MITOCHONDRIAL GENOMIC PHYLOGENY OF GADID FISH: IMPLICATIONS FOR BIOGEOGRAPHIC ORIGINS AND TAXONOMY

MARK W. COULSON

MITOCHONDRIAL GENOMIC PHYLOGENY OF GADID FISH: IMPLICATIONS FOR BIOGEOGRAPHIC ORIGINS AND TAXONOMY

by

© Mark W. Coulson

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Master of Science

Department of Biology/Faculty of Science
Memorial University of Newfoundland

St. John's Newfoundland

Your file Votre référence
ISBN: 0-612-99066-4
Our file Notre référence
ISBN: 0-612-99066-4

ABSTRACT

The phylogeny of ten species of gadine fishes was assessed with complete mitochondrial DNA (mtDNA) genome sequences from the following species: *Microgadus proximus*, *Pollachius virens, Melanogrammus aeglefinus, Merlangius merlangus, Boreogadus saida, Arctogadus glacialis, Theragra chalcogramma* and the three species of *Gadus* (*G. morhua, G. macrocephalus* and *G. ogac*). As most prior mitogenomic studies have addressed resolving basal or very divergent relationships, this study uses mitochondrial genomes among a closely related group of taxa to address taxonomic relationships with a focus on the biogeography of recently diverged species. Maximum parsimony, neighbour-joining and maximum likelihood all produced the same relationships when using mtDNA genomic sequences representing 14036 base pairs and consistently supported at least 8 of 10 nodes. Of these nodes, at least 6 were supported with 95% or greater bootstrap support.

Among the individual mtDNA protein-coding genes, subunits of the ND complex included both the most successful (e.g. ND1 and ND5) and some of the least successful (e.g. ND6) genes for resolving phylogenetic relationships among these taxa. Amino acid sequences supported at least 6 of the nodes in common with nucleotide data and even suggested a functional evolutionary difference for ND5 among *Theragra*, relative to *Gadus*.

The phylogenetic analysis identified the following relationships: *Melanogrammus* and *Merlangius* as sister taxa, a clade composed of *Boreogadus, Arctogadus*, *Theragra* and *Gadus* with *Theragra* and *G. morhua* as sister taxa, and a close relationship between

one of two *G. macrocephalus* individuals and the two *G. ogac* individuals. *Microgadus proximus* was used to root the tree, and *Pollachius* was resolved as the outgroup to the other above mentioned species. These results suggest that *Theragra chalcogramma* should now be included as *Gadus chalcogrammus*, and that *G. ogac* represents a northward and eastward extension of the Pacific *G. macrocephalus* and should be included as a subspecies, *G. macrocephalus ogac*. The close evolutionary relationship between *Theragra* and *G. morhua* may therefore explain their common ability to have sustained the two largest fisheries in the world. The data also supports separate invasions of the Pacific basin by species endemic to these waters, and suggests that there has been at least one occurrence of secondary Atlantic Ocean contact with the introduction of *G. macrocephalus ogac*. The evidence also supported *Melanogrammus aeglefinus* and *Merlangius merlangus* as sister taxa which was surprising considering these species represent very different ecological niches and prior morphological work has failed to yield such a relationship.

Acknowledgements

First, and foremost, I would like to thank my supervisor Dr. Steven M. Carr for his supervision and support during this project. In addition, I would like to thank Dr. Dawn Marshall for her technical expertise and editing on previous drafts. Also, thanks to Dr. Pierre Pepin from the Department of Fisheries and Oceans, for provision of specimens and helpful comments and information on fish biology. I would also like to thank the other people with whom I have shared the lab: Kim Johnstone, Angela Pope, Payal Sipahimalani, Leanne Thorne, and Anne-Marie Gale, all of whom provided discussion and technical advice as well as entertainment in the lab. I would like to acknowledge the Department of Biology, Memorial University for financial assistance and the Department of Fisheries and Oceans (DFO) St. John's for financial support for this project. I would also like to thank fellow graduate students and friends for their support and company. Lastly, I would like to thank my parents, Wally and Jeanne for their constant emotional and financial support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

mtDNA	mitochondrial DNA
PCR	Polymerase Chain Reaction
RFLP	Randomly Amplified Polymorphic DNA
ATP6	adenosine triphosphate subunit 6
ATP8	adenosine triphospate subunit 8
CO1	Cytochrome Oxidase subunit 1
CO2	Cytochrome Oxidase subunit 2
CO3	Cytochrome Oxidase subunit 3
Cyt b	Cytochrome *b*
ND1	nicotinamide adenosine dehydrogenase subunit 1
ND	nicotinamide adenosine dehydrogenase subunit 2
ND3	nicotinamide adenosine dehydrogenase subunit 3
ND4	nicotinamide adenosine dehydrogenase subunit 4
ND4L	nicotinamide adenosine dehydrogenase subunit 4 (light strand)
ND5	nicotinamide adenosine dehydrogenase subunit 5
ND6	nicotinamide adenosine dehydrogenase subunit 6
MP	Maximum Parsimony
NJ	Neighbour-Joining
T-N	Tamura-Nei genetic distance
ML	Maximum Likelihood
γ	gamma (γ) parameter for substitution rate heterogeneity

kbp	kilobase pairs
rRNA	ribosomal RNA subunit
tRNA	transfer RNA
TBR	tree-bisection-and-reconnection
Ts/Tv	Transition/Transversion ratio
d_T	topological distance
A	adenine
C	cytosine
G	guanine
T	thymine
U	uracil
ddH_2O	distilled, deionized water

1.0 INTRODUCTION

1.1 Overview and Classification

Fishes are the most diverse group of living vertebrates, with more than 24,600 extant species currently described (Stepien & Kocher, 1997) and represent a wide array of life histories and morphological diversity. The teleosts represent 96% of all extant fishes and are distributed among 38 orders, 426 families, and 4064 genera with an origin dating to 220-200 million years ago (Miya *et al.*, 2003). Biologists have been studying aspects of their internal and external morphology in an attempt to understand their taxonomic status and evolutionary relationships. The development of molecular techniques has helped to further research in this field. Scientists now routinely use the genetic information in biological macromolecules to address numerous aspects of the behaviours, life histories, and evolutionary relationships of organisms. Molecular data are integrated with information from numerous other fields such as ecology, ethology, comparative morphology, systematics, and palaeontology. These disciplines remain very active today, but have been strengthened by the field of molecular evolution. Morphological studies have generally been successful in defining species and organizing them into genera, and these groupings have usually been confirmed when examined with molecular approaches (Stepien & Kocher, 1997).

The order Gadiformes is a morphologically diverse group of paracanthopterygian fishes, and include some of the world's most important and commercially exploited species, representing more than one-tenth of the total fish caught worldwide (Cohen *et al.*, 1990). The order is currently considered to be composed of eight families:

Bregmacerotidae, Euclichthyidae, Gadidae, Macrouridae, Melanonidae, Merlucciidae, Moridae, and Muraenolepididae. Gadiformes are distributed throughout the world's oceans and are predominantly marine benthopelagic fishes of cool waters, found either in shallow waters of high and mid-latitudes or deep waters in the tropics (Marshall & Cohen, 1973). Historical events as well as temperature have shaped present-day distribution patterns of these fishes, as many patterns are characteristic for particular taxonomic categories (Cohen *et al.*, 1990).

The family Macrouridae is the largest in the order with approximately 300 species. It contains more species than all other gadiformes combined, and is composed of the grenadiers, rattails, and whiptails, all of which are found in all oceans except the high Arctic. Several of these families are extremely small. The Euclichthyidae contains a single rare species of Eucla cod found off New Zealand and southern Australia. The family Melanonidae contains two species of pelagic cods found in subantarctic, temperate, and tropical waters, but they are of no real interest to fisheries. The family Muraenolepididae also contains a single genus with four species currently described. These are found in far southern seas, mainly around Antarctica, and are more commonly referred to as moray cods. The family Bregmacerotidae contains approximately 15 species all belonging to a single genus of codlets. Most of these species are small fishes living in tropical and sub-tropical oceanic waters to depths over 1000 m; however some are also found in shallow coastal areas and even in estuaries. The family Moridae contains about 100 species, most of which are poorly known, and there is little agreement as to how many genera should be recognized by this family. Members of Moridae are

found in all oceans from shallow coastal areas to deep (> 2,500 m) waters, however they are less numerous and have less potential value to fisheries than members of Gadidae, Merlucciidae or Macrouridae (Cohen *et al.*, 1990). The family Merlucciidae consists of the hakes and has a wide distribution. Aside from the 13 species of hakes and whiting belonging to the genus *Merluccius*, there is no great measure of agreement about the assignment of other genera to the family or subfamily (Cohen *et al.*, 1990).

The most commercially important family within Gadiformes is the Gadidae, a morphologically diverse group containing approximately 50 species in 21 genera. With the exception of the burbot (*Lota lota*), most of these fish inhabit coastal zones, continental shelves, and slopes, primarily in the northern oceans at depths up to 1,300 m (Scott & Scott, 1988). The family is currently classified into three subfamilies: Phycinae (27 species in six genera), comprising the rocklings and non-merluccid hakes, Lotinae (four species in three genera), the cusks and lings, and Gadinae (22 species in 12 genera). Gadinae includes some of the most abundant and important fishes in the sea, the true cods (*Gadus* spp.), the haddock (*Melanogrammus*), the Alaskan pollock (*Theragra*), and the tomcods (*Microgadus* spp.) (Cohen *et al.*, 1990). Phycines and lotines are believed to represent the ancestral form of the gadids, with one or two dorsal fins, a single caudal fin, and an elongate, laterally compressed body. Gadines are the more morphologically derived members of the family, with greater differentiation among species (Cohen *et al.*, 1990).

Most gadiformes are benthopelagic and are therefore taken in bottom trawls when fished. A number of species (primarily in the Gadidae and Merlucciidae) are the targets

of very large, directed fisheries which have been operating for over a century. Other species, such as those belonging to the family Macrouridae, have become accessible to fisheries only in the course of the last few decades and are not yet the targets of intense fishery activity. Historically, the global catch of Gadiform fishes (ca. 13,700,000 metric tons in 1987) has represented about 17% of the total world landings from marine waters, and over 95% of these landings have corresponded to a single family, the Gadidae (containing the true cods). While these are historical figures, contemporary catch data have been reduced for many of these species due to decreasing abundance and overexploitation.

All species examined in this study belong in the subfamily Gadinae, whose species are distributed among 12 genera: *Arctogadus, Boreogadus, Eleginus, Gadiculus, Gadus, Melanogrammus, Merlangius, Microgadus, Micromesistius, Pollachius, Theragra,* and *Trisopterus*. Two of these genera, *Arctogadus* and *Boreogadus*, are predominantly found in Arctic regions, while the majority of species are confined to the Northern Hemisphere with the largest concentration in the North Atlantic. Several species include *Gadus macrocephalus* (Pacific cod), *Microgadus proximus* (Pacific tomcod), and *Theragra chalcogramma* (walleye or Alaskan pollock) that are endemic to Pacific waters.

The present study focuses on resolving the phylogenetic relationships among the genera *Arctogadus, Boreogadus, Gadus, Melanogrammus, Merlangius, Microgadus, Pollachius*, and *Theragra* using complete mitochondrial genome sequences. The purpose of establishing the phylogeny of these genera is to address questions relating to the

biogeographic origins of the species of gadines endemic to the Pacific Ocean (*Microgadus proximus, Theragra chalcogramma*, and *Gadus macrocephalus*), and to address the systematic nomenclature of these 10 species, which includes the determination of the species composition with regards to *Gadus macrocephalus* and *Gadus ogac*. Prior studies (e.g., Carr *et al.*, 1999) have indicated a close relationship, however, these two species are found in separate oceans. This study is the first to use mitochondrial genomic sequences to address the taxonomy and biogeography of a closely related group of vertebrates. This study represents the first use of mitochondrial genome sequences to resolve the phylogeny of a closely related group of vertebrates. For the purpose of this analysis, closely related refers to taxa that are all members of the same subfamily (Gadinae). Several genera of gadines were not included in the present study, primarily due to a lack of access to specimens (e.g. *Gadiculus* and *Micromesistius*), or the fact that their phylogenetic position has been well supported in previous molecular analyses (e.g. *Trisopterus,* and *Eleginus*) (Carr *et al.*, 1999). However, it is not expected that this will affect the phylogeny of the current study, as these four aforementioned genera are believed to fall outside of the genera being investigated here. While Arnason *et al.*, (2000) used complete mtDNAs and resolved close relationships among several species of whales, this study included a number of more distant taxa with the focus of estimating more ancient eutherian divergences. As well, the relative efficiency of the various mtDNA gene regions compared to the genomic phylogeny will be examined, including an analysis of the different rates of molecular evolution across loci.

1.2 Geographic Distribution of Study Species

In order to discuss phylogenetic relationships as they pertain to biogeographic origins of a particular group, it is important to discuss their present-day distributions and possible patterns of migrations. Such information may provide useful insights in relating historical evolutionary relationships to contemporary lineages. This section outlines the general distribution for each of the species considered in the present study.

Arctogadus glacialis (Arctic cod) is widely distributed in the western Arctic basin, along both northwest and northeast coasts of Greenland. Other populations occur north of the Beaufort Sea and off the Arctic coasts of Siberia. *Arctogadus* occupies both ice-free and ice-covered seas with preferred temperatures around 0°C (Sufke *et al.*, 1998). Trawl catches conducted off Northeast Greenland by Sufke *et al.* (1998) showed that *A. glacialis* was by far the most abundant species caught, accounting for 44% of all fish. Reports of mass occurrences of *Arctogadus* indicate that these fishes play an important role in the Arctic ecosystem (Walters, 1961; Andriashev *et al.*, 1980; Møller *et al.*, 2002). Traditionally, another species has been recognized, *Arctogadus borisovi* Dryagin or the East Siberian cod. *A. borisovi* has an overlapping range with *A. glacialis* but is also found somewhat further south around the southern coast of Greenland, and in more coastal areas of the Beaufort Sea, also with a more continuous distribution along the Arctic coast of Siberia. This species was not included in this study; however recent meristic and molecular data suggests that these two nominal species are in fact conspecific with *A. glacialis* as the senior synonym (Jordan, 2002; Møller *et al.*, 2002).

Boreogadus saida (Polar cod) is found throughout the entire north polar basin with a circumpolar distribution in Arctic seas, and ranges farther north than any other fish (to latitude 84°42'N, Scott & Scott, 1988). It is found in coastal habitats around Greenland, through the northern Canadian Archipelago, the Beaufort and Bering seas and along the northern coasts of Siberia. Also present off the northern coast of Russia, in the Chukchi and Barents seas, and off Scandinavia. In the Canadian Atlantic region its range extends from Arctic waters, including the Baffin Island region and Davis Strait into Hudson Strait and Hudson Bay down the Labrador coast into the Strait of Belle Isle and the Gulf of St. Lawrence and around northern Newfoundland to the Grand Bank. In the Gulf of St. Lawrence and other parts of its southern range it is rare, usually present only during fall and winter (Scott & Scott, 1988). It has little commercial use due to its small body size and their associated environment makes fishing of this species difficult. *Boreogadus saida* is the principal plankton consumer in Arctic waters and in turn becomes an important seasonal food supply for marine mammals, seabirds, and fishes (Scott & Scott, 1988).

Gadus macrocephalus (Pacific cod) is widely distributed in continental shelf and coastal waters of the North Pacific. They range from southern California through Alaska and the Bering Sea and west along the Aleutian Islands. They are also present off the Kuril Islands to Kamchatka, Okhotsk Sea, Sea of Japan, Korea and the Yellow Sea (Hart, 1973). Studies of protein loci indicate the presence of two major groups throughout their range: a western Pacific (Asian) group and an eastern North Pacific (including the Bering Sea) group (Grant *et al.*, 1987). This fish has a high growth rate and a high natural

mortality, factors that can promote heavy exploitation. The Japanese catch has traditionally accounted for the largest component of the total landing of this species, however since the mid 1970s heavy exploitation has caused decreased catches of these stocks (localized populations). Total annual catches for 1987 amounted to 441,107 metric tons. Pacific cod is among the most important of the trawl-caught bottom fishes of British Columbia.

Gadus morhua (Atlantic cod) occurs on both sides of the North Atlantic. In the eastern North Atlantic it extends from Iceland to the Norwegian Sea, to the Baltic Sea, Barents Sea, and Bay of Biscay. In the western North Atlantic, it is present from Greenland and southern Baffin Island, into Hudson's Bay, along the continental slope off Labrador, Newfoundland, the Gulf of St. Lawrence, the Grand Bank and Scotian Shelf, and throughout the Bay of Fundy and Gulf of Maine, south to Cape Hatteras, NC (Scott & Scott, 1988). Population structure among northwestern Atlantic populations has been debated. Geographic surveys (Hutchings *et al.*, 1993), vertebral data (Templeman, 1981; Lear & Wells, 1984) and tag recovery (Templeman, 1974, 1979; Lear, 1984) data all suggest that northern cod (NAFO Divisions 2J, 3K, and 3L) are divided into several distinct offshore spawning units. Evidence of population structure from genetic studies has also been mixed. Allozyme (Mork *et al.*, 1985) and mitochondrial (Smith *et al.*, 1989; Carr *et al.*, 1995) studies have failed to yield significant differences among populations from the western Atlantic. However significant differences have been detected with small numbers of protein loci (Jamieson, 1975; Cross & Payne, 1978), RFLP analysis (Pogson *et al.*, 1995) and more recently microsatellites (Bentzen *et al.*,

1996; Ruzzante *et al.*, 1996). Some of these studies even failed to resolve significant population structure among trans-Atlantic populations of *G. morhua* (e.g., Smith *et al.*, 1989). The value of Atlantic cod as a prime food fish is enormous; the species has accounted for nearly 30% of the world's total groundfish catch (Cohen *et al.*, 1990). The world catch reported for 1987 totalled 2,054,721 metric tons (Cohen *et al.*, 1990). Overexploitation of this species over the past several decades has resulted in its recent designation by COSEWIC (2003) as an endangered species in part of its range.

Gadus ogac (Greenland cod) has an overlapping range with Atlantic cod with a disjunct population in the White Sea (Cohen *et al.*, 1990). It occurs from Alaska, throughout the Beaufort Sea, along the Canadian Arctic coast and northern archipelago, southward into Hudson's Bay, extending to Greenland. In the Canadian Atlantic it occurs from Ungava Bay and Hudson Strait, southward along the Labrador coast to Newfoundland, and into the Gulf of St. Lawrence, where it is widely distributed (Scott & Scott, 1988). This species inhabits cold temperate to Arctic waters. In some regions it is a more estuarine species with evidence for a tolerance of reduced salinity. Individuals are smaller than Atlantic cod, usually reaching lengths of about 70 cm. Because it generates small catches on the order of 4,000 t (Cohen *et al.*, 1990), this species is of only local importance to the fishery. Stocks throughout their range have been reduced in recent years, possibly due to competition between this species and Atlantic cod, as food sources for the two species are very similar (Nielsen & Andersen, 2001).

Melanogrammus aeglefinus (haddock) occurs on both sides of the North Atlantic from Iceland, to Spitsbergen, westward to the White Sea and Kara Sea, southward in the

Norwegian Sea and from the North Sea to the English Channel. In the western North Atlantic, occasionally found off southwest Greenland, but primarily present along the North American coast from the Strait of Belle Isle, in the eastern and southern Gulf of St. Lawrence, southward to the southern Grand Bank, along the Scotian shelf, Bay of Fundy and Gulf of Maine to Georges Bank (Scott & Scott, 1988). Haddock have been an important target for North Atlantic fisheries and generates about 400,000 t of annual catches (Cohen *et al.*, 1990). It remains the most economically important commercial fin-fishery in Atlantic Canada (cf. O'Reilly *et al.*, 2002). However, stocks of haddock were overexploited in the 1960s and stock biomass in most regions has declined with accompanying changes in age-structure, reductions in size, condition, and maturity (Department of Fisheries and Oceans, 2000; O'Reilly *et al.*, 2002).

Merlangius merlangus (whiting) is found in the eastern North Atlantic from the southeastern Barents Sea, around Iceland and Scandinavia, also found in the Black, Aegean and Adriatic seas in the eastern Mediterranean (Cohen *et al.* 1990). Kabata (1967) and Hislop & Mackenzie (1976) proposed the existence of several stocks of whiting in the North Sea, however, limited molecular studies have provided unclear results (Rico *et al.*, 1997). These stocks are believed to be due to a geographical barrier at the Dogger Bank running across the middle of the North Sea north-east to south-west, however limited information is available regarding the dispersal of whiting, so it is difficult to determine if this differentiation reflects a lower level of gene flow (Rico *et al.*, 1997). Additionally, the biology and demography of whiting in the Irish Sea are poorly documented (Gerritsen *et al.*, 2003). Limited tag-recapture data suggest only limited

interchange with surrounding regions, however, there is no published evidence for more than one genetically distinct stock within the Irish Sea (Gerritsen *et al.*, 2003), however whiting from the Black Sea and adjacent waters are treated as separate subspecies by most ichthyologists (Cohen *et al.*, 1990). Whiting are most commonly found at depths from 30 to 100 m. Fish size is usually under 70 cm (Cohen *et al.*, 1990). Whiting is a relatively common fish taken in European waters, with catches of approximately 150,000 t (Cohen *et al.*, 1990).

Microgadus proximus (Pacific tomcod) extends from central California as far north as the Gulf of Alaska, and the southeastern Bering Sea. This species is not abundant off British Columbia but distributed along the whole coast (Hart, 1973). The species generally reaches about 30 cm in length and is found at depths of 25 to 120 m. It is of limited value since it is not believed to be very abundant, however it is a highly esteemed fish among anglers. The Atlantic tomcod (*Microgadus tomcod*) is the only other species in the genus, and is a popular sports and food fish. It was not included in the present study.

Pollachius virens (saithe) occurs on both sides of the North Atlantic from Iceland, Spitsbergen, Bear Island, and Barents Sea, southward to Bay of Biscay. In the western North Atlantic from southwestern Greenland, off Labrador and around Newfoundland and the Grand Bank, present in the Gulf of St. Lawrence, along the Scotian Shelf, Bay of Fundy and Gulf of Maine to about Cape Hatteras, NC (Scott & Scott, 1988). Saithe reach nearly 130 cm in length and is an important commercial species, similar to cod and haddock, with catches of about 400,000 to 500,000 t (Cohen *et al.*, 1990). It occurs in

inshore and offshore waters at about 200 m depth. A sister species, *Pollachius pollachius*, was not included in this study.

Theragra chalcogramma (Walleye or Alaskan pollock) has a distribution similar to that of *G. macrocephalus*, found on both sides of the North Pacific, extending from central California through the Bering Sea to St. Lawrence Island on the Asian coast to Kamchatka, Okhotsk Sea, and southern Sea of Japan (Hart, 1973). Several populations have been recognized as species or subspecies around the North Pacific Basin. Analysis led to the conclusion that these distinctions were not justified and only one species is recognized in the Pacific (Hart, 1973). Genetic evidence suggests that *T. chalcogramma* (like *G. macrocephalus*) is primarily divided into Asian and North American populations (Olsen *et al.*, 2002). These fish inhabit basin, slope, and shelf waters at depths from 30 to over 400 m. Their abundance is greatest in the eastern Bering Sea. They reach sizes of approximately 80 cm and the species contributes to the largest of all demersal fish resources with an annual catch of 6,703,868 t reported in 1987 (Cohen *et al.*, 1990). Another species belonging to the genus, *Theragra finnmarchica* or Norwegian pollock, is found only off the northern tip of Norway. This species is of no interest to fisheries (Cohen *et al.*, 1990) and was not included in this study.

1.3 Phylogenetic Relationships Among Gadiform Fishes

Despite their abundance and relative importance, there has been little agreement among biologists about the classification of and phylogenetic relationships among the gadiformes, and even discussions regarding which species should be included in this

group. Descriptions of morphological or life-history characteristics (Marshall & Cohen, 1973) that apply to virtually all species thought to belong to the order have been written, but not all of these characteristics apply to, and only to the included species (Cohen *et al.*, 1990). Several studies (Marshall & Cohen, 1973; Rosen, 1973; Cohen, 1984) have failed to characterize unique traits of gadiformes or included few diagnostic features of adult gadiformes. Evolutionary relationships among gadiform taxa have been analyzed with a variety of developmental and/or morphological approaches including early life history stages (Dunn & Matarese, 1984; Fahay & Markle, 1984; Dunn, 1989), otoliths (Nolf & Steurbaut, 1989), musculature (Howes, 1989, 1991), and osteological characters (Cohen, 1984; Dunn, 1989; Markle, 1989). More recently, molecular approaches involving DNA sequence data have been used for phylogenetic analyses within families of Gadidae (Carr *et al.*, 1999) and Macrouridae (Morita, 1999).

The first comprehensive classification that addressed phylogenetic relationships among the Gadidae was devised by Svetovidov (1948) and based upon osteological characteristics, primarily the structure of the median fins. He proposed a graded series of 22 genera which can be interpreted as a phylogenetic hypothesis of successive taxa from most primitive to the most advanced. He assigned these 22 genera to three subfamilies: Lotinae, Merlucciinae, and Gadinae. Svetovidov placed Lotinae as the outgroup of the family, with Merluccinae being the sister group to the Gadinae (Figure 1). Among the Gadinae, he placed the genus *Trisopterus* as the sister group to the remaining taxa and indicated a close relationship between *Microgadus* and *Eleginus* with *Gadus* as an intermediate to *Melanogrammus* and *Gadiculus*. Finally, he recognized *Boreogadus,*

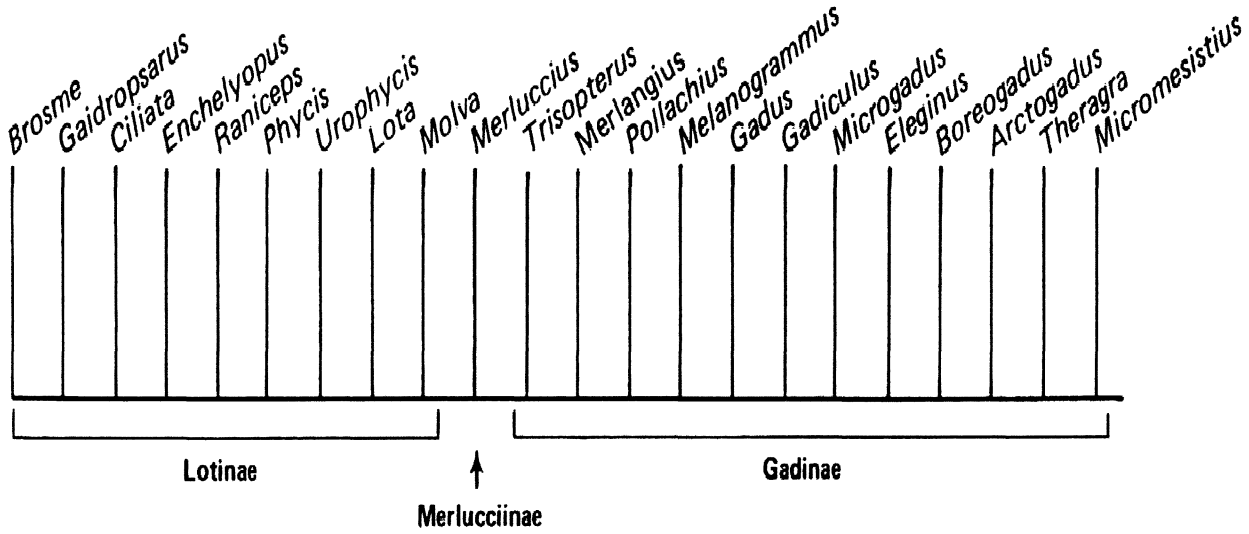


Figure 1. Relationships of the family Gadidae as proposed by Svetovidov (1948).

Arctogadus, and *Theragra* as a closely related group, with *Micromesistius* as the most derived gadine. Since Svetovidov's (1948) work, there has been considerable interest in the systematics and classification of Gadiformes (Rosen & Patterson, 1969; Marshall & Cohen, 1973; Markle, 1982; Dunn & Matarese, 1984; Fahay & Markle, 1984; Cohen, 1989; Dunn, 1989; Carr *et al.*, 1999; Møller *et al.*, 2002).

Markle (1982) addressed Svetovidov's subfamilies Lotinae and Gadinae from an early life history perspective to draw attention to early developmental characteristics that might be diagnostic for gadid subfamilies, genera, or species. He used larval and juvenile characteristics, and these led him to include three subfamilies in the family Gadidae: Phycinae, Lotinae, and Gadinae. The status of Merluccius and its relatives was not addressed because Markle considered the group a family distinct from Gadidae. However, he noted that many problems existed, in that autapomorphies have consistently defined gadines, and that many synapomorphies unite gadines and lotines. These include a similar egg diameter, loss of the X/Y bones, and the apomorphic state of certain meristic characters such as high number of caudal fin rays, low number of primary caudal fin rays, and high number of precaudal vertebrae (Markle, 1982). Convincing evidence of autapomorphies defining phycines was lacking, and in conclusion Markle (1982) provided what could be considered a tentative summary of characteristics of gadid subfamilies based on early developmental characteristics, rather than a phylogenetic interpretation.

The most recent analysis of gadid relationships using morphological/ developmental characteristics was by Dunn (1989). He conducted a cladistic analysis of

11 genera of gadid fishes, using osteological characters and found 28 of 42 examined characteristics to be decisive (Figure 2). In contrast to Svetovidov, Dunn (1989) placed *Gadiculus* as the sister group to all other members of the subfamily Gadinae. As well, Dunn found a close relationship among *Gadus, Theragra,* and *Boreogadus* (*Arctogadus* was not included in his study), while Svetovidov (1948) separated *Gadus* from *Theragra* and *Boreogadus*. Finally, Dunn separated the close relationship between *Microgadus* and *Eleginus* as depicted by Svetovidov (1948). In contrast to Svetovidov's series of relationships, Dunn (1989) included *Micromesistius* as the most derived taxon. However, Dunn also acknowledged that analyses of gadiform morphological data are troubled with a high degree of homoplasy (parallel change) and difficulties in determining character polarity. Thus morphological similarities between groups may have arisen independently of one another and therefore not necessarily correspond to a shared ancestry. As well, if character polarity cannot be determined, then one cannot establish whether a trait is ancestral or derived. This, along with a high level of reversals and losses of character states that Dunn observed, indicates the need for studies of additional gadid characters.

It seems generally agreed that the Gadidae is the most derived family in the order. However, several families have been suggested to be the sister group of Gadidae: Lotidae (Markle, 1989), Phycidae (Nolf & Steurbaut, 1989), and Merlucciidae (Dunn & Matarese, 1984; Howes, 1991). The family Merluccidae is considered by some to be a subfamily of the Gadidae while others recognize it as an entirely separate family. Some authors even suggest that the three most recognized subfamilies of Gadidae (Phycinae, Lotinae, and Gadinae) be given family status (Cohen, 1984; Dunn, 1989; Markle 1989;

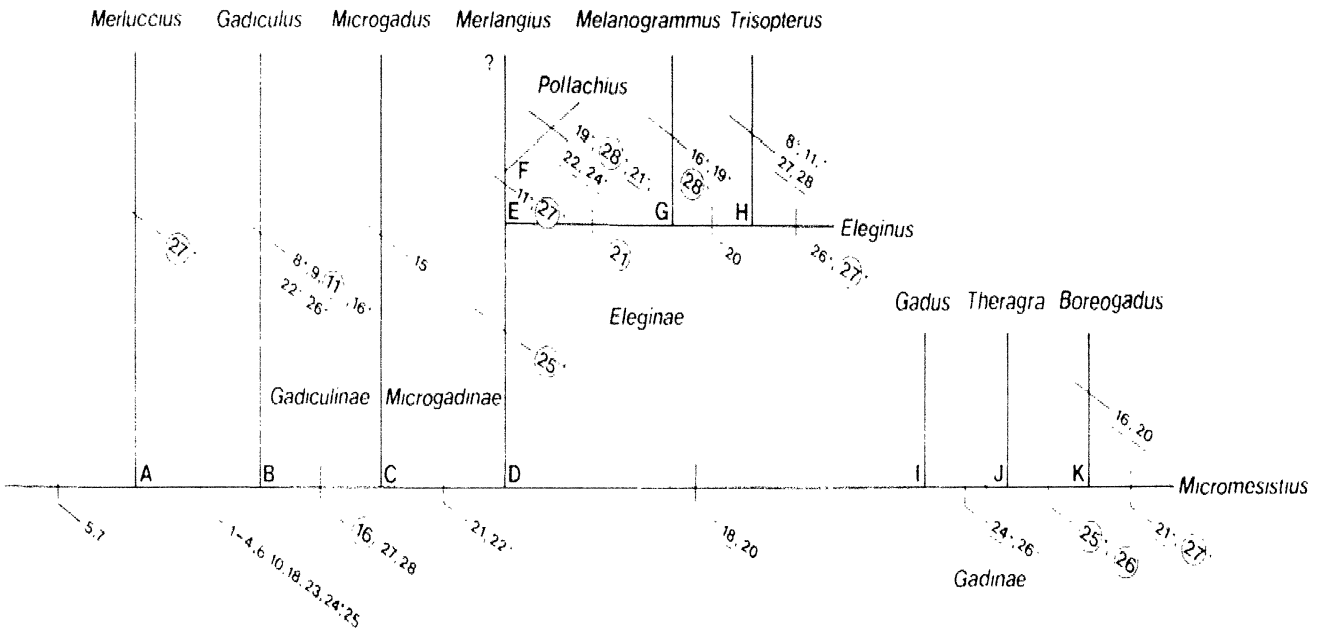
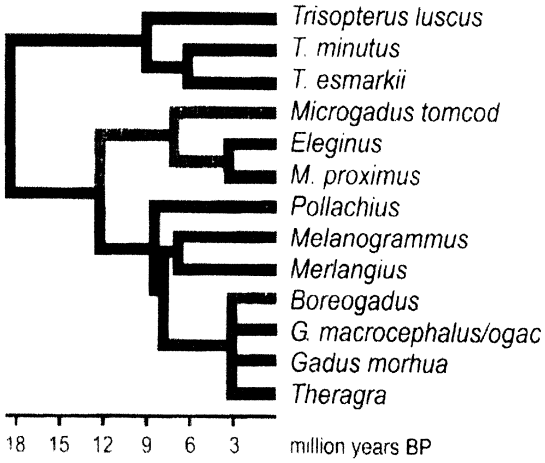


Figure 2. Dunn's cladogram of *Merluccius* and 11 gadid genera (Figure 29 of Dunn 1989). Asterisks represent homoplastic characters.

Howes, 1991). Molecular work using a 397 bp region of the 12S rRNA and a 401 bp region of the cytochrome *b* mitochondrial gene seems to support Merlucciidae as a family separate from Gadidae within the order (Warren, 1999). As well, this study placed the phycines as the sister group to the other two subfamilies, the lotines and the gadines. This is in accordance with the hypothesis put forth by Markle (1982).

To date, the only molecular analysis conducted on gadid relationships was that by Carr *et al.* (1999) who investigated phylogenetic relationships among 14 species of gadid fishes (including all of those studied by Dunn, (1989)) using two mitochondrial genes, a 401 bp segment of the cytochrome *b* gene, and a 495 bp segment of the cytochrome oxidase I gene. Their results identified 4 main clusters: (i) three species of pouting, *Trisopterus*, (ii) two species of tomcod (*Microgadus*) plus *Eleginus navaga*, (iii) *Melanogrammus aeglefinus* and *Merlangius merlangus* and (iv) three species of *Gadus* (cods) plus *Boreogadus saida* (polar cod) plus *Theragra chalcogramma* (Alaskan pollock) (Figure 3a). Maximum Likelihood (ML), Maximum Parsimony (MP) and Neighbour-Joining (NJ) methods all produced trees with similar topologies, however there was a lack of consistent bootstrap support for the deeper branches among the different analyses. Despite the fact that 12 of 16 internal nodes were supported in more than 50% of bootstraps, only two exceeded 95% (Carr *et al.*, 1999). As well, the relationships among the three species of *Gadus*, plus *Theragra* and *Boreogadus* were unresolved (Figure 3b). Like Svetovidov (1948), Carr *et al.* (1999) identified *Trisopterus* as the outgroup to the remaining taxa, and observed a close relationship between *Melanogrammus* and *Merlangius*. The placement of *Pollachius* was indeterminate. In

(A)



(B)

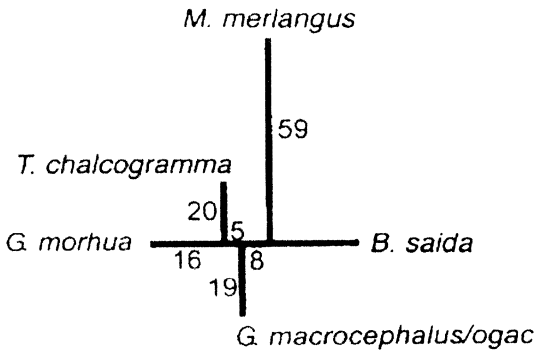


Figure 3. (A) Phylogenetic relationships among 14 gadid genera analyzed by Carr *et al.* (1999) among 896 bp of mtDNA (B) Number of differences among an unresolved clade composed of *Boreogadus, Theragra* and the three nominal species of *Gadus*.

contrast to Svetovidov (1948), Carr *et al.* (1999) recognized a group including *Boreogadus, Gadus,* and *Theragra* but the relative branching order within this group could not be resolved. Several species, including *Gadiculus*, *Micromesistius*, and *Arctogadus*, were not included by these authors. A main point of interest reported by Carr *et al.* (1999) was that the sequences of both gene segments of *G. ogac* and *G. macrocephalus* were identical. Carr *et al.* (1999) suggested that *G. ogac* may simply be a northward and eastward extension of the range of *G. macrocephalus*, and that gene flow may continue throughout this region.

When the molecular data was compared to Dunn's (1989) 28 morphological characters, there is little or no similarity between Dunn's cladogram and the evolutionary relationships implied by the molecular data (with the exception of the close relationship among *Gadus, Theragra* and *Boreogadus*). Carr *et al.* (1999) concluded that the distribution of nucleotides among their species could be explained with no more than 450 substitutions, whereas the minimum-length morphology tree (equivalent to Dunn's hypothesis) would require at least 502 additional changes. They concluded that although morphological data are compatible with the molecular tree, the reverse is not true (Carr *et al.*, 1999).

A number of unresolved questions pertaining to the phylogeny of this order remain. Classification and taxonomic studies in the past have largely relied on morphological and developmental characteristics, however neither internal nor external characters exist that uniquely define the order Gadiformes. As well, there is no great

measure of agreement with respect to the hierarchical arrangement of genera into suborders, families, and subfamilies (Cohen *et al.*, 1990).

1.4 Molecules vs. Morphology in Phylogenetic Analysis

Some of the most important diagnostic characters that have been used for species identification and for resolving phylogenetic relationships are internal features (such as the presence or absence of an oil globule in the egg, the shapes and positions of various bones, the shape of the swimbladder and the shape and number of vertebrae), and/or external features (such as the number and shape of dorsal and anal fins, the presence or absence of chin barbel, and the shape of the caudal fin) (Nelson, 1984). Analysis of internal features often requires careful dissection and observation. Obtaining information from external characteristics may also be problematic. For example, determination of the number of dorsal and anal fins may seem simple, however, for many genera care must be exercised as the prominent fin lobes of some genera (e.g., *Merluccius* and *Mora*) are separated by very short rays and the lobes may actually appear to be separate fins. As well, in genera such as *Trisopterus* and *Molva*, fins that are structurally separate are immediately adjacent at their bases and may appear to be lobes of a single fin (Cohen *et al.*, 1990).

Eggs and larvae have also been a potential source of information useful to the study of population dynamics and systematics. For some species, there is considerable information of this type, however the taxonomy of the early life history stages of gadiformes is known for less than a third of the species (Cohen *et al.*, 1990), and

therefore some species are not identifiable at early larval stages. For example, *Gadus ogac* and *Gadus morhua* are difficult to distinguish where they co-occur in coastal habitats as recently settled juveniles. Hovgard & Lehman (1986) reported difficulty in identifying Atlantic cod from Greenland cod on the basis of colouration and lateral line patterns in fish under 200 mm standard length (SL) without the use of protein electrophoresis. It is believed that ichthyoplankton surveys and surveys which target demersal juvenile cod (*G. morhua*) less than 2-3 years of age may be biased owing to such misidentifications. Many of the small cod that are common around wharves and harbours in northeast Newfoundland that are thought to be *G. morhua* are likely a mixture of age 0 and age 1 Greenland and Atlantic cod (Methven & McGowan, 1998). Methven & McGowan (1998) used protein electrophoresis to identify *G. morhua* and *G. ogac* individuals and then assayed the ability of 14 meristic and morphometric measures to correctly identify each species. These authors found that only eight of the fourteen measurements studied were helpful in distinguishing between these two species, but only for lengths greater than 87 mm (Methven & McGowan, 1998). Several of these characters were of no diagnostic use.

Another problem with using strictly morphological/developmental data is determining the relative importance of characters. For example, Markle (1982) noted that phycine eggs are the smallest, while gadine and lotine eggs are similar in size, but differ in the presence (lotines) or absence (gadines) of oil globules. As well, the internal structure of the first and second dorsal fins suggests a phycine-gadine relationship, while the loss of X and Y bones suggest a lotine-gadine relationship. Thus, difficulty arises as

to which of these traits, dorsal fins or X and Y bones, is a better indication of shared history. Marshall & Cohen (1973) noted that gadiform families manifest a mosaic of primitive and advanced characters. The contribution made by Markle (1982) leaves the subfamily relationships in the same predicament and he suggests that instead one can use early developmental stages and characteristics as a tentative summary of diagnostic characters for gadid subfamilies.

It has been thought that morphological studies have generally been successful in defining the species composition of genera; however it has been rare to find studies that present hypotheses of relationships above this level, primarily due to a lack of correspondence of characters (Stepien & Kocher, 1997). However, among the gadines, Carr *et al.* (1999) found several cases where morphology has misplaced species and erected genera where they were not found to occur based upon molecular data. Therefore, for some groups, morphology may be more successful at defining monophyletic families than monophyletic genera. This is one of the strong points of molecular analysis. DNA sequence analysis allows identification at all developmental stages, and would be of great use in conjunction with developmental and morphological data. However, most molecular studies on gadiformes have addressed intraspecific relationships among populations comprising a species (Grant *et al.*, 1987; Dahle, 1991; Carr *et al.*, 1995; Arnason *et al.*, 1998; Fevolden *et al.*, 1999; Mattiangeli *et al.*, 2000), rather than interspecific relationships. Recent molecular work (Bakke *et al.*, 1999; Bakke & Johansen, 2002) using both mitochondrial DNA sequence data from protein-coding genes, as well as secondary structure analysis of mitochondrial ribosomal subunits has

largely focused on resolving relationships among families and higher level phylogenies with limited resolution or focus on species relationships comprising a family.

In some studies of mtDNA, restriction fragment length polymorphisms (RFLPs) have been analyzed (e.g., Dahle, 1991). However in applying this approach to species and higher-level systematic questions, the homology of restriction site characters becomes less certain. Carr & Marshall (1991) compared sequence data for a 307-bp region of the cytochrome *b* gene from *G. morhua*. Among 400 cod, 23 variant genotypes were observed. Of the 24 known variable nucleotides within this region, only four fall within restriction sites such that they would appear as RFLPs. As well, when *G. morhua* was compared with *G. ogac*, none of the 15 nucleotide differences between these species would be detectable as RFLPs. Therefore one loses a substantial amount of potential phylogenetic signal. Other studies have examined allozyme electrophoretic data, however, only substitutions which result in an amino acid replacement may present themselves as variants. Silent substitutions that result in no amino acid replacement are not detected, although a mutational event has occurred. Both RFLPs and allozymes are indirect approaches of determining sequence variation, and therefore a better approach involves the direct analysis of DNA sequences.

It is well established that molecules, particularly mtDNA, are well suited for reconstructing evolutionary relationships. For species or species groups with little or no phylogenetically informative morphological variation, molecules can provide a taxonomy that can be easily and immediately placed in a phylogenetic context. Thus molecules provide an objective measure of the geographic scale over which phylogenetically

informative differentiation is occurring (Bermingham *et al.*, 1997). Typically, morphological studies use tens of characters, however for DNA sequence data, each nucleotide is considered as a separate character and any one nucleotide is more or less like any other, thereby allowing for the assessment of thousands of potential characters in a phylogenetic context. As well, patterns of nucleotide substitutions are well known, so the problem of character importance is less consequential. The large number of characters available in molecular analyses is useful for studying relationships among both close and distant relatives. By confining attention to silent substitutions, which accumulate rapidly, one can focus on recent branches of the tree, whereas with more slowly evolving substitutions, such as ones which cause amino acid replacements, one can focus on ancient times of divergence as well as the rooting of the tree (Wilson *et al.*, 1985).

Considering the enormous species diversity of fish, their ancient origin and wide-ranging variations not only in morphology but also in behaviour, ecology and physiology, it is no wonder that comparative anatomical approaches have faced a number of difficulties in determining phylogeny. However, the same can be said for traditional molecular phylogenetic studies that have employed only short segments of complete genes or partial gene sequences (generally < 1,000 bp). By using the complete mtDNA genome, one will access data relevant to both recent and ancient divergences. Problems associated with homoplasy and low levels of significance for internal branches will be ameliorated.

1.5 Molecular Phylogenetics

Three factors have made the construction of phylogenetic trees more reliable: the availability of molecular data, the use of computers to handle large bodies of data, and the development of a new theory of classification. Phylogenetic trees can be useful in determining the mechanism of speciation. This can tell us about the way in which geographical variation has arisen within a species and sometimes can reveal incipient speciation (Smith, 1999).

Since the rate of sequence evolution varies extensively with the gene or DNA segment, one can study the evolutionary relationships across many levels of classification of organisms by using different genes or DNA segments. In principle, a DNA sequence splits into two descendant sequences at the time of speciation or gene duplication. Therefore, phylogenetic trees are usually bifurcating. Many different tree-building methods are available for use with molecular data. Each of these methods has advantages and disadvantages and the relative efficiencies in constructing correct phylogenetic trees is still controversial (Nei, 1991).

Reconstruction of phylogenetic trees by using statistical methods was initiated in numerical taxonomy using morphological traits (Sokal & Sneath, 1963) and in population genetics using allele frequency data (Cavalli-Sforza & Edwards, 1964; Nei & Kumar, 2000). DNA sequence data are well-suited for phylogenetic analyses, as DNA evolves in a relatively regular manner and various mathematical models can be applied to the evolution of its character states. Characters used in phylogenetic analyses must be homologous among all included taxa, thereby making these characters related through

common ancestry. Homologous DNA sequences can be related by orthology (common ancestry traced back to a speciation event) or paralogy (common ancestry traced back to a gene duplication event). While homologous DNA sequences can arise via either of these methods, only orthologous sequences are useful for establishing phylogenies. The most common methods for estimating phylogenies fall into three main classes: maximum parsimony methods, distance methods, and maximum likelihood methods. Each of these types of approaches is discussed below.

1.5.1 Maximum Parsimony

In maximum parsimony (MP) methods, each nucleotide site is evaluated for its use as an indicator of phylogeny. A number of trees are constructed based on nucleotide sites which are phylogenetically informative and the tree or trees requiring the fewest number of substitutions are chosen to be the most parsimonious. This approach operates on a minimalist principle and states that one should prefer simpler explanations (requiring fewer assumptions), over more complex ones (Steel & Penny, 2000). It has also been argued that another appeal to parsimony methods is that they assume as little as possible about any underlying model or mechanism for evolution. Sober (1988) stated that the less we need to know about the evolutionary process to make a phylogenetic inference, the more confidence we can have in our conclusion.

If there are no parallel mutations (homoplasy), then MP methods are expected to produce the correct tree. However, parallel mutations do occur and MP can then tend to give incorrect results. To avoid this scenario, weighted parsimony can be performed,

giving a higher score for a transversion which occurs at a lower frequency than a transition. Felsenstein (1978) however, has shown that when the rate of substitution varies extensively among evolutionary lineages, MP methods may generate incorrect topologies even if an infinite number of nucleotides are examined (Nei & Kumar, 2000).

1.5.2 Distance Methods

In distance methods, evolutionary distances are computed for all pairs of taxa and a phylogenetic tree is constructed based on these distance values. The simplest and most common distance methods are the unweighted pair-group method using arithmetic averages (UPGMA), and the neighbour-joining (NJ) method. Generally, the accuracy of an inferred tree depends on two factors: (1) the linearity of the relationship between the distance used and the number of substitutions and (2) the standard error or the coefficient of variation of the estimate of the distance measure (Nei & Kumar, 2000). Distance measures for estimating the number of substitutions are based on mathematical models and require many parameters to be estimated, thus increasing the variance on the estimate of d, the distance measure. However, it is possible to choose a model most appropriate for a given set of data using certain criteria (for example, the Tamura-Nei distance measure was invented specifically for mitochondrial sequence data). A drawback to distance measures is that they are more subject to parallel substitutions (homoplasy), however methods exist for minimizing this effect (e.g. Poisson correction for multiple hits).

1.5.3. Maximum Likelihood

In maximum likelihood (ML) approaches, the likelihood of observing a given set of sequence data for a specific model of substitution is assessed for each topology and the topology with the highest likelihood is chosen as the most likely tree. The parameters considered in these models are not the topologies but rather the branch lengths for each topology, and the likelihood is maximized to estimate branch length (Nei & Kumar, 2000). Unfortunately, the likelihood of a given tree requires a series of parameters to be specified. These parameters need to be considered in order to determine the correct topology under a likelihood approach. Parameters that may need to be estimated include substitution rates and ratios, base composition, and how rates of substitution vary across sites (gamma parameter, γ).

Various methods are employed in constructing phylogenies from genetic data and opinions differ as to which methods work best for a given set of relationships. By employing complete mtDNA genomes, differences between methods may not be as obvious and phylogenetic interpretations may be more congruent. By assessing individual gene regions for phylogenetic accuracy, we may be able to identify which methods perform most accurately. If different methods of tree reconstruction return the same (or very similar) topologies, we can have more confidence in the relationships implied by the molecular data.

1.6 Mitochondrial DNA Organization and Content

The mitochondrion is a eukaryotic organelle possessing its own genome and is responsible for providing cells with energy in the form of ATP. It is widely accepted that mitochondria are descendents of an α-proteobacterium which formed a symbiotic relationship with a primitive eukaryotic cell (Gray *et al.*, 1999b; Saccone *et al.*, 2002). The vertebrate mitochondrial genome is a single, double-stranded, circular molecule with a narrow size range of ~16-17 kbp. Some exceptions include varying sizes from 14 kbp in the nematode, *Caenorhabditis elegans*, to 42 kbp in the scallop, *Placopectin megallanicus* (Scheffler, 1999). There are also a few exceptions of linear and multiple linear molecules in some metazoans. The first mtDNA genome completely sequenced was that of humans (Anderson *et al.*, 1981), and now complete mitochondrial DNA sequences have been reported for more than 60 species representing all major groups of vertebrates (Saccone *et al.*, 1999).

Mitochondrial genomes contain several unique properties. Among metazoans these genomes have been found to contain the same set of genes, with very few exceptions. These include two species of ribosomal RNA (rRNA) genes, 22 distinct types of transfer RNA genes (tRNAs), and 13 protein-coding genes. The latter all encode products of the inner mitochondrial membrane and are involved in oxidative phosphorylation and the electron transport chain. Even slight alterations to this conserved gene content are rare and limited to a few nematodes, molluscs and cnidarians (Boore, 1999). Explanations for the maintenance of the same gene order among very different lineages include: (1) the redox potential of the organelle may regulate the

expression of the genes involved in the electron transfer itself, and (2) that due to the toxicity of some mitochondrial components, they need to be kept in a separate cellular compartment (Race *et al.*, 1999; Saccone *et al.*, 2000).

Despite the conservative nature of gene content, gene order varies, with substantial gene rearrangements occurring among more distantly related lineages (Saccone *et al.*, 1999). Gene rearrangements may occur by tandem duplication of gene regions because of slipped-strand mispairing, followed by the deletion of genes (Macey *et al.*, 1997). In this case, a segment containing two or more genes is duplicated and subsequent random loss of the now supernumerary genes may or may not result in exchange of position (Boore, 1999). In cases where the size range deviates from the typical size, this is a result of increases in non-coding, repetitive regions rather than additional sets of genes. In many major groups, it appears that gene rearrangements are not stable within those groups and therefore it is suggested that phylogenetic studies based on gene rearrangement comparisons (e.g., Boore & Brown, 1998) may be very misleading and may be subject to parallel evolution due to mechanistic constraints (Mindell *et al.*, 1998; Saccone *et al.*, 1999). Gene rearrangements appear to be more common in protostomes, with limited gene arrangements observed in echinoderms and only a few reported transpositions documented within the phylum Chordata (Saccone *et al.*, 1999). The mitochondrial genome in metazoans lacks introns, although most genes have a few intergenic nucleotides and other genes may overlap by several base pairs. A possible explanation for the conservation of mitochondrial gene organization seen within animal phyla may be that the loss of introns and intron-encoded DNA recombinases led

to the loss of genetic recombination in mitochondrial genomes of metazoans (Kotylak *et al.*, 1985). As for the relatively compact, constant size of the mitochondrial genome, it has been hypothesized that a small genome would replicate faster and be more successful than a large genome. This is however very difficult to demonstrate experimentally (Saccone *et al.*, 1999). It is also known that many sequences present in the early mitochondrial genome have been lost or transferred to the nuclear genetic component of the cell. The asexual mode of transmission observed in mitochondria may lead to a faster accumulation of deleterious mutations, which could increase the mutational load on genes in the mitochondria by reducing both the effective population size and the intensity of purifying selection for mitochondrial genomes. If this bias were to exist, this might drive the transfer of sequences from the high-load (mitochondria) to the low-load compartment (nucleus) with consequent improvement in cell fitness (Berg & Kurland, 2000).

The mitochondrial genome has many properties which make it useful for reconstructing phylogenetic history. In addition to being a small, easily accessible molecule, it generally follows a clonal, maternal mode of inheritance and is therefore haploid and effectively nonrecombining. The high degree of conservation provides homologous genes for comparison among diverse groups of organisms. The evolution of the molecule therefore corresponds exactly to the model of bifurcating evolutionary trees (Stepien & Kocher, 1997).

Secondly, mtDNA evolves more quickly than most nuclear genes, allowing for the identification of informative phylogenetic characters among even closely related species and populations. This may be because the mtDNA does not code for any proteins

associated with its own replication, transcription, or translation. Therefore replication errors are more tolerable than in nuclear DNA and fewer mechanisms exist for the repair of such errors (Wilson *et al.*, 1985). This results in a higher effective rate of mutation, and thus rate of evolution, than in nuclear DNA. As well, this increase in mutation rate may be due to reduced functional constraints on mitochondrial gene products (Asakawa *et al.*, 1991). An estimate of the initial rate of sequence divergence is 20×10^{-9} per site per year, or about 5-10 times faster than the highest rates in nuclear DNA (Smith, 1999). It is also observed that mtDNA evolves at slower rates in ectotherms as compared to endotherms (Rand, 1994). For example, amino acid substitutions in fish mtDNA encoded proteins are about five times lower than among mammals (Kocher *et al.*, 1989). Although the salmonid and cod fishes are very distantly related, a high degree of homology can be found in the amino acid sequences of several mtDNA protein-coding genes. This homology is much higher than among mammals (Thomas & Beckenbach, 1989). The base composition among the Metazoa is also highly variable. The GC content is extremely low in the genomes of insects and nematodes, and increases in vertebrates, reaching the highest value in birds, mammals and teleostei (Saccone *et al.*, 1999). Although mtDNA has an increased rate of mutation compared to its nuclear counterpart, this rate of mutation varies throughout the genome depending on structural and functional properties of the gene products.

rRNA genes

There are two ribosomal RNA (rRNA) genes found in vertebrate mitochondrial genomes – a small ribosomal subunit (12S) and a large ribosomal subunit (16S). The rRNA molecules encoded by these genes form complex secondary structures consisting of paired, helical stem regions and unpaired loop regions. These molecules are folded into specific secondary structures and along with ribosomal proteins make up the mitochondrial ribosomes. Mitochondrial rRNAs are smaller and simpler than nuclear rRNAs but are estimated to evolve 5-20 times faster than those found in the nucleus (Saccone *et al.*, 1999). They are, however, among the slowest evolving mitochondrial genes, as their sequence evolution is greatly influenced by their secondary and tertiary level structure. The rRNA genes are more susceptible to insertion/deletion (indel) events than tRNA and protein-coding regions. In divergent taxa, rDNA alignment may be difficult due to the presence of such length variants. Nonetheless, sequences from rRNA genes have been used to assess phylogenetic relationships of more divergent taxa due to their slower rate of evolution (Bakke, 2001).

tRNA genes

The mitochondrial genome codes for a complete but reduced set of transfer RNA genes (tRNAs), which function to translate the mitochondrially-encoded proteins. These tRNA genes have a cloverleaf secondary structure and are much less conserved than nuclear tRNAs, having a rate of sequence evolution approximately 100 times more rapid (Saccone *et al.*, 1999). Like rRNA genes, however, they are some of the slowest

evolving mitochondrial genes and the substitution rate varies within and among these genes according to their levels of functional constraint. These tRNA genes are most often involved in gene rearrangements within the mitochondrial genome as discussed above. It has even been hypothesized that mitochondrial genes flanked by two tRNAs can be considered very similar to a transposable element, with the tRNAs corresponding to long terminal repeats each having a short inverted sequence at their ends (Saccone *et al.*, 1999). Mitochondrial tRNA genes have been used less often than rRNA or protein-coding genes for phylogenetic studies, possibly due to their high degree of conservation and small size (< 100 bp each). They have in some cases proven useful for resolution of deep branch phylogenies (Kumazawa & Nishida, 1993; Miya & Nishida, 2000; Bakke, 2001).

Protein coding genes

Vertebrate mtDNA genomes contain 13 protein-coding genes distributed as follows: two subunits of ATP synthase (ATP6 and ATP8), three subunits of the cytochrome oxidase complex (CO1, CO2, & CO3), a cytochrome *b* gene, and seven nicotinamide adenosine dehydrogenase (NADH) subunits (ND1, ND2, ND3, ND4, ND4L, ND5 and ND6). Length variation is relatively uncommon in most of the mtDNA genome, as it would result in a frameshift mutation in a protein coding gene. The pattern of molecular evolution in the protein-coding genes is strongly affected by codon position. Synonymous sites (3^{rd} position of codons and some 1^{st} positions) are among the fastest evolving components of the mtDNA genome, and evolve about 22 times faster than

nuclear synonymous substitutions (Saccone *et al.*, 1999). Due to this rapid rate of evolution, third codon positions are often useful for studies of recent divergences, however in more ancient divergences they quickly become saturated and may therefore show much higher levels of homoplasy. Non-synonymous substitutions (those causing amino acid changes) occur less frequently, and are affected by the structure and function of the polypeptide. They are among the slowest evolving components of the mitochondrial genome and have a substitution rate more comparable with non-synonymous nuclear protein-coding genes (Saccone *et al.*, 1999).

Among these protein-coding genes, ATP8 and ATP6 are the most variable, followed by the genes for the proteins from the NADH dehydrogenase complex. A lower degree of variability is found in genes encoding cytochrome *b*, while the subunits of the cytochrome *c* oxidase complex have the lowest amount of sequence variation (Pesole *et al.*, 1999; Saccone *et al.*, 1999). Among the vertebrates, the cytochrome *b* gene has been used quite extensively in phylogenetic analysis, partially due to the development of "universal" primers for this region (Kocher *et al.*, 1989). However, several studies have also suggested that ND4 and ND5 perform best in determining phylogenies among vertebrates (Russo *et al.*, 1996; Miya & Nishida, 2000).

Control region

The compact organization found in mtDNA usually allows for only one major non-coding region, the control region, which contains the displacement loop (D-loop). It is in this D-loop that replication of the mtDNA is initiated at the origin of heavy strand

replication (Johansen *et al.*, 1990). In vertebrates, the control region varies from 850-1,500 bp (Lee *et al.*, 1995; Sbisa *et al.*, 1997; Bakke, 2001). Length variation is common due to the presence of the repeated sequences that have been documented in a number of species (Hoelzel *et al.*, 1994). The most accepted hypothesis for the mechanism involved in generating length variation is strand slippage and mispairing during replication (Bakke, 2001).

The control region can be divided into three sections. The central domain of the control region, thought to play a role in the regulation of heavy-strand replication (Clayton, 1982), is highly conserved and has a rate of evolution comparable to that of rRNA and tRNA genes. The regions flanking the central domain are hypervariable regions that can even be used to assess intraspecific phylogenies (Bakke, 2001).

1.7 Mitochondrial genomics

Much phylogenetic work has employed mitochondrial DNA. Historically, however, these studies have largely relied upon a single gene or parts of genes, with different genes being used depending on the estimated time of divergence. For intraspecific or very recent divergences, the control region has been used. For example, Faber & Stepien (1997) used the mtDNA control region to assess phylogenetic relationships among populations, species, and genera of the Percidae. They concluded that mtDNA control-region sequences are valuable for elucidating within-species and congeneric relationships, however higher level phylogenies that rely solely on control region data will probably be affected by homoplasy and may suggest unrealistic

associations among taxa. Among the protein-coding mtDNA genes, cytochrome *b* has historically received much attention. With the characterization of "universal" primers (Kocher *et al.*, 1989) for cytochrome *b*, it became possible to survey sequence variation in a number of taxa across all major groups. Lydeard & Roe (1997) studied the phylogenetic utility of the mitochondrial cytochrome *b* gene among actinopterygian fishes. Their data suggests that cytochrome *b* offers substantial phylogenetic information for assessing relationships over a broad taxonomic range, however many criticisms exist regarding the utility of the gene based upon limited and/or improperly analyzed data, base compositional bias, rate variation among lineages, saturation of third codon positions, and limited variation in first and second codon positions (Meyer, 1994). These problems are not unique to cytochrome *b*, but are general concerns regarding any single mitochondrial gene (Meyer, 1994).

Molecular systematists may need to address questions that are difficult to answer with morphological data or small molecular data sets. When a gene tree yields a poorly resolved taxonomic tree, the conclusion is often that the particular gene being studied is uninformative for that particular question. However, an alternative possibility is that the gene may in fact reveal true relationships. For example, when speciation is rapid, little time may be available for synapomorphies to accrue between nodes and the species branching may be polytomous. This has been the case within poeciliids (lampreys) (Parenti, 1981), birds (Avise, 1994), and bovine mammals (Kraus & Miyamoto, 1991). An explanation such as rapid speciation should not be concluded until sufficient morphological/molecular data have been gathered to establish the pattern. The lack of

resolution and weak support at many nodes may simply imply that no single gene will answer all these questions and that a more comprehensive, genome-based approach is required to yield the desirable resolution and identify patterns above the level of individual genes.

Both morphological and molecular studies have had particular difficulty in resolving higher-level relationships; the central problem is identifying homologous characters with a sufficient number of synapomorphies to identify lineages with statistical confidence (Stepien & Kocher, 1997). This problem is thought to be overcome with the use of large strings of nucleotides, such as complete mitochondrial genomes. Inoue *et al.* (2001b) used complete mtDNA sequences to identify the relationships among the five major basal teleostean lineages (Osteoglossomorpha, Elopomorpha, Clupeomorpha, Ostariophysi, and Protacanthopterygii). Their work revealed a topology which differed from a number of previously existing theories, however with the exception of one clade, all internal branches were resolved with high statistical support (95-100%). They therefore identified this approach as a valid one for elucidating higher-order relationships (e.g. classes, orders, and families). However, this approach can conceivably be applied at various levels of taxonomy. Ballard (2000) used complete mtDNA among several members of the *Drosophila melanogaster* subgroup (*D. mauritiana, D. simulans, D. sechellia* and *D. yakuba*) to investigate forces that shape evolutionary change (i.e. selection, recombination) among this group and to differentiate variation resulting from processes acting at a higher level (i.e. genome) from those acting on single genes. Additionally, Rokas *et al.* (2003) used a series of 106 widely distributed orthologous

nuclear genes to assess phylogenetic accuracy and power among 7 species of the yeast, *Saccharomyces*. They found that after about 20 genes were included in the phylogenetic analysis, further resolution was not possible with the addition of more genes.

Practically, a larger number of nucleotides are necessary to estimate trees which aim to resolve recent, bifurcating divergences, whereas more independent loci are needed to resolve reticulate evolutionary histories. The desirability and feasibility of employing complete mtDNA genomes rather than partial sequences lies in obtaining increased resolution of intraspecific phylogeography, increased accuracy of interspecific phylogenies, and increased understanding of variable rates and patterns of molecular evolution (S.M. Carr, personnel communication, 2001). This is because complete genome sequencing yields sets of gene sequences that can be concatenated, providing many more informative sites than individual gene sequences (Gray *et al.*, 1999a). Relative evolutionary rates of DNA sequences can only be elucidated through multispecies comparisons and such comparisons will undoubtedly help determine phylogenetic signal and utility for other regions of the mitochondrial and nuclear genomes.

Phylogenetic history and genetic diversity should be used in biodiversity to emphasize the phylogenetic and genetic distinctiveness of some groups compared to others. The resulting taxic diversity measures, when coupled with detailed knowledge of organisms' distribution patterns, can be used to identify priority areas for conservation (Vane-Wright *et al.*, 1991). As well, taxonomy itself is important because a correct understanding of phylogenetic relationships reflecting a system of classification is

essential to understanding morphological, behavioural, physiological and biogeographic evolution. By using macromolecules to infer phylogeny, and improve taxonomy, we may better be able to establish morphological and/or developmental characteristics shared by species or groups of species.

1.8 Objectives of Current Study

The aim of this study is to use mitochondrial genome sequences to resolve the interrelationships among gadine genera encompassing *Arctogadus, Boreogadus, Gadus, Microgadus, Melanogrammus, Merlangius, Pollachius* and *Theragra*. The resolution of the phylogeny of this particular group will be to address questions relating to the origins of the three endemic Pacific species, the resolution of the relationship between *Gadus macrocephalus* and *G. ogac* as well as the placement of several of the other species (e.g. *Pollachius* and *Arctogadus*). In addition to these objectives, the use of mtDNA genomic sequences, will allow for the efficiency of phylogeny reconstruction among the various gene regions to be assessed, including an analysis of variable rates and patterns of molecular evolution across loci. The use of such a large data set will allow for the resolution of significant differences in mtDNA among a set of closely related taxa, which has previously been unresolved with partial DNA sequences from one or two gene studies.

2.0 MATERIALS AND METHODS

2.1 Collection of Specimens

Specimens of *Boreogadus saida, Gadus ogac* and *Gadus morhua* were collected on the Newfoundland shelf (between 1991 and 1993) by the Department of Fisheries and Oceans (DFO), St. John's NL. *Theragra chalcogramma, Microgadus proximus* and *Gadus macrocephalus* were collected off the west coast of British Columbia (DFO, Nanaimo, BC). *Arctogadus glacialis* was collected off the coast of Greenland (DFO, St. John's). *Melanogrammus aeglefinus* and *Pollachius virens* were collected off George's Bank (National Marine Fisheries Service, Woods Hole, Mass., USA). *Merlangius merlangus* was collected by The Centre for Environment, Fisheries and Aquaculture Science, Lowestoft, England. Collecting locations for all specimens are shown in Figure 4.

2.2 DNA Extraction and Polymerase Chain Reaction (PCR)

DNA was extracted from frozen muscle or heart tissue using the QIAamp DNA Mini Kit Tissue Protocol (Qiagen Inc.). Tissue samples were digested in 180 μL of Buffer ATL (Qiagen) with 20 μL of proteinase K and incubated at 56°C for 2-4 hours or until all tissue was completely lysed. In some cases, lysis was continued overnight. For removal of DNA from the lysis mixture, the manufacturer's protocol was followed. DNA was eluted in 200 μL of Buffer AE (Qiagen Inc.) and subsequently stored at -20°C.

Primer sequences for complete mitochondrial genomes specific to *G. morhua* (GenBank accession number NC002081) were chosen by comparison to three outgroup

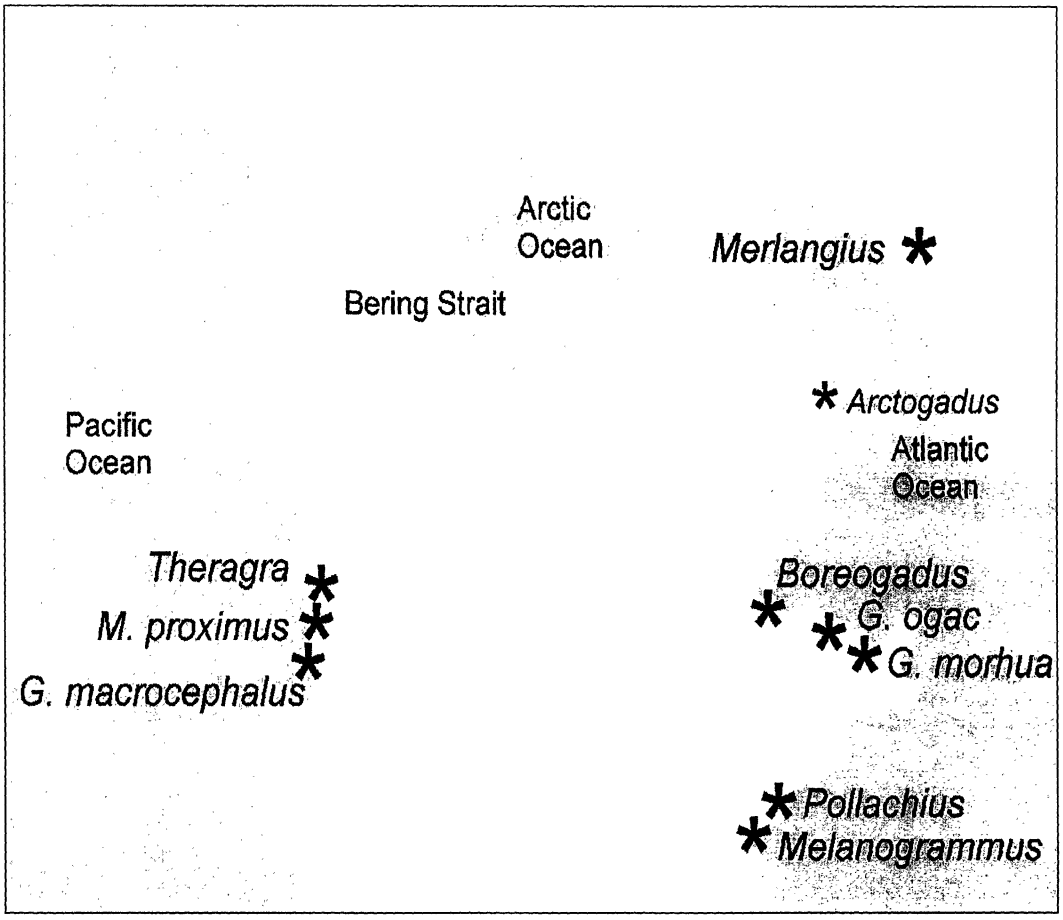


Figure 4. Polar projection map showing the collection localities for the 10 species of gadines sampled in the current study.

orders, a pleuronectiform (*Paralichthys olivaceus* – accession number NC002386), a salmoniform (*Salmo salar* – accession number NC001960), and a perciform (*Trachurus japonicus* – accession number NC002813). Identification of appropriate strings of nucleotides conserved across these four taxa were chosen as primer sites. As an initial criterion, 25 bp regions with more than 96% homology between ingroup species and more than 88% homology with outgroup species were chosen as primer sites. A set of 20 primer pairs were identified to amplify the complete genome in fragments of 750-1,400 bp, with overlapping intervals between fragments of 80-300 bp. The 40 primers are listed in Table 1 along with the region amplified and the product size (bp) expected for each amplification.

Polymerase chain reaction (PCR) amplifications were performed in 25 μL volumes containing the following ingredients: 2.5 μl of 10x PCR buffer (Qiagen, Inc.), 0.5 μL of 10mM dNTPs (2.5 mM each of dATP, dCTP, dGTP and dTTP), 1.0 μL of a 10 μM stock of each primer (Operon Technologies, Inc.), 0.2 μL of 5 U/μL *Taq* DNA polymerase (Qiagen, Inc.) and 19 μL of ddH_2O. Following an initial denaturation at 95°C for 3 min, the amplification profile consisted of the following steps: denaturation at 93°C for 30 s, annealing at x°C (depending on primer pair) for 30 s and extension at 72°C for 1 min. This was repeated for 40 cycles. A final incubation was carried out at 72°C for 10 min. PCR reactions were carried out on a GeneAmp PCR 9600 thermal cycler (Perkin Elmer).

Visualization of PCR products was accomplished by electrophoresing samples on a 2% LE agarose gel containing ethidium bromide. Five microlitres of PCR product was

Table 1. Sequences of the 20 primer pairs used to amplify entire mitochondrial genomes in overlapping regions, the expected size (bp) of the region, and the size of the overlap with the preceding region.

Primer region	Primer sequence (5' – 3')	Length of PCR product (bp)	Overlap with preceding region (bp)
g01	F: 5'-CTGAAGATATTAGGATGGACCCTAG-3' R: 5'-CTAGTCCCTACTTACTGCTAAATCC-3'	841	356
g02	F: 5'-CCAAAAACGTCAGGTCGAGGTGTAG-3' R: 5'-CTATTCATTTCACAGGCAACCAGCT-3'	750	80
g03	F: 5'-ACCCCGAAACTGAGCGAGCTACTCC-3' R: 5'-TAAGCCCTCGTGATGCCATTCATAC-3'	800	91
g04	F: 5'-TTTACCAAAAACATCGCCTCTTG-3' R: 5'-TGAACCTCTGTAGAAAGGGCTTAGG-3'	801	82
g05	F: 5'-GGAGTAATCCAGGTCAGTTTCTATCTATG-3' R: 5'-ATGTTCGGGGTATGGGCCCAAGAGC-3'	1357	133
g06	F: 5'-GGTTAAAGTCCCTTCAACTCCTTAG-3' R: 5'-AGCTTAATTAAAGTATTTGTTTTGC-3'	1337	86
g07	F: 5'-AAACTAGACCAAGGGCCTTCAAAGC-3' R: 5'-GCTAATCAGCTAAAGACTTTTACACC-3'	1341	112
g08	F: 5'-ATGGGTATAGTCTGAGCTATGATGG-3' R: 5'-TAACCCACAATTCTGCCTTGACAAG-3'	865	112
g09	F: 5'-TAACTAACGTTGAGTGACTCCAC-3' R: 5'-ACCCATATTAGCTTCTTAGTGAGG-3'	954	161
g10	F: 5'-TCCCGGAGTTTTCTACGGACAATG-3' R: 5'-AGAGGGCGAATGAATAAACTAATTG-3'	827	100
g11	F: 5'-TAGCAACTGTCCTTATCGGCATACG-3' R: 5'-TAATACTGTGGTGAGCTCAGGTTAC-3'	805	83

Table 1 (*continued*).

g12	F: 5'-CCCCGAACTAGGTGGCTGCTGACC-3' R: 5'-AGGACAAGTAGAGGTGTTGATAGGG-3'	1441	79
g13	F: 5'-CTTTCTCCGCTTGTGAAGCAAG-3 R: 5'-CAATTAGAGATTTCAGGTCAGTTTG-3'	961	256
g14	F: 5'-CTGTTGCAGGCTCAATAGTTCTTGC-3' R: 5'-TTCGAGGGAGCCTTGGGGTCTAACC-3'	847	139
g15	F: 5'-TAACCAAGACATTAGATTGTGATTC-3' R: 5'-TGGTAGTCATGGGTGAAGTCCAAAC-3'	923	77
g16	F: 5'-GGTGATGACACGGCCGAGCAGATG-3' R: 5'-AATAATTGCATCTTTGGAGAAGAAGC-3'	723	190
g17	F: 5'-ATTCATAGCCTAAACGATGAACAAG-3' R: 5'-GTCGTTTTTCATATCATTAGTCCTG-3'	1331	98
g18	F: 5'-GCTACTAAGACCAGTCCTAAAGCAG-3' R: 5'-CTGTGGGATTATTTGAGCCTGTTTC-3'	806	86
g19	F: 5'-GAGGAGGTTTCTCAGTAGATAATGC R: 5'-GTTTAATTTAGAATTCTAGCTTTGG	851	85
g20	F: 5'-GAATGAAACTGCCCTAGTAGCTCAG-3' R: 5'-GGCAGGACATTAAGGGCATTCTCAC-3'	1383	162

mixed with 2 μL of bromophenol blue dye in glycerol. As a standard, 2 μL of the plasmid φX-174-RF cut with *HaeIII* restriction enzyme was loaded into each gel to estimate the total size of each PCR product for verification of target regions. Successful amplification products were purified using a QIAquick PCR Purification Kit (Qiagen, Inc) as per the manufacturer's instructions. Purified DNA was eluted in 50 μL of elution buffer (10 mM Tris-Cl, pH 8.5) and stored at –20°C.

2.3 DNA Sequencing and Alignment

PCR products were then sequenced using ABI Prism BigDye fluorescent dye terminator chemistry on a 377 automated DNA sequencer (Applied Biosystems Inc), using the same primers that were used for PCR amplification. Sequencing reactions were carried out in both directions as follows: 4 μL BigDye terminator v.2.0, 3.2 μL of 2 μM primer (either forward or reverse) and 4.4 μL of PCR product. Alternatively, for weaker PCR products, 5-10 μL of DNA (depending on the relative strength of the band on agarose gel electrophoresis) was dried in a vacuum centrifuge for 5-10 min before addition of the sequencing cocktail, and 4.4 μL of ddH_2O was added to make up the total required volume. Sequencing reactions were carried out on a T1 DNA thermal cycler (Perkin Elmer) as follows: an initial denaturation at 96°C for 2 min, followed by 35 cycles consisting of 96°C for 30 s, annealing of primers at 50°C for 15 s and extension of products at 60°C for 4 min.

Sequencing reactions were purified with an isopropanol precipitation by the addition of 40 μL of 75% isopropanol to the reaction and subsequent precipitation at

room temperature for 30 min. This was followed with a 20 min centrifugation at 12 000 rpm in an Eppendorf 5804 R centrifuge. After centrifugation, samples were aspirated to remove the supernatant, being careful not to disturb the pellet. A wash with 250 μL 75% isopropanol was precipitated at room temperature for 10 min followed by a 10 min centrifugation as described above. Again samples were aspirated and dried in an Eppendorf Speedvac for 5-10 min or left to air dry for ~ 1 h. Dried samples were resuspended in 5 μL of 5 parts deionized formamide : 1 part 25 mM Na_2EDTA with 50 mg/ml blue dextran dye. Samples were then snap-cooled on a T1 thermal cycler by heating at 95°C for 2 min and immediately cooled to 5°C before loading onto the ABI 377 DNA sequencer. Sequencing runs lasted about 9 hours.

2.4 Phylogenetic Analysis

Sequences were aligned and edited using the computer software program Sequencher™ version 4.1.2 (Applied Biosystems). DNA sequences reported here are the consensus sequences of the two complementary strands. Consensus sequences were achieved by sequencing both complementary strands for each region for each individual. Overlapping regions were aligned to assemble contiguous fragments. Individual mtDNA genome sequences were exported for phylogenetic analysis. The Norway individual of *G. morhua* refers to the reference sequence obtained from GenBank (accession number NC002081) and was used as a an estimate of the amount of sequence divergence among trans-Atlantic individuals, as compared with the *G. morhua* 002, which was obtained off the coast of Newfoundland. Maximum parsimony (MP), Neighbour-joining (NJ)

distance analyses, and maximum likelihood (ML) were performed with the Phylogenetic Analysis Using Parsimony (PAUP*) program (Version 4.0) (Swofford, 1998). For MP analysis a heuristic search was done with transversion-transition weightings of 1:1, 3:1, 10:1 and transversion only, with 10 random taxon additions and the tree-bisection-and-reconnection branch-swapping option. Bootstrap support of parsimony trees was estimated with 1000 replicates each of 10 random taxon additions and the tree-bisection-and-reconnection (TBR) branch-swapping option. For the distance analysis, pairwise genetic distances were calculated with Tamura-Nei (T-N) distance (Tamura & Nei, 1993) using a neighbour-joining search and the gamma parameter γ, as estimated via the maximum likelihood approach. Bootstrap support was estimated using 1000 replicates of this algorithm. Maximum likelihood analyses were conducted with an empirical estimate of the transition/transversion (Ts/Tv) ratio of 7.0 and gamma parameter, γ, of 0.956 using a heuristic search with five random taxon additions and the TBR branch-swapping option. The tree was bootstrapped with a heuristic search via stepwise addition with for each of 1,000 replicates. In addition to complete genomes, phylogenies were obtained using the three algorithms for each separate protein-coding gene. For maximum parsimony analysis, unweighted as well as weighted analyses with transversion/transition weightings of 3:1, 10:1 and Tv only were performed for each gene region. Separate gene trees were also resolved using T-N distances with a neighbour-joining search and individually estimated gamma parameters. Finally, ML analyses were performed for each gene region with separate estimates of Ts/Tv ratios, gamma parameter, base composition, and proportion of invariable sites. These estimated values were then used

in a maximum likelihood approach via distance analysis with a heuristic search with 10-random taxon additions and the tree-bisection-and-reconnection branch-swapping option.

Individual gene trees under each approach were assessed for the accuracy of the tree topology as compared to the tree obtained from the genomic data set (14,036 bp). For this analysis, d_T was measured (following Russo *et al.* (1996) and based on Robinson & Foulds (1981) and Penny & Hendy (1985)) as follows: $d_T = 2[\min(q_r,q_t) - r] + |q_r - q_t|$, where q_r and q_t represent the number of interior branches reconstructed from a given tree (gene tree) and the genomic (assumed) tree, respectively and r is the number of interior branches that are identical for the two trees. For bifurcating trees, d_T is equal to twice the number of incorrect interior branches of the reconstructed gene tree (Russo *et al.*, 1996). Therefore a value of d_T equal to zero means that the gene tree has the same topology as the genomic tree and as this value increases, the worse the correspondence between the two trees.

Finally, all protein-coding regions were translated into their amino acid sequences and a phylogeny based on the complete set of amino acid sequences was generated with both unweighted maximum parsimony analysis and a neighbour-joining analysis on the pairwise number of amino acid substitutions among taxa.

3.0 RESULTS

3.1 Alignment of Consensus Sequences

The complete mitochondrial genome of *Gadus morhua* is 16,696 bp (Johansen & Bakke, 1996). The aligned set across all taxa (excluding the D-loop) was 15,720 bp. Of the 15,720 bp, 14,036 bp were common to all of the 13 individuals (excluding missing and ambiguous data), representing 10 species. Of these 14,036 nucleotide positions, 11290 positions were constant across all taxa. Therefore, 2,746 sites were variable of which 1,631 nucleotides were parsimony informative, that is shared by at least two taxa. Control region sequences were not included in the analysis. Table 2 lists the associated gaps in sequence data for each species, with the positions of nucleotides as well as the gene or genes for which there are missing data.

3.2 Phylogenetic Reconstruction from Genomic Sequences

Table 3 shows the genetic distances calculated for the complete data set of 14,036 bp as absolute nucleotide differences and Tamura-Nei (T-N) distances (1993). The most similar sequences are the two *G. ogac* individuals, with a T-N distance of 0.0011 and which differ by 16 nucleotide substitutions. The next most similar pair of taxa is *G. macrocephalus* 002 and either of the two *G. ogac* individuals (001 and 002) with T-N distances of 0.0032 and 0.0039 and 45 and 55 nucleotide substitutions, respectively. As seen in Table 3, this *G. macrocephalus* individual is more similar to the two *G. ogac* individuals than it is to the other *G. macrocephalus* (003) individual (T-N distance 0.0049 and 69 nucleotide substitutions). The distance between the two *G.*

Table 2. Positions of missing base pairs and associated coding region for each species[†].

species	positions of missing base pairs	part of region
A. glacialis	3 938 – 4 032	$tRNA^{Gln}/tRNA^{Met}$
	4 834 – 5 143	$ND2/tRNA^{Trp}$
	14 929 – end	Cyt *b*/$tRNA^{Thr}/tRNA^{Pro}$/Control[*]
B. saida	15 666 – end	$tRNA^{Pro}$/Control
G. macrocephalus 002	5 076 – 5 191	$tRNA^{Trp}/tRNA^{Ala}$
	15 791 – end	Control
G. macrocephalus 003	1 – 53	$tRNA^{Phe}$
	15 666 – end	$tRNA^{Pro}$/Control
G. morhua 002	15 666 – end	$tRNA^{Pro}$/Control
G. ogac 001	15 770 – end	Control
G. ogac 002	15 666 – end	$tRNA^{Pro}$/Control
M. aeglefinus	1 – 54	$tRNA^{Phe}$
	2 811 – 2 872	$tRNA^{Leu}$/ND1
	3 889 – 4 015	$tRNA^{Gln}/tRNA^{Met}$
	10 235 – 10 339	ND4L/ND4
	15 685 – end	$tRNA^{Pro}$/Control
M. merlangus	1 – 53	$tRNA^{Phe}$
	5 101 – 5 135	$tRNA^{Trp}/tRNA^{Ala}$
M. proximus	1 – 53	$tRNA^{Phe}$
	15 734 – end	Control
P. virens	15 648 – end	$tRNA^{Pro}$/Control
T. chalcogramma	15 681 – end	$tRNA^{Pro}$/Control

[†]Missing base pairs were excluded for phylogenetic analysis across all taxa.
[*]Control refers to the control region or D-loop.

Table 3. Pairwise number of nucleotide differences (lower half of the matrix) and Tamura-Nei distances (upper half of the matrix) calculated from mitochondrial gene sequences for 10 gadine species.

	G. morhua 002	*G. morhua* Norway	*G. ogac* 001	*G. ogac* 002	*G. macro-cephalus*002	*G. macro-cephalus*003	*T. chalco gramma*
G. morhua 002	---	0.0036	0.042	0.042	0.043	0.042	0.040
G. morhua Norway	51	----	0.043	0.043	0.044	0.043	0.042
G. ogac 001	567	578	---	0.0011	0.0039	0.0044	0.043
G. ogac 002	570	581	16	---	0.0032	0.0040	0.043
G. macrocephalus 002	577	588	55	45	---	0.0049	0.043
G. macrocephalus 003	572	585	62	56	69	---	0.042
T. chalcogramma	545	562	576	575	579	574	---
A. glacialis	613	623	584	586	588	592	614
B. saida	729	745	697	696	695	706	693
M. aeglefinus	1166	1165	1174	1174	1183	1191	1147
M. merlangus	1088	1091	1096	1095	1098	1110	1064
P. virens	1168	1175	1155	1154	1147	1158	1146
M. proximus	1332	1345	1280	1284	1287	1299	1278

Table 3 (*continued*).

	A. glacialis	*B. saida*	*M. aeglefinus*	*M. merlangus*	*P. virens*	*M. proximus*
G. morhua 002	0.045	0.054	0.090	0.084	0.091	0.104
G. morhua Norway	0.046	0.056	0.090	0.084	0.091	0.105
G. ogac 001	0.043	0.052	0.091	0.084	0.089	0.099
G. ogac 002	0.043	0.052	0.091	0.084	0.089	0.100
G. macrocephalus 002	0.044	0.052	0.091	0.084	0.088	0.100
G. macrocephalus 003	0.044	0.053	0.092	0.085	0.089	0.101
T. chalcogramma	0.046	0.052	0.089	0.082	0.088	0.100
A. glacialis	---	0.045	0.088	0.080	0.088	0.099
B. saida	604	---	0.090	0.082	0.090	0.101
M. aeglefinus	1139	1165	----	0.074	0.097	0.113
M. merlangus	1033	1075	969	---	0.093	0.109
P. virens	1135	1168	1242	1200	---	0.100
M. proximus	1275	1296	1427	1392	1283	---

morhua individuals (Newfoundland shelf vs. Norway) is 0.0036 with 51 nucleotide substitutions between them, which is on the same level of differentiation as *G. macrocephalus* and *G. ogac*. These trans-Atlantic *G. morhua* individuals contain fewer nucleotide differences than the two *G. macrocephalus* individuals from the Pacific. Aside from differences within *Gadus*, the next most similar pair of taxa are *G. morhua* and *T. chalcogramma* (T-N distance 0.040-0.042 and 545-562 nucleotide substitutions). However, the distances between *Theragra* and all species of *Gadus* are on the same order. The least similar pair of taxa are *M. proximus* and *M. aeglefinus* with a T-N distance of 0.113 and 1,427 nucleotide substitutions.

Maximum likelihood analysis estimated base frequencies as follows: A (28.9%), C (25.9%), G (15.7%) and T (29.5%). As is typical of mtDNA, there is a lower G+C content, with a particular bias against G as it is usually avoided in 3rd codon positions (Saccone *et al.*, 1999). As well, this analysis estimated the transition/transversion (Ts/Tv) ratio for the entire data set as 7.0, and the shape of the gamma parameter (γ) as 0.956.

Maximum parsimony, neighbour-joining (based on T-N distances) and maximum-likelihood methods all produced similar results and the same consensus tree topology (Figures 5, 6, and 7 respectively) when using the consensus sequences (14,036 bp) for each species. *Microgadus proximus* was used as the outgroup to the remaining species following Carr *et al.* (1999). All three analyses identified the following: (1) *Pollachius* as the outgroup to the remaining taxa, (2) a clade composed of *Merlangius merlangus* and *Melanogrammus aeglefinus*, (3) a clade composed of *T. chalcogramma* and the three

Figure 5. Phylogenetic tree produced by unweighted maximum parsimony (MP) analysis on 14,036 consensus bases from 13 individuals representing 10 species of gadines: heuristic search with 10 random taxon additions and the TBR branch-swapping option. Bootstrap analysis (bold numbers) performed with 1,000 replicates.

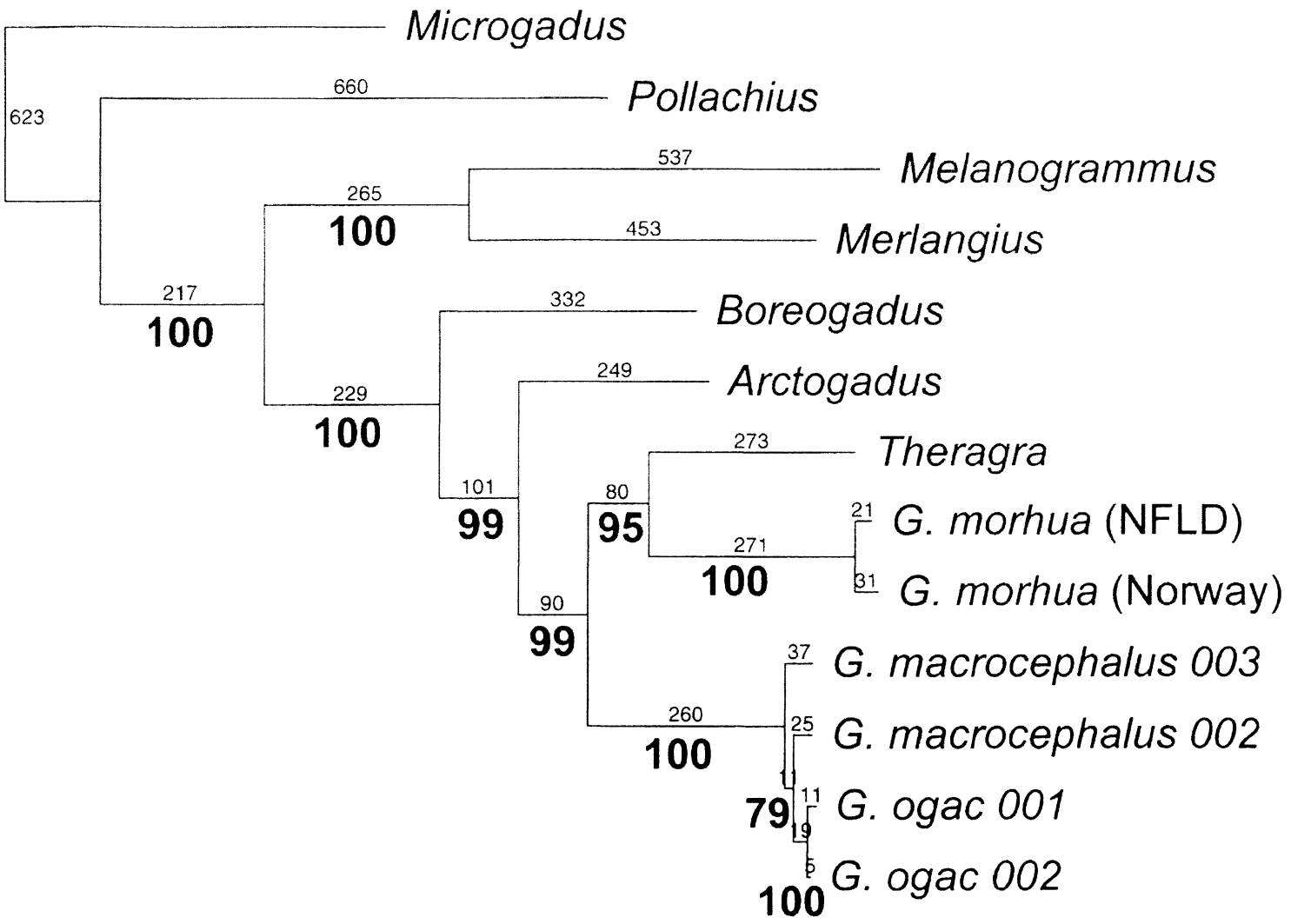

Microgadus
Pollachius
Melanogrammus
Merlangius
Boreogadus
Arctogadus
Theragra
G. morhua (NFLD)
G. morhua (Norway)
G. macrocephalus 003
G. macrocephalus 002
G. ogac 001
G. ogac 002
623
660
537
265
100
453
217
100
332
249
229
100
273
101
99
80
95
21
271
100
31
90
99
37
260
25
100
79
11
19
5
100

Figure 6. Phylogenetic tree produced by Neighbour-joining (NJ), based on Tamura-Nei (T-N) distances of 14,036 consensus bases from 13 gadines with a shape parameter of gamma = 0.965. Bootstrap values (bold numbers) estimated with 1,000 replicates.

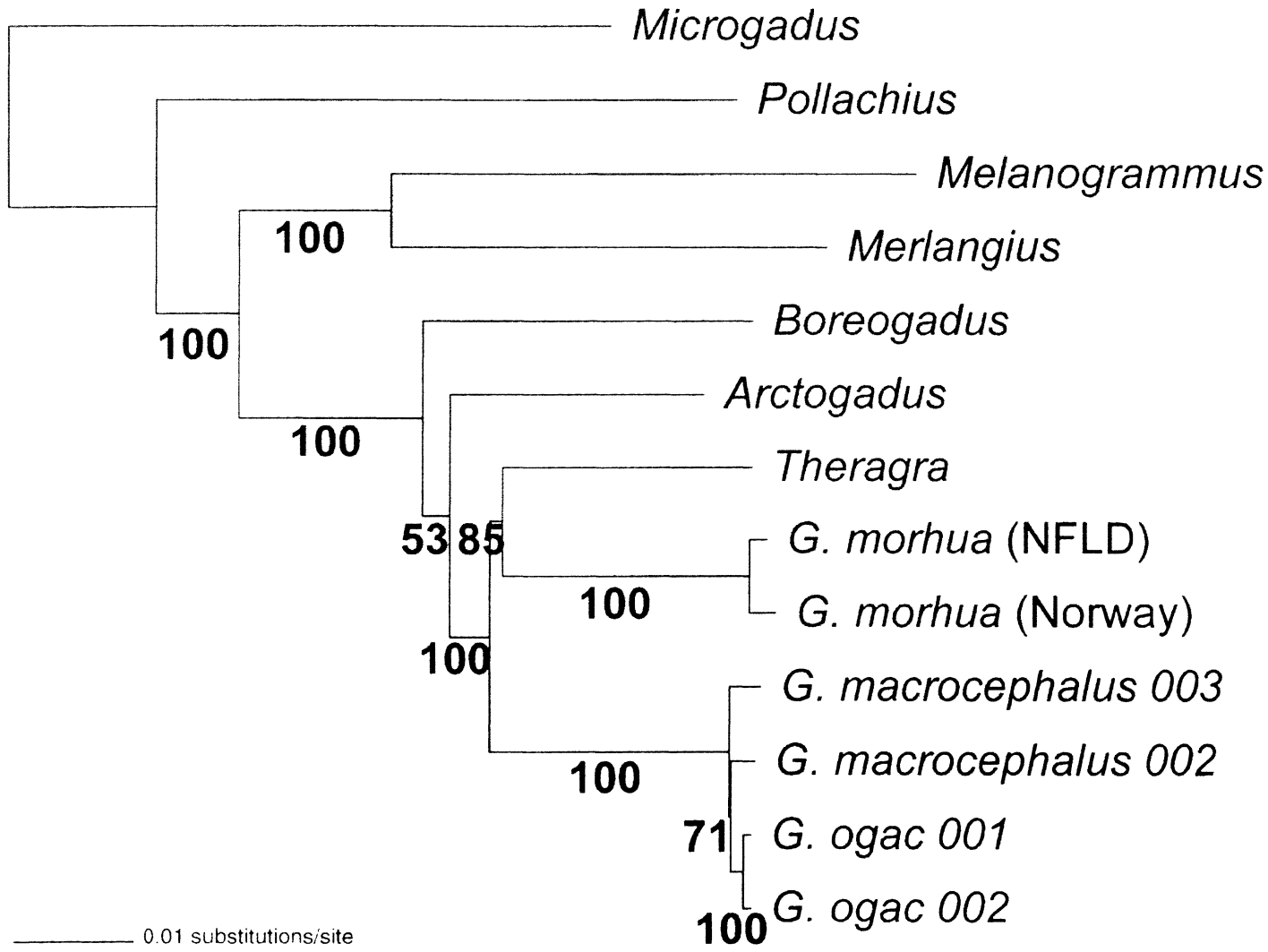
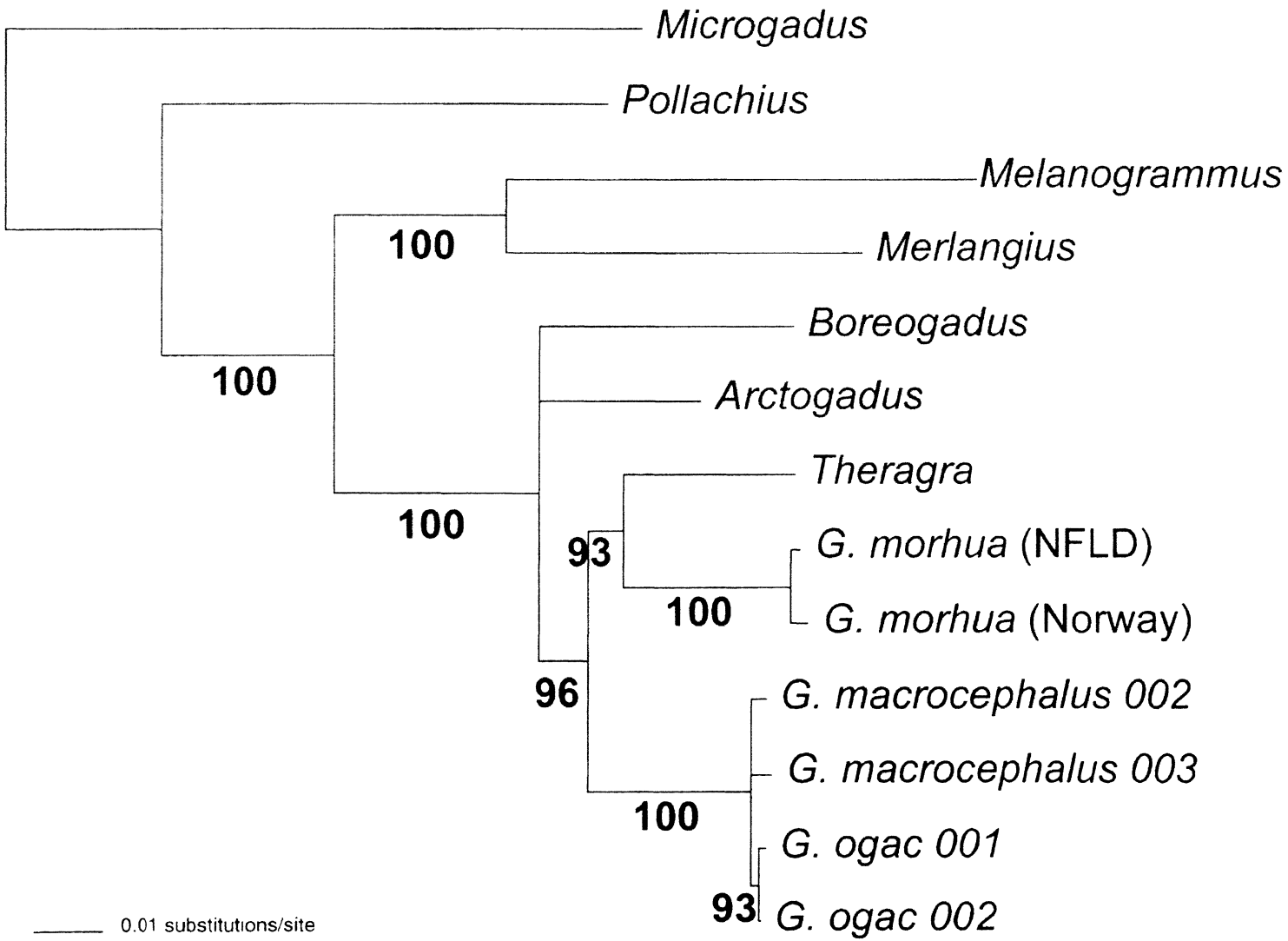

Microgadus
Pollachius
Melanogrammus
Merlangius
Boreogadus
Arctogadus
Theragra
G. morhua (NFLD)
G. morhua (Norway)
G. macrocephalus 003
G. macrocephalus 002
G. ogac 001
G. ogac 002
100
100
100
53
85
100
100
71
100
0.01 substitutions/site

Figure 7. Phylogenetic tree produced for 13 gadines for 14,036 consensus bases of mtDNA by maximum likelihood (ML) with an estimated gamma parameter of 0.965 and a heuristic search done with three random taxon additions and the tree bisection and reconnection branch-swapping option. Bootstrap values in bold numbers are based on 1,000 replicates.


Microgadus
Pollachius
Melanogrammus
100
Merlangius
Boreogadus
100
Arctogadus
Theragra
100
93
G. morhua (NFLD)
100
G. morhua (Norway)
96
G. macrocephalus 002
G. macrocephalus 003
100
G. ogac 001
93
G. ogac 002
0.01 substitutions/site

species of *Gadus*, with *Theragra* and *G. morhua* as sister taxa, and (4) a clade within that mentioned in (3) comprising the two *G. macrocephalus* and two *G. ogac* individuals. Among this latter grouping in both the MP and NJ analyses, one *G. macrocephalus* individual (002) was more similar to the two *G. ogac* individuals than it was to its conspecific. In the ML analysis the two *G. macrocephalus* individuals were collapsed at a polytomy. Weighted parsimony with transversion/ transition (Tv/Ts) weightings of 3:1, 10:1 and Tv only, differed only with respect to the relative branching order among the four individuals in the *G. macrocephalus/ogac* clade.

The only topological difference among the various modes of analyses, was the relative branching order of *Boreogadus* and *Arctogadus* with respect to the clade composed of the three species of *Gadus* plus *Theragra.* In all three analyses, *Boreogadus* and *Arctogadus* were outside of the clade composed of *Theragra* and the three species of *Gadus*. In the unweighted MP analysis, *Boreogadus* falls outside of *Arctogadus,* with extremely high bootstrap support (99%), while in the ML analyses *Boreogadus* and *Arctogadus* come out as sister taxa, however bootstrap analysis collapsed these two species at a polytomy as there was virtually equal bootstrap support for the two alternatives. The topology of the neighbour-joining analysis showed the two as sister taxa, however, when the bootstrap is performed, a low support (53%) resolves *Boreogadus* outside. Table 4 shows the different modes of phylogenetic analysis used and which of the alternative relationships for *Boreogadus* and *Arctogadus* each supports. The transversion-only analysis identified two equally parsimonious trees, one supporting each of the two arrangements between *Boreogadus* and *Arctogadus*. Ultimately, the only

Table 4. Modes of analyses supporting possible relationships between *Boreogadus* and *Arctogadus* with ≥50% bootstrap support based on 1000 bootstrap replicates.

	Boreogadus outside *Arctogadus*	*Boreogadus* + *Arctogadus*
Unweighted maximum parsimony	✓ (99%)	---
3:1 weighted parsimony	✓(99%)	---
10:1 weighted parsimony	✓(87%)	---
Tv only parsimony*	✓(50%)	✓ (50%)
Neighbor-joining	✓(53%)	---
Maximum likelihood	✓ (50%)	✓ (50%)
Average	**73%**	**50%**

*produced two equally parsimonious trees supporting both arrangements.

strong support is for *Boreogadus* outside of *Arctogadus*, as no other alternative was supported in more than 50% bootstrap replicates.

Bootstrap support for the grouping of the two *G. ogac* individuals as well as one of the *G. macrocephalus* individuals (002) differed among the three analyses. In MP and NJ analyses, there was 79% and 71% bootstrap support for this group, respectively (Figures 5 & 6), whereas in the ML approach, there was no resolution between the two *G. macrocephalus* individuals (Fig 7). When these three analyses were repeated, without *Microgadus, Pollachius, Melanogrammus,* and *Merlangius*, and with *Boreogadus* as the outgroup, unweighted parsimony, NJ and ML analyses showed higher bootstrap supports for this node with 86%, 78% and 54% (results not shown). As well, although all three approaches defined a clade composed of *G. morhua* and *T. chalcogramma* as sister taxa, their relative support for this grouping varied with 95%, 85% and 86% for MP, NJ, and ML approaches, respectively. All other branches were statistically significant, in all three approaches, generally with 100% bootstrap support (Figures 5, 6, & 7).

3.3 Phylogenetic Analysis on Separate Gene Regions

Information on the separate gene regions is presented in Table 5, which shows the number of phylogenetically informative sites as well as estimates of the transition/ transversion ratio and the estimated gamma parameter for each protein-coding gene as well as the two species of rRNA. The shortest gene examined, ATP8 contains only 168 bp. Of these, 11 of the 19 variable positions are phylogenetically informative. ND5 was the largest gene examined (1839 bp) with 330 of the 520 variable sites phylogenetically

Table 5. Statistical properties of 13 mitochondrial protein-coding genes and the two rRNA genes with the number of base pairs sequenced, numbers of variable sites and phylogenetically informative sites, transition/transversion ratio and the estimated gamma parameter.

Gene	# bp	# variable sites	% variable sites	# of informative Sites	Ts/Tv	gamma (γ) parameter
12S	953	47	4.9	22	9.8	0.62
16S	1672	102	6.1	31	3.1	0.64
ATP6	684	164	24.0	98	6.6	2.01
ATP8	168	19	11.3	11	11.5	0.69
CoxI	1551	313	20.2	170	5.7	0.57
CoxII	699	103	14.7	63	7.7	6.39
CoxIII	786	152	19.3	82	7.0	0.57
ND1	975	256	26.3	148	6.5	2.11
ND2	1047	291	27.8	164	6.8	0.48
ND3	351	96	27.4	58	8.6	3.97
ND4	1386	394	28.4	269	8.5	1.22
ND4L	297	61	20.5	22	3.8	0.72
ND5	1839	520	28.3	330	7.9	1.73
ND6	522	133	25.5	76	7.6	0.37
Cyt b	1185	267	22.5	155	7.9	2.04

informative. There was a strong correlation between size of gene region (bp) and the number of phylogenetically informative sites ($r = 0.936$). Among the protein coding regions, the transition/transversion ratio averaged 7.4, with a range from 3.8 (ND4L) to 11.5 (ATP8), while the average percentage of variable sites was 22.8% (range 11.3% for ATP8 to 28% for ND4 and ND5). The two species of rRNA exhibited lower variation with 5-6% of variable sites, while collectively the tRNAs exhibited 5% sequence variation.

Table 6 shows the topological differences (d_T) for each of the protein coding genes under various methods of phylogeny reconstruction in comparison to the consensus trees based on the 14,036 bp data set (according to the method presented in Russo *et al.*, 1996; see Methods). For the individual gene regions, under MP, the consensus tree was as shown in Figure 5. For the individual gene analyses under either distance or ML, the consensus was considered as shown in the respective genomic phylogenies (Figures 6 & 7). However as long *Boreogadus* and *Arctogadus* were resolved as either of the two scenarios (sister taxa or *Boreogadus* outside), these were both assumed to count as an identical node in relation to the consensus trees since there was no strong support for one of these scenarios over the other. The individual gene tree is compared to the consensus tree, based on genomic sequences, by identifying the number of branches in common between the two trees. The raw score refers to the number of events from which the gene tree differs from the consensus tree. Therefore, a score of zero means the gene tree had the identical topology to the consensus tree. As the raw score increases, this indicates an increase in the number of topological differences of the gene tree compared to the

Table 6. Topological distances (d_T) of reconstructed gene trees from the consensus tree for mtDNA nucleotide sequence data.

Genes	ND1	ND2	Cox1	Cox2	ATP8	ATP6	Cox3	ND3	ND4L	ND4	Nd5	Nd6	Cytb*	All	Sum
MP															
Unweighted	0	6	5	1	4	6	5	3	8	1	2	10	6 (3)	0	57
Tv/Ts 3:1	0	5	5	1	4	6	5	2	10	1	2	7	6 (3)	0	54
Tv/Ts 10:1	0	1	5	3	4	10	5	6	12	3	1	7	7 (8)	2	66
Transversion only	3	6	5	10	11	10	6	9	11	5	4	9	10 (8)	5	104
NJ															
T-N	0	3	4	3	6	4	6	4	8	5	2	12	4 (4)	0	61
ML	4	6	4	2	5	13	6	12	12	3	2	10	8 (6)	2	87
Totals	**7**	**27**	**28**	**20**	**34**	**49**	**33**	**36**	**61**	**18**	**13**	**55**	**41** (32)		
Rank	1	5	6	4	8	11	7	9	13	3	2	12	10 (7)		

*missing 50% of the sequence for *A. glacialis*.
Note: numbers in parentheses represent *d*T when *A. glacialis* is removed from the analysis.

consensus tree. Therefore a higher score reflects a more inaccurate gene tree either due to incorrect branching patterns or lack of resolution for certain nodes. Across all modes of analyses, the raw scores were added for each gene which allowed the individual genes to be ranked relative to one another. The most 'accurate' gene among all types of reconstruction was ND1 (raw score 7), followed by ND5 (raw score 13) and ND4 (raw score 18). The least accurate gene in obtaining the overall phylogeny was ND4L (raw score 61). A weak negative correlation is observed between the size of the gene region (bp) and the raw score in topological differences (r = -0.59; Figure 8). A regression analysis gave the regression equation $d_T = 49.2 - 0.0189 \times L$ ($p = 0.032$), where L represents the length of the gene in base pairs. [It should be noted that the efficiency of cytochrome *b* may be underestimated, as more than half of the data for this gene are missing for *Arctogadus* and therefore may not reflect its true accuracy as a phylogenetic indicator. In the MP analysis for this region, the positions of *Arctogadus* and *Theragra* are reversed, however, when *Arctogadus* is removed from this analysis, the resulting tree is as would be expected based on the phylogeny derived from the consensus trees (14,036 bp).]

Table 6 also shows which of the methods of construction may be best suited for a particular phylogenetic analysis. For this, topological distances were summed across all protein-coding genes for a given mode of tree reconstruction. Among all 13 protein-coding genes, the methods that showed the fewest topological re-arrangements compared to the consensus trees were a parsimony analysis with a Tv/Ts weighting of 3:1 and the unweighted maximum parsimony approach, followed by the distance analysis based on

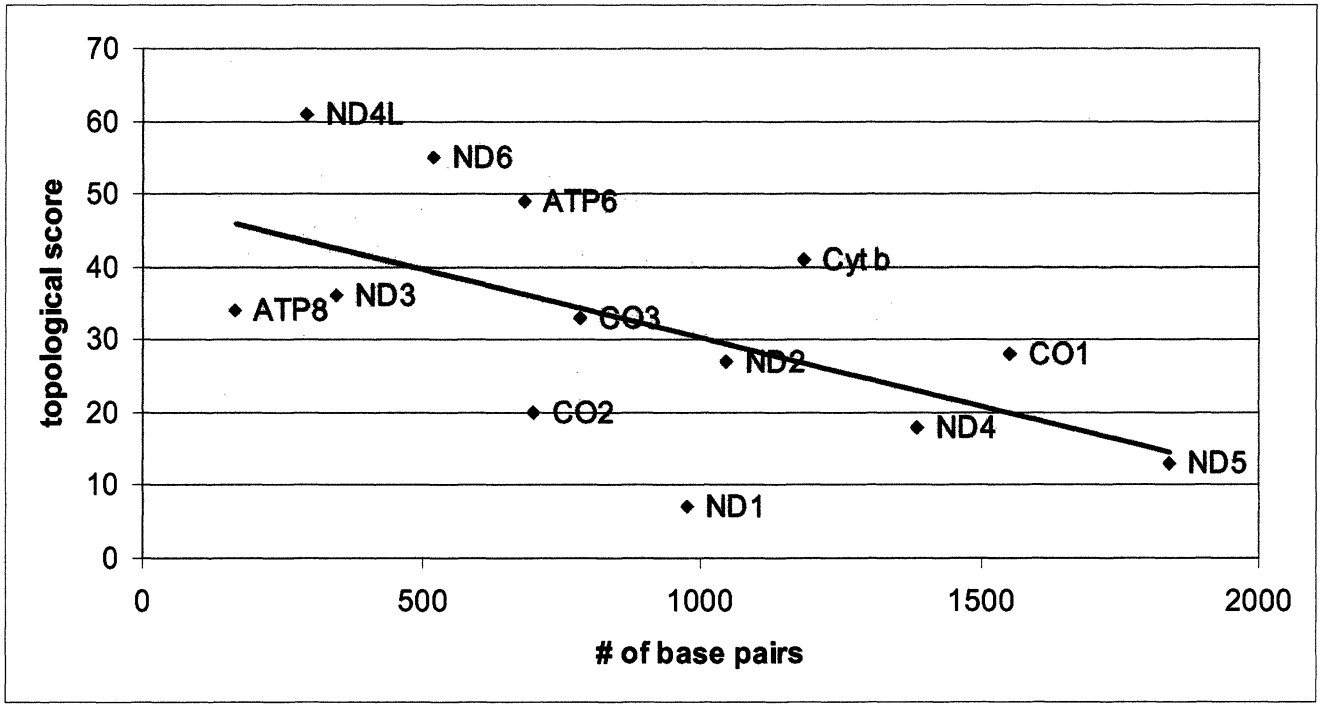


Figure 8. Regression plot of the raw topological score for each protein-coding gene versus its length in base pairs ($P = 0.032$).

T-N distances. The 10:1 Tv/Ts parsimony analysis was slightly less accurate than the NJ while Tv only and maximum likelihood were the least accurate methods of construction. The ranking of the relative efficiency of each gene region for the present study is given in Table 6. As well, a ranking was done based upon the data set comprising the same 13 protein-coding genes, generated by Russo *et al.* (1996) to compare between the two data sets. Russo *et al.* (1996) looked at very divergent taxa (including two species of cetaceans, three species of teleosts, two species of rodentia, an opossum, a chicken, and a frog). Figure 9 shows a comparison of the relative ranking for each mtDNA protein-coding gene between these two studies for both an ancient set of divergences (Russo *et al.* 1996) as well as a very recent series of relationships (this study).

Table 7 shows the number of nodes for each gene region that are supported with 50%, 75% and 95% bootstrap support. Numbers in parentheses represent a simulated data set for each of the three bootstrap criteria. This simulated data set was to control for gene length. For this, a bootstrap analysis was done separately for each region that resampled (with replacement) the data until 1,839 characters had been sampled. This length is equivalent to the largest gene in the mtDNA genome (ND5) and therefore we can assess what the phylogeny would look like if every gene were of equal length. Figure 10 shows a plot of the number of nodes supported by bootstrap values of 50% and 95% for both the actual gene length as well as the simulated 1839 bp data set. This procedure of bootstrapping to estimate the variance associated with a particular parameter for which the sampling distribution is not obvious has been conducted in prior investigations (e.g., Huelsenbeck, 1991; Buckley & Cunningham, 2002). Bootstrapping

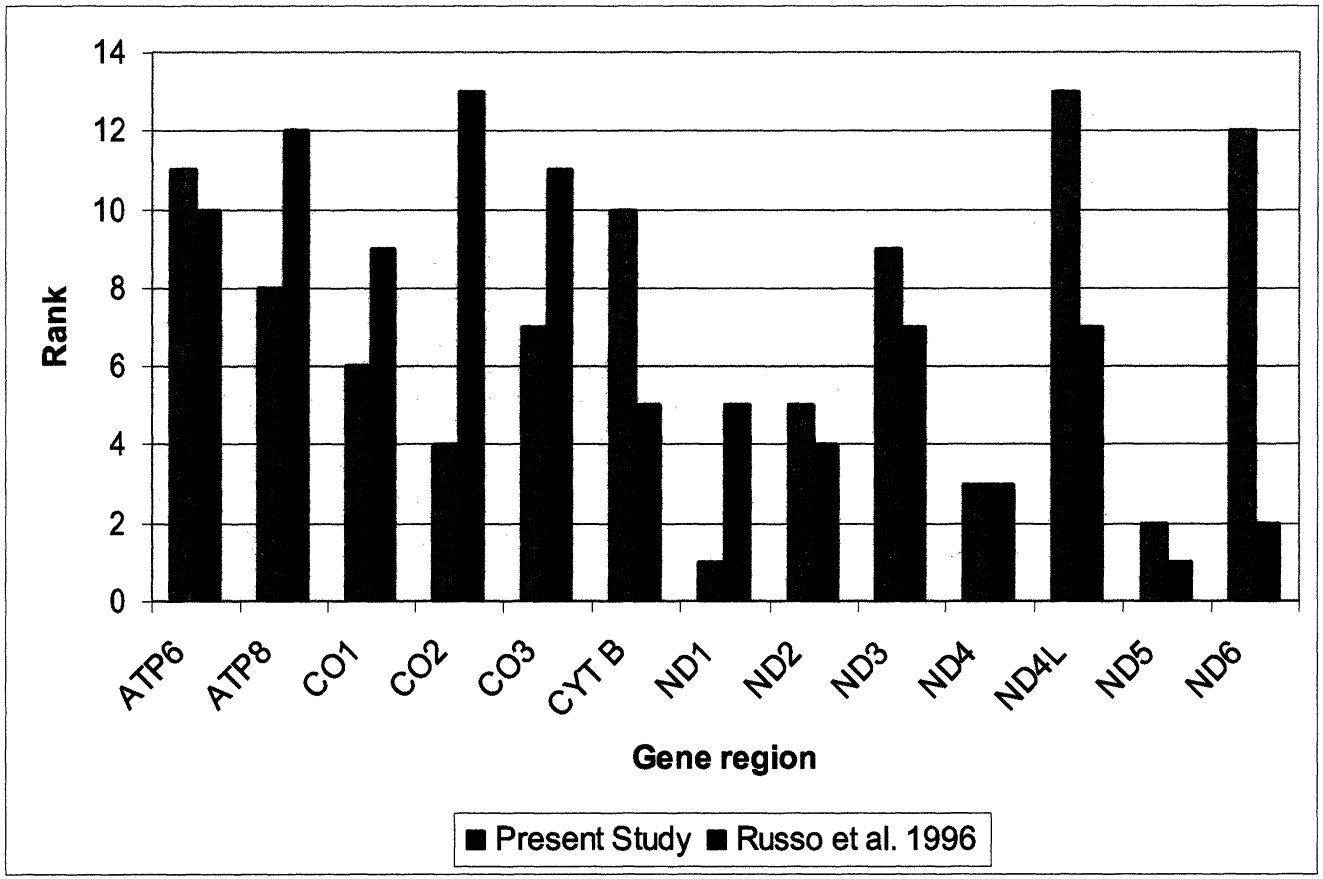


Figure 9. Comparison of the relative ranking for the topology of each mtDNA protein-coding gene from the present study in comparison to Russo *et al.* (1996). Correlation coefficient between the two data sets calculated as $r = 0.16$.

Table 7. Counts of the number of nodes supported by three values of bootstrapping for each gene region, and for a simulated data set of 1839 bp for each gene region (in parentheses).

Gene	# bootstrap ≥50%	# bootstrap ≥ 75%	# bootstrap ≥ 95%
ATP6	7 (8)	3 (5)	2 (2)
ATP8	5 (6)	2 (6)	1 (6)
Co1	7 (7)	4 (4)	4 (4)
Co2	9 (9)	6 (9)	2 (6)
Co3	6 (8)	3 (6)	2 (2)
Cytb	7 (8)	7 (7)	3 (5)
ND1	9 (10)	9 (9)	2 (5)
ND2	8 (9)	5 (7)	5 (5)
ND3	5 (7)	2 (5)	2 (3)
ND4	7 (7)	7 (7)	3 (3)
ND4L	2 (6)	0 (3)	0 (1)
ND5	7	5	3
ND6	6 (6)	3 (6)	2 (4)

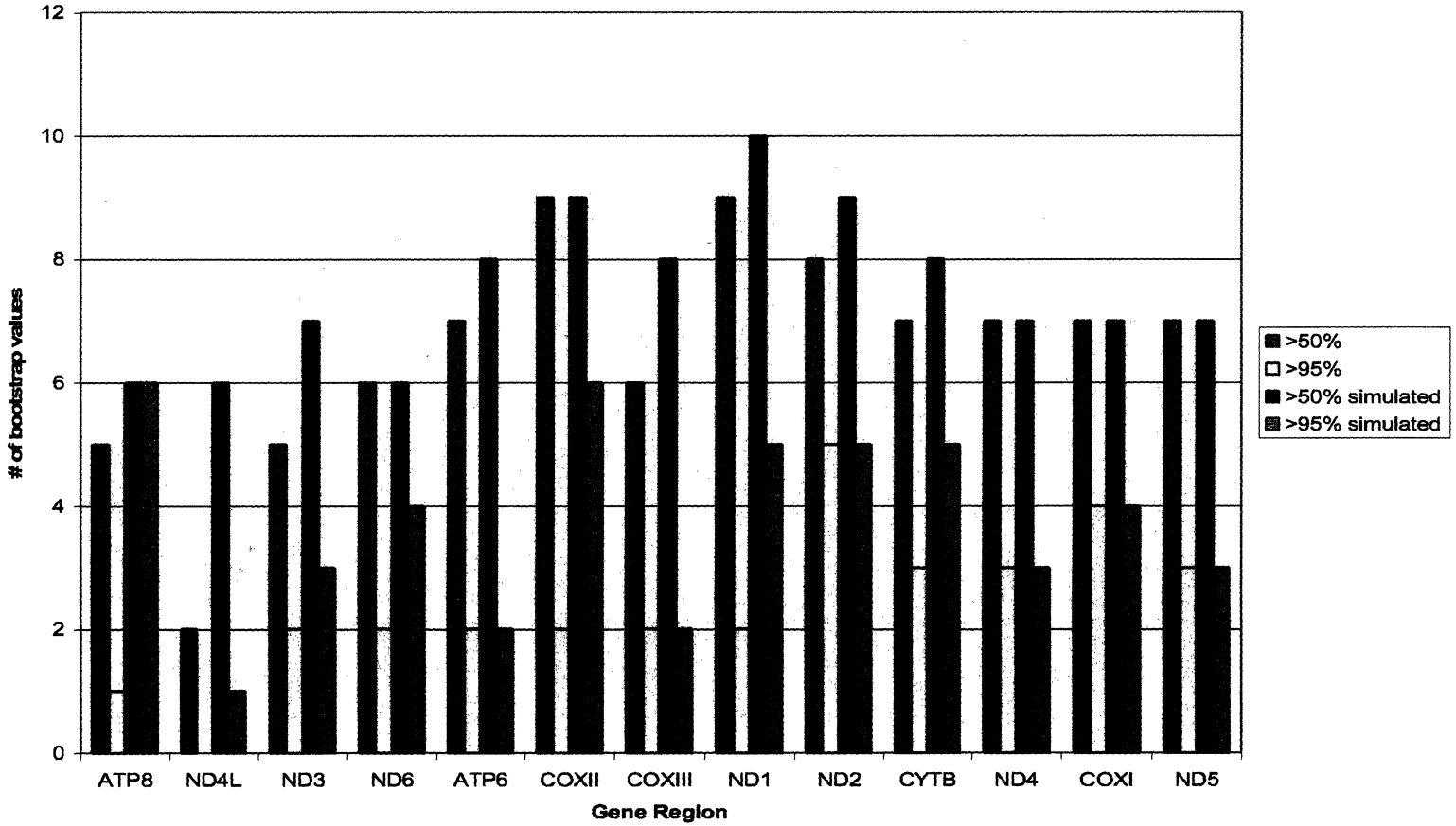


Figure 10. Plot of the number of nodes supported by three levels of bootstrap for each gene region sampled for both the actual gene length and a simulated resampling of 1839 bp.

across genes, controlling for length allows shorter genes to be assessed to see whether their lack of resolution is due to the pattern of mutation or rather simply the fact that they contain relatively fewer base pairs.

Only one length variant occurred among the protein-coding genes, and this was found among the cytochrome *b* sequences. Any such length variants would cause a frameshift mutation and are therefore expected to be infrequent. Interestingly, however, for *Microgadus proximus* and *Merlangius merlangus*, the cytochrome *b* transcript was longer than for the other species. Normally, the stop codon in the mRNA is UAG, corresponding to TAG in the coding strand of the DNA. Interestingly, however, in these two species the T has undergone a transitional point mutation (as opposed to an indel) to a C thus causing the stop codon to now code for the amino acid glutamine. Twenty-four bases downstream of the usual stop codon there is an AGG (UGG in the mRNA) sequence, which presumably now functions as a stop codon and terminates polypeptide synthesis. The only difference is where the transcript terminates. The cytochrome *b* gene and the adjacent $tRNA^{Thr}$ overlap and both regions contain the AGG sequence naturally. This point mutation at the stop codon of cytochrome *b* causes it to be eight amino acids longer than is expected for this group. Thus, even though the transcript is longer, the DNA sequence has not changed in size.

Individual tRNAs are short and were therefore analyzed collectively as very little phylogenetic signal was observed in each of them separately. Of 1,403 bp of tRNA data, 71 bases were variable, and of these only 20 were phylogenetically informative.

Excluding missing and/or ambiguous sites, the total number of comparable sites across all taxa is 1,014 bp. None of the methods resolved the relationship between *Boreogadus* and *Arctogadus*. In all methods, a clade composed of *Pollachius, Merlangius,* and *Melanogrammus* was identified, usually with *Merlangius* and *Melanogrammus* as sister taxa. As well, in all cases a clade was identified composed of the three species of *Gadus* as well as *Theragra*, however *Theragra* usually came out in an unresolved clade with *G. macrocephalus/ogac* rather than as a sister to *G. morhua*.

Amino acid sequences generated for each protein-coding gene were also assessed for phylogenetic signal. Table 8 shows the pairwise number of amino acid differences among all taxa. As well, Table 9 shows the number of amino acids (i.e., codons) for each protein along with the number of variable and the number of phylogenetically informative amino acids per gene. In most cases, fewer than 5 phylogenetically informative sites were available. However, ND5 and ND4 contained 19 and 7, respectively. None of the individual amino acid sequences were able to resolve the complete phylogeny. An unweighted parsimony analysis of the 3,830 amino acids taken together produced three equally parsimonious trees. Bootstrap support identified a clade of *Merlangius* and *Melanogrammus* (76%) and the clade composed of *Arctogadus, Boreogadus, Theragra* and the three species of *Gadus* (96%). The clade composed of *G. macrocephalus* and *G. ogac* was resolved with 87% support. However, *Theragra* and *G. morhua* were not resolved as sister taxa, and in fact the former species was collapsed with *Boreogadus* at a polytomy (Figure 11). A neighbour-joining search on the total number of amino acids generated the tree shown in Figure 12. This approach identified more

Table 8. Pairwise number of amino acid differences calculated from 13 mitochondrial protein coding genes for 10 gadine species.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. *G. morhua* 002	-												
2. *G. morhua* Norway	14	-											
3. *G. ogac* 001	20	28	-										
4. G. *ogac* 002	10	18	10	-									
5. *G. macrocephalus* 002	15	23	19	9	-								
6. *G. macrocephalus* 003	17	25	16	11	14	-							
7. *T. chalcogramma*	22	30	30	20	25	27	-						
8. *A. glacialis*	26	34	33	24	28	30	32	-					
9. *B. saida*	32	40	42	32	37	39	36	35	-				
10. *M. aeglefinus*	54	60	61	54	60	62	61	58	67	-			
11. *M. merlangus*	56	62	67	57	60	60	60	60	61	60	-		
12. *P. virens*	42	49	51	43	45	48	46	47	54	50	60	-	
13. *M. proximus*	54	61	62	54	57	59	57	59	67	69	68	46	-

Table 9. Information on amino acid sequences for each gene region including the number of amino acids, the number of variable sites in each protein, and the number of phylogenetically informative sites for each protein.

Gene region	# amino acids	# variable sites	# phylogenetically informative sites
ATP6	227	8	3
ATP8	55	2	1
Cox1	516	9	2
Cox2	232	2	0
Cox3	261	10	3
Cytb	407	19	4
ND1	324	16	3
ND2	348	18	5
ND3	116	5	3
ND4	461	16	7
ND4L	98	11	2
ND5	612	45	19
ND6	173	13	2

Figure 11. Unweighted maximum parsimony analysis based on 3,830 amino acids spanning the 13 protein coding genes of the mitochondrial genome. Bootstrap values are shown in boldface and based on 1,000 replicates.

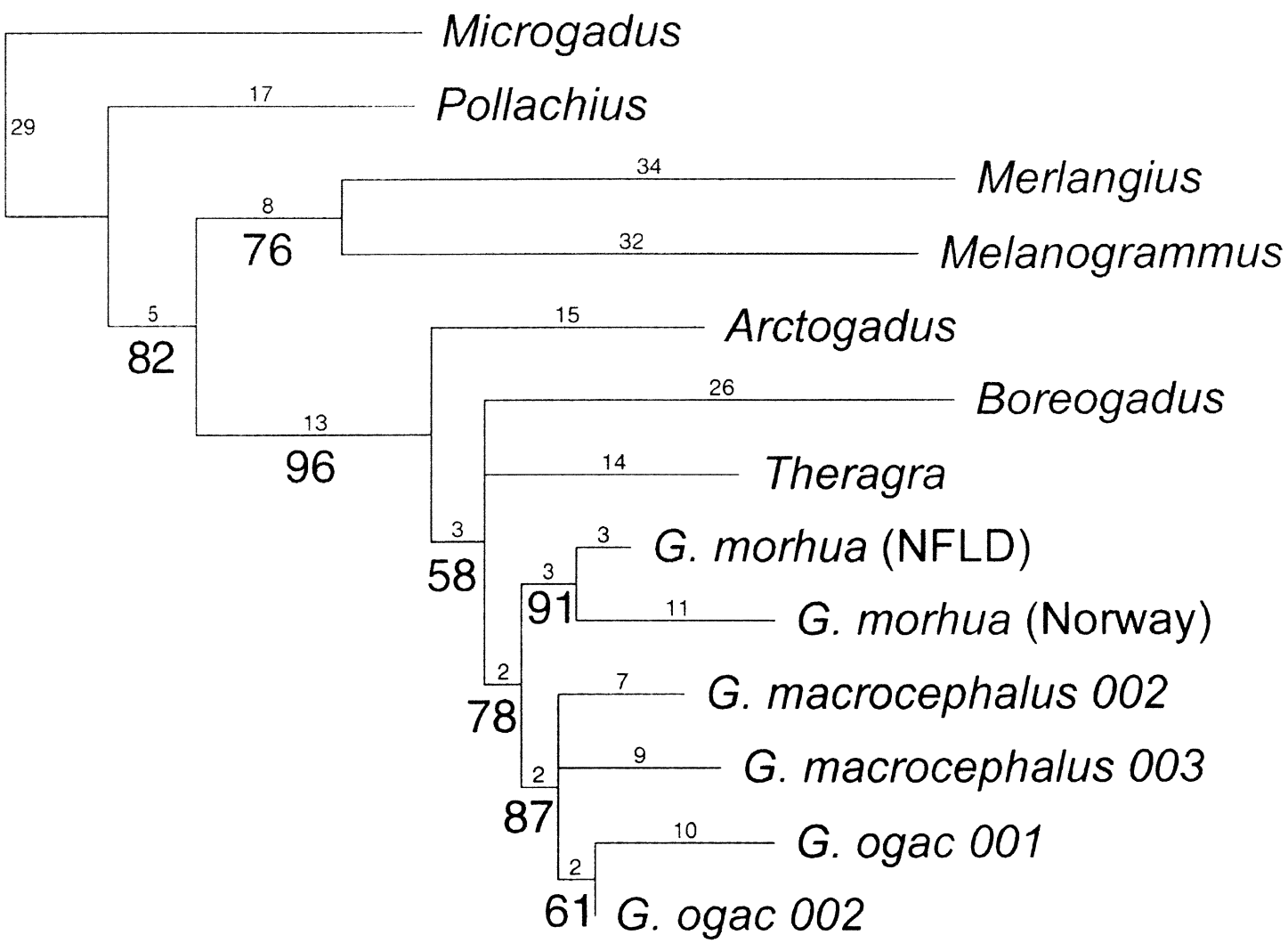
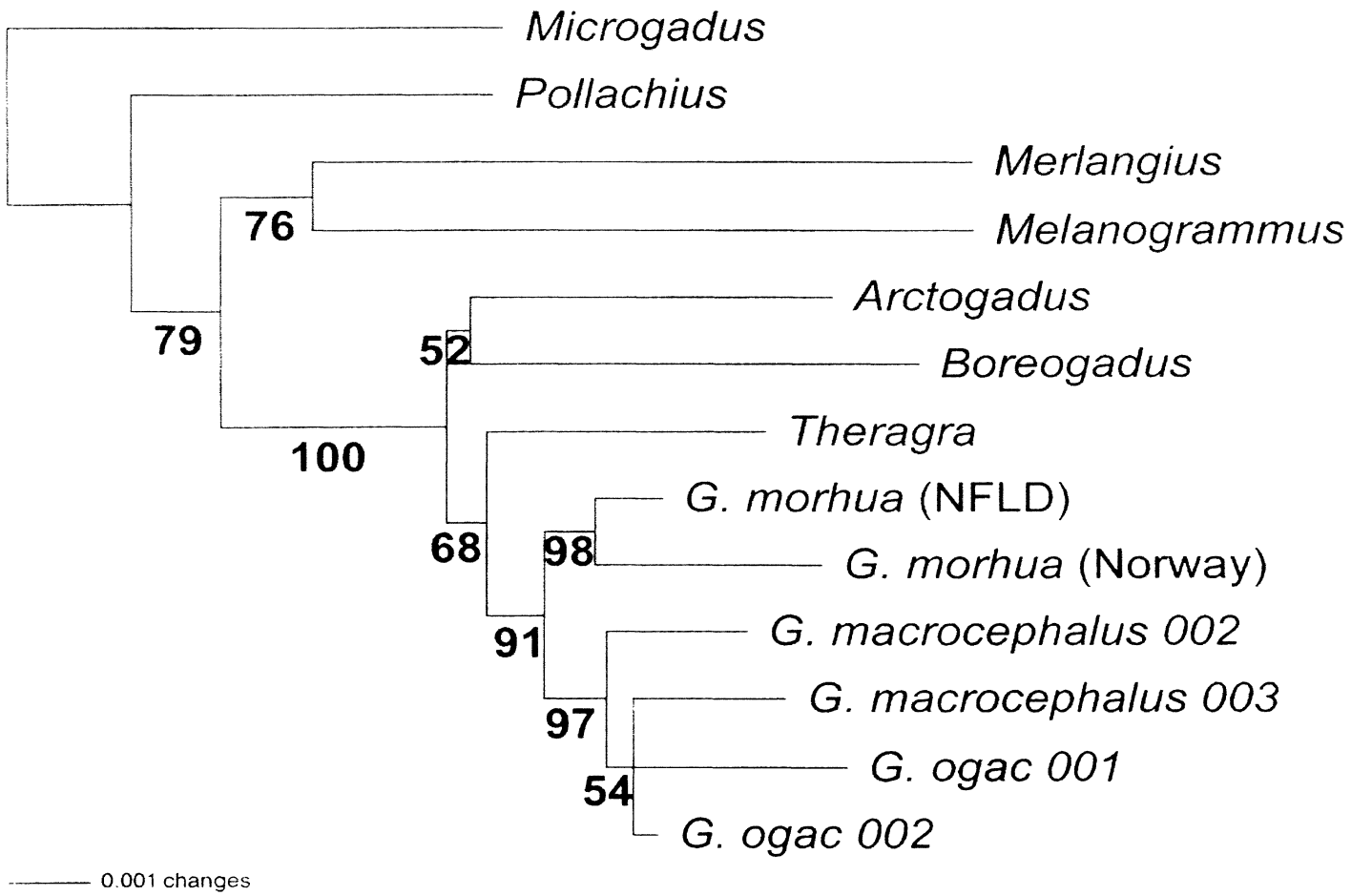

Microgadus
Pollachius
Merlangius
Melanogrammus
Arctogadus
Boreogadus
Theragra
G. morhua (NFLD)
G. morhua (Norway)
G. macrocephalus 002
G. macrocephalus 003
G. ogac 001
G. ogac 002
29
17
34
8
76
32
5
82
15
26
13
96
14
3
58
3
3
91
11
2
78
7
2
87
9
10
2
61

Figure 12. Neighbour-joining tree based on 3,830 amino acids spanning the 13 protein coding genes of the mitochondrial genome. Bootstrap values shown in boldface are based on 1,000 replicates.


Microgadus
Pollachius
Merlangius
Melanogrammus
Arctogadus
Boreogadus
Theragra
G. morhua (NFLD)
G. morhua (Norway)
G. macrocephalus 002
G. macrocephalus 003
G. ogac 001
G. ogac 002
76
79
52
100
68
98
91
97
54
0.001 changes

supported nodes via bootstrapping than the MP analysis, and gave a topology closer to that obtained with nucleotide data. NJ analysis identified *Boreogadus* and *Arctogadus* as sister taxa with only 52% bootstrap support. It also supported a close relationship between the two *G. ogac* and one of the *G. macrocephalus* individuals, however, contrary to nucleotide data, it identified *G. macrocephalus* 003 as more closely related to the two *G. ogac* individuals. Bootstrap values with the NJ approach were generally higher than with MP. However, *Theragra* was not recognized as a sister taxa to *G. morhua*, as the latter species as well as *G. macrocephalus* and *G. ogac* formed a monophyletic clade, excluding *Theragra* (contrary to the nucleotide data) with 91% bootstrap support (Figure 12).

4.0 DISCUSSION

4.1 Phylogenetic Relationships Among the Gadinae

The present day distribution of gadines suggests that their original area of endemism was the eastern North Atlantic (Svetovidov, 1948) and molecular work also supports this hypothesis (Carr *et al.*, 1999; Møller *et al.*, 2002). The study by Carr *et al.* (1999) identified a clade composed of the two species of *Microgadus* (*M. proximus* and *M. tomcod*), as well as *Eleginus navaga* as an outgroup to the remainder of the species considered here, and allowed for the rooting of trees presented in this study with *M. proximus*. All resolved branches across the three modes of phylogeny reconstruction identified the same series of relationships, generally with statistically significant bootstrap support (minimum 6 of 10 nodes supported with $\geq 95\%$).

Concerning the origins of the Pacific gadines, the phylogeny of Carr *et al.* (1999) indicated that *Microgadus proximus* represents an invasion of the Pacific separate from that for *G. macrocephalus* and *Theragra chalcogramma*. Because a clade containing the latter two species could not be resolved in that study, it could not be determined whether *Theragra* and *G. macrocephalus* represent a single invasion which has since diverged into these two species, or whether each of these species represents a separate invasion into the Pacific basin on their own. The fact that *Theragra* and *G. macrocephalus* are not each other's closest relative supports the theory that each of these species represents a separate invasion into Pacific waters, with *G. macrocephalus* having a more recent origin than *Theragra* due to its shorter branch length (Figures 5, 6, and 7). It was proposed by Carr *et al.* (1999) that each of these Pacific species represented a separate yet

simultaneous invasion into the Pacific around the time of the opening of the Bering Strait ~3-4 MYBP. From Carr et al. (1999), it can be hypothesized that the common ancestor to *M. proximus* and *M. navaga* split and one lineage (that leading to *M. proximus*) entered the Pacific basin from Arctic waters (Figure 3a). It therefore follows that the lineage leading to *Theragra* entered the Pacific basin later than the entrance of *M. proximus*, but before that of *G. macrocephalus*, according to the phylogenies presented here.

The fact that *G. ogac* is most closely related to *G. macrocephalus*, yet is found in the Atlantic Ocean, would therefore suggest that this is a secondary invasion into the Atlantic, where it has come into contact with the sympatric *G. morhua*. This re-introduction of *G. ogac* into the Atlantic could have been done by either of two routes. One would involve migration through the northern Canadian archipelago and into the Atlantic basin. The second scenario would be a migration route through the Polar basin and down the east coast of Greenland. A Seattle oceanographer tracking floating debris moving with the ice in the Pacific, to study ocean currents discovered that some of the tracking devices were ending up in the Atlantic and had drifted there by a major ocean current running directly through the Bering Strait into the North Pole and down the eastern coast of Greenland (Dr. Cutis Ebbesmeyer, personal communication) (Figure 13). Therefore it is possible that either of these passages could act as interoceanic routes for marine species and that a founding population of Pacific cod entered the Atlantic basin and differentiated into *G. ogac*. Taking into consideration the environment and the species biology, the first scenario (a northwest passage through the northern islands)

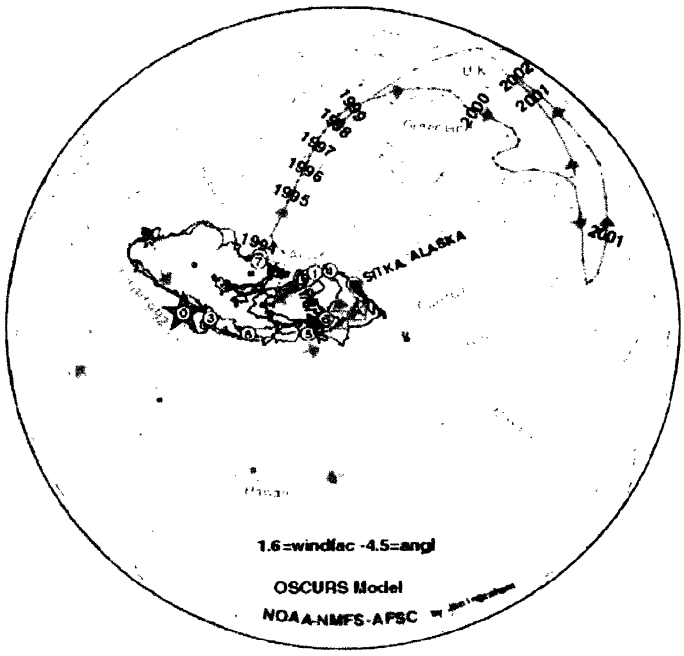


Figure 13. Polar projection map showing the route of a Pacific Ocean current passing through the North Pole and into the Atlantic Ocean (diagram provided by C. Ebbesmeyer)

seems more plausible. Due to impeding ice flows and a greater distance required to travel, the passage through the North Pole would seem to require considerably more time. Despite this fact, the Arctic Ocean is several thousand meters deep, which is much deeper than the normal isobaths in which these fish are found. This route would therefore require that these fish would have to become pelagic for an extended period, which is in contrast to their demersal life-history. Therefore, this would require a major change in their biology. Additionally, contemporary studies have shown that *G. ogac* can be found in and around the Hudson's Bay area and throughout the Bering and Beaufort seas (Mikhail & Welch, 1989; Morin *et al.*, 1991). If passage therefore occurred through the northern islands, it is not surprising that remnant populations in this area continue to exist. The passage through the North Pole would have by-passed this area, and contemporary populations would therefore have had to re-invade the Hudson's Bay area after Atlantic Ocean contact. This movement from the Pacific to the Atlantic through the northern archipelago could, therefore, have occurred as a more gradual progression in the distribution of fish from one ocean to the other and would have allowed them to remain in more coastal, shallower areas, and therefore maintain their demersal habitat.

Historically, it has been assumed that *G. ogac* is more closely related to the sympatric *G. morhua* than to *G. macrocephalus*. However, Carr *et al.* (1999) observed identical nucleotide sequences across 896 bp of mtDNA between *G. ogac* and *G. macrocephalus*. The expanded data set of 14,036 bp of mtDNA resolved 45-56 nucleotide differences between *G. ogac* and *G. macrocephalus*, and indicates that they are each other's closest relatives. Renaud *et al.* (1986) also showed with protein

electrophoretic comparisons that *G. macrocephalus* and *G. ogac* were genetically indistinguishable as they were fixed for identical allele frequencies at 17 of 21 protein loci, and the other 4 loci only differed by minor alleles at low frequencies. Additionally, Sick (1965) studied hemoglobin types among the three nominal species of *Gadus*, and noticed a closer similarity between *G. macrocephalus* and *G. ogac* than either of these to *G. morhua*. Another similarity between *G. macrocephalus* and *G. ogac* is that they both produce a negatively or neutrally buoyant egg, in contrast to most other gadines which have positively buoyant eggs. As well, similarity between *G. ogac* and *G. macrocephalus* was indicated by Schultz & Welander (1935) who were unable to find any morphological differences in the shape of the swimbladders, counts of fin rays and vertebrae, and numerous body and fin measurements between these two species (Møller *et al.*, 2002).

G. ogac is distinguished from *G. macrocephalus* by the colouration of the peritoneum and by a difference in size with *G. ogac* not usually longer than 70 cm whereas *G. macrocephalus* reaches lengths over one meter. This size difference could simply be a result of *G. ogac* occupying a colder, more northerly environment, as it is known that *G. macrocephalus* grow much faster in the south and spawning occurs at a younger age. Carr *et al.* (1999) suggested that *G. ogac* therefore represents a northward and eastward extension of the range of *G. macrocephalus*. The fact that mitochondrial genomic differences between *G. ogac* and *G. macrocephalus* are on the same order as differences between the two *G. morhua* individuals suggests that gene flow may still occur throughout this region. However, it is also likely that the lack of divergence simply

reflects the recent origin of the separation between the two nominal species and the lower amount of sequence divergence among the two *G. ogac* individuals may additionally be associated with a bottleneck experienced by a founding population in the Atlantic. Further work is necessary to determine whether contemporary gene flow still occurs between these two ocean basins. Since the present study only contained two individuals each of *G. ogac* and *G. macrocephalus*, the clustering of the two *G. ogac* individuals together may not necessarily indicate that gene flow has ceased with the nominal *G. macrocephalus* and that phylogenetic relationships may be less pronounced between these two nominal taxa. In either case, the data indicate that the two *G. ogac* individuals are more closely related to one of the two *G. macrocephalus* individuals. It is therefore suggested that *G. ogac* be included as a subspecies of *G. macrocephalus* and therefore be recognized as *G. macrocephalus ogac*. *G. macrocephalus* (Tilesius 1810) has nomenclatorial priority over *G. ogac* (Richardson 1836).

Another hypothesis, however, could explain the distributions of these pairs of Atlantic and Pacific species. Since biogeographical movements from the Atlantic into the Pacific are less common than the reverse, a more parsimonious hypothesis may include fewer individual lineages entering the Pacific and more incidents of re-invasion into the Atlantic basin. Among invertebrate species, the number of Pacific immigrants into the Atlantic has been estimated to exceed the number of Atlantic immigrants into the Pacific by a factor of eight (Svitoch & Taldenkova, 1994). These authors results were based on the origins of invertebrate fossil fauna dating back as far as the Miocene-Pliocene boundary (~ 5.5-6.5 million years ago) (Svitoch & Taldenkova, 1994). They

noticed an almost complete lack of Arctic and boreal-Arctic species considered to be of Atlantic origin in sediments from the Bering Strait and Pacific coasts and a one-directional migration of Pacific boreal species through the Bering Strait into the Beaufort Sea and through the Canadian archipelago to the North Atlantic (Svitoch & Taldenkova, 1994). They suggested that such historical faunal distribution patterns support the existence of a warm current originating in the Bering Sea, similar to more recent (~2,000-3,000 years ago) currents, but more powerful. It was also noted that some of these molluscs lacked a planktonic larval stage and therefore migrated as adults rather than being carried on ocean currents (Svitoch & Taldenkova, 1994). Under this hypothesis, the *Gadus* clade (including *Theragra*) would have entered the Pacific basin from the *Boreogadus-Arctogadus* lineage originating in the Arctic. At some point after entering the Pacific, there was a split resulting in two lineages, one represented by present-day *G. morhua* and *Theragra*, and the other by a more recent *G. macrocephalus*. Assuming this scenario implies that not only does *G. macrocephalus ogac* represent a re-invasion of the Atlantic, but so does *G. morhua*. If the *G. morhua-Theragra* lineage split in the Pacific, then *G. morhua* also would have followed the more common trend of immigration from the Pacific to the Atlantic. Recent studies have suggested an earlier opening of the Bering Strait than the previous age of 3-4 MYBP. Marincovich & Gladenkov (1999) placed the Strait's first opening at 4.8 ~ 7.4 Myr ago based on the presence of Atlantic-Arctic fossil diatoms in the Pacific prior to the appearance of Pacific fossils in the Atlantic ~3.5 MYBP. Other studies suggest that a more accurate estimate for the first opening is 4.8-5.5 MYBP (Marincovich Jr. & Gladenkov, 2001; Gladenkov *et al.*, 2002)

based upon fossil diatoms and fossil molluscs. Additionally, evidence exists for a dominant southward flow through the Bering Strait during its first opening (~4.8-5.5 Myr), that continued until sometime after 4.6 MYBP. At this point a critical threshold in the closure of the Isthmus of Panama is believed to have caused a marked reorganization of Northern Hemisphere ocean circulation, including the onset of the present northward flow through the Bering Strait (Marincovich Jr. & Gladenkov, 1999). Therefore, during this southward flow through the Bering Strait, it may have been more likely (temporarily) for Atlantic to Pacific migration to occur, which would have allowed at least two separate invasions of gadines to take place. One of these invasions would have been the *Microgadus* invasion and the second invasion represented by the *Gadus* lineage. Upon the closure of the Isthmus of Panama, the change in direction would have contributed to the current pattern of inter-ocean migration allowing *Theragra chalcogramma* and *G. ogac* to invade the Atlantic. This first opening of the Bering Strait allowed marine organisms to migrate between the Arctic and North Pacific oceans for the first time since the middle Cretaceous period, ~105 MYBP (Marincovich Jr. & Gladenkov, 1999). The time of occurrence for this event has never been precisely dated and makes it difficult to incorporate it into models of biogeography and oceanography (Marincovich Jr. & Gladenkov, 1999).

Studies of allozyme allele frequencies in Pacific and Atlantic species of *Gadus* not only support recognition of *G. morhua* and *G. macrocephalus* as separate species, but also show a significant reduction in heterozygosity in *G. macrocephalus* relative to *G. morhua* (Grant & Ståhl, 1988). These authors explained this by proposing that Pacific

cod experienced a population bottleneck upon entering the Pacific Ocean, and that insufficient time has elapsed to reach mutation-drift equilibrium and to regain levels of heterozygosity consistent with current population size. From the present analysis, however, the number of nucleotide differences between the two *G. macrocephalus* individuals is greater than that between the two *G. morhua* individuals (73 versus 52, respectively). Differences in genetic diversity within each of these species has not been attributed to genetic structuring as neither of these species occur as multiple, genetically distinct local stocks (Grant *et al.*, 1987; Grant & Ståhl, 1988; Carr *et al.*, 1995; Bentzen *et al.*, 1996; Carr & Crutcher, 1998). It is recognized that this comparison (amino acids *vs.* nucleotides) is not for a parallel data set. However, the discrepancy between the protein data implying reduced heterozygosity and the nucleotide data which indicate *G. macrocephalus* mtDNA is as variable, if not more so, than that for *G. morhua*, may simply reflect that *G. macrocephalus* has existed for only a "relatively" short time and perhaps insufficient mutations have occurred to create more protein diversity. As well, Grant & Utter (1980) showed that *Theragra chalcogramma* is also genetically depauperate compared to *G. morhua*. Again, this would favour the origin of the genus *Gadus* in the Atlantic and two separate Pacific invasions for each of *Theragra* and *G. macrocephalus*. Further molecular studies will help to resolve the directionality of these inter-ocean migrations.

The fact that *Theragra chalcogramma* and *Gadus morhua* are sister taxa (Figures 5, 6, and 7) raises some important evolutionary considerations. *Theragra* and *G. morhua* have supported the largest fisheries in the world. The ability of these two species to

support such intense, heavily exploited fisheries for many years may therefore be the result of a common evolutionary history. For example, the fact that both species produce large numbers of individuals may be attributable to their evolutionary background. Both of these species are among the most fecund marine fishes in the world, with average egg production on the order of 1 million eggs per female for *G. morhua* (with a 10 kg female producing 5 million eggs) and 500 000 to 15 million eggs, depending on age for *Theragra chalcogramma* (Cohen *et al.*, 1990). The interesting aspect of this relationship is that these two species are both large, fecund fish that appear to have similar ecological roles in their respective oceans, which has allowed them to be abundant. Additionally, decreases in biomass have limited their exploitation, and responses that have lead to such changes in their abundances may be a result of common evolutionary history. It has been hypothesized that the divergence between them is on the order of 3-4 million years and therefore that each of these species has evolved in similar ways in their respective environments in the two separate oceans.

In light of this relationship, the taxonomic status of *Theragra* requires revision in order to keep the genus *Gadus* monophyletic. Interestingly, *Theragra chalcogramma* was initially described as *Gadus chalcogrammus* (Pallas, 1811). Most, if not all, the gadines were originally described as *Gadus*, because it was recognized that they were all codfish. As more differences became apparent, the nomenclatorial status of many of them changed, including the splitting of *Gadus chalcogrammus* to *Theragra fucensis* (Jordan & Evermann, 1898) and subsequently as currently described – *Theragra chalcogramma* (Svetovidov, 1948). Keeping this lineage monophyletic is more

parsimonious than having it paraphyletic, as would be the case if the genus name remains *Theragra*. As well, since scientific names reflect evolutionary relationships, calling this species *Theragra* incorrectly implies that *Gadus morhua* is more closely related to *Gadus macrocephalus* and *Gadus ogac*, since these are congeneric species. Therefore, a name change to *Gadus chalcogrammus* for the walleye pollock would reflect its close evolutionary relationship to *G. morhua*, a relationship supported by the molecular data.

Carr *et al.* (1999) hypothesized at least one invasion of gadines into Arctic waters has occurred. This invasion from the eastern Atlantic into the western Atlantic and Arctic is represented by the lineage leading to the two species of *Microgadus* plus *Eleginus navaga*. Carr *et al*. (1999) argued that *E. navaga* should be included in *Microgadus* (as *M. navaga* Pallas 1811) in order to keep that genus monophyletic. *Microgadus tomcod*, which came out as the sister to the other two species is endemic to the eastern Atlantic, while *Eleginus* is found in the Arctic and *M. proximus* is one of the three endemic Pacific species. A second Arctic invasion, suggested by the present study, is represented by the lineage containing *Boreogadus*, *Arctogadus*, and *Gadus* (including *Theragra*). *Boreogadus* is one of the most northerly distributed species of bony fish (Scott & Scott, 1988), and it along with *Arctogadus* and *Eleginus* represent the three species of gadines endemic to Arctic waters. The placement of *Boreogadus* as the outgroup to *Arctogadus* and *Gadus* (including *Theragra*) suggests that this particular clade had its origins in the Arctic Basin. As the outgroup to the *Boreogadus*-inclusive clade (*Pollachius, Melanogrammus* & *Merlangius*) are represented on either or both sides

of the Atlantic, the *Boreogadus* lineage may have split off from the eastern and western Atlantic lineage to enter Arctic waters.

The exact speciation scenario of *Boreogadus* and *Arctogadus* is not easily explained by geographic or thermal events (Møller *et al.*, 2002). Although the distribution of *Boreogadus* in Arctic waters is more continuous than the fragmented, disjunct population of *Arctogadus*, these two species are sympatric in parts of their range. It has been hypothesized that this may be a result of a secondary invasion of *Boreogadus* into the Arctic Ocean (Møller *et al.*, 2002). As seen in Table 4, across all six methods used to resolve the relationship between *Boreogadus* and *Arctogadus*, the only strong support is for *Boreogadus* outside of *Arctogadus* and the rest of the clade. Møller *et al.* (2002) suggested that *Arctogadus* be synonymized either with the genera *Boreogadus* or *Gadus*. These authors found that the genetic differences between *Arctogadus* and *Boreogadus* and between *Arctogadus* and *Gadus* (including *Theragra*) were less than the difference within other genera (such as *Trisopterus*). However, basing this reclassification on sequence divergence alone is not warranted. As different lineages have different evolutionary rates and histories, looking at percent sequence divergence may be misleading. Møller *et al.'s* (2002) comparison of *Arctogadus-Boreogadus* to that of *Trisopterus* is not valid, as *Trisopterus* is an older lineage and may therefore exhibit a larger amount of sequence divergence among its species. Additionally, *Arctogadus* should not go into *Boreogadus* or *Gadus* if they are not sister taxa. *Arctogadus* differs from all other codfishes in having dentigerous palatine bones and elliptical imbricate scales (Nielsen & Jensen, 1967; Møller *et al.*, 2002). Additionally, *Arctogadus* differs

from *Boreogadus* by having larger eyes, the number of rakers on the anterior gill arch (26-38 vs. 37-45), and a strait *vs.* a wavy lateral line in *Boreogadus*. Therefore, at present, *Arctogadus* should remain as currently classified.

One of the more surprising relationships found in this group is the occurrence of *Merlangius* and *Melanogrammus* as sister taxa. *Pollachius* is the immediate outgroup and occupies either the eastern Atlantic (*P. pollachius* – not included here) or is trans-Atlantic (*P. virens*). *Merlangius merlangus* occupies the eastern Atlantic and is divided into two subspecies by most authors (Cohen *et al.*, 1990), while *Melanogrammus aeglefinus* has a trans-Atlantic distribution. The eastern Atlantic distribution of *Melanogrammus* is almost entirely sympatric with *Merlangius* except for the former's absence in the Mediterranean Sea. Re-analysis of Dunn's (1989) morphological characters fails to identify any unique synapomorphies uniting the two, and on general inspection of morphological data, the only unique feature appears to be a dark spot above the pectoral fin, which is less pronounced in *Merlangius*. It is known that whiting and haddock occupy relatively different ecological niches, with whiting generally feeding higher in the water column (Cohen *et al.*, 1990). It may be possible that the evolution of these two species has occurred through ecological selection and subsequent reproductive isolation. Such occurrences are more typical of anadromous or freshwater species (i.e. stickleback, and various salmonids) (Ferguson & Mason, 1981; McPhail, 1992; Taylor & Bentzen, 1993; Schluter, 1996b, a; Pigeon *et al.*, 1997) and seem to be relatively rare in marine fishes. Alternatively, if some oceanographic conditions (such as those involved with the reorganization of ocean currents upon formation of the Bering Strait and Isthmus

of Panama) had separated the common ancestor into an eastern (*Merlangius*) and western (*Melanogrammus*) lineage, then the presence of *Melanogrammus* in the eastern Atlantic could represent a secondary invasion after an extended period of time of allopatry. This seems less likely, as many gadiforms have an eastern Atlantic distribution and others tend to be trans-Atlantic. This scenario would therefore have to involve a complex shifting of the oceanographic patterns, possibly multiple times, as this pattern does not appear to coincide with the occurrence of a particular lineage. Whatever the case, in all analyses presented here, these two species were resolved as sister taxa with 100% bootstrap support, and had more synapomorphies between their mtDNA sequences than did the *Boreogadus* clade (265 *vs.* 229).

In summary, the data presented here support the hypothesis of multiple invasions of the Pacific basin by the three endemic species. Carr *et al.* (1999) resolved an invasion of *M. proximus* into the Atlantic separate from that of *Theragra* and *G. macrocephalus*. The present study, further resolves the invasions of the latter two species as separate, invasions into the Pacific and suggests that there has been at least one occurrence of secondary contact with the introduction of *G. macrocephalus ogac* into the Atlantic Ocean. The hypothesis put forward for the relationship between *Melanogrammus* and *Merlangius* as sister taxa also requires further attention, as this relationship has not been documented from prior work, and therefore further study is required in light of the evolutionary relationships implied by the mtDNA data. Further analyses will help to clarify distributional patterns among this family as well as other families of Gadiformes.

Phylogenetic history and genetic diversity should be used in biodiversity indices in order to emphasize both the phylogenetic and genetic distinctiveness of certain groups of species compared to others. Correct taxonomic names are important to ensure correct thinking about the evolutionary relationships among species. If the taxonomic status of a group of species is misapprehended, questions regarding almost any aspect of their biology may be confounded by improper understanding of their evolutionary relationships implied by their classification. Thus, by understanding the correct names of species, we are more confident about the validity of addressing questions related to evolutionary biology.

4.2 Utility of Individual mtDNA Genes in Phylogeny reconstruction

Although phylogenetic trees based on mtDNA are predicted to be good estimators of species trees, gene tree topologies may not accurately reflect species trees due to historical lineage sorting and/or horizontal gene transfer (Pamilo & Nei, 1988; Avise, 1989; Moore, 1995; Creer *et al.*, 2003). Most past phylogenetic studies have historically employed single genes or parts of selected genes. More recent phylogenetic studies have been employing complete mitochondrial genomes (Rasmussen & Arnason, 1999; Takezaki & Gojobori, 1999; Saitoh *et al.*, 2000; Inoue *et al.*, 2001a; Miya *et al.*, 2003; Murata *et al.*, 2003; Phillips & Penny, 2003) or at least all mtDNA protein-coding loci (e.g., Corneli & Ward, 2000; Inoue *et al.*, 2003). Most of these 'mitogenomic' approaches have been used to resolve ancient lineages and basal taxonomic relationships that had previously been poorly resolved or unrealistic with smaller data sets. It is noted,

however, whole mtDNA genomes are still subject to lineage sorting as all loci are effectively linked The current study is an example of this approach, however, on a relatively recent evolutionary lineage. Therefore, by sampling mitochondrial genomic sequences, discrepancies between topologies reflected by individual gene regions were avoided, and a more consistent topology was generated. This resulted in little difference across the various modes of analysis due to the large amount of data, thereby reducing the discrepancy between topologies which may arise as a consequence of which tree-building method was employed. This increased sampling also allowed for a sufficient number of substitutions to accrue throughout the genome, which in most cases was statistically significant.

Since only 13 genes were studied, statistical evaluation of the relative efficiencies of the different tree-building methods may require additional sampling of nuclear gene sequences (Russo *et al.*, 1996). However, MP weighting transversion only was the least accurate method of tree construction. This fact is not surprising given the recent divergence times of this group, because less time has passed for transversions to accumulate. It was evident in the individual gene trees for the MP transversion-only analysis that the high scores contributing to the increased amount of topological differences were primarily due to a lack of resolution of interior branches, caused by the relatively few transversional synapomorphies among taxa.

Maximum-likelihood was the next least accurate mode of tree topology reconstruction. This may be due to the fact that this method requires a number of assumptions regarding substitution pattern and rate of evolution. These parameters may

not have been accurately estimated given the different rates and modes of evolution among the various gene regions. Additionally, the consensus sequences for this mode of construction may have been more problematic as it may have been more difficult to estimate certain parameters with the different gene regions contributing various amounts and degrees of variation.

When the extent of sequence divergence is high and the pattern of nucleotide substitution remains nearly the same for all sites over time, ML methods are expected to give better results in topology estimation than other methods. However, as the amount of sequence divergence among the taxa in the present analysis varies and may be less than typical studies of more ancient divergences, one can expect that this advantage of ML over other methods to decline because other methods require less rigid assumptions about substitution pattern and rates of evolution (Nei, 1996). It is also known that MP methods tend to be more prone to give biased estimates of branch lengths than other methods, and there is no natural way to compute means and variances of minimum numbers of substitutions. These factors may, at least in part, explain the discrepancy in bootstrap support among the various analyses for the resolution of *Boreogadus* and *Arctogadus*.

The slightly higher bootstrap support for the paraphyly of *G. macrocephalus* in the analysis of all taxa *versus* the analysis of only the *Boreogadus*-inclusive clade may be explained by levels of homoplasy. In the reduced data set, bootstrap values increased by as much as 12%. As more distant taxa are added to an analysis, the likelihood of homoplasy (parallel evolution) increases. Therefore, the addition of more divergent taxa (such as *M. proximus*) may have acted to reduce the confidence placed on the *G.*

macrocephalus node. Aside from this, it is apparent that as a sufficiently large number of nucleotides or amino acids are used, the discrepancy between modes of analyses (i.e. MP, NJ, & ML, Figures 5, 6, and 7) becomes less and topologies become more congruent. This also implies that for larger data sets, simpler models of evolution may perform more accurately than more complex models (Nei, 1996).

As indicated among the individual gene regions, there was a strong correlation between the size of the gene region and the number of phylogenetically informative sites ($r = 0.936$) (data not shown). This is not surprising, since a greater number of nucleotides would yield more 3rd codon positions and therefore proportionally more variable and more phylogenetically informative sites. The negative correlation ($r = -0.59$, Figure 8) between the size of the gene and the accuracy of its tree topology as compared to the consensus tree (as measured by Russo *et al.*, 1996) suggests that there is a slight tendency for the larger genes to be more phylogenetically accurate. However, it also suggests that phylogenetic accuracy depends heavily as well on the intrinsic properties of the individual gene regions and their rate of evolution and substitutional patterns.

A comparison of the relative ranking of efficiency for the 13 protein coding genes from the present study as well as that of Russo *et al.* (1996) (Figure 10) showed that some genes behaved very differently between the two studies. For example, the present study found that ND1 and CO2 ranked fairly high in phylogenetic accuracy (1st and 4th, respectively), while Russo *et al.* (1996) found these genes to be less efficient or among the least accurate for their taxa (5th and 13th, respectively). These same authors found that ND6 ranked 2nd, while for this study it was 12th. As well, Russo *et al.* (1996) found

cytochrome *b*, to be an average indicator of phylogenetic accuracy (5^{th}), while here it was one of the least accurate (10^{th}), however as mentioned half of the data are missing for *Arctogadus*, which may cause cytochrome *b* to appear less accurate. Other genes received nearly identical rankings between the two analyses. ND5 was either 1^{st} or 2^{nd} among the two studies, while ND4 ranked 3^{rd} in both situations (Figure 10). As well, for both studies, ATP6 performed relatively poor, being ranked either 10^{th} (Russo *et al.*, 1996) or 11^{th} (this study). Similarities in the efficiency of particular genes in resolving phylogeny across studies with different taxonomic groups and times of divergence may reflect some intrinsic properties for a given gene in assessing phylogenetic signal, however, this comparison between only two studies does not allow for firm conclusions to be drawn. On the other hand, differences in performance among genes reflects that certain genes have properties that make them better for phylogeny reconstruction, depending on the time since divergence and/or the taxonomic lineage involved. This indicates that suggestions that any particular gene (e.g. CO1 – Hebert *et al.*, 2003a,b) is best suited for phylogeny reconstruction universally at any given level and for any given set of taxa is unfounded, and could lead to erroneous conclusions if assumed to apply to all taxonomic groups at various degrees of evolutionary divergence. The current study and that of Russo *et al.* (1996) both suggest that it is apparent that some genes give an accurate topology and other genes consistently give a very inaccurate topology regardless of the mode of analysis. However, this is not necessarily a characteristic of the gene itself and as of yet there is no way to predict, for a given series of taxa, which gene will

be the most accurate. However as sequencing becomes more routine and less expensive, this can be circumvented by using the mitogenomic approach.

Genes of the cytochrome oxidase complex and cytochrome *b* have been used extensively in the past for phylogeny reconstruction. However, genes belonging to the NADH complex may also be appropriate and in some cases preferable. Both ND4 and ND5 performed relatively well in both of these studies, and Russo *et al.* (1996) suggested that NADH gene complexes should be used more often in single- or few-gene phylogenetic analyses. Their relative efficiency and the fact that some of these genes are among the largest mitochondrial-encoded genes may make them excellent candidates as a exploratory genes (i.e. as a preliminary assessment of phylogeny based on a more limited data set).

The results presented in Table 7 and Figure 9 indicate that phylogenetic accuracy of certain gene regions may be limited primarily by their size. For example, for ND4L which is among the smallest mtDNA protein-coding genes, only two nodes were supported with slightly more than 50%. The simulated data set of 1,839 bp increased this to six nodes at 50% and even began resolving some nodes at the 75% and 95% level. Therefore, for this gene region, as size was increased more branches became resolved. Alternatively, an increase in other gene regions did not change the number of resolved branches, but rather increased the support on pre-existing nodes. For example, ND6 resolved six nodes with more than 50% support ($3 \geq 75\%$ and $2 \geq 95\%$). The increased sampling of ND6 did not resolve any further nodes, but increased the support of these nodes such that four were greater than 75% and two greater than 95%. These results

suggest that while some genes may be limited due to a relatively small size of the gene, others may be more restricted for phylogenetic use by inherent patterns of their molecular evolution.

The amino acid phylogenies generated by MP and NJ analyses revealed an interesting aspect regarding the placement of *Theragra*. In both modes of analysis, *Theragra* is resolved outside of a clade containing the nominal species of *Gadus* (Figures 11 & 12). The null hypothesis for amino acid substitutions is that they are not functionally significant, and are more an epiphenomenon of an accumulation of nucleotide substitutions. However, the relatively high support for the phylogenetic placement of *Theragra* suggests that there could be some functional considerations, since the protein sequences give a different relationship than the nucleotide sequences. Individual amino acid sequences among these taxa were investigated in an attempt to identify the 2 synapomorphies separating *Theragra* from *Gadus*. Both of these events were found to occur in the amino acid sequences for ND5. This suggests that this protein may have an adaptive function selected for the particular environment associated with *Theragra*. This would suggest that there may be kinetic properties associated with this protein and that studies addressing differences in metabolism may be warranted. It has been demonstrated that the protein lysozyme among leaf-eating primates is more similar to ruminant lysozyme than would be expected on the basis of the associated nucleotide sequence (Messier & Stewart, 1997), suggesting that selection among proteins may permit the exploitation of new niches.

Clearly, increasing the amount of molecular data for phylogenetic studies helps to resolve conflict surrounding the analysis and allows for resolution of relationships that previously have been unresolved or troubled with high levels of homoplasy. Additionally, the need for rigourous choice among modes of analysis is eliminated. Previous molecular phylogenetic studies have largely focused on specific genes or parts of genes. However, with advancing technology, it is reasonable to apply large data sets, such as the one presented here to answer phylogenetic questions. Currently, there is a lot of debate surrounding molecular phylogenetics and which approaches to use as well as the relative merits and demerits of various modes of tree reconstruction and analysis. As more data become available we will be in a better position to understand phylogeny reconstruction and apply it to issues regarding organisms distribution patterns, biogeographic history, species diversity and factors involved in speciation. All of these are heavily debated topics and are fundamental issues in evolutionary biology.

5.0 Literature Cited

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457-465.

Andriashev AP, Mukhomediyarov BF, Pavshtiks EA (1980) On the mass gathering of cryopelagic cod fishes (*Boreogadus saida* and *Arctogadus glacialis*) in circumpolar areas of the arctic. National Museum of Canada.

Arnason E, Henry P, Palsson S (1998) Mitochondrial cytochrome b DNA sequence variation of Atlantic cod, *Gadus morhua*, from the Baltic and the White Seas. *Hereditas*, **129**, 37-43.

Arnason U, Gullberg A, Gretarsdottir S, Ursing B, Janke A (2000) The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. *Journal of Molecular Evolution*, **50**, 569-578.

Asakawa S, Kumazawa Y, Araki T, Himeno H, Miura K, Watanabe K (1991) Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. *Journal of Molecular Evolution*, **32**, 511-520.

Avise JC (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, **43**, 1192-1208.

Avise JC (1994) Molecular Markers, Natural History and Evolution. Chapman and Hall, New York, N.Y.

Bakke I (2001) Characterization of mitochondrial DNA sequences and molecular phylogenetics of the family Gadidae and related taxa. PhD Thesis, Department of Molecular Biotechnology, University of Trömso, Trömso, Norway.

Bakke I, Johansen S (2002) Characterization of mitochondrial ribosomal RNA genes in gadiformes: sequence variations, secondary structural features, and phylogenetic implications. *Molecular Phylogenetics and Evolution*, **25**, 87-100.

Bakke I, Shields G, Johansen S (1999) Sequence characterization of a unique intergenic spacer in Gadiformes mitochondrial DNA. *Marine Biotechnology*, **1**, 411-415.

Ballard JWO (2000) Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *Journal of Molecular Evolution*, **51**, 48-63.

Bentzen P, Taggart CT, Ruzzante DE, Cook D (1996) Microsatellite polymorphism and the population structure of Atlantic cod (*Gadus morhua*) in the northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 2706-2721.

Berg OG, Kurland CG (2000) Why mitochondrial genes are most often found in nuclei. *Molecular Biology and Evolution*, **17**, 951-961.

Bermingham E, McCafferty SS, Martin AP (1997) Fish Biogeography and Molecular Clocks: Perspectives from the Panamanian Isthmus. In: *Molecular Systematics of Fishes*. Kocher TD and Stepien CA (eds). Academic Press, San Diego, CA, pp. 113-128.

Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Research*, **27**, 1767-1780.

Boore JL, Brown WM (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Current Opinions in Genetics and Development*, **8**, 668-674.

Buckley TR, Cunningham CW (2002) The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. *Molecular Biology and Evolution*, **19**, 394-405.

Carr SM, Crutcher DC (1998) Population genetic structure in Atlantic cod (*Gadus morhua*) from the North Atlantic and Barents Sea: contrasting or concordant patterns in mtDNA sequence and microsatellite data? In: Hunt von Herbing I, Kornfield I, Tupper M, Wilson J (eds). Northeast Regional Agricultural Engineering Service, Ithaca, New York, pages 91-103.

Carr SM, Kivlichan DS, Pepin P, Crutcher DC (1999) Molecular systematics of gadid fishes: implications for the biogeographic origins of Pacific species. *Canadian Journal of Zoology*, **77**, 19-26.

Carr SM, Snellen AJ, Howse KA, Wroblewski JS (1995) Mitochondrial DNA sequence variation and genetic stock structure of Atlantic cod (*Gadus morhua*) from bay and offshore locations on the Newfoundland continental shelf. *Molecular Ecology*, **4**, 79-88.

Cavalli-Sforza LL, Edwards AWF (1964) Analysis of human evolution. Proceedings of the 11th International Congress in Genetics, Pergamon, New York.

Clayton DA (1982) Replication of animal mitochondrial DNA. *Cell*, **28**, 693-705.

Cohen DM (1984) Gadiformes: overview. In: *Ontogeny and Systematics of Fishes*. Moser HG, Richards WJ, Cohen DM, Fahay MP, Kendall Jr. AW, Richardson

SL (eds). American Society of Ichthyologists and Herpetologists,Allen Press, Lawrence, KS, pp. 259-265.

Cohen DM (1989) Papers on the Systematics of Gadiform Fishes. Scientific Publications Committee, Natural History Museum of Los Angeles County, Los Angeles, CA.

Cohen DM, Inada T, Iwamoto T, Scialabba N (1990) FAO Species Catalogue Vol.10 Gadiform Fishes of the World. Food and Agriculture Organization of the United Nations, Rome.

Corneli PS, Ward RH (2000) Mitochondrial genes and mammalian phylogenies: increasing the reliability of branch length estimation. *Molecular Biology and Evolution*, **17**, 224-234.

Creer S, Malhotra A, Thorpe RS (2003) Assessing the phylogenetic utility of four mitochondrial genes and a nuclear intron in the Asian pit viper genus, *Trimeresurus*: separate, simultaneous, and conditional data combination analyses. *Molecular Biology and Evolution*, **20**, 1240-1251.

Cross TF, Payne RH (1978) Geographic variation in Atlantic cod, *Gadus morhua*, off eastern North America: a biochemical systematics approach. *Journal of the Fisheries Research Board of Canada*, **35**, 117-123.

Dahle G (1991) Cod, *Gadus morhua* L., populations identified by mitochondrial DNA. *Journal of Fish Biology*, **38**, 295-303.

Department of Fisheries and Oceans C (2000) Updates on selected Scotian Shelf ground stocks in 2000. In: *Stock Status Report A*. Department of Fisheries and Oceans, Canada, pp. 3-35.

Dunn JR (1989) A Provisional Phylogeny of Gadid Fishes Based on Adult and Early Life-history Characters. In: *Papers on the Systematics of Gadiform Fishes*. Cohen DM (ed). Scientific Publications Committee, Natural History Museum of Los Angeles County, Los Angeles, pp. 209-235.

Dunn JR, Matarese AC (1984) Gadidae: Development and Relationships. In: *Ontogeny and systematics of fishes*. Moser HG, Richards WJ, Cohen DM, Fahay MP, Kendall Jr. AW, Richardson SL (eds). American Society of Ichthyologists and Herpetologists, and Allen Press, Lawrence, Kansas, pp. 283-299.

Faber JE, Stepien CA (1997) The Utility of Mitochondrial DNA Control Region Sequences for Analyzing Phylogenetic Relationships among Populations, Species, and Genera of the Percidae. In: *Molecular Systematics of Fishes*. Kocher TD, Stepien CA (eds). Academic Press, San Diego, CA, pp. 129-144.

Fahay MP, Markle DF (1984) Gadiformes: Development and relationships. In: *Ontogeny and systematics of fishes*. Moser HG, Richards WJ, Cohen DM, Fahay MP, Kendall Jr. AW, Richardson SL (eds). American Society of Ichthyologists and Herpetologists, and Allen Press, Lawrence, Kansas, pp. 265-283.

Ferguson A, Mason FM (1981) Allozyme evidence for reproductively isolated sympatric populations of brown trout *Salmo trutta* L. in Lough Melvin, Ireland. *Journal of Fish Biology*, **18**, 629-642.

Fevolden S-E, Martinez I, Christiansen JS (1999) RAPD and scnDNA analyses of polar cod, *Boreogadus saida* (Pisces, Gadidae), in the North Atlantic. *Sarsia*, **84**, 99-103.

Gerritsen HD, Armstrong MJ, Allen M, McCurdy WJ, Peel JAD (2003) Variability in maturity and growth in a heavily exploited stock: whiting (*Merlangius merlangus* L.) in the Irish Sea. *Journal of Sea Research*, **49**, 69-82.

Gladenkov AY, Oleinik AE, Marincovich Jr. L, Barinov KB (2002) A refined age for the earliest opening of Bering Strait. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **183**, 321-328.

Grant S, Stahl G (1988) Evolution of Atlantic and Pacific Cod: Loss of Genetic Variation and Gene Expression in Pacific Cod. *Evolution*, **42**, 138-146.

Grant WS, Utter FM (1980) Biochemical genetic variation in walleye pollock, *Theragra chalcogramma*: population structure in the southeastern Bering Sea and the Gulf of Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, **37**, 1093-1100.

Grant WS, Zhang CI, Kobayashi T, Stahl G (1987) Lack of genetic stock discretion in Pacific cod (*Gadus macrocephalus*). *Canadian Journal of Fisheries and Aquatic Sciences*, **44**, 490-498.

Gray MW, Burger G, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Lang BF (1999a) A genomics approach to mitochondrial evolution. *Biological Bulletin*, **196**, 400-403.

Gray MW, Burger G, Lang BF (1999b) Mitochondrial Evolution. *Science*, **283**, 1476-1481.

Hart JL (1973) Pacific Fishes of Canada. Bulletins of the Fisheries Research Board of Canada, Ottawa, Canada.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London*, **270**, 313-321.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the. Royal Society of London* (In Press)

Hislop JRG, Mackenzie K (1976) Population studies of the whiting *Merlangius merlangus* (L.) of the northern North Sea. *Journal du Conseil International pour l'Exploration de la Mer*, **26**, 68-72.

Hoelzel AR, Lopez JV, Dover GA, O'Brien SJ (1994) Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. *Journal of Molecular Evolution*, **39**, 191-199.

Howes GJ (1989) Phylogenetic Relationships of Macrouroid and Gadoid Fishes Based on Cranial Myology and Arthrology. In: *Papers on the Systematics of Gadiform Fishes*. Cohen DM (ed). Scientific Publications Committee, Natural History Museum of Los Angeles County, Los Angeles, CA, pp. 113-128.

Howes GJ (1991) Biogeography of gadoid fishes. *Journal of Biogeography*, **18**, 595-622.

Huelsenbeck JP (1991) Tree-length distribution skewness: an indicator of phylogenetic information. *Systematic Zoology*, **40**, 257-270.

Hutchings JA, Myers RA, Lilly GR (1993) Geographic variation in the spawning of Atlantic cod, *Gadus morhua*, in the northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, **50**, 2457-2467.

Inoue JG, Miya M, Tsukamoto K, Nishida M (2001a) Complete mitochondrial DNA sequence of *Conger myriaster* (Teleostei: Anguilliformes): Novel gene order for vertebrate mitochondrial genomes and the phylogenetic implications for Anguilliform families. *Journal of Molecular Evolution*, **52**, 311-320.

Inoue JG, Miya M, Tsukamoto K, Nishida M (2001b) A mitogenomic perspective on the basal teleostean phylogeny: resolving higher-level relationships with longer DNA sequences. *Molecular Phylogenetics and Evolution*, **20**, 275-285.

Inoue JG, Miya M, Tsukamoto K, Nishida M (2003) Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the "ancient fish". *Molecular Phylogenetics and Evolution*, **26**, 110-120.

Jamieson A (1975) Enzyme type of Atlantic cod stocks on the North American banks. In: *Genetics and evolution. Isozymes. IV.* Markert CL (ed). Academic Press, New York, pp. 491-515.

Johansen S, Bakke I (1996) The complete mitochondrial DNA sequence of Atlantic cod (*Gadus morhua*): relevance to taxonomic studies among codfishes. *Molecular Marine Biology and Biotechnology*, **5**, 203-214.

Johansen S, Guddal PH, Johansen T (1990) Organization of the mitochondrial genome of Atlantic cod, *Gadus morhua*. *Nucleic Acids Research*, **18**, 411-419.

Jordan AD (2002) *Arctogadus*: A Single Species? Studies of Taxonomy, Phylogeny and Physiology. MSc Thesis, University of Copenhagen, Copenhagen, The Netherlands.

Kabata Z (1967) Whiting stocks and their gall-bladder parasites in British waters. *Marine Research*, **2**, 1-11.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Science USA*, **86**, 6196-6200.

Kraus F, Miyamoto MM (1991) Rapid cladogenesis among the pecoran ruminants: evidence from mitochondrial DNA sequences. *Systematic Zoology*, **40**, 117-130.

Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *Journal of Molecular Evolution*, **37**, 380-398.

Lear WH (1984) Discrimination of the stock complex of Atlantic cod (*Gadus morhua*) off Southern Labrador and Eastern Newfoundland, as inferred from tagging studies. *Journal of the Northwest Atlantic Fisheries Sciences*, **5**, 143-159.

Lear WH, Wells R (1984) Vertebral averages of juvenile cod, *Gadus morhua*, from coastal waters of eastern Newfoundland and Labrador as indicators of stock origin. *Journal of the Northwest Atlantic Fisheries Sciences*, **5**, 23-31.

Lee W, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**, 54-66.

Lydeard C, Roe KJ (1997) The Phylogenetic Utility of the Mitochondrial Cytochrome *b* Gene for Inferring Relationships among Actinopterygian Fishes. In: *Molecular Systematics of Fishes*. Kocher TD, Stepien CA (eds). Academic Press, San Diego, CA, pp. 285-304.

Macey JR, Larson A, Ananjeva NB, Fang Z, Papenfuss TJ (1997) Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Molecular Biology and Evolution*, **14**, 91-104.

Marincovich Jr. L, Gladenkov AY (1999) Evidence for an early opening of the Bering Strait. *Nature*, **397**, 149-151.

Marincovich Jr. L, Gladenkov AY (2001) New evidence for the age of Bering Strait. *Quaternary Science Reviews*, **20**, 329-335.

Markle DF (1982) Identification of larval and juvenile Canadian Atlantic gadoids with comments on the systematics of gadid subfamilies. *Canadian Journal of Zoology*, **60**, 3420-3438.

Markle DF (1989) Aspects of Character Homology and Phylogeny of the Gadiformes. In: *Papers on the Sytematics of Gadiform Fishes*. Cohen DM (ed). Scientific Publications Committee, Natural History Museum of Los Angeles County, Los Angeles, pp. 59-88.

Marshall NB, Cohen DM (1973) Order Anacanthini (Gadiformes): Characters and synopsis of families. *Memoirs of the Sears Foundation for Marine Research*, **1**, 479-495.

Mattiangeli V, Bourke EA, Ryan AW, Mork J, Cross TF (2000) Allozyme analyses of the genus *Trisopterus*: taxonomic status and population structure of the poor cod. *Journal of Fish Biology*, **56**, 474-494.

McPhail JD (1992) Ecology and evolution of sympatric sticklebacks (*Gasterosteus*): morphological and genetic evidence for a species pair in Paxton Lake, British Columbia. *Canadian Journal of Zoology*, **70**, 361-369.

Messier W, Stewart CB (1997) Episodic adaptive evolution of primate lysozymes. *Nature*, **385**, 151-154.

Methven DA, McGowan C (1998) Distinguishing small juvenile Atlantic cod (*Gadus morhua*) from Greenland cod (*Gadus ogac*) by comparing meristic characters and discriminant function analyses of morphometric data. *Canadian Journal of Zoology*, **76**, 1054-1062.

Meyer A (1994) Shortcomings of the cytochrome b gene as a molecular marker. *Trends in Ecology and Evolution*, **9**, 278-280.

Mikhail MY, Welch HE (1989) Biology of Greenland cod, *Gadus ogac*, at Saqvaqjuac, northwest coast of Hudson Bay. *Environmental Biology of Fishes*, **26**, 49-62.

Mindell DP, Sorenson MD, Dimcheff DE (1998) Multiple independent origins of mitochondrial gene order in birds. *Proceeding of the National Academy of Sciences of USA*, **95**, 10693-10697.

Miya M, Nishida M (2000) Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. *Molecular Phylogenetics and Evolution*, **17**, 437-455.

Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M (2003) Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete

mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **26**, 121-138.

Moller P, Jordan AD, Gravlund P, Steffensen JF (2002) Phylogenetic position of the cryopelagic codfish genus *Arctogadus* Drjagin, 1932 based on partial mitochondrial cytochrome b sequences. *Polar Biology*, **25**, 342-349.

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

Morin B, Hudon C, Whoriskey F (1991) Seasonal distribution, abundance, and life-history traits of Greenland cod, *Gadus ogac*, at Wemindji, eastern James Bay. *Canadian Journal of Zoology*, **69**, 3061-3070.

Morita T (1999) Molecular phylogenetic relationships of the deep-sea fish genus *Coryphaenoides* (Gadiformes: Macrouridae) based on mitochondrial DNA. *Molecular Phylogenetics and Evolution*, **13**, 447-454.

Mork J, Ryman N, Stahl G, Utter F, Sundnes G (1985) Genetic variation in Atlantic cod (*Gadus morhua*) throughout its range. *Canadian Journal of Fisheries and Aquatic Sciences*, **42**, 1580-1587.

Murata Y, Nikaido M, Sasaki T, Cao Y, Fukumoto Y, Hasegawa M, Okada N (2003) Afrotherian phylogeny as inferred from complete mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **28**, 253-260.

Nei M (1991) Relative Efficiencies of Different Tree-Making Methods for Molecular Data. In: *Phylogenetic Analysis of DNA Sequences*. Miyamoto MM, Cracraft J (eds). Oxford University Press, New York, pp. 90-128.

Nei M (1996) Phylogenetic analysis in molecular evolutionary genetics. *Annual Reviews in Genetics*, **30**, 371-403.

Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Nelson JS (1984) Fishes of the World. John Wiley and Sons, New York.

Nielsen JG, Jensen JM (1967) Revision of the Arctic cod genus, *Arctogadus* (Pisces, Gadidae). *Medd Gron*, **184**, 1-26.

Nielsen JR, Andersen M (2001) Feeding habits and density patterns of Greenland cod, *Gadus ogac* (Richardson 1836), at west Greenland compared to those of the coexisting Atlantic cod, *Gadus morhua* L. *Journal of the Northwest Atlantic Fisheries Sciences*, **29**, 1-22.

Nolf D, Steurbaut E (1989) Evidence from otoliths for establishing relationships between gadiforms and other groups. In: *Papers on the Systematics of Gadiform Fishes*. Cohen DM (ed). Scientific Publications Committee, Natural History Museum of Los Angeles County, Los Angeles, pp. 37-45.

Olsen JB, Merkouris SE, Seeb JE (2002) An examination of spatial and temporal genetic variation in walleye pollock (*Theragra chalcogramma*) using allozyme, mitochondrial DNA, and microsatellite data. *Fishery Bulletin*, **100**, 752-764.

O'Reilly PT, McPherson AA, Kenchington E, Taggart C, Jones MW, Bentzen P (2002) Isolation and characterization of tetranucleotide microsatellites from Atlantic haddock (*Melanogrammus aeglefinus*). *Marine Biotechnology*, **4**, 418-422.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution*, **11**, 426-435.

Parenti LR (1981) A phylogenetic and biogeographic analysis of cyprinodontiform fishes. *Bulletin of the American Museum of Natural History*, **1658**, 341-557.

Penny D, Hendy MD (1985) The use of tree comparison metrics. *Systematic Zoology*, **34**, 75-82.

Pesole G, Gissi C, De Chirico A, Saccone C (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. *Journal of Molecular Evolution*, **48**, 427-434.

Phillips MJ, Penny D (2003) The root of the mammalian tree inferred from whole mitochondrial genomes. *Molecular Phylogenetics and Evolution*, **28**, 171-185.

Pigeon D, Chouinard A, Bernatchez L (1997) Multiple modes of speciation involved in parallel evolution of sympatric morphotypes of lake whitefish (*Coregonus clupeaformis*, Salmonidae). *Evolution*, **51**, 196-205.

Pogson GH, Mesa KA, Boutilier RG (1995) Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics*, **139**, 375-385.

Race HL, Herrmann RG, Martin W (1999) Why have organelles retained genomes? *Trends in Genetics*, **15**, 364-370.

Rand DM (1994) Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends in Ecology and Evolution*, **9**, 125-131.

Rasmussen A-S, Arnason U (1999) Phylogenetic studies of complete mitochondrial DNA molecules place cartilaginous fishes within the tree of bony fishes. *Journal of Molecular Evolution*, **48**, 118-123.

Renaud CB, Speers LI, Qadri SU (1986) Biochemical evidence of speciation in the cod genus *Gadus*. *Canadian Journal of Zoology*, **64**, 1563-1566.

Rico C, Ibrahim KM, Rico I, Hewitt GM (1997) Stock composition in North Atlantic populations of whiting using microsatellite markers. *Journal of Fish Biology*, **51**, 462-475.

Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Mathematical Bioscience*, **53**, 131-147.

Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, **425**, 798-804.

Rosen DE (1973) Interrelationships of higher euteleostean fishes. In: *Interrelationships of fishes*. Greenwood PH, Miles RS, Patterson C (eds). Academic Press, London, pp. 397-513.

Rosen DE, Patterson C (1969) The structure and relationships of the paracanthopterygian fishes. *Bulletin of the American Museum of Natural History*, **141**, 357-474.

Russo CAM, Takezaki N, Nei M (1996) Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. *Molecular Biology and Evolution*, **13**, 525-536.

Ruzzante DE, Taggart CT, Cook D, Goddard SV (1996) Genetic differentiation between inshore and offshore Atlantic cod (*Gadus morhua*) off Newfoundland:

microsatellite DNA variation and antifreeze level. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 634-645.

Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene*, **238**, 195-209.

Saccone C, Gissi C, Lanave C, Larizza A, Pesole G, Reyes A (2000) Evolution of the mitochondrial genetic system: an overview. *Gene*, **261**, 153-159.

Saccone C, Gissi C, Reyes A, Larizza A, Sbisa E, Pesole G (2002) Mitochondrial DNA in metazoa: degree of freedom in a frozen event. *Gene*, **286**, 3-12.

Saitoh K, Hayashizaki K, Yokoyama Y, Asahida T, Toyohara H, Yamashita Y (2000) Complete nucleotide sequence of Japanese flounder (*Paralichthys olivaceus*) mitochondrial genome: structural properties and cue for resolving teleostean relationships. *The American Genetic Association*, **91**, 271-278.

Sbisa E, Tanzariello F, Reyes A, Pesole G, Saccone C (1997) Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene*, **205**, 125-140.

Scheffler I (1999) Mitochondria. Wiley-Liss, New York.

Schluter D (1996a) Ecological causes of adaptive radiation. *The American Naturalist*, **148** (Suppl. 1), S40-S64.

Schluter D (1996b) Ecological speciation in postglacial fishes. *Proceedings of the Royal Society of London Series B*, **351**, 807-814.

Schultz LP, Welander AD (1935) A review of the cods of the northeastern Pacific with comparative notes on related species. *Copeia*, **1935**, 127-139.

Scott WB, Scott MG (1988) Atlantic Fishes of Canada. University of Toronto Press, Toronto.

Sick K (1965) Haemoglobin polymorphism of cod in the North Sea and the North Atlantic Ocean. *Hereditas*, **54**, 49-69.

Smith PJ, Birley AJ, Jamieson A, Bishop CA (1989) Mitochondrial DNA in the Atlantic cod, *Gadus morhua*: lack of genetic divergence between eastern and western populations. *Journal of Fish Biology*, **34**, 369-373.

Sokal RR, Sneath PHA (1963) Principles of Numerical Taxonomy. Freeman, San Francisco, CA.

Steel M, Penny D (2000) Parsimony, Likelihood, and the role of models in molecular phylogenetics. *Molecular Biology and Evolution*, **17**, 839-850.

Stepien CA, Kocher TD (1997) Molecules and Morphology in Studies of Fish Evolution. In: *Molecular Systematics of Fishes*. Kocher TD, Stepien CA (eds). Academic Press, San Diego, CA, pp. 1-11.

Sufke L, Piepenburg D, von Dorrien CF (1998) Body size, sex ratio and diet composition of *Arctogadus glacialis* (Peters, 1874) (Pisces: Gadidae) in the Northeast Water Polynya (Greenland). *Polar Biology*, **20**, 357-363.

Svetovidov AN (1948) Gadiformes. In: Fauna of the USSR. Vol. IX. No. 4. Fishes. Zoological Institute of the Academy of Sciences of the USSR. [Translated by the Israel Program for Scientific Translations for the National Science Foundationand Smithsonian Institution as OTS 63-11071.]

Svitoch AA, Taldenkova EE (1994) Recent history of the Bering Strait. *Oceanology*, **34**, 400-404.

Swofford L,(1998). PAUP: phylogenetic analysis using parsimony, version 4.0d63. Smithsonian Institution, Washington, DC.

Takezaki N, Gojobori T (1999) Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences. *Molecular Biology and Evolution*, **16**, 590-601.

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

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

Templeman W (1974) Migrations and intermingling of Atlantic cod (*Gadus morhua*) stocks of the Newfoundland area. *Journal of Fisheries Research Board of Canada*, **31**, 1073-1092.

Templeman W (1979) Migrations and intermingling of stocks of Atlantic cod, *Gadus morhua*, of the Newfoundland and adjacent areas from tagging 1962-1966. *International Commission of Northwest Atlantic Fisheries Research Bulletin*, **14**, 5-50.

Templeman W (1981) Vertebral numbers in Atlantic cod, *Gadus morhua*, of the Newfoundland and adjacent areas, 1947-1971, and their use for delineating cod stocks. *Journal of Northwest Atlantic Fisheries Science*, **2**, 21-45.

Vane-Wright RI, Humphries CJ, Williams PH (1991) What to protect? Systematics and the agony of choice. *Biology Conservation*, **55**, 235-254.

Walters V (1961) Winter abundance of *Arctogadus glacialis* in the polar basin. *Copeia*, **1961**, 236-237.

Warren HM,(1999). Molecular Systematics of Gadiform Fishes: Implications for the Evolutionary Origins of Gadid Subfamilies. BSc Hon. Thesis, Department of Biology, Memorial University of Newfoundland, St. John's, Newfoundland.

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helm-Bychowski KM, Higuchi RG, Palumbi SR, Prager EM, Sage RD, Stoneking M (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, **26**, 375-400.