughout this region. Recent phylogenetic analysis of mtDNA and morphology (Mattern 2004; Mattern and McLennan 2004) yielded a well-resolved phylogeny of the family, although, as with

the Salmonidae, divergence dates and biogeographic questions have yet to be addressed. Large genetic databases generated by Crespi and Fulton (2004), Mattern (2004), and Mattern and McLennan (2004) provide a great opportunity for comparative Holarctic biogeographic analysis.

Although there are apparently no comparable marine Holarctic studies to that of the Osmeridae through the Cenozoic, a considerable amount of biogeographic research has focused on two areas: (1) diversification in the North Pacific, especially in the Miocene, and (2) comparisons between North Pacific and North Atlantic faunas resulting from the trans-Arctic connection made available with the opening of the Bering Seaway in the late Miocene. As discussed in Chapter 3 (Ilves and Taylor 2007), the mid-Miocene cooling period and associated sea-level changes were hypothesized to have been important in generating trans-Pacific distributions seen across many taxa (e.g., Collins et al. 1996; Kai et al. 2003; Lee 2003). Furthermore, the western Pacific is regarded as the centre of origin in many North Pacific groups, which then subsequently dispersed northward and eastward (MacNeil 1965; Gladenkov 1994; Oleinik 2001; Hyde and Vetter 2007). Biogeographic results for *Hypomesus* diversification in the Miocene also conformed to this pattern (Table 5.3, Fig. 5.3). Recent systematic and biogeographic work on the pan-Pacific rockfish genus *Sebastes* (Hyde and Vetter 2007) suggested that changes in upwelling patterns and sea-level were important drivers of diversification within this group. Hyde and Vetter (2007) noted that periods of intense upwelling in the eastern Pacific beginning in the late Miocene (~6.5 mya) coincided with *Sebastes* diversification and they hypothesized that increased productivity allowed increased ecological

specialization, leading to diversification. Hyde and Vetter (2007) also correlated low sea-levels with increased speciation events from the late-Miocene through the Pleistocene. Divergences among the sympatric eastern Pacific osmerid genera (*Thaleichthys*, *Allosmerus* and *Spirinchus* [excluding the western Pacific *S. lanceolatus*]) in the mid-Miocene (Table 5.2) predated these events, raising the question of what drove speciation among these lineages? McAllister (1963) postulated that life-history differences might have allowed similarly distributed osmerid species to coexist. Interestingly, the eastern Pacific species of *Thaleichthys*, *Allosmerus* and *Spirinchus* show variation across their life-histories. *T. pacificus* is anadromous, *A. elongatus* is strictly marine, as is *S. starksi*, while *S. thaleichthys* has both anadromous and freshwater populations. This partitioning of life-histories, and therefore of important limited resources, such as spawning habitat, may have been an important factor in the divergences among these taxa.

Understanding the processes underlying relationships among taxa in the North Pacific and Atlantic Oceans has long been of interest to biogeographers (e.g., Ekman 1967; Briggs 1970, 1974; Vermeij 1991). A particularly perplexing question has been why the North Atlantic is remarkably species-poor relative to the North Pacific. The difference in species richness applied to numerous taxonomic groups (e.g., Briggs 1970; Vermeij 1991) and is also evident in the Osmeridae, where only three of 14 species are found in the North Atlantic (and only two exclusively, as *Mallotus villosus* has a circumpolar distribution). In relation to the Osmeridae, the question becomes more puzzling as fossil osmerids dating to the Oligocene have been found in Europe (Gaudant 1985, cited in Wilson and Williams 1991). One

explanation for the disparity in diversity between the two oceans is that the extinction rate has been higher in the North Atlantic due to its connection with the Arctic and because this region experienced dramatically larger temperature fluctuations during periods of global cooling, such as the late Tertiary and Pleistocene, owing to its relatively small size (Briggs 1970, and references therein). This hypothesis was supported by an analysis of 295 molluscan species showing that the extinction rate in the North Atlantic since the opening of the Bering Seaway was greater than that in the North Pacific (Vermeij 1991). In addition to periods of cooling associated with the Pleistocene glaciations, dramatic temperature drops also occurred in the Oligocene and mid-Miocene (Zachos et al. 2001). The Oligocene osmerid taxa may have been driven to extinction in the Atlantic as a result of these climatic changes, explaining the lack of extant osmerids in the North Atlantic that predate the opening of the Bering Strait. Further exploration of this hypothesis requires phylogeographic analysis of the circumpolar *Mallotus villosus*, or support from additional Atlantic osmerid fossils.

The opening of the Bering Seaway in the late Miocene was an important driver of diversification for numerous organisms, including molluscs (Vermeij 1991, 2005), sea urchins (Lee 2003; Addison and Hart 2005), fishes (Grant 1987) and algae (van Oppen et al. 1995; Lindstrom 2001). A large body of evidence also supports biased dispersal from the North Pacific to the North Atlantic (e.g., Briggs 1970; Vermeij 1991; Rawson and Hilbish 1995; Collin 2003; Väinölä 2003); however, recent work on sea urchins (Addison and Hart 2005) studying more recent

Pleistocene dispersals suggests that back colonization from the Atlantic to the Pacific could also occur.

The Bering Strait connection has also long been hypothesized to have had an important role in the evolution of the osmerid genus *Osmerus* (McAllister 1963; Klyukanov 1975, cited in Taylor and Dodson 1994; Taylor and Dodson 1994). Based on mtDNA RFLP and cytochrome *b* sequence divergence, Taylor and Dodson (1994) dated divergences within the genus to the Plio- and Pleistocene. The biogeographic results from the current study also supported the importance of the opening of the Bering Seaway for the evolution of *Osmerus* species; however, the divergence estimates from my more extensive nDNA sequence data suggest slightly early dates of diversification in the late Miocene/early Pliocene (Table 5.2). The alternative biogeographic hypotheses put forth by McAllister (1963) [later supported by Taylor and Dodson (1994)] and Klyukanov (1975, cited in Taylor and Dodson 1994) cannot be assessed with the current data as the phylogenetic relationships among *Osmerus* species remain unresolved (Chapter 4). Further sequencing of independent nDNA loci and phylogeographic analysis of the widespread *O. dentex* may help clarify the interrelationships of these species and permit more explicit testing of biogeographic hypotheses, such as the area of origin (Pacific, Atlantic, or Arctic) and the direction of dispersal.

5.4.4 Conclusions

Divergence dates in the Osmeridae largely coincide with climatic and geologic events that have had well-documented effects on the diversification within other

temperate marine taxa. The origin of the family was estimated to be in the early Oligocene, when cold-temperate conditions were known to exist, and the mid-Miocene cooling period and opening of the Bering Seaway in the late Miocene appear to have been particularly important in the subsequent evolution of osmerid taxa. Parsimony biogeographic reconstruction of extant species suggested a North Pacific Ocean origin for the family, while ML modeling resulted in a hypothesis of a widespread ancestor in the northern Pacific and Atlantic Oceans. The parsimony model was considered to best conform to the geologic evidence through the Cenozoic; however, inclusion of fossil taxa requires a reassessment of these biogeographic conclusions. With Atlantic osmerid fossils dating to the Oligocene, the origin of the osmerid lineage must predate the Cretaceous connection between eastern Asia and western North America to account for the distribution of extant and extinct taxa within this group. Future osmerid fossil discoveries dating to the Cretaceous, from the Arctic region in particular, would provide the most compelling evidence for this hypothesis. Phylogeographic work on the circumpolar *Mallotus villosus* and *Osmerus dentex* would also help clarify the area of origin and dispersal pathways for these widely distributed species. Most importantly, comparative biogeographic analysis of other Holarctic taxa with a similar timeframe of evolution, and of taxa distributed in and around the Sea of Japan, will permit further exploration of the biogeographic hypotheses discussed in this chapter.

Table 5.1 Parsimony step-matrix between areas occupied by Osmeridae species. Lower left symmetrical with upper right triangle.

From		WP	EP	NP	A	WA	EA	NP, A	WP, EP, NP, A	WP, EP, NP, A, WA, EA
To	WP	0	2	1	2	3	3	1.5	1.5	2.5
	EP		0	1	2	3	3	1.5	1.5	2.5
	NP			0	1	2	2	0.5	1.5	2.5
	A				0	1	1	0.5	1.5	2.5
	WA					0	2	1.5	2.5	2.5
	EA						0	1.5	2.5	2.5
	NP, A							0	1	2
	WP, EP, NP, A								0	1
	WP, EP, NP, A, WA, EA									0

Table 5.2 Estimated divergence times (millions of years ago) for nodes of the nDNA Osmeridae phylogeny with *Mallotus villosus* as sister to the OTAS clade and as sister to all Osmeridae genera, using ML (r8s; Sanderson 2002) and Bayesian (Beast; Drummond and Rambaut 2006) reconstruction methods. Numbers in brackets following Bayesian estimates represent lower and upper 95% confidence intervals.

Estimated divergence dates				
	nDNA phylogeny		*Mallotus* sister	
	ML	Bayes	ML	Bayes
Root	65.0	59.9 (55.0-64.5)	65.0	59.9 (55.0-64.5)
Osmeridae	30.0	31.9 (30.0-35.4)	30.0	32 (30.0-35.5)
Hypomesus	22.3	24.2 (17.5-30.4)	16.9	21.8 (15.6-27.7)
Ho+Hj+Hp+Ht	17.6	19.3 (13.0-26.6)	13.3	17.5 (11.7-23.8)
Hj+Hp+Ht	12.0	12.7 (7.2-18.4)	9.0	11.5 (6.3-16.5)
Hp+Ht	3.6	4.2 (1.5-7.5)	2.7	4.0 (1.5-6.8)
MOTAS/HOTAS	27.2	28.3 (23.8-33.0)	26.8	30.0 (25.4-34.4)
OTAS	17.7	20.8 (17.2-25.1)	16.3	21.3 (17.2-25.6)
Osmerus	5.5	6.9 (2.9-11.8)	4.3	6.7 (2.9-11.0)
Od+Oe	5.1	5.8 (2.2-10.1)	4.0	5.7 (2.2-9.6)
TAS	15.0	17.3 (14.4-20.6)	14.1	17.6 (14.4-21.1)
AS	13.6	15.1 (12.9-17.7)	13.1	15.2 (13.0-18.1)
Spirinchus	12.0	12.5 (12.0-13.6)	12.0	12.6 (12.0-13.7)
Ss+St	8.9	9 (5.1-12.1)	8.0	9.0 (5.0-12.2)

Table 5.3 Biogeographic scenario for nodes of the Osmeridae nDNA phylogeny as estimated by LAGRANGE (Ree et al. 2005) using divergence date estimated by the Bayesian approach in Beast (Drummond and Rambaut 2006). Ancestral area shown represents most likely scenario with the number of other scenarios falling within two log-likelihoods in square brackets. Inherited ranges followed by lineage number or taxon names in brackets. It is assumed that speciation occurs in a single geographic area.

Node (#)	Node age	Ancestral area	Range inherited by lineage (x)	Range inherited by lineage (y)
Root (1)	59.9	WP, EP, EA, WA [82]	WP, EP, EA, WA (2)	WP (*Plecoglossus altivelis*)
Osmeridae (2)	31.9	WP, EP, NP, EA, WA [20]	WP, EP, NP, EA, WA (7)	WP *(Hypomesus)*
Hypomesus (3)	24.2	WP [4]	WP (4)	WP (*H. nipponensis*)
Ho+Hj+Hp+Ht (4)	19.3	WP, NP [7]	WP, NP (5)	NP (*H. olidus*)
Hj+Hp+Ht (5)	12.7	WP, EP [5]	EP (6)	WP (*H. japonicus*)
Hp+Ht (6)	4.2	EP [1]	EP (*H. pretiosus*)	EP (*H. transpacificus)*
MOTAS (7)	28.3	WP, EP, NP, EA, WA [99]	WP, EP, NP, EA, WA (8)	NP (*Mallotus villosus*)
OTAS (8)	20.8	WP, EP, NP, A, EA, WA [20]	EP (11)	WP, EP, NP, A, EA, WA (9)
Osmerus (9)	6.9	WP, EP, NP, A, EA, WA [1]	WP, EP, NP, A, EA (10)	WA (*Osmerus mordax*)
Od+Oe (10)	5.8	WP, EP, NP, A, EA [1]	WP, EP, NP, A (*O. dentex*)	EA (*O. eperlanus*)
TAS (11)	17.3	EP [1]	EP (12)	EP (*Thaleichthys pacificus*)
AS (12)	15.1	EP [2]	EP (13)	EP (*Allosmerus elongatus*)
Spirinchus (13)	12.5	EP, NP [3]	EP (14)	NP (*Spirinchus lanceolatus*)
Ss+St (14)	9.0	EP [1]	EP (*S. starksi*)	EP (*S. thaleichthys)*

Table 5.4 Osmerid molecular clock rates per million years for cytb, 16S, 12S, ITS2, S71 and RAG1 calculated from corrected pairwise distances and the estimated divergence dates for each node (Table 5.3). Standard deviation shown in brackets.

Gene	Sequence divergence (% per million years)
Cyt*b*	1.06 [0.36]
16S	0.11 [0.08]
12S	0.22 [0.16]
ITS2	0.30 [0.17]
S71	0.24 [0.08]
RAG1	0.21 [0.06]

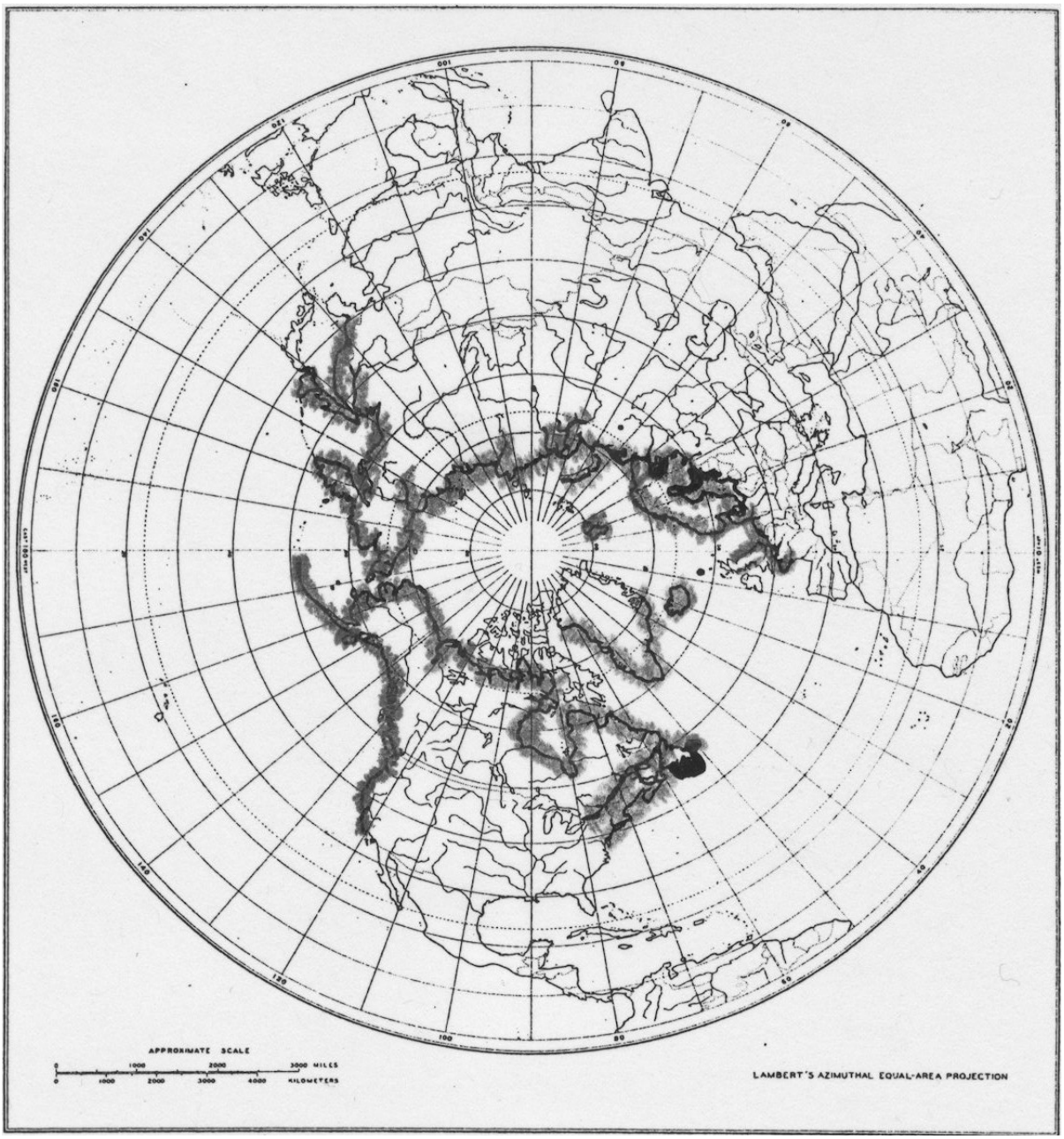


Figure 5.1 Holarctic distribution of the Osmeridae (McAllister 1963).

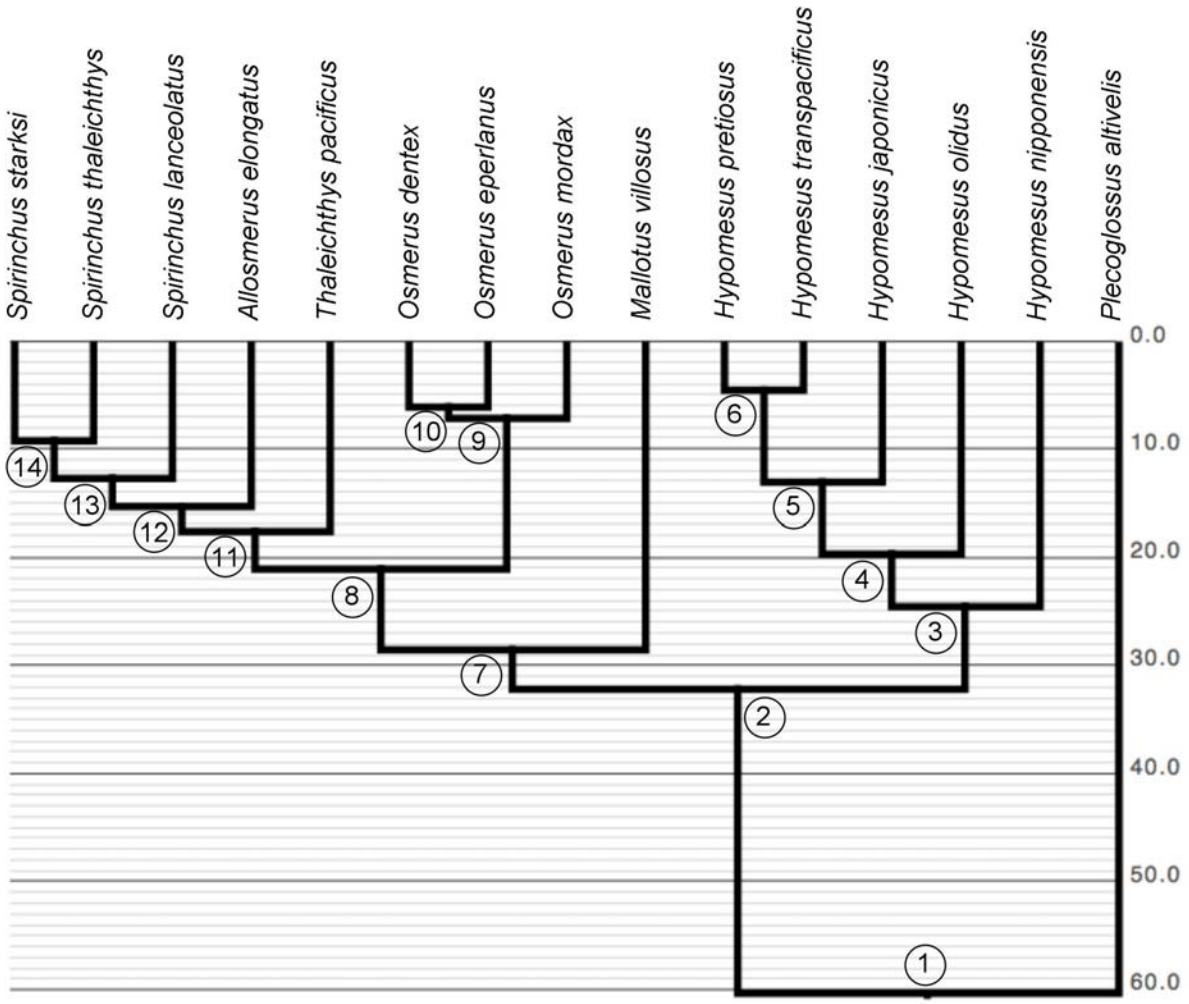


Figure 5.2 Ultrametric Osmeridae phylogeny from Bayesian estimation of divergence dates using Beast (Drummond and Rambaut 2006). Numbers below nodes refer to node numbers in Table 5.3.

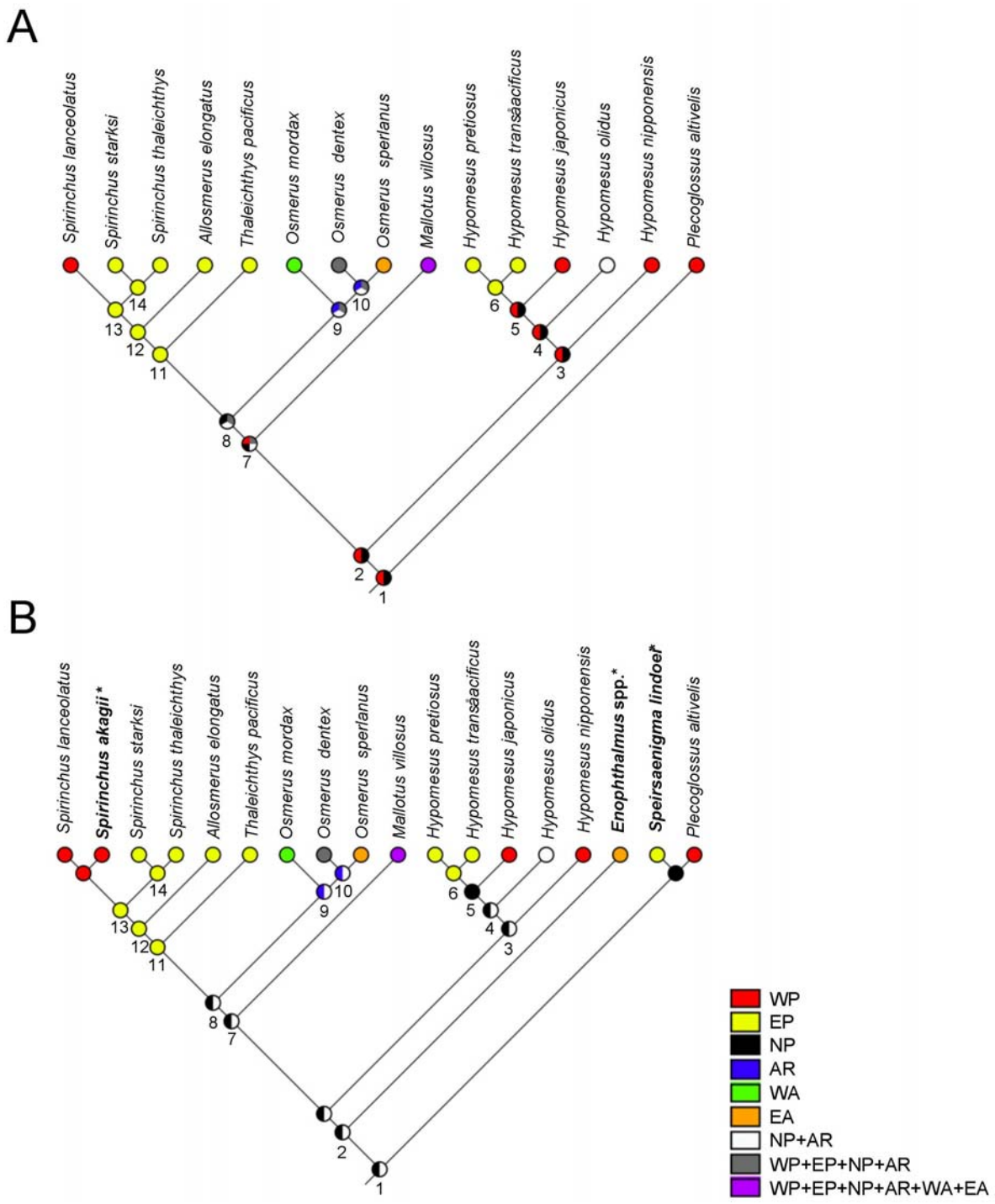


Figure 5.3 Parsimony reconstruction of ancestral areas based on step-matrix costs in Table 5.1. A: Osmeridae nDNA/allDNA phylogeny B: Osmeridae nDNA/allDNA phylogeny with three fossil taxa added, indicated by bold font and *. Geographic area abbreviations as follows: WP = western Pacific, EP = eastern Pacific, NP = northern Pacific, AR = Arctic, WA = western Atlantic, EA = eastern Atlantic. Numbers below nodes refer to node numbers in Table 5.3.

5.5 REFERENCES

Addison, J. A., and M. W. Hart. 2005. Colonization, dispersal, and hybridization influence phylogeography of North Atlantic sea urchins (*Stronglyocentrotus droebachiensis*). Evolution 59:532-543.

Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002 Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annu. Rev. Ecol. Syst. 33:707-740.

Arnason, U., A. Gullberg, A. Janke, M. Kullberg, N. Lehman, E. A. Petrov, and R. Väinölä. 2006. Pinniped phylogeny and a new hypothesis for their origin and dispersal. Mol. Phylogenet. Evol. 41:345-354.

Audzijonyte, A., and R. Väinölä. 2006 Phylogeographic analyses of a circumarctic coastal and a boreal lacustrine mysid crustacean, and evidence of fast postglacial mtDNA rates. Molec. Ecol. 15:3287-3301.

Bowen, G. J., W. C. Clyde, P. L. Koch, S. Ting, J. Alroy, T. Tsubamoto, Y. Wang, and Y. Wang. 2002. Mammalian dispersal at the Paleocene/Eocene boundary. Science 295:2062-2065.

Briggs, J. C. 1970. A faunal history of the North Atlantic Ocean. Syst. Zool. 19:19-34.

———. 1974. Marine Zoogeography. McGraw-Hill Inc., Toronto.

Brunner, P. C., M. R. Douglas, A. Osinov, C. C. Wilson, and L. Bernatchez. 2001. Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. Evolution 55:573-586.

Burbrink, F. T., and R. Lawson. 2007. How and when did Old World ratsnakes disperse into the New World? Mol. Phylogenet. Evol. 43:173-189.

Casier, E. 1966. Faune ichthyologique du London Clay. British Museum, London.

Collin, R. 2003. Phylogenetic relationships among calyptreid gastropods and their implications for the biogeography of marine speciation. Syst. Biol. 52:618-640.

Collins, T. M., K. Frazer, A. R. Palmer, G. J. Vermeij, and W. M. Brown. 1996. Evolutionary history of northern hemisphere *Nucella* (Gastropoda, Muricidae): molecular, morphological, ecological, and paleontological evidence. Evolution 50:2287-2304.

Crespi, B. J., and M. J. Fulton. 2004. Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. Mol. Phylogenet. Evol. 31:658-679.

Devlin, R. H. 1993. Sequence of sockeye salmon type 1 and 2 growth hormone genes and the relationship of rainbow trout with Atlantic and Pacific salmon. Can. J. Fish. Aquat. Sci. 50:1738-1748.

Drummond, A. J., and A. Rambaut. 2006. BEAST v1.4, Available from http://beast.bio.ed.ac.uk/

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLOS 4:e88.

Ekman, S. 1967. Zoogeography of the sea. Translated by E. Palmer. Sidgwick and Jackson, London.

Fu, C., J. Luo, J. Wu, J. A. López, Y. Zhong, G. Lei, and J. Chen. 2005. Phylogenetic relationships of salangid fishes (Osmeridae, Salanginae) with comments on phylogenetic placement of the salangids based on mitochondrial DNA sequences. Mol. Phylogenet. Evol. 35:76-84.

Gaudant, J. 1985. Mise en évidence d'Osmeridae (poissons téléostéens, Salmoniformes) dans l'Oligocène lacustre d'Europe occidentale. Comptes Rendus de l'Académie des Sciences Paris 300:79-82.

Gladenkov, Y. B. 1994. Cenozoic paleogeography and climatic change in the North Pacific Ocean. Palaeogeogr., Palaeoclimatol., Palaeoecol. 108:311-318.

Golikov, A. N. and N. L. Tzvetkova. 1972. The ecological principle of evolutionary reconstruction as illustrated by marine animals. Mar. Biol. 14:1-9.

Grant, W. S. 1987. Genetic divergence between congeneric Atlantic and Pacific Ocean fishes. Pp. 225-246 in N. Ryman and F. M. Utter, eds. Population genetics and fishery management. University of Washington Press, Seattle.

Grant, W. S., and R. W. Leslie. 2001. Inter-ocean dispersal is an important mechanism in the zoogeography of hakes (Pisces: Merluccius spp.). J. Biogeogr. 28:699-721.

Hainey, R. A., and D. J. Taylor. 2003. Testing paleolimnological predictions with molecular data: the origins of Holarctic *Eubosmina*. J. Evol. Biol. 16:871-882.

Hellberg, M. E. 1998. Sympatric sea shells along the sea's shore: the geography of speciation in the marine gastropod *Tegula*. Evolution 52:1311–1324. 1790–1805.

Hendrixson, B. E., and J. E. Bond. 2007. Molecular phylogeny and biogeography of an ancient Holarctic lineage of mygalomorph spiders (Araneae: Antrodiaetidae: *Antrodiaetus*). Molec. Phylogenet. Evol. 42:738-755.

Ho, S. Y. W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. J. Avian. Biol. 38:409-414.

Hyde, J. R., and R. D. Vetter. 2007. The origin, evolution, and diversification of rockfishes of the genus *Sebastes* (Cuvier). Mol. Phylogenet. Evol. 44:790-811.

Ilves, K. L., and E. B. Taylor. 2007. Evolutionary and biogeographic patterns within the smelt genus *Hypomesus* (Pisces: Osmeridae) in the North Pacific Ocean. J. Biogeogr., in press.

Itoh, Y., N. Nakajima, and A. Takemura. 1997. Neogene deformation of the back-arc shelf of Southwest Japan and its impact on the palaeoenvironments of the Japan Sea. Tectonophysics 281:71-82.

Kai, Y., K. Nakayama, and T. Nakobo. 2003. Molecular phylogenetic perspective on speciation in the genus *Sebastes* (Scorpaenidae) from the northwest Pacific and the position of Sebastes within the subfamily Sebastinae. Ichthyol. Res. 50:239-244.

Kennett, J. P. 1982. Marine geology. Prentice-Hall, Englewood Cliffs, NJ.

Kontula, T., and R. Väinölä. 2003. Relationships of Palearctic and Nearctic ‘glacial relict’ *Myoxocephalus* sculpins from mitochondrial DNA data. Molec. Ecol. 12:3179-3184.

Klyukanov, V. A. 1975. Taxonomy and evolutionary relationships between the smelts of the genera *Osmerus* and *Hypomesus* and their dispersion. Zool. J. 54: 590-596.

Lang, P., F. Dane, T. L. Kubisiak, and H. Huang. 2007. Molecular evidence for an Asian origin and a unique westward migration of species in the genus *Castanea* via Europe to North America. Mol. Phyl. Evol. 43:49-59.

Langley, C. H., and W. Fitch. 1974. An estimation of the constancy of the rate of molecular evolution. J. Mol. Evol. 3:161-177.

Lee, Y.-H. 2003. Molecular phylogenies and divergence times of sea urchin species of Strongylocentrotidae, Echinoida. Mol. Biol. Evol. 20:1211-1221.

Lindberg, G. V. 1953. Principles of the distribution of fishes and the geological history of the far-eastern seas. Akademia Nauk SSR, Ikhthiolog. Komm., Moscow-Leningrad. [In Russian].

Lindstrom, S. C. 2001. The Bering strait connection: dispersal and speciation in boreal macroalgae. J. Biogeogr. 28:243-251.

MacNeil, F. S. 1965. Evolution and distribution of the genus *Mya*, and Tertiary migrations of Mollusca. Prof. Pap. U.S. Geol. Surv. 483G:1-51.

Maddison, D. R., and W. P. Maddison. 2003. MacClade 4.06: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA.

Maddison, W. P., and D. R. Maddison. 2006. Mesquite: a modular system for phylogenetic analysis, Version 1.12. http://mesquiteproject.org

Marincovich, L., Jr., and A. Y. Gladenkov. 1999. Evidence for an early opening of the

Bering Strait. Nature 397:149-151.

Martin, A. P., and S. R. Palumbi. 1993. Body size, metabolic rate , generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA 90:4087-4091.

Mattern, M. Y. 2004. Molecular phylogeny of the Gasterosteidae: the importance of using multiple genes. Mol. Phylogenet. Evol. 30:366-377.

Mattern, M. Y., and D. A. McLennan. 2004. Total evidence phylogeny of Gasterosteidae: combining molecular, morphological and behavioral data. Cladistics 20:14-22.

McAllister, D. E. 1963. A revision of the smelt family, Osmeridae. Bull. Nat. Mus. Can. 191:1-53.

McPhail, J. D. 1997. The origin and speciation of *Oncorhynchus* revisited. Pp. 29-38 in D. J. Stouder, P. A. Bisson, R. J. Naiman, eds. Pacific salmon and their ecosystems: status and future options. Chapman and Hall, New York.

Murata, S., N. Takasaki, M. Saitoh, H. Tachida, and N. Okada. 1996. Details of retropositional genome dynamics that provide a rationale for a genetic division: the distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). Genetics 142:915-926.

Neave, F. 1958. The origin and speciation of *Oncorhynchus*. Trans. Roy. Soc. Canada. 52:25-39.

Oakley, T. H., and R. B. Phillips. 1999. Phylogeny of Salmonine fishes based on growth hormone introns: Atlantic (*Salmo*) and Pacific (*Oncorhynchus*) salmon are not sister taxa. Mol. Phylogenet. Evol. 11:381-393.

Oleinik, A. E. 2001. Eocene gastropods of western Kamchatka – implications for high-latitude North Pacific biostratigraphy and biogeography. Palaeogeogr., Palaeoclimatol., Palaeoecol. 166:121-140.

Ortí, G., M. A. Bell, T. E. Reimchen, and A. Meyer. 1994. Global survey of mitochondrial DNA sequences in the threespine stickleback: evidence for recent migrations. Evolution 48:608-622.

Palumbi, S. R., and B. D. Kessing. 1991. Population biology of the trans-Arctic interchange: mtDNA sequence similarity between Pacific and Atlantic sea urchins. Evolution 45:1790-1805.

Phillips, R. B., and T. H. Oakley. 1997. Phylogenetic relationships among the Salmoninae based on their nuclear and mitochondrial sequences. Pp. 145-162 in T. D. Kocher and C. A. Stepien, eds. Molecular systematics of fishes. Academic Press, San Diego, CA.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817-818.

Rambaut, A., and A. J. Drummond. 2004. Tracer v1.3, Available from http://beast.bio.ed.ac.uk/Tracer

———. 2007. Beauti: Bayesian evolutionary analysis utility. Available from http://beast.bio.ed.ac.uk/BEAUti

Ravelo, A. C., D. H. Andreasen, M. Lyle, A. O. Lyle, and M. W. Wara. 2004. Regional climate shifts caused by gradual global cooling in the Pliocene epoch. Nature 429:263-267.

Rawson, P. D. , and T. J. Hilbish. 1995. Evolutionary relationships among the male and female lineages in the *Mytilus edulis* species complex. Mar. Biol. 12:893-901.

Ree, R. H., B. R. Moore, C. O. Webb, and M. J. Donoghue. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59:2299-2311.

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

———. 2003. r8s: inferring absolute rates of evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301-302.

———. 2004. r8s, version 1.70 user's manual. Available from http://loco.biosci.arizona.edu/r8s/

Sanmartín, I., H. Enghoff, and F. Ronquist. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol. J. Linn. Soc. 73:345-390.

Schweitzer, C. E. 2001. Paleobiogeography of Cretaceous and Tertiary decapod crustaceans of the North Pacific Ocean. J. Paleontol. 75:808-826.

Schultheis, L. M., and M. J. Donoghue. 2004. Molecular phylogeny and biogeography of *Ribes* (Grossulariaceae), with an emphasis on Gooseberries (subg. Grossularia). Syst. Bot. 29:77-96.

Smith, S. A., P. R. Stephens, and J. J. Wiens. 2005. Replicate patterns of species richness, historical biogeography, and phylogeny in holarctic treefrogs. Evolution 59:2433-2450.

Stamford, M. D., and E. B. Taylor. 2004. Phylogeographical lineages of Arctic grayling (*Thymallus arcticus*) in North America: divergence, origins and affinities with Eurasian *Thymallus*. Molec. Ecol. 13:1533-1549.

Stearley, R. F., and G. R. Smith. 1993. Phylogeny of the Pacific trouts and salmons (*Oncorhynchus*) and genera of the family Salmonidae. Trans. Am. Fish. Soc. 122:1-33.

Stepien, C. A., A. K. Dillon, and A. K. Patterson. 2000. Population genetics, phylogeography, and systematics of the thornyhead rockfishes (*Sebastolobus*) along the deep continental slopes of the North Pacific Ocean. Can. J. Fish. Aquat. Sci. 57:1701-1717.

Swofford, D. L. 2002. PAUP*:phylogenetic analysis using parsimony. Ver. 4.0b10. Sinauer Associates, Sunderland, MA.

Taylor, E.B., and J.J. Dodson. 1994. A molecular analysis of relationships and biogeography within a species complex of Holarctic fish (genus *Osmerus*). *Molec. Ecol*. **3**: 235-248.

Uyeno, T., and K. Sakamoto. 1999. *Spirinchus akagii*, a new Miocene smelt from Tottori Prefecture, Japan (Pisces: Osmeriformes: Osmeridae). Bull. Natn. Sci. Mus., Tokyo, Ser. C 25:143-150.

Väinölä , R. 2003. Repeated trans-Arctic invasions in littoral bivalves: Molecular zoogeography of the *Macoma balthica* complex. Mar. Biol. 143:935–946.

Van Houdt, J. K., B. Hellemans, and F. A. M. Volckaert. 2003. Phylogenetic relationships among Palearctic and Nearctic burbot (*Lota lota*): Pleistocene extinctions and recolonization. Mol. Phylogenet. Evol. 29:599-612.

Van Houdt, J. K. J., L. de Cleyn, A. Perretti, and F. A. M. Volckaert. 2005. A mitogenic view on the evolutionary history of the Holarctic freshwater gadoid, burbot (*Lota lota*). Molec. Ecol. 14:2445-2457.

Van Oppen, M. J. H, S. G. A. Draisma, J. L. Olsen, and W. T. Stam. 1995. Multiple trans-Arctic passages in the red alga *Phycodrys rubens*: evidence from nuclear rDNA ITS sequences. Mar. Biol. 123:179-188.

Vermeij, G. J. 1989. Geographical restriction as a guide to the causes of extinction: the case of the cold northern oceans during the Neogene. Paleobiology 15:335-356.

———. 1991. Anatomy of an invasion: the trans-Arctic interchange. Paleobiol. 17:281-307.

Wilson, M.V. H., and R. R. G. Williams. 1991. New Paleocene genus and species of smelt (Teleostei: Osmeridae) from freshwater deposits of the Paskapoo Formation, Alberta, Canada, and comments on osmerid phylogeny. J. Vertebr. Paleontol. 11:434-451.

Wolfe, J. A. 1975. Some aspects of plant geography of the northern hemisphere during the late Cretaceous and Tertiary. Ann. Miss. Bot. Gard. 62:264-279.

Yang, Z., and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23:212- 226.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups. 2001. Trends, rhythms, and aberrations in global climate 65 ma to present. Science 292:686-693.

CHAPTER 6: GENERAL DISCUSSION

6.1 GENERAL OVERVIEW OF CHAPTERS 2-5

The overall aim of my thesis was to address longstanding issues in the systematics and biogeography of the northern hemisphere smelts (Osmeridae) and place the results in a Holarctic biogeographic context. Molecular phylogenetic analysis of mtDNA and nDNA sequences was used to examine the genetic relationships of two putative species of *Hypomesus* (Chapter 2), reconstruct the phylogeny of *Hypomesus* (Chapter 3) and the Osmeridae (Chapter 4), and estimate divergence dates within the Osmeridae (Chapter 5). Two methods of biogeographic reconstruction (parsimony step-matrices and maximum likelihood modeling [ML]) were applied to the phylogenetic results of Chapters 3 and 5 to examine biogeographic patterns and processes in the North Pacific genus *Hypomesus*, and Holarctic family Osmeridae, respectively, which were then compared to patterns seen in other similarly distributed taxa. As seen in many other lineages, biogeographic analysis of the molecular phylogeny suggested the mid-Miocene cooling period was important in the generation of trans-Pacific distributions in *Hypomesus* and *Spirinchus*, and the opening of the Bering Seaway contributed to the diversification of *Osmerus*.

The results from Chapter 2 showed that individuals assigned to the newly described freshwater-resident *Hypomesus chishimaensis* (Saruwatari 1997) and

those of the anadromous *H. nipponensis* do not form reciprocally monophyletic clades. In combination with recent morphological work (Sidorov and Pichugin 2004), which examined individuals of *H. chishimaensis* samples from two of the same lakes from which samples were collected for the original identification and found no discernable differences between those samples and representatives of *H. nipponensis*, I concluded that the freshwater populations should not be attributed species-level recognition.

One possible objection to the conclusions from this chapter is that the divergence between *H. chishimaensis* and *H. nipponensis* was so recent that the genes examined have not yet had time to evolve reciprocal monophyly. As discussed in Chapter 2, the freshwater populations on the southernmost Kuril Islands could not have originated prior to the formation of these islands in the Pliocene and may be as recent as the late Pleistocene (10, 000 – 30, 000 years ago) when sea-level regressions could have isolated individuals of *H. nipponensis* in freshwater habitats on the Kuril Islands (Pietsch et al. 2001). The genes sequenced for this study are generally used in studies above the species level and are perhaps not sufficiently sensitive for addressing the question posed in Chapter 2 of whether recently diverged populations are true species. That being said, assigning a species designation is not a trivial matter, and given the lack of both morphological evidence for differences between *H. chishimaensis* and *H. nipponensis* (Sidorov and Pichugin 2004), and ecological or behavioural evidence that differentiates the two taxa, the genetic data collected in Chapter 2 cast further doubt on the validity of this species identification, a conclusion that is echoed by Chereshnev et al. (2001) and Sidorov

and Pichugin (2004) based on morphology alone. Future genetic work examining more sensitive markers, such as microsatellites that can provide an indication of the differentiation and amount of gene flow between populations, in combination with ecological or experimental studies examining reproductive isolation may show the freshwater and anadromous forms do in fact show significant differences. With this type of evidence available I would be more amenable to considering *H. chishimaensis* a biological species; however, the possibility that each freshwater population evolved independently must also be taken into consideration if species are defined as monophyletic lineages (cf. Taylor and Bentzen 1993).

The systematics and biogeography of *Hypomesus*, the most species-rich genus of the Osmeridae, were examined in Chapter 3. A generally well-supported phylogeny resulted from combined analysis of two mtDNA and three nDNA genes ((*H. nipponensis, H. olidus*), (*H. japonicus*, (*H. pretiosus*, *H. transpacificus*))); however, analysis of the mtDNA data alone suggested that *H. japonicus* grouped as sister to the *H. nipponensis – H. olidus* clade. Even though this relationship received a high degree of bootstrap (ML) and posterior probability (Bayes) support, statistical comparison of the two topologies using the mtDNA dataset showed they were not significantly different; however, using the nDNA dataset the nDNA topology was significantly better with the nDNA data. This suggests that the differences seen between the datasets were a false conflict and that a misleading picture of *Hypomesus* would have emerged if only mtDNA were analyzed. There remains uncertainty in the placement of *H. olidus* as either sister to *H. nipponensis* or sister

to the (*H. japonicus*, (*H. pretiosus*, *H. transpacificus*)) clade. Although an *H. olidus* – *H. nipponensis* sister relationship is well supported, the alternative topology is not significantly different. Additional sampling of *H. olidus* from the Arctic range of its distribution may yet resolve its placement in the Hypomesus phylogeny.

Parsimony reconstruction of ancestral areas suggested that a western or northern Pacific ancestor for *Hypomesus*, while ML modeling resulted in a more widespread ancestor. The small size of the phylogeny and high degree of homoplasy pose difficulties for the biogeographic reconstruction of this genus in isolation from the other Osmeridae taxa. Divergence date estimations calibrated by assuming the divergence of *H. japonicus* and the ancestor of the two eastern Pacific species was initiated by the formation of the Sea of Japan in the mid-Miocene (Itoh et al. 1997) suggested the genus arose in the early-mid Miocene and that the divergences between *H. nipponensis* and *H. olidus* and *H. japonicus* and the eastern Pacific species were coincident with the mid-Miocene cooling period, an event suggested to have been important in generating the distributions of numerous North Pacific marine taxa (e.g. Estes and Steinberg 1988; Collins et al. 1996; Kai et al. 2003).

The calibration point used in dating this phylogeny was merely a hypothesis, as no fossil *Hypomesus* have been identified and there was no osmerid-specific molecular clock available. The possibility that this hypothesis could be rejected based on future fossil evidence or divergence estimation based on a wider phylogeny was discussed in Chapter 3; however, the divergence times estimated from the Osmeridae phylogeny in Chapter 5 are at least consistent with this

estimate. The results from Chapter 5 suggest *H. japonicus* and (*H. pretiosus*/*H. transpacificus*) diverged around 12 mya, which slightly post-dates the origin of the Sea of Japan. There is no evidence that this event initiated the divergence between the lineages, although the timing of the split leaves open the possibility that the ancestral *H. japonicus* invaded this basin as it became available habitat, thereby facilitating allopatric divergence. At the present time this is merely speculation. The divergence time is also synchronous with the mid-Miocene cooling, which in combination with the opening of the Japan Sea basin, or alone, may explain the trans-Pacific disjunction seen in current distributions.

Short of finding *H. japonicus* or *H. japonicus*-like fossil evidence from the Sea of Japan that date to shortly after its formation, it is difficult to speculate what types of evidence could determine the role of this basin in the trans-Pacific distribution within *Hypomesus*. If the Japan Sea's formation was a pivotal event in *Hypomesus* evolution, then presumably it would have played a similar role for other marine taxa in the region. To my knowledge no such data are available, although a comparative biogeographic approach could provide insight into the relative role of this basin in the evolution of North Pacific taxa.

The aim of Chapter 4 was to reconstruct the phylogeny of the Osmeridae from sequence data from six genes (three mtDNA, three nDNA) and compare the results to the eight previous morphology-based hypotheses (Johnson and Patterson 1996, and references therein). The mitochondrial data yielded generally unresolved phylogenies; however, analysis of the nuclear and all data combined datasets

produced well-supported phylogenies at the intergeneric level. Statistical tests comparing the molecular topology to the earlier hypotheses based on morphological data showed the molecular tree fit the molecular data significantly better than all earlier proposals.

There is still remaining uncertainty about the relationships in several parts of the Osmeridae phylogeny. First, *Mallotus villosus*, the most geographically widespread species, was placed either sister to all other Osmeridae genera or as sister to the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) clade with different genes and reconstruction methods. Additional geographic and genetic sampling and more rigorous phylogeographic analysis of this Holarctic taxon may clarify its relationship to the other genera. Second, the interrelationships of the three polytypic genera are not fully resolved. As also seen in Chapter 3, within *Hypomesus*, the position of *H. olidus* as either sister to *H. nipponensis* or sister to the (*H. japonicus*, (*H. pretiosus*, *H. transpacificus*)) clade remains unclear. Although there is generally high support for the former topology from analysis of the nDNA and all DNA datasets, the two alternative topologies are not significantly different. The relationships among the species of *Spirinchus* and *Osmerus* were also not well supported by analysis of the genes included in this study.

These remaining questions in Osmeridae systematics may be resolved by additional geographic sampling of the Holarctic *M. villosus*, *O. dentex* and from Arctic populations of *H. olidus*. Additional sampling of the genome by sequencing other unlinked nuclear loci would help confirm the intergeneric relationships generated in Chapter 4 and may also clarify uncertain relationships within the

phylogeny. The divergences in *Osmerus* in particular appear to have occurred over a relatively short time period (~ 1 million years; Chapter 5); therefore, the rate of evolution of any additional loci chosen for study should be high enough such that synapomorphies could accrue over this timeframe.

Mapping the morphological characters from several earlier studies of Osmeridae interrelationships (McAllister 1963; Begle 1991; Wilson and Williams 1991; Johnson and Patterson 1996; Patterson and Johnson 1997) onto the molecular phylogeny revealed numerous synapomorphies that define the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) clade, five traits are shared by (*Thaleichthys*, (*Allosmerus*, *Spirinchus*)), and one trait defines (*Mallotus*, (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*)))). Of the included characters, none is uniquely shared by *Allosmerus* and *Spirinchus*.

A closer examination of Johnson and Patterson's (1996) and Patterson and Johnson's (1997) analyses, which differed from my molecular results in the placement of *Plecoglossus* and the Salangidae and positioned *Spirinchus* and *Thaleichthys* as sister to *Allosmerus*, revealed (1) no unambiguous synapomorphies separate *Hypomesus* from the other taxa, (2) *Mallotus* and the Salangidae share three character states, while *Spirinchus* and *Thaleichthys* share two character states, and (3) when *Plecoglossus* and the Salangidae as sister taxa, as suggested by my molecular analyses, they do not share any unique traits. From this examination I concluded that homoplasy is a common phenomenon in Osmeridae character evolution, an observation also made by previous authors and suggested that further morphological examination of *Allosmerus*, *Spirinchus*, *Plecoglossus* and

the salangids may yet reveal shared traits that support the sister relationships from the molecular phylogeny. Additional independent nuclear gene sequences of salangid species will aid in confirming the sister relationship between the Salangidae and *Plecoglossus*.

In Chapter 5, I utilized the phylogenetic information from Chapter 4 to estimate divergence times and reconstruct the biogeographic history of the Osmeridae. Three fossils were used to calibrate the Osmeridae phylogeny and the ML and Bayesian methods employed yielded similar divergence estimates for most nodes. Biogeographic reconstruction of extant taxa suggested a western or northern Pacific Osmeridae ancestor (parsimony step-matrix) or a widespread ancestor found in the eastern and western regions of the Pacific and Atlantic oceans, but absent from the intervening northern Pacific and Arctic regions (ML modeling). This latter result was surprising given the model of area connections, which precluded dispersal between the Arctic and Atlantic until the Oligocene and between the Arctic and northern Pacific until the late Miocene. I discussed the possibility that there may be a shortcoming in the analysis in that it does not permit areas to come into existence at a particular time. From the analysis without fossil taxa, *Hypomesus* was inferred to have a western or northern distribution from the parsimony model and a western Pacific origin using the ML method. Overall, my analyses suggest a Pacific basin origin for the Osmeridae.

Including fossil taxa into the parsimony reconstruction method yielded different results: both the Osmeridae and *Hypomesus* had a northern Pacific or

northern Pacific and Arctic distribution (it was not possible to implement the ML model as DNA data were not available for divergence date estimation). Two possible explanations to account for the Holarctic distribution with evidence of North Atlantic fossils dating to the Oligocene when trans-Arctic dispersal was prevented by the Bering landbridge: dispersal either occurred along the southern margin of North America through the Panama Seaway, or the Osmeridae lineage is much older than current data suggest.

Several estimated divergence dates corresponded to climatic and geologic events known to have had significant effects on the evolution of other Holarctic taxa, including the development of a cool water environment by the Eocene, the mid-Miocene cooling period, and the late-Miocene opening of the Bering Seaway. Trans-Pacific sister relationships in *Hypomesus* and *Spirinchus* are coincident with the mid-Miocene cooling period, which is consistent with McAllister's (1963) compression hypothesis to explain such disjunctions. The divergences within the Holarctic *Osmerus* occurred soon after the opening of the Bering Seaway, supporting the hypothesis of Taylor and Dodson (1994) that this event was significant in the evolution of this genus.

Further understanding of the biogeography of the Osmeridae may be obtained by phylogeographic analysis of individual species, particularly the widespread *Mallotus villosus* and *Osmerus dentex* and additional testing of biogeographic hypotheses of *Osmerus* requires resolution of their phylogenetic relationships. A recurrent theme in Chapter 5 is the need for more comparative data on Holarctic marine taxa that predate the Cenozoic opening of the Bering Seaway.

Available genetic databases for the Salmonidae and Gasterosteidae in particular provide an excellent starting point for direct comparison with the Osmeridae.

6.2 COMMENTS ON OSMERIDAE TAXONOMY

With a dated osmerid phylogeny available (Chapter 5), a more explicit examination of the alternative biological nomenclature systems discussed in Chapter 1 is now possible. Figure 6.1 depicts the osmerid phylogeny scaled to time using Bayesian estimates of divergence dates (Chapter 5). Avise and John's (1999) temporal-banding scheme for identifying standardized taxon ranks is indicated alongside the phylogeny. In order to conform to this scheme, almost all species would require a name change. The only divergence that falls into the 'genus' timeframe (2 – 5 million years ago [mya]) is that between *Hypomesus pretiosus* and *H. transpacificus* (Fig. 6.1). All of the polytypic genera predate this time period, meaning that their species' names would need revision. *Hypomesus* would be elevated to subfamily, and most other monophyletic groupings diverged during the 'tribe' period of the Miocene, meaning that numerous additional ranks would be required in order to categorize each node. The situation is even more convoluted for the (*Osmerus*, (*Thaleichthys*, (*Allosmerus*, *Spirinchus*))) [OTAS] clade due to the number of divergences that fall within the 'tribe' timeframe (Fig. 6.1). Overall, the classification scheme would become extremely complicated and require multiple name changes, which clearly illustrates one of the main objections to a nomenclature system with ranked names: name changes are required even when

monophyletic groups remain constant. In the molecular osmerid phylogeny all of the polytypic genera are monophyletic, yet their species would all require taxonomic revision as the genera are now considered superfamilies or tribes. This point was also demonstrated by Hibbett and Donoghue (1998) using a phylogeny of two fungal genera under Hennig's (1966) proposal of defining rank by absolute age. By contrast, using the PhyloCode system (Cantino and de Queiroz 2006), all of the osmerid species names would remain as they are because names are attached to nodes that define monophyletic groups and all Osmeridae genera are monophyletic based on my molecular analyses. Further, additional nodes, such as those joining *H. pretiosus* and *H. transpacificus* and the OTAS clade, respectively, could be given names. Because names should only be attached to well supported monophyletic groups, it would be unwise to name the node at the *Mallotus* – OTAS junction as the placement of *Mallotus* in the phylogeny remains uncertain.

With respect to the question of whether the Salangidae and Plecoglossidae should be nested within the Osmeridae, according to the scheme in Fig. 6.1 the Osmeridae (31.9 mya) just barely postdate the cutoff for family-level designation (33 mya); however, the divergence between the Osmeridae and Plecoglossidae dates to the mid-Paleocene, which is at the rank of superfamily. Thus, under this scheme, the Salangidae and Plecoglossidae should not be nested within the Osmeridae as suggested by Fu et al. (2005). Again, using the PhyloCode, this question is irrelevant. Each of the 'families' are monophyletic according to current data, therefore, they would all retain their respective names. Additional names could be attached to the Plecoglossidae-Salangidae node and the node joining this clade to

the Osmeridae, with sufficient evidence to support these relationships; however, because names are rank-free and do not require particular suffixes, the level of classification does not need to be taken into consideration.

With everything considered, the likely reason for why the temporal-banding scheme suggested by Avise and Johns (1999) has not been implemented for many taxa is that it further complicates the classification and naming of organisms. Although any change to the current system of biological classification will be an enormous undertaking, the PhyloCode is a workable alternative that will cause relatively little nomenclatural disruption, especially in relation to the Linnaean alternatives suggested by Hennig (1966) and Avise and Johns (1999), and it also more accurately represents the evolutionary perspective adopted by most systematists. One component lacking from the current PhyloCode draft is the ability to compare diversity across different groups, which was the main motivation underlying Avise and John's (1999) proposal. This could be remedied by combining the system with the timeclip procedure suggested by Avise and Mitchell (2007), which would simply involve fixing a divergence date (if available) to every named node in a phylogeny. With a well-designed database it would be possible to search for nodes that date to a particular time period and directly compare species diversities or number of divergences postdating that point in time.

6.3 FUTURE RESEARCH

Multiple suggestions for future research have been discussed in section 6.1 with respect to each chapter. Additional possibilities include more detailed phylogeographic analysis of each Osmeridae species. In this thesis, samples were only available from a limited portion of often wide geographic ranges. Thus, further sampling of each species would not only function to corroborate (or call into question) the phylogenetic and biogeographic results in the preceding chapters, but may also help address the areas of uncertainty and provide more fine-scale data about the evolutionary history of the individual species, such as area of origin and direction of dispersal. There is also a clear need for ecological data on most Osmeridae species. Several species have undergone marked population declines over the last half-century (*Thaleichthys pacificus* along the western coast of North America, and *Spirinchus thaleichthys* and *H. pretiosus* in the San Francisco estuary), and it is possible that other species are facing similar yet unknown threats due to the lack of basic data.

Additional unresolved questions in Osmeridae evolution include the status of freshwater populations of the longfin smelt (*Spirinchus thaleichthys*) and post-glacial dispersal patterns of the European smelt (*Osmerus eperlanus*). The longfin smelt is primarily anadromous; however, there are apparent freshwater-resident populations in two British Columbia lakes (Pitt and Harrison) and Lake Washington, WA, USA. The populations in British Columbia are of particular interest as they have evolved into a 'dwarf' ecotype relative to the 'normal' sized anadromous fishes, and presumably invaded these basins in proglacial lakes at the end of the Pleistocene

glaciations. By contrast, the Lake Washington population is thought to be of relatively recent origin (~100 years), colonizing the lake following the construction of canals connecting the lake to Puget Sound (T. Quinn, personal communication). Numerous issues with regard to these populations have yet to be addressed: (1) Do the populations in Pitt and Harrison lakes represent independent evolution of the dwarf phenotype, or did this phenotype have a single origin followed by dispersal into each basin?; (2) What is the ancestral lineage of these populations (e.g. Fraser River or dispersal from a more northern refugium)?; (3) What is the relationship between the freshwater British Columbia and Washington populations and the anadromous populations ranging from Alaska to California?; (4) Is the dwarf phenotype a heritable trait, or is it an example of phenotypic plasticity resulting from the freshwater environment?; and, (5) Are the populations truly freshwater residents? The first three questions could be addressed through comparative sequence and/or microsatellite analysis of individuals spanning the geographic range of the species. A complete answer for question (1) would require identification of the 'dwarf' gene(s); however, comparative genetic analysis as described could elucidate whether the lake populations had a single or independent origin. The fourth question would require rearing of 'pure' freshwater and anadromous individuals and crosses between the types in a common environment, and the final suggested question could be answered by chemical profiling of otoliths from the freshwater populations.

A phylogeographic analysis of the European smelt, which also has anadromous and freshwater populations, would nicely complement the vast

literature on the post-glacial colonization of primary freshwater (e.g. Nesbø et al. 1999; Koskinen et al. 2000; Kontula and Väinölä 2001; Gum et al. 2005) and anadromous (e.g. García-Marín et al. 1999; Koljonen et al. 1999; Nilsson et al. 2001) fishes in Europe, as well as work on the colonization of the Baltic Sea region by marine fishes (Jørgensen et al. 2005). This could be accomplished through mitochondrial sequence and microsatellite analyses of samples spanning the species' range to address the question of whether the anadromous *O. eperlanus* had a different post-glacial dispersal pattern in Europe than the patterns seen in freshwater organisms. Similar to the longfin smelt, there are numerous freshwater populations, at least one of which also appears as a dwarf form relative to the anadromous populations (Lake Peipsi, Estonia), about which questions 2 – 5 above are also unanswered.

One of the main conclusions of my thesis was that there is a great need for additional comparative biogeographic data. As such, in addition to a focus on the unresolved issues surrounding Osmeridae systematics and biogeography, I could imagine a research programme with a goal of generating a multilocus genetic sequence database for a suite of marine Holarctic taxa that could then be examined for patterns of simultaneous vicariance (or dispersal) as well as for processes unique to individual taxa. This could help resolve basin-wide questions in systematics and biogeography for a diverse set of taxa and contribute to global biogeographic understanding.

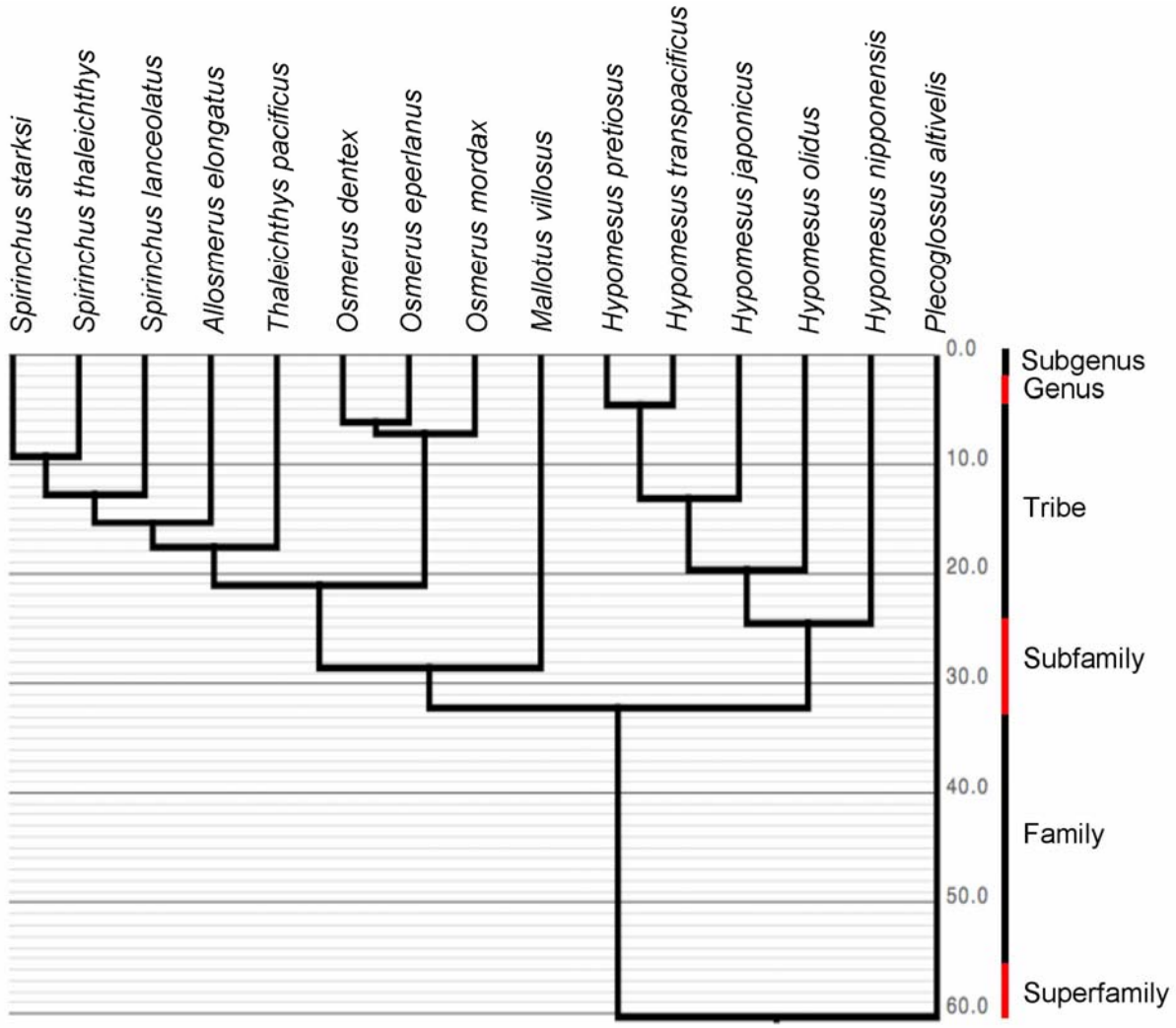


Figure 6.1 Dated Osmeridae phylogeny based on Bayesian and maximum likelihood analysis of sequences from six genes (three mtDNA, three nDNA) showing the temporal-banding scheme for identifying standardized taxonomic rank (Avise and Johns 1999). Divergence dates estimated using Beast (Drummond and Rambaut 2006).

6.4 REFERENCES

Avise, J. C., and G. C. Johns. 1999. Proposal for a standardized temporal scheme of biological classification for extant species. Proc. Natl. Acad. Sci. 96:7358-7363.

Avise, J. C., and D. Mitchell. 2007. Time to standardize taxonomies. Syst. Biol. 56:130-133.

Begle, D. P. 1991. Relationships of osmeroid fishes and the use of reductive characters in phylogenetic analysis. Syst. Zool. 40:33-53.

Cantino, P. D., and K. de Queiroz. 2006. *PhyloCode*. http://ohiou.edu/phylocode

Chereshnev, I. A. A. V. Shestakov, and S. V. Frolov. 2001. On the systematics of species of the genus *Hypomesus* (Osmeridae) of Peter the Great bay, Sea of Japan. Russ. J. Mar. Biol. 27:296-302.

Collins, T. M., K. Frazer, A. R. Palmer, G. J. Vermeij, and W. M. Brown. 1996. Evolutionary history of northern hemisphere *Nucella* (Gastropoda, Muricidae): molecular, morphological, ecological, and paleontological evidence. Evolution 50:2287-2304.

Drummond, A. J., and A. Rambaut. 2006. BEAST v1.4, Available from http://beast.bio.ed.ac.uk/

Estes, J. A. and P. D. Steinberg. 1988. Predation, herbivory, and kelp evolution. Paleobiol. 14:19-36.

Fu, C., J. Luo, J. Wu, J. A. López, Y. Zhong, G. Lei, and J. Chen. 2005. Phylogenetic relationships of salangid fishes (Osmeridae, Salanginae) with comments on phylogenetic placement of the salangids based on mitochondrial DNA sequences. Mol. Phylogenet. Evol. 35:76-84.

García-Marín, J.-L., F. Utter, and C. Pla. 1999. Postglacial colonization of brown trout in Europe based on distribution of allozyme variants. Heredity 82:46-56.

Gum, B., R. Gross, and R. Kuehn. 2005. Mitochondrial and nuclear DNA phylogeography of European grayling (*Thymallus thymallus*): evidence for secondary contact zones in central Europe. Molec. Ecol. 14:1707-1725.

Hennig, W. 1966. Phylogenetic systematics. translated by D.D. Davis and R. Zangerl. University of Illinois Press, Urbana.

Hibbett, D. S., and M. J. Donoghue. 1998. Integrating phylogenetic analysis and classification in fungi. Mycologia 90:347-356.

Itoh, Y., N. Nakajima, and A. Takemura. 1997. Neogene deformation of the back-arc shelf of Southwest Japan and its impact on the palaeoenvironments of the Japan Sea. Tectonophysics 281:71-82.

Johnson, G. D., and C. Patterson. 1996. Relationships of lower Euteleostean fishes, Pp. 251-332 in M. Stiassny, L. J. Parenti, L. R. Johnson, and G. David, eds. Interrelationships of fishes. Academic Press, Toronto.

Jørgensen, H. B., M. H. Hansen, D. Bekkevold, D. E. Ruzzante, and V. Loeschcke. 2005. Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. Molec. Ecol. 14:3219-3234.

Kai, Y., K. Nakayama, and T. Nakobo. 2003. Molecular phylogenetic perspective on

speciation in the genus *Sebastes* (Scorpaenidae) from the northwest Pacific and the position of Sebastes within the subfamily Sebastinae. Ichthyol. Res. 50:239-244.

Koljonen, M.-L., H. Jansson, T. Paaver, O. Vasin, and J. Koskiniemi. 1999. Phylogeographic lineages and differentiation pattern of Atlantic salmon (*Salmo salar*) in the Baltic Sea with management implications. Can. J. Fish. Aquat. Sci. 56:1766-1780.

Kontula, T., and R. Väinölä. 2001. Postglacial colonization of Northern Europe by distinct phylogeographic lineages of the bullhead, *Cottus gobio*. Molec. Ecol. 10:1983-2002.

Koskinen, M. T., E. Ranta, J. Piironen, A. Veselov, S. Titov, T. O. Hauden, J. Nilsson, M. Carlstein, and C. R. Primmer. 2000. Genetic lineages and postglacial colonization of grayling (*Thymallus thymallus*, Salmonidae) in Europe, as revealed by mitochondrial DNA analyses. Molec. Ecol. 9:1609-1624.

McAllister, D. E. 1963. A revision of the smelt family, Osmeridae. Bull. Nat. Mus. Can. 191:1-53.

Nesbø, C. L., T. Fossheim, L. A. Vøllestad, and K. S. Jakobsen. 1999. Genetic divergence and phylogeographic relationships among European perch (*Perca fluviatilis*) populations reflect glacial refugia and postglacial colonization. Molec. Ecol. 8:1387-1404.

Nilsson, J., R. Gross, T. Asplund, O. Dove, H. Jansson, J. Kelloniemi, K. Kohlmann, A. Löytynoja, E. Nielsen, T. Paaver, C. R. Primmer, S. Titov, A. Vasemägi, A. Veselov, T. Ost, and J. Lumme. 2001.Matrilinear phylogeography of Atlantic salmon (*Salmo salar* L.) in Europe and postglacial colonization of the Baltic

Sea area. Molec. Ecol. 10:89-102.

Patterson, C., and G. D. Johnson. 1997. The data, the matrix, and the message: comments on Begle's "Relationships of osmeroid fishes". Syst. Biol. 46:358-365.

Pietsch, T. W. K. Amaoka, D. E. Stevenson, E. L. MacDonald, B. K. Urbain, and J. A. López. 2001. Freshwater fishes of the Kuril Islands and adjacent regions. Species Diversity 6:133-164.

Saruwatari, T. J. A. López, and T. W. Pietsch. 1997. A revision of the osmerid genus *Hypomesus* Gill (Teleostei: Salmoniformes), with the description of a new species from the Southern Kuril Islands. Species Diversity 2:59-82.

Sidorov, L. K., and Y. Pichugin. 2004. Morphological traits of lacustrine forms of smelts of the genus *Hypomesus* (Salmoniformes) from the Southern Kurils. J. Ichthyol. 44:433-443.

Taylor, E. B., and P. Bentzen. 1993. Evidence for multiple origins and sympatric divergence of trophic ecotypes of smelt (*Osmerus*) in northeastern North America. Evolution 47:813-832.

Taylor, E.B., and J.J. Dodson. 1994. A molecular analysis of relationships and biogeography within a species complex of Holarctic fish (genus *Osmerus*). *Molec. Ecol*. **3**: 235-248.

Wilson, M.V. H., and R. R. G. Williams. 1991. New Paleocene genus and species of smelt (Teleostei: Osmeridae) from freshwater deposits of the Paskapoo Formation, Alberta, Canada, and comments on osmerid phylogeny. J. Vertebr. Paleontol. 11:434-451.

Appendix 1 Voucher numbers and sample localities for Chapter 2.

Voucher numbers and sample localities for specimens used in molecular analyses. Institutional abbreviations follow Leviton et al. (1985). Numbers in brackets indicate more than one individual from a particular lot. The specimens referred to as *Hypomesus chishimaensis* below were considered as conspecific to *H. nipponensis* in the present study based on a lack of morphological (Sidorov and Pichugin 2004) and genetic (Ilves and Taylor 2007) divergence among the individuals assigned to each taxonomic designation.

Hypomesus chishimaensis, UW 043710 (4), Iturup Island, Kuybyshevskoe Lake, Japan, (cyt*b*: DQ010175-DQ01078; 16S: DQ0101202-DQ0101205; ITS2: DQ010234-DQ0101237; S71: DQ010262-DQ010265; RAG1: DQ010293-DQ010296). *Hypomesus chishimaensis*, UW 041862 (3), Kunashir Island, Lake Serebryanoye, Japan, (cyt*b*: DQ010179-DQ010180; 16S: DQ010206-010208; ITS2: DQ010238-DQ010240; S71: DQ010266-DQ010268; RAG1: DQ010297). *Hypomesus chishimaensis*, UW 041869 (3), Zelionyi Island, stream of Lake Srednoye, Japan, (16S: DQ010209-DQ010211; ITS2: DQ010241-DQ010243; S71: DQ010269-DQ010270). *Hypomesus chishimaensis*, UW 046336 (3), Sakhalin Island, Shlyuzovka River, Russia, (cyt*b*: DQ010181-DQ010182; 16S: DQ010212-DQ010214; ITS2: DQ010244-DQ010246; S71: DQ010271-DQ010273; RAG1: DQ010298). *Hypomesus nipponensis*, BC uncat. (6), Harutori River, Hokkaido, Japan, (cyt*b*: DQ010189-DQ010194; 16S: DQ010221-DQ010226; ITS2: DQ010247-

DQ010250; S71: DQ010274-DQ010279; RAG1: DQ010299-DQ010301). *Hypomesus olidus*, UW 043724 (4), Yavinskoye Lake, Kamchatka, Russia, (cyt*b*: DQ010195-DQ010198; 16S: DQ010227-DQ010230; ITS2: DQ010251-DQ010253; S71: DQ010280-DQ010283; RAG1: DQ010302-DQ010304). *Mallotus villosus*, BC uncat., Trevor Channel, BC, Canada, (cyt*b*: DQ397093; 16S: DQ397094; ITS2: DQ397095; S71: DQ397096; RAG1: DQ397097)

REFERENCES

Leviton, A. E., R. H. Gibbs, E. Heal, and C. E. Dawson. 1985. Standards in herpetology and ichthyology. 1. Standard symbolic codes for institutional resource collections in herpetology and ichthyology. Copeia 1985:802-832.

Ilves, K. L., and E. B. Taylor. 2007. Are *Hypomesus chishimaensis* and *H. nipponensis* (Pisces: Osmeridae) distinct species? A molecular assessment using comparative sequence data from five genes. Copeia 2007:180-185.

Sidorov, L. K., and Y. Pichugin. 2004. Morphological traits of lacustrine forms of smelts of the genus *Hypomesus* (Salmoniformes) from the Southern Kurils. Journal of Ichthyology, 44, 433-443.

Appendix 2 *Hypomesus* GenBank accession numbers.

GenBank accession numbers for the DNA sequences analysed in Chapter 3 for species of *Hypomesus* and the outgroup *Mallotus villosus*. The alphanumeric indicator in parentheses identifies the individual sample, + indicates specimens used for the final analyses and * identifies sequences from Ilves and Taylor (2007). [1] indicates the RAG1 sequence substituted for sample HJ5, [2] the RAG1 substitution for HPBMS2, [3] the RAG1 substitution for HT1, [4] the RAG1 substitution for HT3, and [5] the ITS2 substitution for HO1. [6] indicates the individual sample used to represent each species for the LAGRANGE analysis.

	DNA Region				
Species	Cytb	16S	ITS2	S71	RAG1
Hypomesus chishimaensis [HC1]	DQ010175*	DQ010202*	DQ010234*	DQ010262*	DQ010293*
Hypomesus chishimaensis [HC2]	DQ010176*	DQ010203*	DQ010235*	DQ010263*	DQ010294*
Hypomesus chishimaensis [HC3]	DQ010177*	DQ010204*	DQ010236*	DQ010264*	DQ010295*
Hypomesus chishimaensis [HC4]	DQ010178*	DQ010205*	DQ010237*	DQ010265*	DQ010296*
Hypomesus chishimaensis [HC5]	DQ010179*	DQ010206*	DQ010238*	DQ010266*	---
Hypomesus chishimaensis [HC6]	DQ010180*	DQ010207*	DQ010239*	DQ010267*	DQ010297*
Hypomesus chishimaensis [HC7]	---	DQ010208*	DQ010240*	DQ010268*	---
Hypomesus chishimaensis [HC10]	---	DQ010209*	DQ010241*	DQ010269*	---

Hypomesus chishimaensis [HC11]	---	DQ010210*	DQ010242*	DQ010270*	---
Hypomesus chishimaensis [HC12]	---	DQ010211*	DQ010243*	---	---
Hypomesus chishimaensis [HC18]	DQ010181*	DQ010212*	DQ010244*	DQ010271*	DQ010298*
Hypomesus chishimaensis [HC19]	DQ010182*	DQ010213*	DQ010245*	DQ010272*	---
Hypomesus chishimaensis [HC20]	---	DQ010214*	DQ010246*	DQ010273*	---
Hypomesus japonicus [HJ5]+[6]	DQ010183	DQ010215	DQ010254	DQ010284	---
Hypomesus japonicus [HJ6]	DQ010184	DQ010216	DQ010255	DQ010285	DQ010305
Hypomesus japonicus [HJ7]	DQ010185	DQ010217	DQ010256	DQ010286	DQ010306
Hypomesus japonicus [HJ8]+	DQ010186	DQ010218	DQ010257	DQ010287	DQ010307
Hypomesus japonicus [HJ9]	DQ010187	DQ010219	DQ010258	DQ010288	DQ010308[1]
Hypomesus japonicus [HJ10]+	DQ010188	DQ010220	DQ010259	DQ010289	DQ010309
Hypomesus nipponensis [HN1]+[6]	DQ010189*	DQ010221*	DQ847513	DQ010274*	DQ847515
Hypomesus nipponensis [HN2]	DQ836403	DQ836424	DQ836442	DQ836460	DQ836479
Hypomesus nipponensis [HN3]	DQ836404	DQ836425	DQ836443	DQ836461	DQ836480
Hypomesus nipponensis [HN5]+	DQ010190*	DQ010222*	DQ010247*	DQ010275*	DQ010299*
Hypomesus nipponensis [HN6]	DQ010191*	DQ010223*	DQ010248*	DQ010276*	DQ010300*
Hypomesus nipponensis [HN7]+	DQ010192*	DQ010224*	DQ010249*	DQ010277*	DQ010301*
Hypomesus nipponensis [HN8]	DQ010193*	DQ010225*	DQ010250*	DQ010278*	---

Hypomesus nipponensis [HN9]	DQ010194*	DQ010226*	DQ847514	DQ010279*	---
Hypomesus nipponensis [HN11]	DQ836399	DQ836420	DQ836438	DQ836456	DQ836477
Hypomesus nipponensis [HN12]	DQ836400	DQ836421	DQ836439	DQ836457	DQ836478
Hypomesus nipponensis [HN13]	DQ836401	DQ836422	DQ836440	DQ836458	---
Hypomesus nipponensis [HN14]	DQ836402	DQ836423	DQ836441	DQ836459	---
Hypomesus olidus [HO1]+[6]	DQ010195*	DQ010227*	---	DQ010280*	DQ010302*
Hypomesus olidus [HO2]	DQ010196*	DQ010228*	DQ010251*[5]	DQ010281*	---
Hypomesus olidus [HO3]	DQ010197*	DQ010229*	DQ010252*	DQ010282*	DQ010303*
Hypomesus olidus [HO4]+	DQ010198*	DQ010230*	DQ010253*	DQ010283*	DQ010304*
Hypomesus olidus [HOAK1]	DQ836405	DQ836426	DQ836444	DQ836462	---
Hypomesus olidus [HOAK2]	DQ836406	---	---	DQ836463	---
Hypomesus olidus [HOAK3]	DQ836407	---	---	DQ836464	---
Hypomesus olidus [HOAK4]	DQ836408	---	---	DQ836465	---
Hypomesus pretiosus [HP1]+[6]	DQ836409	DQ836427	DQ836445	DQ836466	DQ836481[2]
Hypomesus pretiosus [HP3b]	DQ010199	DQ010232	DQ010260	DQ010290	DQ010310
Hypomesus pretiosus [HPBMS2]+	DQ836412	DQ836430	DQ836448	DQ836469	---
Hypomesus pretiosus [HPJB1]	DQ836410	DQ836428	DQ836446	DQ836467	---
Hypomesus pretiosus [HPJB2]+	DQ836411	DQ836429	DQ836447	DQ836468	---

Hypomesus transpacificus [HT1]+[6]	DQ836413	DQ836431	DQ836449	DQ836470	---
Hypomesus transpacificus [HT2]	DQ010200	DQ010231	DQ010261	DQ010291	DQ010311[3]
Hypomesus transpacificus [HT3]+	DQ836414	DQ836432	DQ836450	DQ836471	---
Hypomesus transpacificus [HT4]	DQ836415	DQ836433	DQ836451	DQ836472	DQ836482
Hypomesus transpacificus [HT5]	DQ836416	DQ836434	DQ836452	DQ836473	DQ836483
Hypomesus transpacificus [HT7]	DQ836417	DQ836435	DQ836453	DQ836474	DQ836484[4]
Hypomesus transpacificus [HT8]+	DQ836418	DQ836436	DQ836454	DQ836475	DQ836485
Mallotus villosus [MV2]	DQ836419	DQ836437	DQ836455	DQ836476	DQ836486
Mallotus villosus [MVTC2]+ [6]	DQ397093*	DQ397094*	DQ397095*	DQ397096*	DQ397097*

REFERENCES

Ilves, K. L., and E. B. Taylor. 2007. Are *Hypomesus chishimaensis* and *H. nipponensis* (Pisces: Osmeridae) distinct species? A molecular assessment using comparative sequence data from five genes. Copeia 2007:180-185.

Appendix 3 List of sample substitutions for reduced and one individual per species datasets.

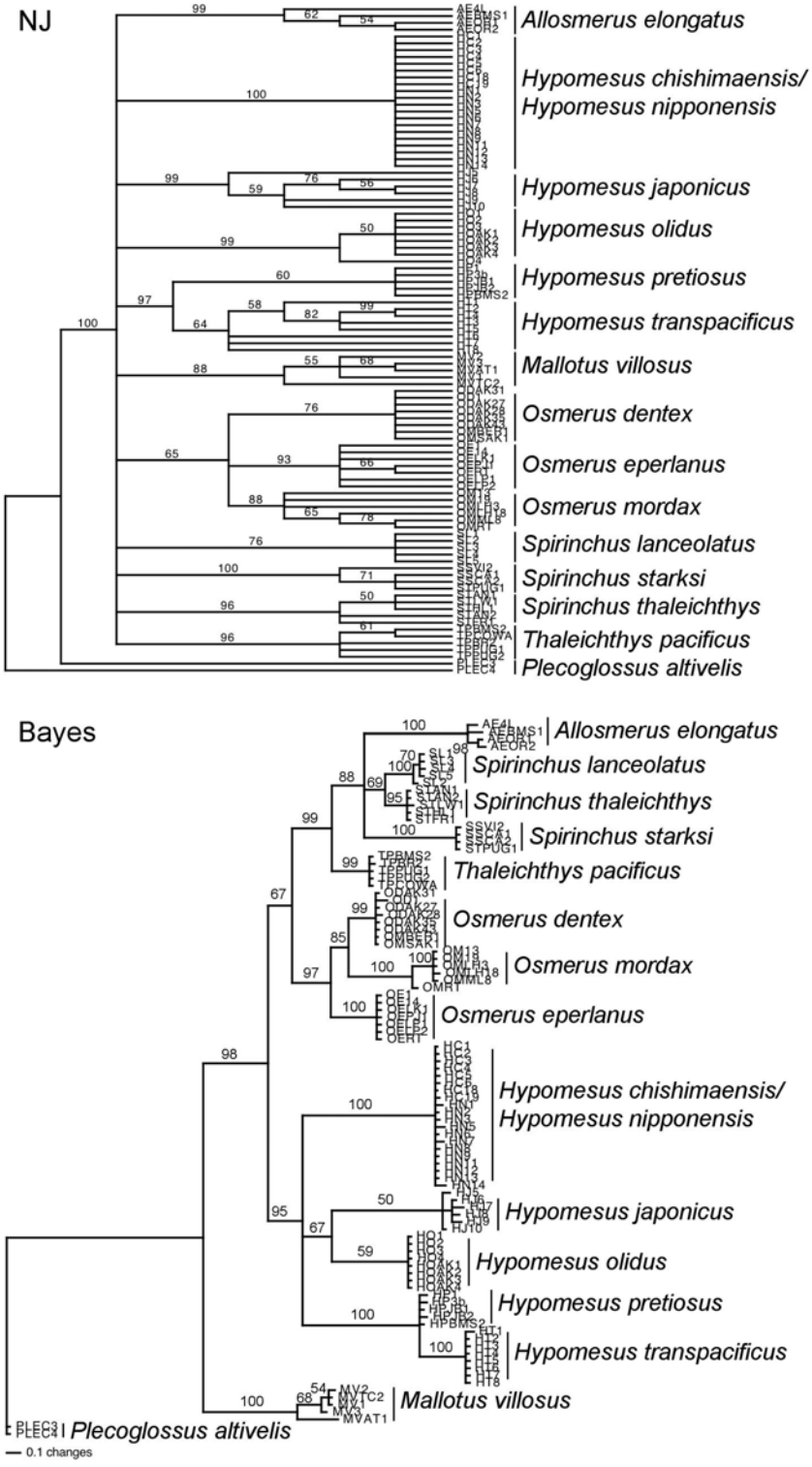
Reduced dataset

Species	Gene	Sample	Substitution
Hypomesus pretiosus	RAG1	HPBMS2	HP3b
H. transpacificus	RAG1	HT1	HT2
H. transpacificus	RAG1	HT3	HT7
Osmerus mordax	16S	OMLH3	OM19
Spirinchus thaleichthys	S71	STAN1	STAN2
Thaleichthys pacificus	ITS2	TPBR2	TPPUG1
T. pacificus	S71	TPBR2	TPPUG2

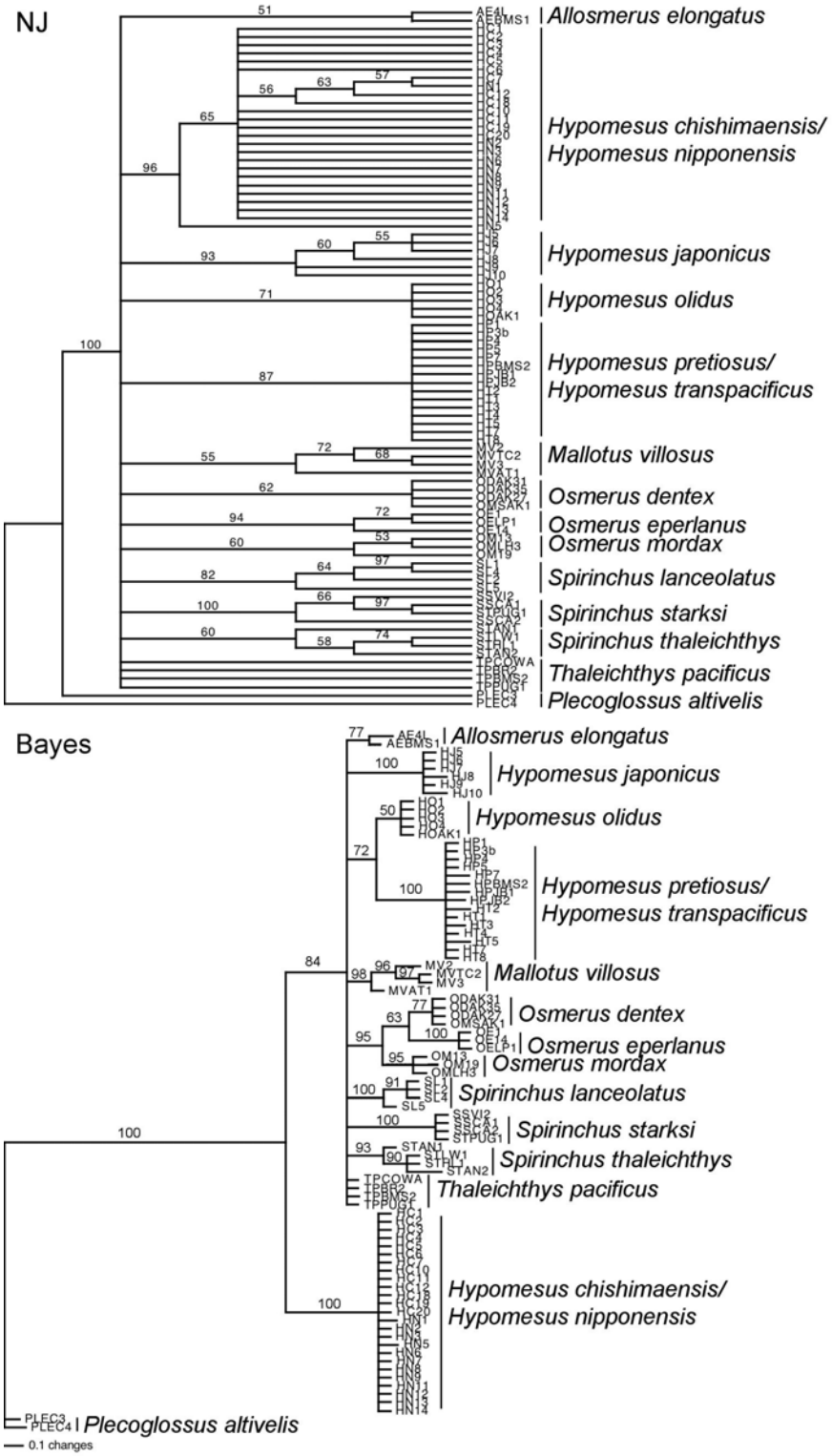
One individual/species dataset

Species	Gene	Sample	Substitution
Allosmerus elongatus	RAG1	AE4L	AEBMS1
Hypomesus pretiosus	S71	HP1	HPJB2
H. transpacificus	RAG1	HT3	HT7
Spirinchus thaleichthys	S71	STAN1	STAN2
Thaleichthys pacificus	S71	TPCOWA	TPBR2

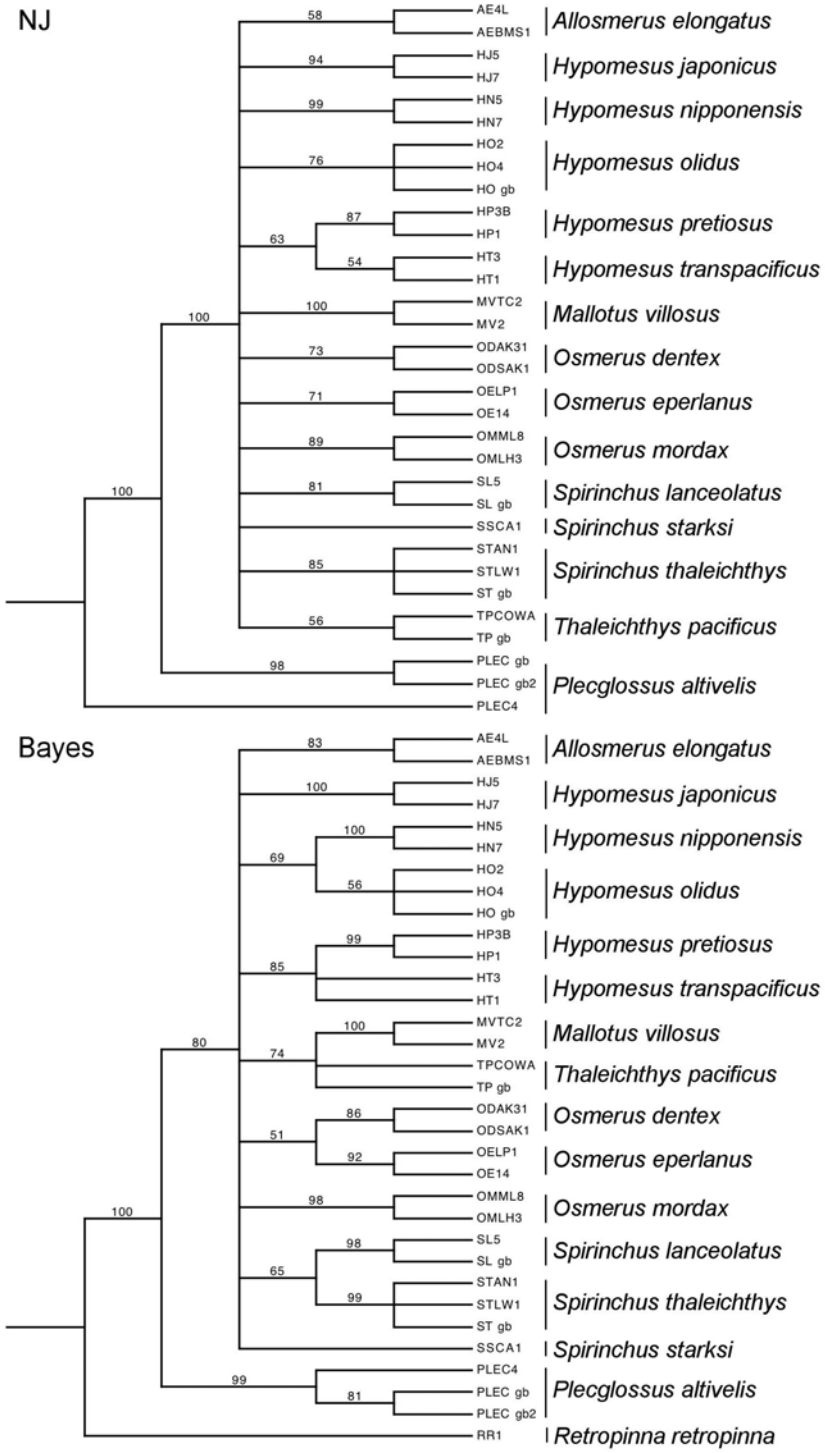
Appendix 4 Cytochrome *b* phylogenies of all Osmeridae samples resulting from neighbour-joining (NJ), and Bayesian (Bayes) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ) or posterior probabilities from a consensus of 18000 trees (Bayes).



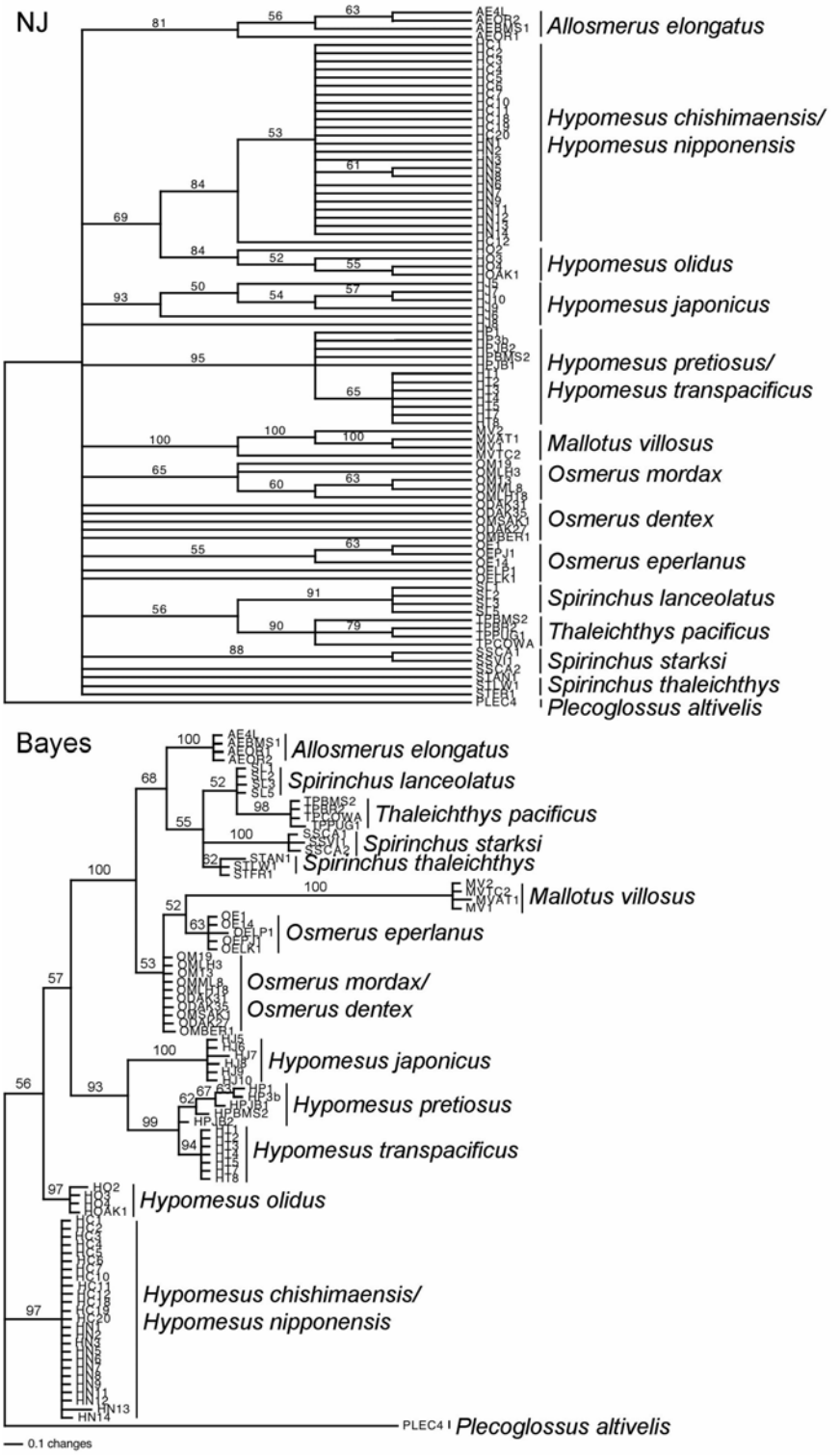
Appendix 5 16S phylogenies of all Osmeridae samples resulting from neighbour-joining (NJ), and Bayesian (Bayes) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ) or posterior probabilities from a consensus of 18000 trees (Bayes).



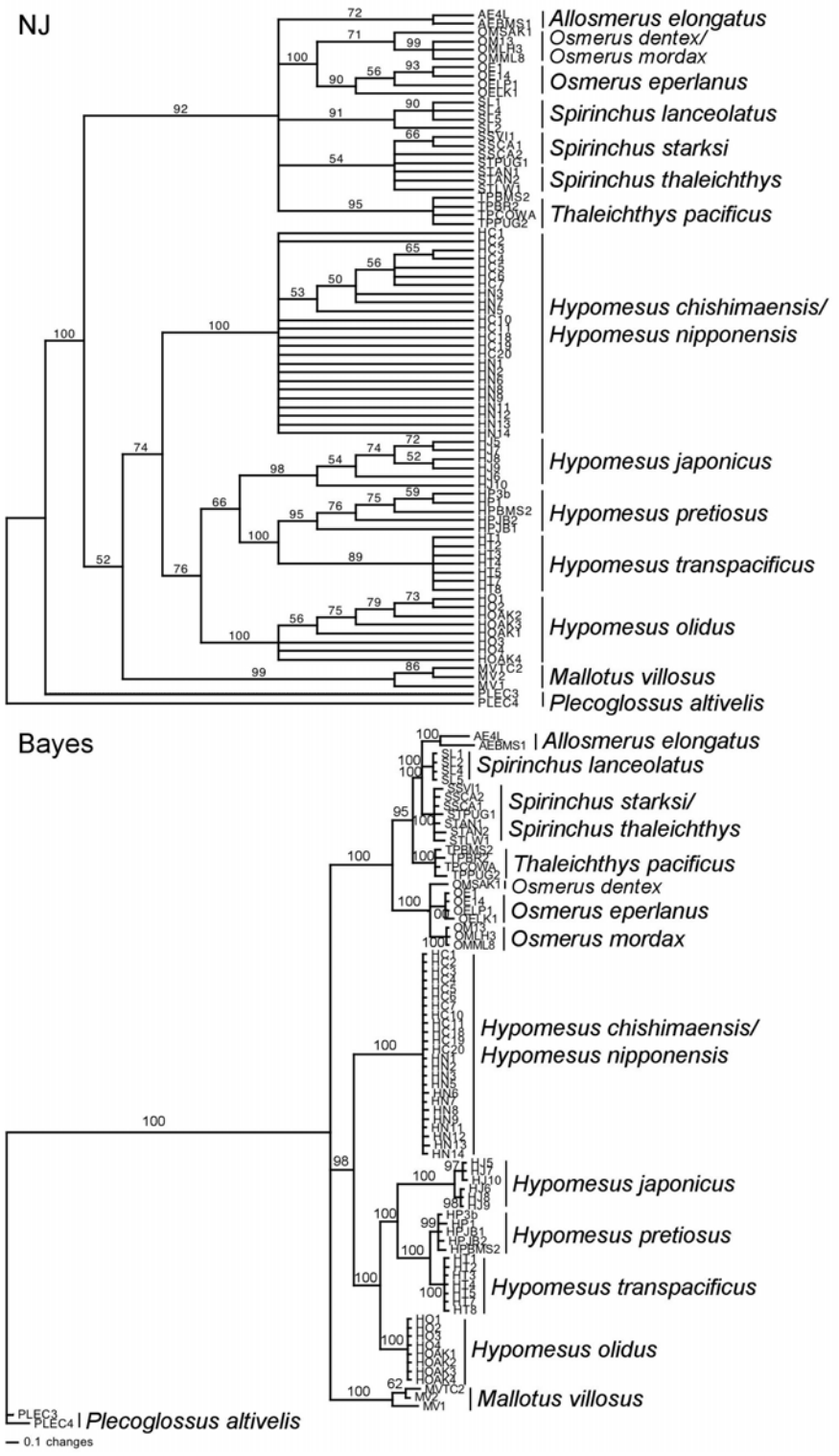
Appendix 6 12S phylogenies of all Osmeridae samples resulting from neighbour-joining (NJ), and Bayesian (Bayes) reconstruction. Bayesian reconstruction includes *Retropinna retropinna* as the outgroup. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ) or posterior probabilities from a consensus of 18000 trees (Bayes).



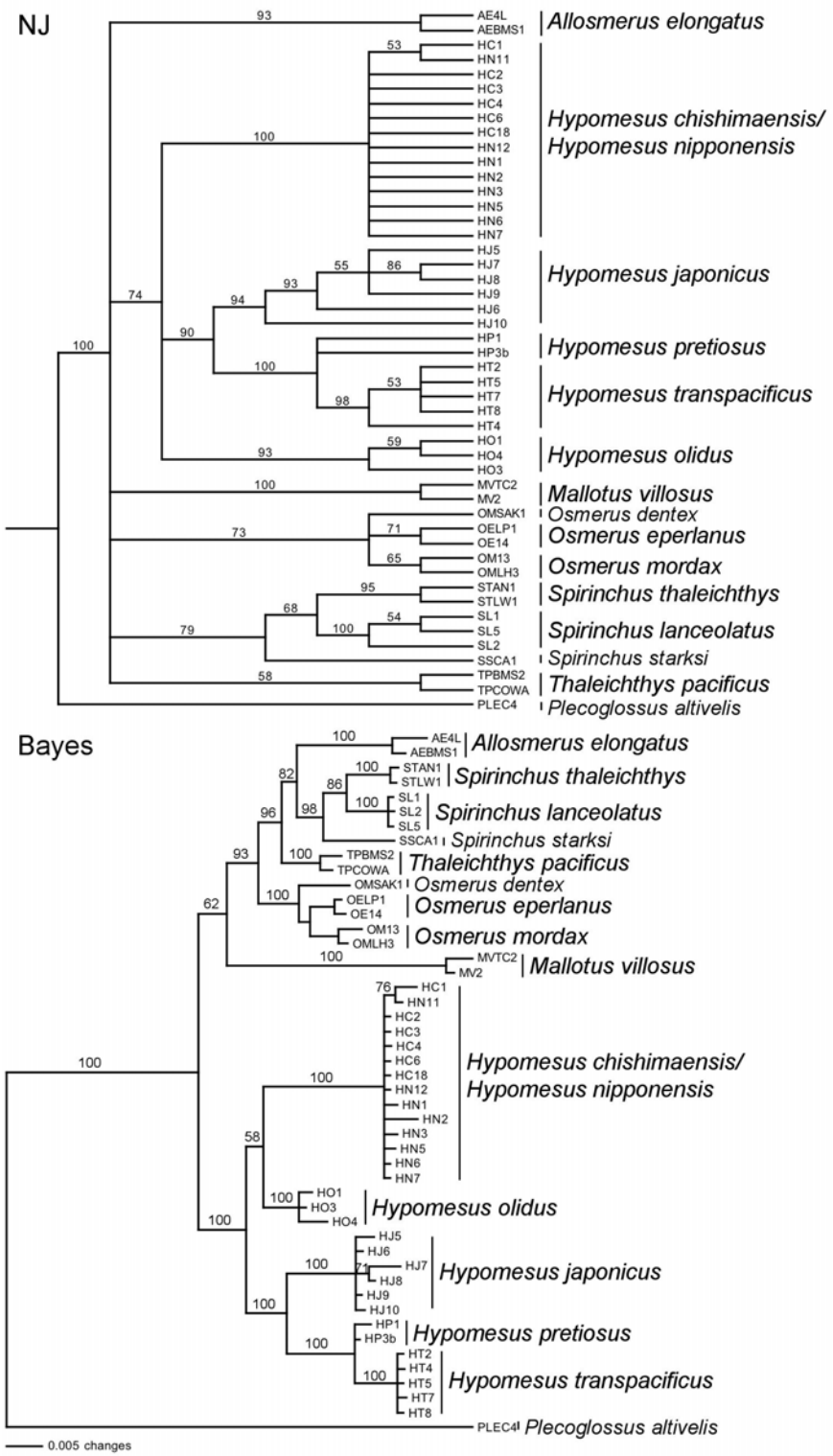
Appendix 7 ITS2 phylogenies of all Osmeridae samples resulting from neighbour-joining (NJ), and Bayesian (Bayes) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ) or posterior probabilities from a consensus of 18000 trees (Bayes).



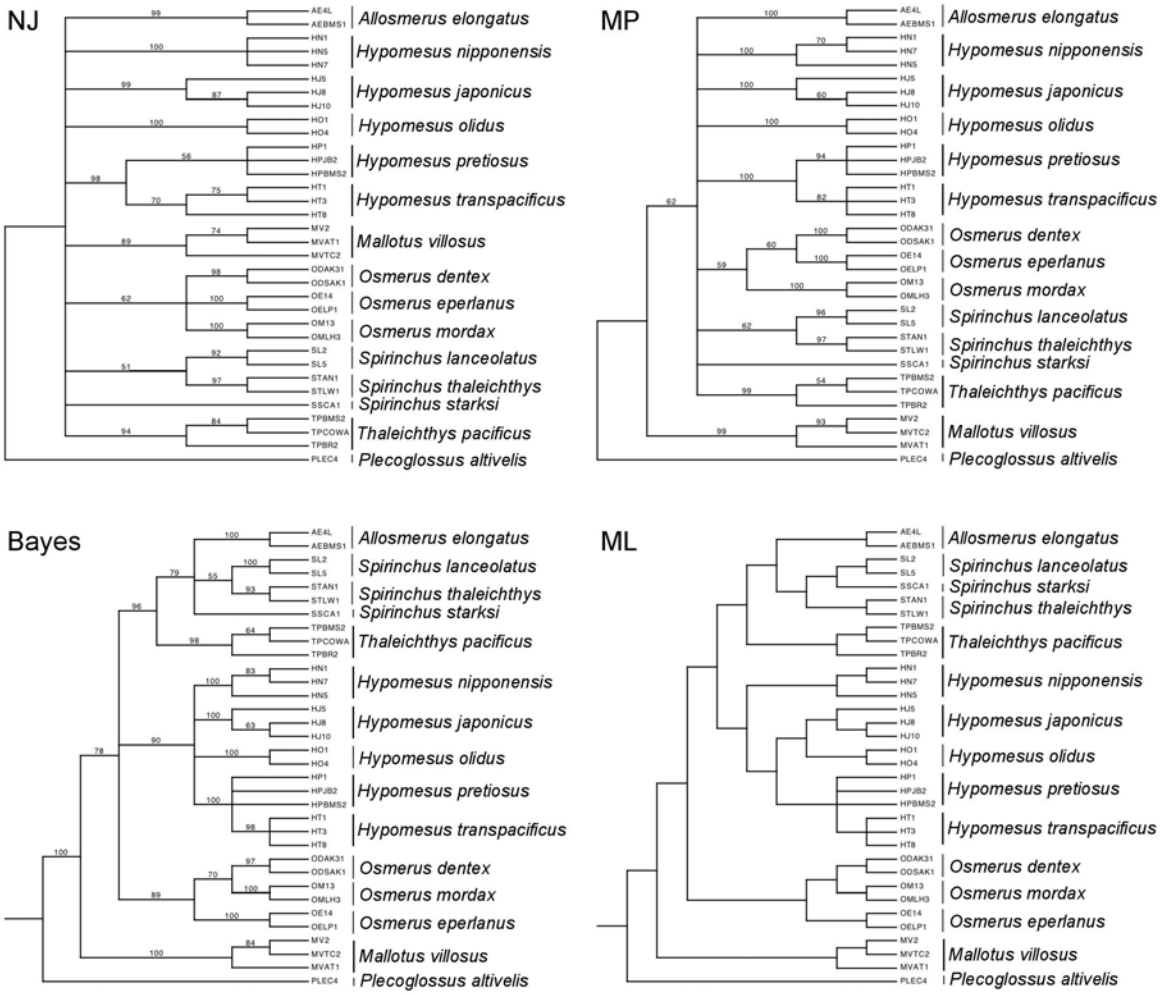
Appendix 8 S71 phylogenies of all Osmeridae samples resulting from neighbour-joining (NJ), and Bayesian (Bayes) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ) or posterior probabilities from a consensus of 18000 trees (Bayes).



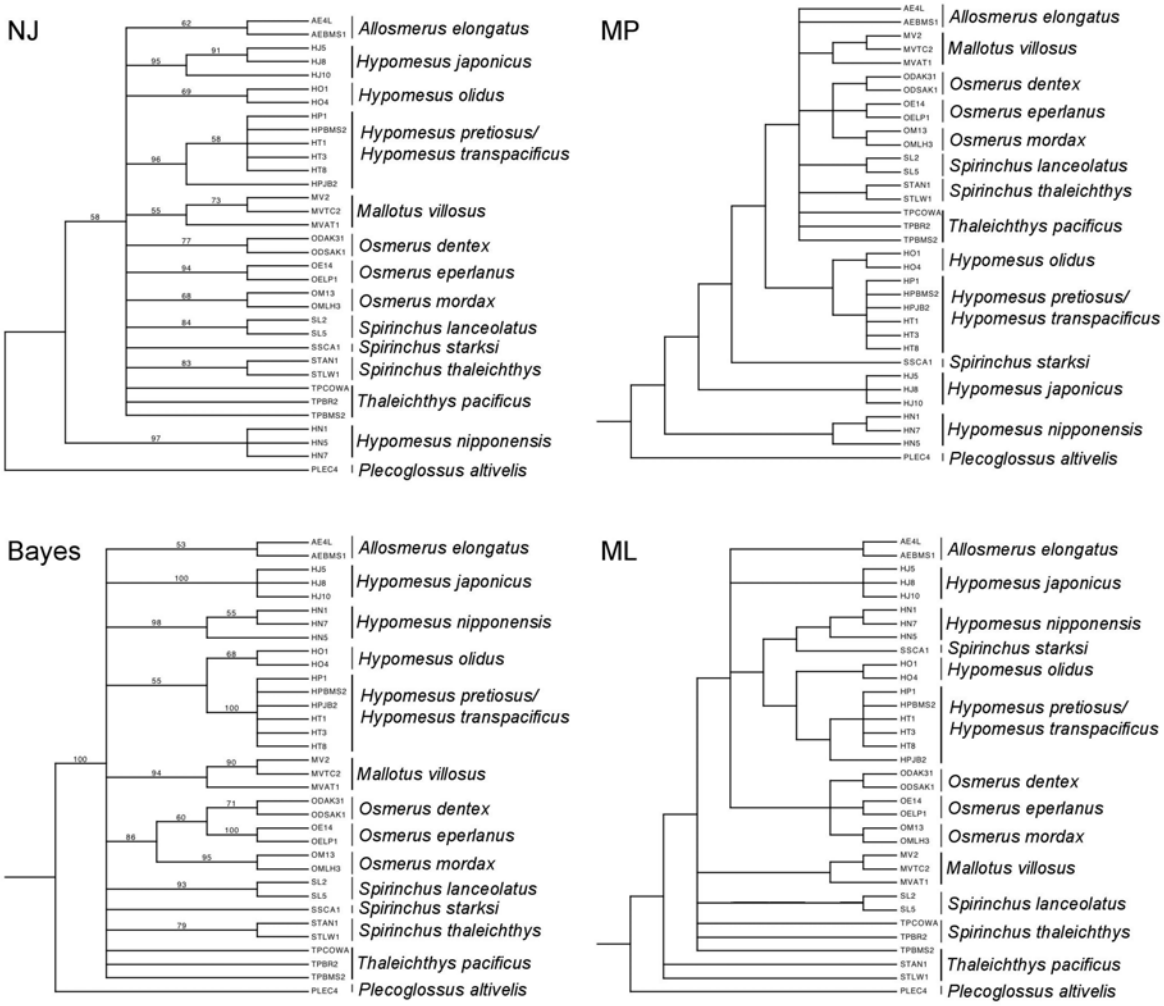
Appendix 9 RAG1 phylogenies of all Osmeridae samples resulting from neighbour-joining (NJ), and Bayesian (Bayes) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ) or posterior probabilities from a consensus of 18000 trees (Bayes).



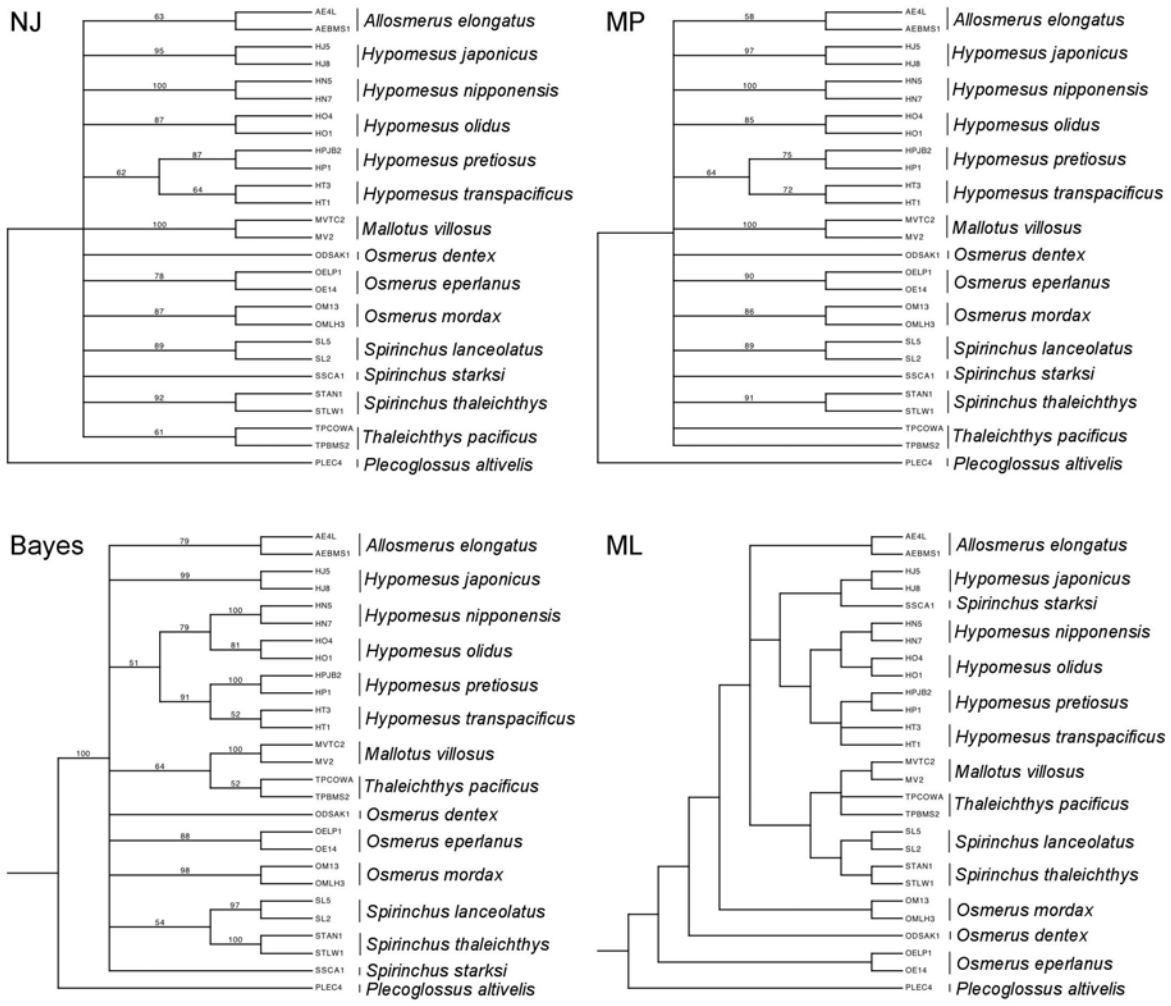
Appendix 10 Reduced dataset Osmeridae cytochrome *b* phylogenies resulting from neighbour-joining (NJ), parsimony (MP), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ, MP) or posterior probabilities from a consensus of 18000 trees (Bayes). ML trees from heuristic searches with 10 replicates of random taxon addition.



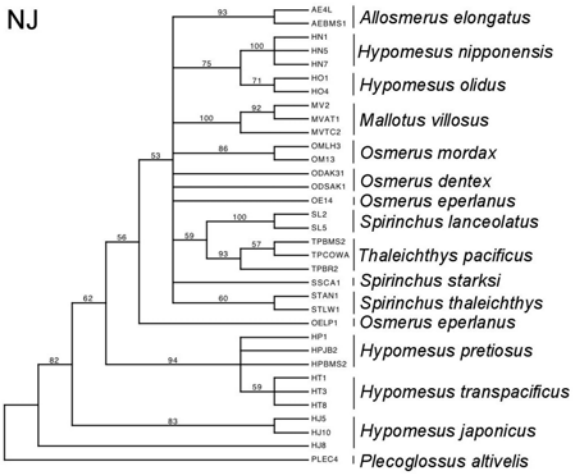
Appendix 11 Reduced dataset Osmeridae 16S phylogenies resulting from neighbour-joining (NJ), parsimony (MP), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ, MP) or posterior probabilities from a consensus of 18000 trees (Bayes). ML tree from heuristic searches with 10 replicates of random taxon addition.

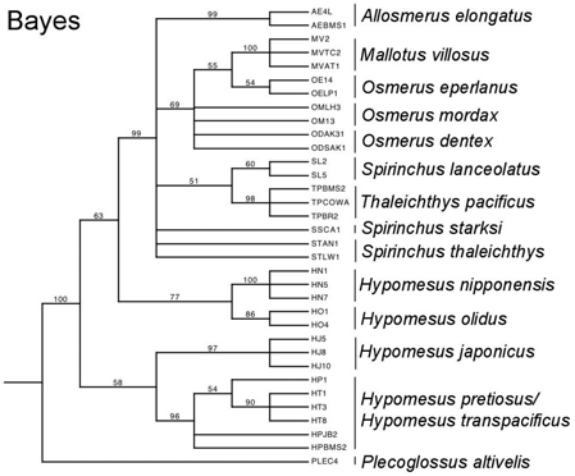


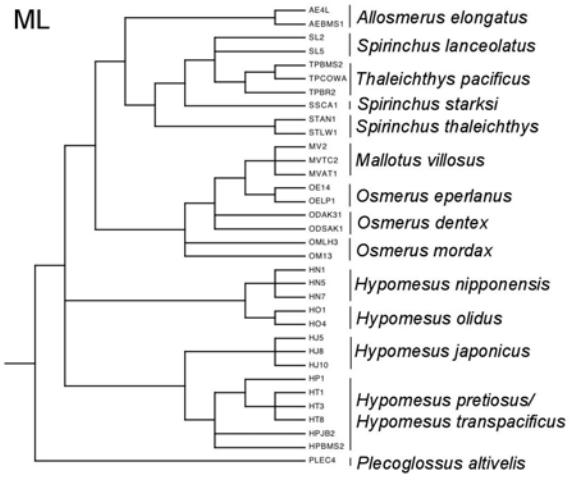
Appendix 12 Reduced dataset Osmeridae 12S phylogenies resulting from neighbour-joining (NJ), parsimony (MP), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ, MP) or posterior probabilities from a consensus of 18000 trees (Bayes). ML tree from heuristic searches with 10 replicates of random taxon addition.



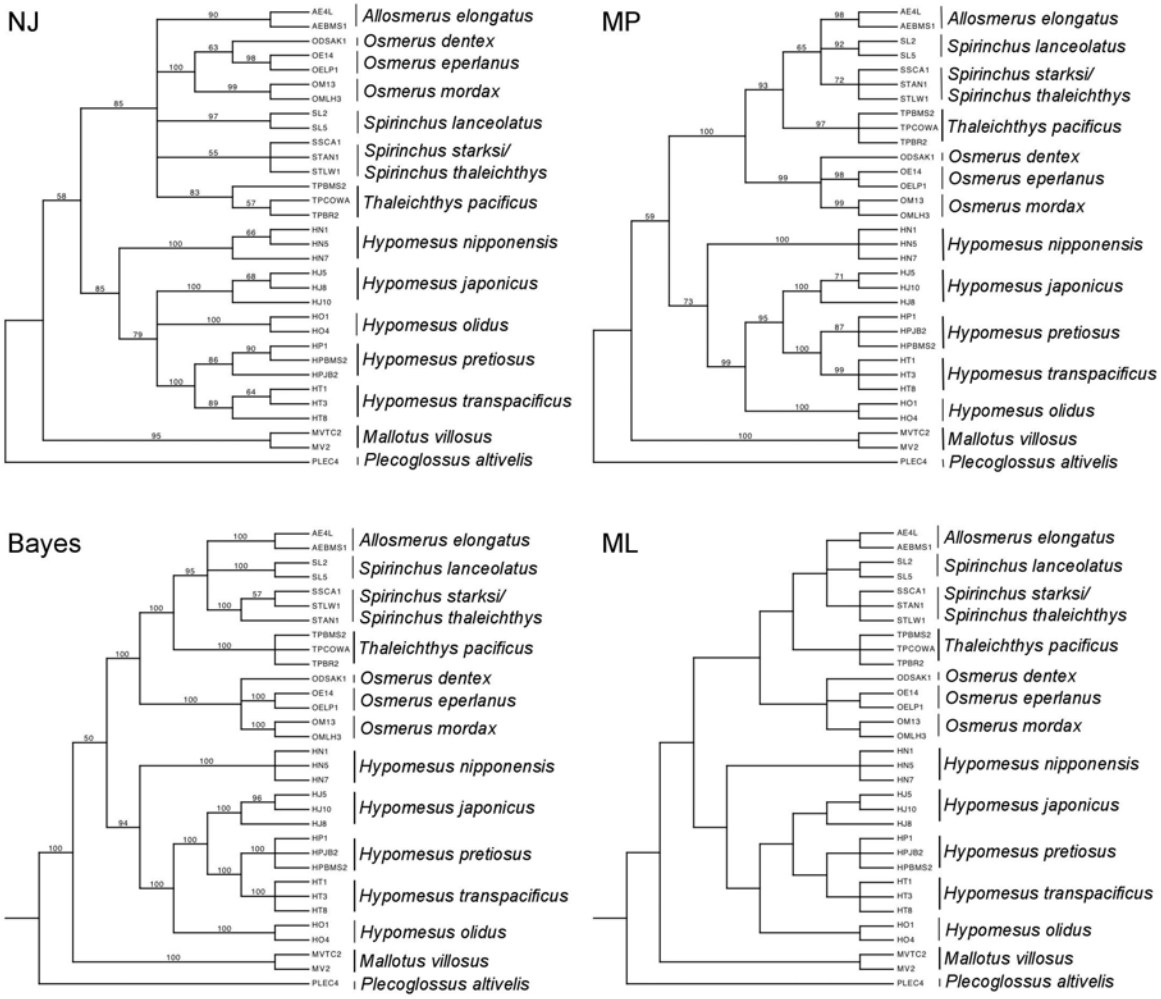
Appendix 13 Reduced dataset Osmeridae ITS2 phylogenies resulting from neighbour-joining (NJ), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ) or posterior probabilities from a consensus of 18000 trees (Bayes). ML trees from heuristic searches with 10 replicates of random taxon addition.



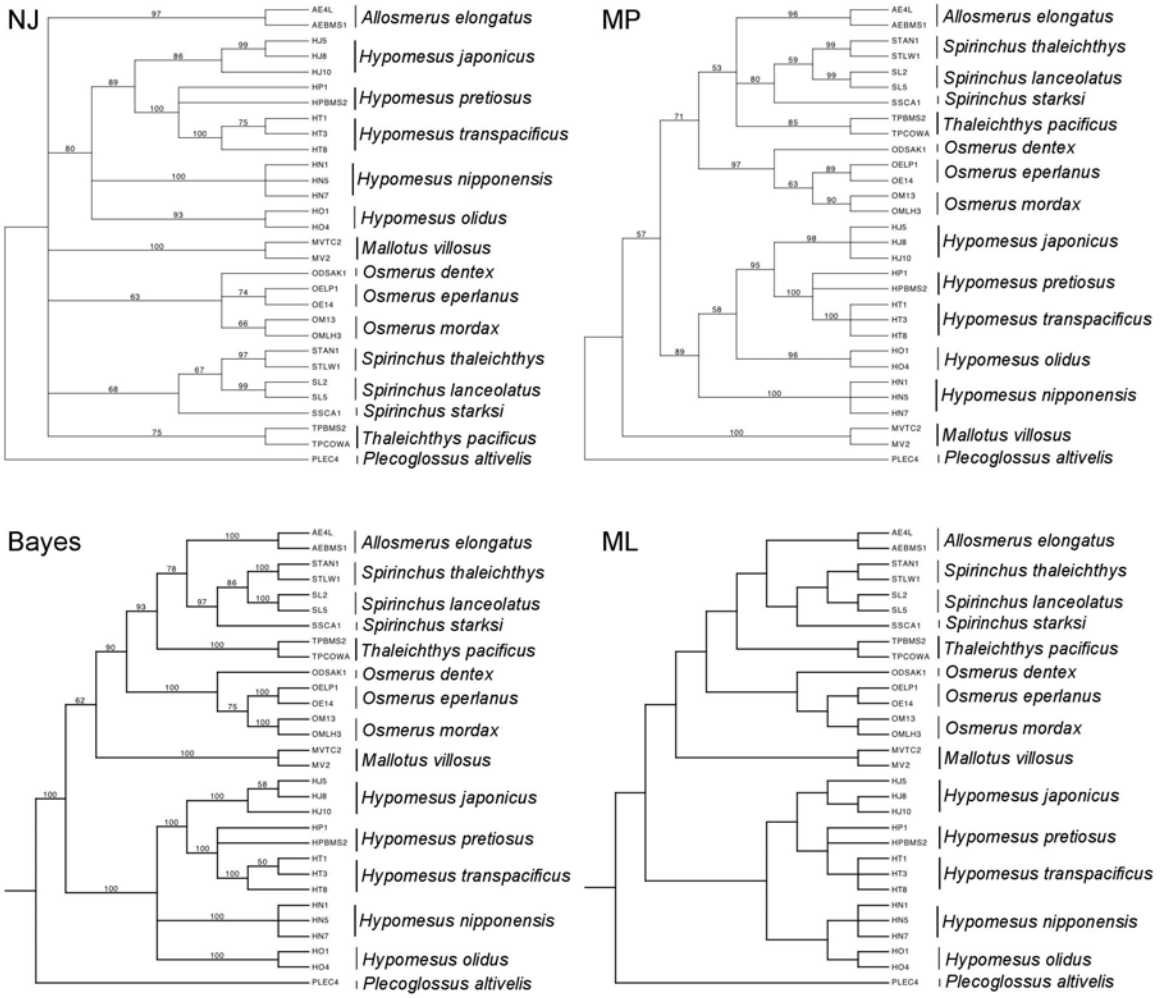




Appendix 14 Reduced dataset Osmeridae S71 phylogenies resulting from neighbour-joining (NJ), parsimony (MP), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ, MP) or posterior probabilities from a consensus of 18000 trees (Bayes). ML trees from heuristic searches with 10 replicates of random taxon addition.



Appendix 15 Reduced dataset Osmeridae RAG1 phylogenies resulting from neighbour-joining (NJ), parsimony (MP), Bayesian (Bayes), and maximum likelihood (ML) reconstruction. Numbers above nodes represent support values from 1000 bootstrap pseudo-replicates (NJ, MP) or posterior probabilities from a consensus of 18000 trees (Bayes). ML tree from heuristic searches with 10 replicates of random taxon addition.



Appendix 16 Character and state names for Osmeridae morphological characters mapped in Fig. 4.8. Numbers in brackets refer to the characters in Fig. 4.9.

1. lateral shelf of ectopterygoid
    0: absent
    1: present
2. dorsal margin of quadrate
    0: convex
    1: straight
3. metapterygoid located
    0: posterodorsal to quadrate
    1: dorsal to quadrate
4. lateral strut of hyomandibular
    0: vertical
    1: lateral
5. preopercular canal
    0: completely open
    1: partly closed
6. ventral arm of preopercular
    0: long
    1: short
7. articular process of premaxilla
    0: prominent
    1: much reduced
8. maxilla
    0: sickle-shaped
    1: straight
9. laterally-projecting dorsal flange on maxilla
    0: absent
    1: present
10. coronoid process of mandible
    0: high
    1: low
11. mandibular sensory canal
    0: open
    1: partly closed
    2: closed
12. anguloarticular shape
    0: deep
    1: elongate
13. proethmoids
    0: absent
    1: present
    2: two, oval
14. prootic shape
    0: deep

1: shallow
15. otic capsules
0: not expanded
1: expanded
2: very expanded
16. parasphenoid
0: winged
1: wingless
17. parasphenoid posteroventrally
0: not concave
1: very concave
18. number of branchiostegals
0: fewer than 8
1: 8-9
19. canal on supracleithrum
0: open
1: closed
20. number of anal fin rays
0: fewer than 23
1: 23 or more
21. scales
0: large, squared
1: small, round
22. articulation of premaxilla with maxilla
0: premaxillary articular process not tightly adhering to maxillary head
1: articular process of premaxilla forming syndesmosis with maxillary head
23. uncinate process of third pharyngobranchial process
0: not extending over second epibranchial
1: extending over second epibranchial
24. levator process on fourth epibranchial
0: wide
1: narrow
25. epihyal
0: without foramen
1: with midlateral foramen
26. lateral spine on sphenotic
0: blunt
1: rodlike
27. midlateral band of silver pigment
0: absent
1: present
28 [1]. Dermethmoid
0: median
1: paired
2: absent
29 [2]. Ethmoid endoskeleton

0: short with one or more perichondral ossifications anterior to the lateral ethmoids
1: short and unossified
2: long and unossified

30 [3]. Vomer
0: with shaft
1: without shaft
2: absent

31 [4]. Pterosphenoid - central process
0: without central process of flange from anterior half of ventral margin
1: with central process of flange from anterior half of ventral margin

32 [5]. Pterospenoid - posteroventral process
0: without posteroventral membrane-bone process towards anterodorsal process from prootic
1: with posteroventral membrane-bone process towards anterodorsal process from prootic

33 [6]. prootic-pterosphenoid contact
0: no contact between anterodorsal process of prootic and pterosphenoid
1: contact between anterodorsal process of prootic and pterosphenoid

34 [7]. Cartilaginous interorbital septum
0: absent in anterior part of orbit
1: present in anterior part of orbit

35 [8]. Otic bulla
0: not inflated and with little or no cartilage in its wall
1: somewhat inflated
2: globose

36 [9]. Parietals
0: in contact medially
1: partially separated by supraoccipital
2: completely separated by supraoccipital

37 [10]. Fontanelles in cartilaginous roof of otic region
0: remaining open during ontogeny
1: closed during ontogeny

38 [11]. Dermatopalatine and autopalatines
0: separate
1: fused
2: dermatopalatine absent

39 [12]. Endopterygoid teeth
0: concentrated along dorsal margin of bone, with a patch of teeth posteriorly
1: in a single row
2: absent

40 [13]. Anterior margin of metapterygoid
0: above quadrate
1: anterior to quadrate

41 [14]. Hyomandibular
0: with a vertically elongate lateral crest

1: with a short vertical crest fitting against the preopercular
2: with a triangular spur
3: with an obliquely orientated spurlike crest
4: with no preopercular crest

42 [15]. Dentary
0: with toothed margin occupying less than half of length of lower jaw
1: with toothed margin occupying more than half of length of lower jaw

43 [16]. Meckelian fossa
0: small and anteriorly placed
1: large and opening beneath the hind end of the dentary tooth row

44 [17]. Dorsal margin of opercular
0: entire and unmodified
1: with an anterodorsal notch
2: with notch and a tongue-like process behind it
3: not extending above articulation with hyomandibular

45 [18]. Basihyal
0: with scattered teeth
1: with marginal fangs
2: toothless

46 [19]. Uncinate process on fourth epibranchial
0: absent
1: present

47 [20]. Epineural bones and/or ligaments
0: originate on neural arch
1: on centrum on several anterior vertebrae
2: absent

48 [21]. Epipleural bones
0: absent
1: present

49 [22]. Supraneurals
0: numerous, ca. 15 or more
1: fewer than ten
2: one

50 [23]. Median keels of laminar bone
0: absent
1: present on distal parts of last few neural and haemal spines

51 [24]. Epurals
0: three
1: two
2: one
3: none

52 [25]. Upper and lower caudal median cartilages
0: present
1: a single cartilage
2: absent

53 [26]. First pectoral radial

0: unmodified
1: enlarged and embracing scapula

54 [27]. Fourth pectoral radial
0: articulating with glenoid
1: tapering proximally and failing to articulate with glenoid

55 [28]. Fourth pectoral radial
0: single
1: multifid distally

56 [29]. Articular surface for pelvic fin
0: short and transverse
1: elongate and oblique

57 [30]. Adipose cartilage
0: beanlike
1: transversely arched, fenestrate plate

58 [31]. Posttemporal
0: penetrated by lateral line
1: with a separate canal-bearing ossicle
2: with no relation to sensory canal

59 [32]. Supracleithrum
0: penetrated by lateral line
1: with a separate canal-bearing ossicle
2: with no relation to sensory canal

60 [33]. Ovaries
0: both present in females
1: only left present in females

61 [34]. Scales on anal fin base of mature males
0: unmodified
1: enlarged

62 [35]. Anal fin skeleton
0: unmodified in mature males
1: anterior anal endoskeleton and central fin rays modified
2: entire anal fin skeleton greatly modified

63 [36]. Pyloric caeca
0: present
1: absent

64 [37]. Cucumber odor
0: absent
1: present

65 [38]. Life cycle
0: entirely marine
1: diadromous (anadromous or amphidromous)
2: entirely freshwater

66 [39]. Skeletal ontogeny
0: unmodified
1: retarded relative to sexual maturity
2: accelerated

67. Glossohyal teeth
    0: canine
    1: not canine
68. maxillary extends past mid-eye
    0: no
    1: yes
69. otic bulla wide anteriorly
    0: no
    1: yes
70. posterior myodome opening narrow
    0: no
    1: yes
71. with parasphenoid wing joining prootic
    0: no
    1: yes
72. mandible shallow
    0: no
    1: yes
73. palatine dumbbell-shaped
    0: no
    1: yes
74. dorsal edge of pterygoid straight
    0: no
    1: yes
75. metapterygoid with dorsal vane over hyomandibular head
    0: no
    1: yes
76. dorsal fork of posttemporal lobe long
    0: no
    1: yes
77. frontal with lateral wings over orbit
    0: no
    1: yes
78. vomerine teeth small
    0: no
    1: yes
79. palatine teeth small
    0: no
    1: yes
80. subopercle and/or opercle with striae
    0: no
    1: yes
81. snout-dorsal length greater than or equal to dorsal-caudal length
    0: no
    1: yes
82. midlateral ridge in males

0: absent
1: present

83. elongate midlateral scales
0: absent
1: present

84. gill rakers 25 or more
0: no
1: yes

85. pyloric caeca never more than 8 (9)
0: no
1: yes

86. blind stomac sac
0: absent
1: present

87. midlateral scales always >70
0: no
1: yes

88. pectoral rays 16-23
0: no
1: yes

89. lateral line
0: incomplete
1: complete

90. length of adipose base never > orbit
0: no
1: yes

91. orbit 2/3 or less caudal peduncle depth
0: no
1: yes

92. mesethmoid simple
0: no
1: yes

93. pterosphenoid reaches parasphenoid wing anteriorly
0: no
1: yes

94. no slit between hyomandibular and preopercle
0: no
1: yes

95. actinosts 4
0: no
1: yes

96. ventrals 8
0: no
1: yes

97. dorsal 10-14
0: no

1: yes
98. vertebrae 64 or more
0: no
1: yes
99. branchiostegals 8-10
0: no
1: yes
100. head 4.7 or less in standard length
0: no
1: yes
101. pectoral always 70% or more of distance to pelvic
0: no
1: yes
102. pelvic origin anterior to dorsal
0: no
1: yes
103. peritoneum silver
0: no
1: yes
104. ductus pneumaticus attaches to anteriormost end of gas bladder
0: no
1: yes
105. gill rakers
0: short
1: long
106. standard length exceeds 200mm
0: no
1: yes
107. mouth horizontal
0: no
1: yes
108. range attains or exceeds 60 N latitude
0: no
1: yes
109 [40]. posterior shaft of vomer
0: long
1: short
110 [41]. lateral hyomandibular spur
0: hyomandibular without lateral projection
1: hyomandibular with lateral spur at level of its articulation with opercle
111 [42]. symphesial cartilages
0: paired cartilages not present at dentary symphysis
1: paired cartilages present at dentary symphysis
112 [43]. adipose cartilage
0: absent
1: present

2: present, pear shaped

113 [44]. dentary symphysis
0: without medial process
1: with medial tusklike process

114 [45]. metapterygoid
0: without lateral shelf
1: with short lateral shelf
2: with prominent diagonal shelf

115 [46]. fusion of 5th epibranchial to 4th to form circular foramen for efferent artery
0: no
1: yes, no levator process on Eb4 and fusion between Eb5 and Eb4 is initiated at its upper end
2: yes, levator process on Eb4 and a vascular notch enclosed between it and Eb5, which fuses with Eb4 at its lower end

116 [47]. sculpture on quadrate
0: absent
1: present

117 [48]. sculpture on metapterygoid
0: absent
1: present

118 [49]. anterodorsal border of opercle
0: without spine, horizontal
1: with notch and spine
2: with deep narrow notch

119 [50]. pterosphenoid posteroventral process
0: absent
1: present

120 [51]. pterosphenoid process directed towards prootic
0: absent
1: present

121 [52]. anterior extent of prootic
0: smooth
1: notched with a dorsal projection
2: more medial, contact with pterosphenoid by interdigitation

Appendix 17 Character matrix for characters displayed in Fig. 4.8. Characters and their state names listed in Appendix 16.

	1								10								20								30					35
Allosmerus	0	0	1	1	0	1	1	1	0	1	0	1	2	1	1	1	0	1	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1
Hypomesus	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	&	1	0	0	0	0
Mallotus	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	1	1	1	1	0	0	2	1	1	0	0	0	1
Osmerus	0	0	1	1	0	1	1	1	0	1	0	1	2	1	1	1	0	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	1	0	1	1
Spirinchus	0	0	1	1	0	1	1	1	0	1	0	1	2	1	2	1	1	0	0	0	0	0	0	1	1	1	1	0	1	0	1	0	1	1	1	2
Thaleichthys	0	0	1	0	0	1	1	1	0	1	0	1	2	1	2	1	1	0	0	1	0	0	1	0	0	0	0	1	0	1	0	1	1	1	1	2
Plecoglossus	1	1	1	0	1	1	0	1	1	1	2	1	0	0	0	0	0	1	1	0	1	1	1	0	1	0	0	0	0	0	1	1	0	0	0	

	36				40								50								60										70
Allosmerus	2	0	1	1	1	2	1	1	2	1	&	1	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	1	1	1	1
Hypomesus	0	0	0	0	0	1	0	0	1	0	0	&	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	&	0	0	0	0	0
Mallotus	1	0	1	1	1	1	0	0	2	0	0	1	1	1	0	1	1	0	1	1	1	1	1	1	0	0	1	1	0	1	0	1	0	0	0	0
Osmerus	1	0	0	1	1	2	1	0	1	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	
Spirinchus	2	1	1	1	1	2	1	1	1	1	1	1	1	0	0	1	0	1	0	0	1	0	1	0	0	0	&	0	0	&	0	1	1	1	1	
Thaleichthys	2	1	1	1	1	2	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	0	1	1	0	0	0	0	1	1	0	1	1	1	1	
Plecoglossus	0	0	0	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	?	?	?	?	

	71								80								90								100					105
Allosmerus	1	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	1	1	0	1	0	1	0	1	0	1	0	
Hypomesus	0	0	0	0	1	0	0	1	1	0	0	0	0	1	1	0	0	0	0	0	?	1	0	1	1	1	0	&	0	0	0	&	&	&	?	
Mallotus	0	0	0	0	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	1	1	1	0	0	1	1	1	0	
Osmerus	1	1	1	1	0	1	1	0	0	0	0	1	0	1	1	1	0	0	0	1	1	1	0	0	1	1	0	1	0	&	0	1	0	1	?	
Spirinchus	1	1	1	1	0	1	1	0	0	&	&	&	0	1	1	&	0	0	0	1	1	0	0	1	1	1	0	0	0	1	1	1	1	1	&	
Thaleichthys	1	1	1	1	0	1	1	0	0	1	1	1	0	0	0	1	1	0	0	1	1	0	1	0	1	0	1	1	1	1	0	0	0	1	0	
Plecoglossus	?	?	1	?	?	?	?	?	?	?	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	

	106				110												121
Allosmerus	1	0	0	1	1	0	2	0	2	2	1	1	2	1	1	2	
Hypomesus	&	0	&	&	1	0	2	0	2	2	0	0	1	1	0	0	
Mallotus	1	1	1	1	1	0	1	0	2	2	1	0	2	1	0	0	
Osmerus	1	0	1	1	1	0	2	0	2	2	0	0	1	0	1	1	
Spirinchus	0	0	&	1	1	0	2	0	2	2	1	1	1	0	1	2	
Thaleichthys	1	0	1	1	0	0	2	0	2	2	1	1	1	0	1	2	
Plecoglossus	?	?	0	0	1	1	2	1	1	0	0	0	1	1	1	2	

Appendix 18 Character matrix for characters displayed in Fig. 4.9. Characters and their state names listed in Appendix 16.

	1								10									20									30					35
Allosmerus	1	0	1	1	1	1	1	1	2	0	1	1	1	2	1	1	2	1	&	1	1	0	0	0	0	1	1	0	1	0	0	0	0	0	0
Hypomesus	0	0	&	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	&	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0
Mallotus	0	2	1	1	0	0	0	1	1	0	1	1	1	1	0	0	2	0	0	1	1	1	0	1	1	0	1	1	1	1	1	0	0	1	1
Osmerus	1	0	1	0	1	0	1	1	1	0	0	1	1	2	1	0	1	1	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0
Spirinchus	1	0	1	0	1	1	1	2	2	1	1	1	1	2	1	1	1	1	1	1	1	0	0	1	0	1	0	0	1	0	1	0	0	0	&
Thaleichthys	1	0	1	0	1	1	1	2	2	1	1	1	1	2	1	1	1	1	1	1	1	0	0	0	0	1	1	0	1	0	1	1	0	0	0
Plecoglossus	0	0	0	1	1	0	0	0	0	0	0	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	0	0	0	0	0	1	0	0
Salangidae	0	2	2	?	?	?	0	?	?	0	?	2	0	4	&	0	3	0	?	1	0	2	0	0	0	0	0	1	1	1	2	2	0	1	1
outgroup	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	36				40										50	52
Allosmerus	1	0	0	0	1	1	0	2	0	2	2	1	1	2	1	1	2
Hypomesus	0	1	&	0	&	1	0	2	0	2	2	0	0	1	1	0	0
Mallotus	0	1	0	1	1	1	0	1	0	2	2	1	0	2	1	0	0
Osmerus	0	1	1	0	1	1	0	2	0	2	2	0	0	1	0	1	1
Spirinchus	0	0	&	0	1	1	0	2	0	2	2	1	1	1	0	1	2
Thaleichthys	0	1	1	0	1	0	0	2	0	2	2	1	1	1	0	1	2
Plecoglossus	0	0	1	0	0	1	1	2	1	1	0	0	0	1	1	1	2
Salangidae	1	0	1	1	2	0	0	&	0	0	2	0	0	0	?	?	0
outgroup	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0