A MITOGENOMIC STUDY OF FOUR AT-RISK MARINE FISH SPECIES: ATLANTIC WOLFFISH, SPOTTED WOLFFISH, NORTHERN WOLFFISH, AND ATLANTIC COD, WITH SPECIAL EMPHASIS ON THE WATERS OFF NEWFOUNDLAND AND LABRADOR

by

© Linda Amy Lait

A Thesis submitted to the

School of Graduate Studies

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

Memorial University of Newfoundland

October 2016

St. John's, Newfoundland and Labrador

ABSTRACT

High-resolution mitogenomics can answer questions as to how species survived the last glacial maximum, and can also address contemporary factors such as physical barriers, isolation, and gene flow. This study examines the population genomic structures of two genera of At-Risk marine fish species found across the North Atlantic Ocean, Atlantic cod (*Gadus morhua*) and wolffish (*Anarhichas* spp.). Despite their sympatric distribution, the two taxa exhibit very different life history characteristics, and have very different patterns of genetic variation and structure. Populations of Atlantic cod show high levels of genomic variation, with eight major clades found across most populations and trans-Atlantic differences. The Arctic lake population (Lake Qasigialiminiq) was significantly different from all other populations (Φ_{ST} = 0.15 to 0.34), and comprised two distinct and essentially monomorphic clades. The Baltic and Barents Sea populations showed high levels of diversity, extensive variation among samples, and significant pairwise differences with many Northwest Atlantic populations (Φ_{ST} = 0.03 to 0.18). Within Atlantic Canada there was no evidence of differentiation in Newfoundland waters or between trans-Laurentian populations (Φ_{ST} = 0.0 to 0.06). High diversity levels, the absence of European fish in the most recent clades, and the presence of basal lineages support a European origin to the postglacial expansion, with a second smaller refugium likely in North America. Atlantic wolffish (*Anarhichas lupus*) and the congeneric spotted and northern wolffish (*A. minor* and *A. denticulatus*) each comprised two or three haplogroups dating back to the Pleistocene glaciations (63 - 220 kya). The haplogroups were not structured geographically: in Atlantic wolffish the overall Φ_{ST} was 0.05 and

pairwise values ranged from 0.0 to 0.24. A similar pattern of distinct but shallow groups was seen in spotted wolffish, while northern wolffish exhibited two deeper lineages. This suggests isolation in multiple glacial refugia – likely three nearby regions in European waters (Atlantic and spotted) or two more distant refugia (northern) – followed by secondary admixture during recolonisation of the Northwest Atlantic. The two taxa show very different patterns of variation and structure, with greater variation in Atlantic cod and greater structure in the sedentary wolffish species.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my family and friends for all of their support over the past four years, especially Ryan for distracting me when I needed it and helping me keep on track when I wavered, Ainsley and Heather for providing a sounding board for ideas and frustrations, and my mum for reading this thesis cover to cover.

I would like to thank my supervisor, Dr. Steve Carr, for all of his help over the past few years. He has helped nurture my interest in teaching, guided me through a maze of data, and pointed me in the right direction for upcoming opportunities. My committee members, Drs. Dawn Marshall and Matt Rise, have provided me with invaluable ideas and encouragement. I would also like to thank the examiners for their insightful comments that contributed to the final product and made me think outside the box.

Many thanks to Fisheries and Oceans Canada for collecting cod and wolffish samples over the past 20 years, and to Drs. Paul Bentzen and Megan McCusker for providing wolffish samples from their study. A special thank you to Ian Winters, Errol Andersen, and George Gear of the Nunatsiavut Government, and Noah and Jackie of the Pangnirtung Hunters and Trappers Organisation, for all of their help with planning and carrying out field work in Labrador and Nunavut; to William Hunter, Errol Andersen, Leon Jacque, and Norman Mike for braving the weather and catching the cod; and to William Hunter for teaching me how to fish for cod and helping me to catch my first fish.

I would like to acknowledge the work of Drs. Carr and Marshall, upon which the cod chapter is built, and their recent paper on the “codmother”. Their preliminary study

was the foundation for the cod mitogenomics chapter, and their data is included alongside new data here.

This research was made possible by grants from Natural Sciences and Engineering Research Council (NSERC) to SMC and LAL, Alberta Innovations Sir James Lougheed Award of Distinction to LAL, and support from Memorial University of Newfoundland and Labrador. The field component was made possible by the generous funding from the Association of Canadian Universities for Northern Studies (ACUNS) W. Garfield Weston Award for Northern Research and the Canadian Northern Studies Trust (CNST) Northern Scientific Training Program Award to LAL.

TABLE OF CONTENTS

CHAPTER FOUR: Mitogenomics of Atlantic cod (*Gadus morhua*): variation in trans-Atlantic, Baffin Island, and Baltic Sea populations

CHAPTER FIVE: Summary

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

°C	degrees Celsius
°E	degrees east of the meridian
°N	degrees north of the equator
°W	degrees west of the meridian
%v	percent variation
Γ	gamma-distributed rate
Φ_{ST}	measure of genetic differentiation
ΔS/N	signal-to-noise ratio
µg	microgram
µL	microlitre
µM	micromolar
π	nucleotide diversity
X^2	chi square
12S	12S rRNA
16S	16S rRNA
3'	3 prime end (of DNA)
5'	5 prime end (of DNA)
A	adenine
ACUNS	Association of Canadian Universities for Northern Studies
AFLP	amplified fragment length polymorphism
Ala	alanine
AMOVA	analysis of molecular variance
Arg	arginine
Asn	asparagine
Asp	aspartic acid
ATP	adenosine triphosphate
ATP6	ATPase 6
ATP8	ATPase 8
BAPS	Bayesian analysis of population structure

BEAST	Bayesian evolutionary analysis sampling trees
bp	base pair
C	cytosine
C&M	Carr & Marshall 2008 paper
CHIP	'Arkchip' microarray sequencing
CITES	Convention on international trade in endangered species
cm	centimetre
CNST	Canadian Northern Studies Trust
COSEWIC	Committee on the Status of Endangered Wildlife in Canada
COX1	cytochrome oxidase subunit 1
COX2	cytochrome oxidase subunit 2
COX3	cytochrome oxidase subunit 3
CR	control region
Cys	cysteine
CYTB	cytochrome *b*
DD	dideoxy
df	degrees of freedom
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DnaSP	DNA Sequence Polymorphism
dNTP	deoxyribonucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
Eur	Europe
ESS	effective sample size
exo	exonuclease I
F	forward primer
FDR	false discovery rate
g	G force
G	guanine
GenAlEx	Genetic Analysis in Excel

Gln	glutamine
Glu	glutamic acid
Gly	glycine
Gp1	BAPS Group 1 (cod)
Gp2	BAPS Group 2 (cod)
GPS	global positioning system
GTR	generalised time reversible model
h	number of haplotypes
H_d	haplotype diversity
HKY	Hasegawa, Kishino, and Yano model
His	histidine
HPD	highest probability density interval
I	invariable site
ID	insertion-deletion event
Ile	isoleucine
indel	insertion-deletion event
IUCN	International union for the conservation of nature
IUPAC	International Union of Pure and Applied Chemistry
K	number of groups
kbp	kilo base pair
kg	kilogram
km	kilometre
km^2	square kilometre
kya	thousand years ago
Leu	leucine
LGM	last glacial maximum
loc	location (of 5' end)
LR	long-range (PCR)
Lys	lysine
m	metre

M	molar (moles/litre)
MCMC	Markov chain Monte Carlo
MEGA	Molecular Evolutionary Genetics Analysis
Met	methionine
$MgCl_2$	magnesium chloride
mg	milligram
mid	mid-Atlantic
min	minute
mL	millilitre
mm	millimetre
mM	millimolar
mRNA	messenger RNA
mtDNA	mitochondrial DNA
MY	million years
mya	million years ago
n	sample size
NA	North America
NaCl	sodium chloride
NADH	nicotinamide adenine dinucleotide (reduced)
NAFO	North Atlantic Fisheries Organization
ND1	NADH dehydrogenase subunit 1
ND2	NADH dehydrogenase subunit 2
ND3	NADH dehydrogenase subunit 3
ND4	NADH dehydrogenase subunit 4
ND4L	NADH dehydrogenase subunit 4L
ND5	NADH dehydrogenase subunit 5
ND6	NADH dehydrogenase subunit 6
N_E	effective population size
NJ	neighbour-joining
ng	nanogram

nm	nanometre
ns	not significant
NS	non-synonymous substitution
NSERC	Natural Sciences and Engineering Research Council
O_L	origin of light strand replication
p	probability value
PanI	Pantophysin
PAUP*	Phylogenetic Analysis Using Parsimony*
PCoA	principal coordinates analysis
PCR	polymerase chain reaction
P_{crit}	critical p value
Phe	phenylalanine
PI	number of parsimony informative sites
pop	population
post	posterior probability
Pro	proline
PSRF	potential scale reduction factor
R	reverse primer
r^2	correlation coefficient (linear regression analysis)
ref	primer source
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal RNA
s	second
S	synonymous substitution
SAMOVA	spatial analysis of molecular variance
SAP	shrimp alkaline phosphatase
SARA	Species at Risk Act
SDS	sodium dodecyl sulfate

Ser	serine
SNP	single nucleotide polymorphism
SP	significant pairwise comparisons
Sub	substitution
t	divergence time
T	thymine
T_A	annealing temperature
Thr	threonine
tRNA	transfer RNA
Trp	tryptophan
Ts	transition
Tv	transversion
Tyr	tyrosine
U	unit
UV	ultra violet
v	version
Val	valine
vs.	versus
VS	number of variable sites
X	times

Populations:

BAL	Baltic Sea, Sopot, Poland
BS	Barents Sea
EG	East Greenland
ESS	Eastern Scotian Shelf
FC	Flemish Cap
FSS	Fundy, Scotian Shelf
GEO	Georges Bank, Gulf of Maine
GIL	Gilbert Bay, Labrador

HC	Hawkes Channel, Labrador
IC	Iceland
LAB	Labrador
LABC	central Labrador
LABS	southern Labrador
LBH	Labrador – Hopedale
LBM	Labrador – Makkovik
LBP	Labrador – Postville
NC	Norwegian Coast, Tromsö
NENL	northeast Newfoundland
NGB	North Cape, Grand Banks
NS	North Sea
QAS	Lake Qasigialiminiq, Baffin Island, Nunavut
RAN	Random Island, Newfoundland
RK	Rockall Bank
SEGB	southeast Grand Banks
SNL	southern Newfoundland
SS	Scotian Shelf, Nova Scotia
SWGB	southwest Grand Banks
WG	West Greenland

LIST OF APPENDICES

CHAPTER ONE - GENERAL INTRODUCTION

Population genetics is the study of genetic variation within a species, how this variation is distributed across space, and how it changes over time (Avise *et al.* 1987; Avise 2004). Genetic variation is influenced by four evolutionary phenomena: mutation, natural selection, drift, and migration. Phylogeography, a branch of population genetics, studies the distribution of genetic variation on a geographic scale. It looks at whether there is any pattern to the variation, how this corresponds to the physical landscape, and what underlying processes may be responsible (Avise *et al.* 1987; Avise 2000). For example, population genetics can identify cryptic species – organisms previously classified as a single species based on morphology or life history can be separated through identification of intraspecific diversity (Montresor *et al.* 2003; Hebert *et al.* 2004). Alternatively, organisms previously classified as separate species based on differences in morphology can be identified as different ecotypes of the same species (Allendorf & Leary 1988; Johannesson *et al.* 1993). Patterns of dispersal and the identity of source populations can also be elucidated – older populations generally contain higher levels of diversity and a greater number of unique genotypes, with decreasing diversity and loss of alleles as the populations disperse due to founder effects and bottlenecks (Figure 1.1; Burg *et al.* 2003; Avise 2004).

1.1 Population Genetics

There are a number of processes, both historical and contemporary, that influence how variation is distributed among populations: major geological events (e.g., the rising of the Isthmus of Panama), major climatic events (e.g., the glaciations of the Pleistocene Epoch), and contemporary physical barriers (e.g., mountain ranges, deserts, and large bodies of water; Avise 2004). For example, the rising of the Isthmus of Panama both connected and separated populations: it created a migration path for terrestrial species between North and South America, and it formed a barrier to gene flow in marine species (Lessios 2008; Bacon *et al.* 2015). The Pleistocene glaciations played a major role in the survival and resulting genetic structure of the majority of contemporary species: much of the Northern Hemisphere was periodically covered by large ice sheets, temperatures were generally a great deal colder than today, and many areas were uninhabitable (Flint 1947; Nilsson 1983; Pielou 1991; Hewitt 2000).

Population genetics is also an important consideration in conservation biology. A number of factors can contribute to the decline, and resulting genetic structure, of a species or population: habitat loss, introduction of non-native species, overharvesting, and environmental change (Wilson 1988; Frankham 1995). Many species of plants and animals have been locally extirpated or have gone extinct as a result of habitat degradation and destruction (Pimm & Raven 2000). The introduction of invasive species, both accidental (e.g., zebra mussels *Dreissena polymorpha*) and deliberate (e.g., purple loosestrife *Lythrum salicaria*), has contributed to the loss of dozens of native species (Mooney & Cleland 2001). Overexploitation has heavily impacted

species as far removed as elephants in Africa (*Loxodonta* spp.), the great auk *Pinguinus impennis* in North America, and the northern elephant seal *Mirounga angustirostris* in the Pacific Ocean. Environmental pressures, such as pollution or climate change, often act in concert with other factors to push an already threatened species to extinction. The species that do manage to survive often suffer population bottlenecks and loss of genetic diversity (Frankham 1995). Severe population bottlenecks may result in extreme reduction of genetic variability within species (e.g., northern elephant seal, Hoelzel 1999; Hawaiian monk seal *Monachus schauinslandi*, Schultz *et al.* 2009). This loss of genetic diversity may result in reduced fitness, reduced fecundity, or the inability to respond to disease (Frankham 1995).

1.2 The Marine Environment

Much of what we know about population genetics has come from terrestrial and freshwater species. This is because both environments contain clear boundaries, connections, and delineations, and support populations that are easily accessed and characterised. To a large extent reproduction occurs in a confined area, and where it doesn't (e.g., fungal spores or flower pollen) it is generally limited in the distance gametes can travel by the heterogeneity of the landscape. The extent of population genetic structure in the marine environment, however, remains much more uncertain.

Marine species, particularly animal species, often have large distributions, extensive dispersal capabilities, and high levels of gene flow (Palumbi 1992; Bradbury *et al.* 2008). Marine invertebrates can undergo dramatic changes from pelagic larvae to

sedentary or benthic adult forms, while fish can experience extreme transformations from larvae to juvenile to adulthood (Thorson 1950; Pechenik *et al.* 1998; Hamilton *et al.* 2008). Species with pelagic or planktonic larvae often show high dispersal rates and as a result little or no population genetic structure (Palumbi 1992; Ward *et al.* 1994). Conversely, more sedentary species, those with strong natal philopatry, or those found in regions with barriers to larval dispersal, can exhibit much higher levels of population genetic structure (Saunders *et al.* 1986; Bowen & Karl 2007; Lee *et al.* 2007; Bradbury *et al.* 2008).

While there are generally fewer physical barriers in the marine environment compared to on land, a number of factors have been shown to influence population connectivity and gene flow. Differences are seen between the Atlantic and Pacific Ocean basins separated by large land masses (Yurick & Gaskin 1987; Chow *et al.* 2000; Bowen & Karl 2007), ocean currents and upwellings have been shown to prevent larval dispersal (Barber *et al.* 2002; Henriques *et al.* 2014), and long distances between suitable habitat has been shown to prevent movement in a number of marine organisms (Lavery *et al.* 1996; King *et al.* 2001).

1.2.1 The North Atlantic Ocean

The North Atlantic Ocean originated ca. 180 - 190 million years ago (mya) when the large supercontinent Pangaea began to break into smaller land masses (Berggren & Hollister 1974). This initial opening was followed by deep-water circulation and the establishment of the North Atlantic Ocean as a pelagic ecosystem ca. 50 - 60 mya. The lower species richness and minimal endemism in the North

Atlantic Ocean supports a colonisation from the Pacific Ocean (Briggs 1970). While continuing continental shifts have changed the conditions and constituents of the ocean, two recent events played a noteworthy role in the oceanic conditions and species make-up: the rise of the Isthmus of Panama, which essentially severed movement between the Atlantic and Pacific Oceans and created the Gulf Stream, and the Pleistocene glaciations, which generated the Labrador Current and shaped the biogeographic history of many marine species (Berggren & Hollister 1974).

More recently, the North Atlantic has undergone more than 400 years of intense commercial fishing, first by small-boat European fisheries and currently by high-tech fishing fleets from all over the world. The advent of more efficient fishing technologies, in combination with a lack of effective management, has led to catastrophic population declines in the majority of marine species (Rose 2007). Changes in sea surface temperatures may have exacerbated the problem (Rose 2007; Hutchinson 2008). Some of the more notable species declines occurred in Atlantic cod *Gadus morhua*, a large predatory fish upon which many fisheries were built, and capelin *Mallotus villosus*, a small pelagic fish that is the main prey of many other Atlantic species (e.g., whales, birds, and other fish). Non-targeted species were also affected as a result of habitat destruction and by-catch.

The North Atlantic ecosystem, particularly the Northwest Atlantic and the waters off Newfoundland and Labrador (Canada), has a number of distinctive characteristics which make it an ideal system for studying population genetics. The location of the Grand Banks and the Flemish Cap, at the confluence of the warm Gulf Stream and the ice-cold waters of the Labrador Current, has created an immense

upwelling of nutrients from deeper waters. This enrichment, coupled with the fact that the shallow banks receive sunlight across the ocean floor, has led to a wealth of phytoplankton, zooplankton, and the marine creatures which they sustain (Rose 2007 and references therein). A number of other areas in the North Atlantic have also benefitted from the upwelling and productivity related to this region: the coast of Labrador, the Gulf of St. Lawrence, and the Bay of Fundy are all rich in marine life (Rose 2007). Similarly, high levels of phytoplankton and species richness are seen where different aquatic fronts meet, whether rivers, ocean currents, or open oceans (Angel 1993; Hao *et al.* 2010; Taylor & Ferrari 2011).

1.3 The Pleistocene Epoch

The Pleistocene Epoch, the first epoch in the (current) Quaternary Period, consisted of a series of glacial and interglacial stages beginning ca. 2.58 mya (Gibbard *et al.* 2010). The most recent glacial period, the Nearctic Wisconsinan or Palaearctic Würm glaciation (ca. 110 - 12 thousand years ago (kya)), affected much of the Northern Hemisphere (Pielou 1991; Barendregt & Irving 1998). While many of the more extreme cases were seen on land, the glaciations also had a profound effect on marine environments. Sea level drops of 120 m or more exposed many continental shelves and slopes including the Baltic Sea, North Sea, and Grand Banks, uncovering 20% more land than today. Ice sheets extended over most of the emerged shelves (as far south as 40°N in North America and 52°N in Europe), ocean currents were

disrupted or ceased altogether, and sea surface temperatures dropped to levels preventing winter ice-pack from melting and reaching average temperatures of 3°C colder than winter sea temperatures today (Flint 1947; Pielou 1991; Hewitt 2004; Shaw 2006). At its greatest extent, the last glacial maximum (LGM) that occurred between 21 and 18 kya, more than two-thirds of North America was covered with ice (Barendregt & Irving 1998; Dyke *et al.* 2002; Clark *et al.* 2009).

1.3.1 Glacial refugia

In order to survive the Pleistocene glaciations, both marine and terrestrial species had to retreat to ice-free regions known as glacial refugia. These were typically found at lower latitudes, such as southern North America (Pielou 1991), the Balkan, Iberian, and Apennine Peninsulas of Europe, and the Caucasus Mountains of Eurasia (Hewitt 2004; Provan & Bennett 2008). However, more recent evidence also supports the occurrence of northern periglacial refugia, such as the Carpathian Mountains in Europe and the Rocky Mountains, between the Laurentide and Cordilleran ice sheets, in North America (Stewart & Lister 2001; Provan & Bennett 2008).

Marine refugia are less well-defined, although phylogeographic evidence suggests seven putative refugia in Europe: three regions south of the ice sheets and four periglacial locations (Maggs *et al.* 2008). To the south, neither the Azores off the coast of northwest Africa, nor the Atlantic coast along the Iberian Peninsula, were heavily impacted by the last glaciation, and, although the Mediterranean Sea underwent significant sea level drops, it likely acted as a marine glacial refugium. Studies of seaweeds support “marine lakes” in the Hurd Deep trench in the English

Channel, and potential habitat near the deep Porcupine Seabight off the partially-unglaciated region of southwest Ireland (Provan *et al.* 2005; Hoarau *et al.* 2007; Provan & Bennett 2008). Farther north still, it has been suggested that both the Icelandic coast and the Lofoten coast of Norway may have supported marine species (Wares & Cunningham 2001; Maggs *et al.* 2008), although the presence of an Icelandic coastal refugium is contested (Ingólfsson 2009). In North America two glacial marine refugia are suggested: a southern region heading from mid-latitude United States south into the Gulf of Mexico, and a periglacial region in Atlantic Canada (Bernatchez 1997; Maggs *et al.* 2008). A number of studies have suggested that the Flemish Cap (Shaw 2006; Bigg *et al.* 2008), a marine seamount southeast of the island of Newfoundland, may have acted as a glacial refugium, while others suggest a Scotian Shelf refuge (Young *et al.* 2002).

1.4 Study Species

Two species found across the North Atlantic Ocean are the Atlantic wolffish *Anarhichas lupus* and the Atlantic cod *Gadus morhua*. Both are classified in the taxon Acanthomorpha, a group of teleosts (bony fishes) that includes over 14,000 species (Chen *et al.* 2003). Within this group Atlantic wolffish are classified as Acanthopterygii and Atlantic cod are Paracanthopterygii, which is basal within Acanthomorpha (Figure 1.2; Nelson 2006). The two groups shared a common ancestor ca. 140 mya.

1.4.1 Anarhichadidae

The Anarhichadidae family (wolffishes), in the order Perciformes, is a group of large predatory marine fish found across the North Atlantic, Arctic, and North Pacific Oceans (Barsukov 1959). The family comprises two genera: the wolfeel in the monotypic *Anarrhichthys ocellatus* (Ayres, 1855), and *Anarhichas* with four species – the Bering wolffish *A. orientalis* (Pallas, 1814), the Atlantic wolffish *A. lupus* (L., 1758), the spotted wolffish *A. minor* (Ólafsson, 1772), and the northern wolffish *A. denticulatus* (Krøyer, 1845) (Figure 1.3a; Barsukov 1959).

In contrast to the majority of marine fish, Anarhichadidae is more speciose in the Atlantic Ocean than in the Pacific; this originally led to the suggestion that Anarhichadidae may have originated in the Atlantic Ocean, similar to Gadidae and Clupeidae (Barsukov 1959). However, the molecular and biogeographic evidence places the Bering wolffish from the North Pacific and Arctic Ocean as outgroup to the other three, and the wolfeel as sister to the other genus, which suggests a Pacific origin (Barsukov 1959; Johnstone *et al.* 2007; McCusker & Bentzen 2010b).

1.4.1.1 Distribution

The wolfeel and Bering wolffish are found in the Pacific Ocean, with ranges extending from Japan and Russia in the west to Canada and the United States in the east (Barsukov 1959; FishBase 2013d, 2013e). The Bering wolffish is found farther north, not extending past Alaska to the south, while the wolfeel is found as far south as California. Although the two species tend to prefer similar habitat – shallow water with a stony bottom – they are generally not found together (Barsukov 1959).

The Atlantic, spotted, and northern wolffish are found in the North Atlantic Ocean, and in a small area of the Arctic Ocean where it meets the Atlantic (FishBase 2013a, 2013b, 2013c). The ranges extend from New England and Atlantic Canada in the west, to the Bay of Biscay and the Barents Sea in the east (Barsukov 1959; Jónsson 1982). Some of the highest densities are seen in Atlantic Canada waters, off the coast of Newfoundland (O'Dea & Haedrich 2002). Spotted and northern wolffish are generally found farther north than Atlantic wolffish, though their ranges are sympatric across much of the North Atlantic. Some exceptions are in the White Sea at the far south of the Barents Sea, where the Atlantic wolffish is the only wolffish present, the North Sea, where spotted and northern wolffish are rare, and the Flemish Cap, where the Atlantic wolffish is not found (Barsukov 1959; FishBase 2013b).

1.4.2 The Atlantic, spotted, and northern wolffish

1.4.2.1 Description and habitat

Wolffish are large fish, growing up to 145 cm long and weighing more than 25 kg (Jónsson 1982; Templeman 1986b, a). They can live for up to 20 years in the wild, although most caught are much younger, and reach sexual maturity anywhere from 2 to 10 years old (Jónsson 1982; Templeman 1986b; Scott & Scott 1988; O'Dea & Haedrich 2002). As adults they have few predators – the main one being the Greenland shark *Somniosus microcephalus* – though young fish are often eaten by Atlantic cod or saithe *Pollachius virens* (Barsukov 1959; Scott & Scott 1988). They are a nocturnal species, generally spending the days sheltering in crevices, under boulders, or in caves, and coming out to feed at night (Jónsson 1982; Keats *et al.* 1985). While the three

species have a similar body form, Atlantic wolffish are characterised by a slight bluish colour, and by dark transverse stripes along their length, spotted wolffish are covered by distinct spotting, and northern wolffish are brownish or grey with no discrete spots or stripes (Barsukov 1959; Templeman 1986c).

The three wolffish species all reside on the continental shelves of North America and Europe (Figure 1.4). Atlantic and spotted wolffish generally inhabit the shallower slope waters up to depths of about 450 m and 550 m respectively, while the northern wolffish range extends to much deeper waters, up to about 1,000 m (Barsukov 1959; Albikovskaya 1982; Jónsson 1982; Dutil *et al.* 2014). Atlantic wolffish occupy the shallowest habitat, and have evolved genes for antifreeze proteins that allow them to cope with the cold winters (Desjardins *et al.* 2006; Desjardins *et al.* 2007). The three species are generally found in slightly different habitats – the Atlantic preferring a harder, stony bottom, the spotted a stony/sandy bottom, and the northern a soft, muddy bottom (Barsukov 1959; Jónsson 1982). This may help to minimise competition among the three sympatric species, and limit interspecific hybridisation.

1.4.2.2 Migration and reproduction

Wolffish tend to be solitary and sedentary creatures, and have even been seen to defend a territory (Jónsson 1982; Keats *et al.* 1985). They do not generally undertake long migrations; the majority of tag recaptures are within 10 nautical miles of the original tagging site (Barsukov 1959; Konstantinov 1961; Albikovskaya 1982; Jónsson 1982; Riget & Messtorff 1988), although a few instances of long-distance migration have been observed in all three species (Jónsson 1982; Templeman 1984;

Riget & Messtorff 1988). Short-distance seasonal migrations between feeding and spawning grounds are commonly observed (Barsukov 1959). In North America, the wolffish move inland in late summer before pairing up and breeding (Templeman 1984; Keats *et al.* 1985). In Iceland and the White Sea (Russia) the movement is reversed; fish move from shallow feeding grounds to deeper spawning grounds in August (Jónsson 1982; Pavlov & Novikov 1993).

Unlike most fish species, wolffish begin the spawning season by forming pairs and finding nesting holes (Keats *et al.* 1985). The females have internal fertilisation and lay large demersal eggs that produce relatively large young (Johannessen *et al.* 1993; Pavlov & Novikov 1993). The egg masses are large, containing from a few hundred to tens of thousands of eggs and measuring 10 - 14 cm in diameter, and are guarded by the males (Jónsson 1982; Keats *et al.* 1985; Templeman 1986b). The eggs hatch in approximately 1,000 degree days (Falk-Petersen *et al.* 1990). Unlike the demersal eggs and adults, the larvae can be pelagic, often being found in open ocean and feeding on common pelagic plankton (Smidt 1981; Templeman 1984; Falk-Petersen *et al.* 1990). In other cases the larvae will remain close to the bottom where they were hatched (O'Dea & Haedrich 2002).

The different wolffish species breed at approximately the same time and in similar locations; as such hybrids have been suggested (Luhmann 1954; Barsukov 1959; Jónsson 1982; Gaudreau *et al.* 2009). Atlantic-spotted hybrids have been created artificially in the lab, showing that the two species are compatible. In these cases the hybrids are morphologically identical to the spotted wolffish, although microsatellite analyses confirm the hybrid parentage (Gaudreau *et al.* 2009). In the wild, individuals

have been found who show intermediate morphology to spotted and northern wolffish – body shape, dentition, vertebral numbers, and fin-ray numbers like that of the northern wolffish, but clearly spotted skin like that of the spotted wolffish (Templeman 1986c). Genetic analysis has shown that these are a result of hybridisation between the male of the spotted wolffish and the female of the northern wolffish (Genge, Lait & Carr, unpublished data).

1.4.2.3 Dentition and feeding behaviour

Wolffish are demersal benthivores, feeding on bottom-dwelling organisms such as molluscs, echinoderms, and crustaceans (Barsukov 1959; Templeman 1985). In order to deal with these hard-shelled organisms, wolffish have an extensive and unusual dentition system, for which the species are named. They have powerful teeth lining their jaws and pharynx; canines to strip organisms off the bottom substrate, molars to crush hard shells, and incisors to catch more mobile prey (Barsukov 1959; Albikovskaya 1983). These teeth are replaced annually, all at once, just after the spawning season; during this time the female wolffish feeding decreases and the males feed little or not at all (Jónsson 1982; Templeman 1986b; Liao & Lucas 2000).

The three species have slightly different dentition, and as a result feed on different organisms. The Atlantic wolffish has larger and stronger teeth, and feeds more on large molluscs and crustaceans such as whelks, scallops, and clams. The spotted wolffish is intermediate in jaw strength, and eats mostly smaller molluscs, crustaceans, and echinoderms including crabs, brittle stars, and sand dollars. The northern wolffish has the weakest dentition, feeding on smaller and more mobile prey

such as echinoderms, ctenophores, and jellyfish (Barsukov 1959; Jónsson 1982; Albikovskaya 1983). All three species eat fish as a lesser part of their diet, particularly Atlantic redfish *Sebastes norvegicus* (Ascanius, 1772) and Atlantic cod (Albikovskaya 1983; Templeman 1985). The wolffish are the top benthic predator in their habitat, and may in fact be a keystone species for benthic invertebrates in the Atlantic Ocean (Liao & Lucas 2000).

1.4.2.4 Fishery and population status

While not targeted commercially in the Northwest Atlantic, there have recently been specific wolffish fisheries in Greenland and Iceland, and previously in Canada, the Soviet Union, and East Germany (O'Dea & Haedrich 2002). In Iceland, wolffish catches increased following a lull during World War I, with the highest catch rates seen in 1962 (Jónsson 1982). In Greenland, the wolffish fishery increased following the decline of the cod fishery, with a high in the 1950s, but has since declined with a shift to shrimp fishing (Smidt 1981). The wolffish were originally targeted for their skins which were tanned and used as a light leather (e.g., for shoes, bags, and cases), while the meat can be used fresh, dried, smoked, or pickled (Barsukov 1959; Smidt 1981). Only the Atlantic and spotted wolffish were caught for food – the meat of the northern wolffish is considered jellied and unpleasant (Barsukov 1959; Templeman 1984). Presently, wolffish are caught as bycatch to a number of trawl fisheries including the cod, haddock, and herring fisheries (Barsukov 1959; Smidt 1981; Albikovskaya 1982).

Despite the fact that wolffish escaped the overfishing seen in other North Atlantic directed fisheries, their populations were heavily impacted as bycatch by the trawlers used in fisheries, and the resulting damage to the habitat (Watling & Norse 1998; O'Dea & Haedrich 2002). Populations began declining in the 1960s and 1970s, with smaller fish being seen and overall biomass decreasing from the mid-1980s onwards (Riget & Messtorff 1988). All three wolffish species have suffered extensive population declines since the late 1970s (COSEWIC 2000, 2001a, b). Between 1978 and 1994 the number of Atlantic wolffish in Newfoundland waters decreased by an estimated 91%; in a single year (from 1978 to 1979) the abundance dropped by over 25% (O'Dea & Haedrich 2002). Even greater drops were seen in the other two species (96% in spotted wolffish, 98% in northern wolffish). Accordingly, wolffish were the first marine species to be designated as At Risk by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) under the Canadian Species at Risk Act (SARA): Atlantic wolffish were assessed as Special Concern in 2000, followed the next year by both spotted and northern wolffish as Threatened (COSEWIC 2000, 2001a, b). These assessments were all reviewed and sustained in 2012 (COSEWIC 2012c, b, a). The species are not listed by the International Union for the Conservation of Nature (IUCN) or the Convention on International Trade in Endangered Species (CITES). The chance of a status upgrade, however, has improved following a recent IUCN assessment of European marine fish species that identifies northern wolffish as Endangered, spotted wolffish as Near-Threatened, and Atlantic wolffish as Data Deficient (Nieto *et al.* 2015). The assessment only looked at fish in European waters.

1.4.2.5 Previous genetic work

A few previous studies have attempted to define the population genetic structure in the three North Atlantic wolffish species, and to a lesser extent the Bering wolffish. Imsland *et al.* (2008) used allozymes and restriction length fragment polymorphisms (RFLPs) to look at spotted wolffish in the North Atlantic. They found significant differences among their three groups (North America, the mid-Atlantic, and Norway), although isolation-by-distance was weak. The greatest differences were seen with the Norwegian (Barents Sea) individuals.

A similar study using microsatellites and amplified fragment length polymorphisms (AFLPs) again saw differences between the Barents Sea and other populations in both spotted and northern wolffish (McCusker & Bentzen 2011). In Atlantic wolffish a general separation of eastern (Europe) and western (North America) Atlantic populations was seen, with isolation-by-distance and low genetic diversity evident (McCusker & Bentzen 2010a). The population structure was more pronounced in North America, with the exception that Rockall Bank off the west coast of the British Isles was distinct. This may be due to physical isolation of this population by deep waters, while the others are all connected along the continental shelves, or it may represent an older lineage (McCusker & Bentzen 2010a). In Icelandic waters no population structuring was found (Pampoulie *et al.* 2012).

A phylogeographic study was also conducted which used two mitochondrial loci (ND1 and the control region) and nuclear AFLPs. Little to no structure was found across the range of each wolffish species, with shared haplotypes between North American and European populations and a star-like phylogeny (McCusker & Bentzen

2010b). The study showed the relationship among the three North Atlantic wolffish species, with Atlantic and spotted wolffish as sister-species (previously identified in Johnstone *et al.* 2007 using mitochondrial genomes), and concluded that the Northwest Atlantic populations were of postglacial origin (i.e., < 10,000 years) from a single glacial refugium in Europe, possibly near Rockall Bank (McCusker & Bentzen 2010b, a).

1.4.3 Gadidae

The Gadidae family ("true cods"), in the order Gadiformes, is a group of bottom-dwelling sea fish found along continental shelves and slopes in temperate and cold water environments around the world. While the phylogeny remains controversial, the current estimate is that the family contains 13 genera and 24 species (Eschmeyer *et al.* 2016), including the commercially-important *Gadus*, *Melanogrammus*, and *Pollachius* (Figure 1.3b). Gadids are second only to Clupeidae (herrings, sardines, and anchovies) in terms of total biomass fished. Gadidae is species-rich in the Atlantic Ocean, several species occur in the Arctic Ocean, and few species are found in the Pacific, with only 3 endemics (Scott & Scott 1988; Carr *et al.* 1999). Both biogeographic and molecular evidence indicate an Atlantic origin for the family (Svetovidov 1948; Carr *et al.* 1999).

1.4.4 Atlantic cod

The Atlantic cod is a large demersal (near-bottom) sea fish in the family Gadidae. It is one of three species in the genus *Gadus*, along with the Alaska (or

walleye) pollock *Gadus chalcogrammus* and the Pacific cod *Gadus macrocephalus* (Figure 1.3b). A fourth *Gadus* species, Greenland cod *Gadus ogac*, is now considered conspecific with *G. macrocephalus* (Carr *et al.* 1999; Coulson *et al.* 2006)

Atlantic cod are the largest of the gadids, reaching up to 1.8 m in length and 100 kg in mass, and they can live up to 25 years (Scott & Scott 1988; Moyle & Cech Jr. 2000; Rose 2007). In many regions Atlantic cod are the apex predatory fish, and may in fact act as a keystone species. As with most gadids, Atlantic cod typically have elongated body shape, small curved scales, no spines, and a closed-off swim bladder (Scott & Scott 1988; Moyle & Cech Jr. 2000; Nelson 2006). They are brown, grey, or greenish in colour, exhibit a distinctive lateral line, and have a chin barbel used to help find food. Atlantic cod are coldwater fish generally found in shallow waters. They prefer water temperatures between 0 and 5 °C, and can produce antifreeze proteins to allow survival in sub-zero temperatures (Hew *et al.* 1981; Rose & Leggett 1988; Goddard *et al.* 1994).

1.4.4.1 Distribution and habitat

Atlantic cod are found in relatively shallow waters across the continental slopes and shelves of the North Atlantic Ocean. Their distribution extends from North Carolina on the east coast of the United States, north through Atlantic Canada to Greenland, east across Iceland and the Faroe Islands, and as far as Svalbard and the White Sea in northern Europe (Figure 1.5; Rose 2007). Atlantic cod are found in the North Sea, Baltic Sea, Danish Belt Sea, and Barents Sea, and also in three coastal Arctic fjords on Baffin Island, Nunavut, Canada: Lake Ogac located near Iqaluit on

Frobisher Bay, and Lakes Qasigialiminiq and Tariujarusiq near Pangnirtung on Cumberland Sound (Hardie 2003). The three Arctic lake populations are unusual; the lakes are a combination of fresh and saltwater (meromictic), and there are currently no marine populations of Atlantic cod nearby. The lake populations are small, estimated at 500 - 1,000 fish, and cod are the only fish in the lakes. This has resulted in a very different life history strategy for these cod, with larger body size and higher rates of cannibalism. The Baltic Sea population is also unusual; the Baltic Sea is a brackish sea with a salinity less than a third that of seawater (< 1% compared to ~3.5% salinity; Howes 1991; Meier & Kauker 2003). It is on the margin of suitable habitat for Atlantic cod, and yet it supports a large and healthy population (Nissling & Westin 1997).

1.4.4.2 Migration, feeding, and reproduction

Atlantic cod are large omnivores, commonly migrating in large shoals of year-class between offshore wintering grounds and inshore feeding grounds (Rose 2007). Long distance migration is not uncommon in cod; tagging studies have found extensive movement within the Northwest Atlantic, and a cod tagged in the North Sea was caught four years and over 3,000 km later on the Grand Banks off Canada (Gulland & Williamson 1962; Jónsson 1996; Robichaud & Rose 2004). Other studies suggest that cod generally stay in a single region (Lear 1984). A review of cod tagging data by Robichaud & Rose (2004) showed that there are four migration strategies employed in Atlantic cod: sedentary (~40%), accurate and inaccurate homing (~40%), and widespread dispersal (~20%). There also exists sedentary populations of cod in

inshore bays around Newfoundland and Labrador (Pinhorn 1969; Ruzzante *et al.* 1996; Rose 2007).

Atlantic cod are generalist predators, feeding opportunistically on any available food source (Moyle & Cech Jr. 2000; Rose 2007). Juvenile and small cod feed primarily on large zooplankton (including amphipods, tunicates, and ctenophores), capelin, and shrimp (Konstantinov 1961; Savenkoff *et al.* 2006). Fish make up over 80% of the diet of adult cod, with capelin their preferred prey. Redfish, herring, crustaceans (particularly crabs), and large zooplankton also contribute to the diet. Populations in the United States feed primarily on herring, sand eels, and invertebrates (Powles 1958; Rose 2007). The Arctic lake populations subsist on invertebrates, such as echinoderms and molluscs, as well as smaller cod (Patriquin 1967; Hardie 2004; Hardie & Hutchings 2011). Cannibalism has also been seen in marine cod populations, accounting for ~5% of the diet (Savenkoff *et al.* 2006).

Atlantic cod mature at anywhere from two to seven years, and spawning occurs continuously from this point onwards (Rose 2007). Reproduction occurs by broadcast spawning, with large numbers of eggs released and low survival rates (Rose 2007). A female cod may hold 1 - 10 million eggs which she releases over a period of weeks; larger and older females carry more and higher quality eggs, and larger males have greater fecundity (Rose 2007). Spawning continues over a period of months in the spring, and continues even as the fish move towards feeding grounds (Hutchings *et al.* 1993; Myers *et al.* 1993). The fertilised eggs are sufficiently buoyant to float near the surface, and they drift for weeks in the currents (Rose 2007). Once hatched the larvae are pelagic and continue to float aimlessly.

1.4.4.3 Population status

Atlantic cod have been a favourite food fish species since at least the 15th century, and were (and in some places still are) the basis for many fisheries (Rose 2007). Many people targeted cod due to their large size, their population abundance, and their ease of capture (Moyle & Cech Jr. 2000). Early stories told of a plethora of Atlantic cod as far as the eye could see; it seemed unthinkable that this resource would ever cease. Today, the cod fisheries in North America are essentially non-existent, while those abroad are greatly reduced (Moyle & Cech Jr. 2000).

Atlantic cod were fished extensively throughout the 19th and 20th centuries using a variety of fishing techniques (e.g., long-lines and otter trawlers; Pinhorn 1969). Severe population declines arose first in Europe then in North America. The population crash in the Northwest Atlantic occurred from the 1960s to 1980s, and a moratorium on commercial cod fishing was introduced in Canada in 1992. Limited fisheries have been reopened in some inshore regions around Newfoundland, and ongoing sentinel and recreational or food-fisheries remain in operation (Rose 2007). Russian and Norwegian cod stocks fared better with effective management regulations allowing both stocks to recover. It has been suggested that Atlantic cod populations have been slow to recover due to a combination of climate and oceanographic conditions (Beaugrand *et al.* 2003; Köster *et al.* 2005; Hiddink *et al.* 2008). Other potential causes include very low starting numbers, reduced prey availability, and increased predation.

By the latter half of the 20th century Atlantic cod populations had declined by > 90% in many regions, and by > 99% in the northeast Newfoundland and Labrador

stock (the "Northern Cod" complex; COSEWIC 2003; Rose 2007). In response, Atlantic cod were classified as “vulnerable” on the IUCN Red List (International Union for Conservation of Nature 2011). Northwest Atlantic cod are currently separated into six regions by COSEWIC; the Laurentian North, Laurentian South, Newfoundland and Labrador, and Southern populations have all been assessed as “Endangered”, the Arctic lakes populations as “Special Concern”, and the marine population as “Data Deficient” (COSEWIC 2010). Despite the fact that Atlantic cod have been assessed under SARA (Species at Risk Public Registry 2016), the recommendation was not accepted by the Minister, and so the species is not currently protected under SARA.

1.4.4.4 Previous population work

Given its importance as a food fish, and the dramatic population changes it has undergone, Atlantic cod have been extensively studied. By the early 20th century scientists were interested in whether Atlantic cod were a large homogeneous group, or whether they separated along geographical or morphological lines. Meristic analyses suggested regional races in Europe (Schmidt 1930), and found discrete differences in the Northwest Atlantic (Templeman 1962, 1979). In particular, cod from southern populations tend to have a lower mean vertebral count than those in colder northern waters, inshore fish have lower numbers than offshore fish, and Flemish Cap cod have fewer vertebrae relative to the nearby Grand Banks (Stanek 1968; Lear *et al.* 1979; Templeman 1981; Lear & Wells 1984; Pepin & Carr 1993). Subsequent tagging studies also supported the idea that cod in the North Atlantic form distinct stocks

despite high dispersal capabilities and observed long-distance migration (Gulland & Williamson 1962; Lear 1984; Swain *et al.* 2001; Robichaud & Rose 2004). Other morphological characteristics, such as colour and otolith shape, have also been used to differentiate races or stocks (Møller 1966, 1968; Love 1970; Jónsdóttir *et al.* 2006).

With the introduction of molecular techniques, studies turned to genetic relationships. Early studies employed protein markers to compare cod populations. Sick (1965b, a) and Frydenberg *et al.* (1965) studied cod from across the North Atlantic using haemoglobin markers. They identified frequency differences across the different populations, and suggested that there were six distinct regions: North America, Greenland-Iceland, Faroes, North Sea-Danish Sea-Belt Sea-Kattegat Fjord-Skagerrak, Baltic Sea, and Norway (Barents Sea). Transferrin studies identified Atlantic Canada as unique from European populations, while a combination of transferrin and phosphoglucose isomerase showed further separation within Canadian waters (into north Laurentian, south Laurentian, and Flemish Cap; Jamieson 1967, 1975; Cross & Payne 1978). Later studies that employed a wide variety of protein markers, however, detected only weak differentiation, and suggested that early findings may have been due to selection at the markers used (Mork *et al.* 1985; Jamieson & Birley 1989; Magnussen 1996).

The late 1980s and early 1990s introduced mitochondrial DNA (mtDNA) studies – first through RFLP analysis, and then to direct sequencing. RFLPs allowed the mitochondrial genome to be scanned for specific changes in the restriction enzyme cut sites. Restriction analyses showed significant differences in Norway, between coastal and Arctic samples (Dahle 1991), and either strong separation consistent with

isolation-by-distance (Pogson *et al.* 1995), or little differentiation among populations studied, even on the large scale (Smith *et al.* 1989; Árnason *et al.* 1992). It was suggested that recent divergence was the likely cause of the low variation observed in RFLPs.

The first population study to sequence cod mtDNA looked at a short fragment of the cytochrome *b* locus (CYTB); 298 base pairs (bp) were sequenced from Norway and Newfoundland (Carr & Marshall 1991a). The study identified allele frequency differences between the two regions, and Newfoundland showed lower diversity than Norway. A follow-up study expanded the sequenced region to 307 bp from across the Northwest Atlantic. A single common haplotype was identified in many individuals, with all other samples one to eight variable sites away (Carr & Marshall 1991b). This short section of mtDNA has since been sequenced in thousands of cod from across the North Atlantic (Árnason *et al.* 1992; Pepin & Carr 1993; Carr *et al.* 1995; Árnason & Pálsson 1996; Árnason *et al.* 1998; Carr & Crutcher 1998; Árnason *et al.* 2000; Sigurgíslason & Árnason 2003; Árnason 2004b). The majority of samples show one of five high-frequency haplotypes, *'A'*, *'C'*, *'D'*, *'E'*, or *'G'*, each only a few bases different. There exists a trans-Atlantic cline in allele frequency with much higher diversity in eastern populations: haplotype *'A'* is the dominant genotype in Newfoundland, with a haplotype frequency of over 80%, and decreases in frequency as we move east to ~20% in the Baltic Sea; similarly, haplotype *'E'* is at low frequency in Newfoundland and increases to more than 50% in the Baltic Sea (Árnason 2004b). MtDNA does not support the separation of inshore and offshore cod stocks in Newfoundland, nor does it support the separation of coastal and Arctic populations in Norway (Carr *et al.* 1995;

Árnason & Pálsson 1996). Overall there is a starburst pattern of variation, indicative of a recent rapid expansion (Árnason 2004b; Avise 2004).

A recent study used the complete mitochondrial genome to evaluate population structure in Atlantic cod. The study looked at 32 samples from four populations (three in Atlantic Canada and one in Norway) and found a unique sequence in each individual (Carr & Marshall 2008a). Six major groups were identified, although there was little correlation to geographical origin. They noted that what had originally been a single widespread haplotype (CYTB haplotype *'A'*) was actually a paraphyletic assemblage. The study showed that there was much more variation in Atlantic cod than we have seen so far in mtDNA, and that the population divergence is much deeper than previously believed.

Nuclear markers tell a slightly different story. One study using minisatellites showed strong population differences, particularly among trans-Atlantic populations, and higher diversity in Europe (Galvin *et al.* 1995). Differences were also seen within regions (e.g., among the different Norwegian fjords), with greater divergence among the European populations than in North America. Some microsatellite markers suggest range-wide isolation-by-distance and structure at the regional scale (Nielsen *et al.* 2001; O'Leary *et al.* 2007). In North America at least five distinct groups have been identified: northeastern Newfoundland and southern Labrador (North), the Grand Banks (South), the Flemish Cap, the Scotian Shelf, and Georges Bank (Bentzen *et al.* 1996; Ruzzante *et al.* 1998). Differences have also been seen among inshore and offshore wintering fish (Ruzzante *et al.* 1996; Beacham *et al.* 2002), between Georges Bank and the southern Nantucket Shoals (Lage *et al.* 2004), and among the three

landlocked Arctic populations (Hardie *et al.* 2006). In contrast, using a different suite of microsatellites Hardie *et al.* (2006) found no differences among the Atlantic Canada populations, while Beacham *et al.* (2002) found no differences in the Newfoundland offshore fish. In Europe, microsatellites have identified differences among spawning sites in the North Sea, English Channel, and Outer Hebrides, but not among those in the Irish or Celtic Seas (Hutchinson *et al.* 2001), a hybrid zone was located between the North Sea / Belt Sea and the Baltic Sea (Nielsen *et al.* 2003), and populations in Norway show significant differences from each other (Knutsen *et al.* 2003; Dahle *et al.* 2006). A recent study suggests that the differences observed with microsatellites, particularly with *Gmo 132*, are likely overvalued as there is evidence of hitchhiking selection where 'neutral' microsatellite markers are linked to loci under selection (Nielsen *et al.* 2006)

A number of other nuclear markers are also being investigated. Studies have looked at the vesicle membrane protein pantophysin I (PanI). Differences have been observed among Norwegian coastal and Arctic cod (Fevolden & Pogson 1995, 1997), Icelandic spawning sites (Jónsdóttir *et al.* 2002), the Flemish Cap, Gilbert Bay (an inshore region in Labrador), and Newfoundland inshore "bay cod" (Beacham *et al.* 2002), and between U.S. shoals (Lage *et al.* 2004). Further research into this marker suggests strong positive selection, particularly in the two intravesicle domains, in Atlantic cod and other gadids (Pogson & Mesa 2004). The pantophysin locus is likely linked to temperature, salinity, and/or depth (Case *et al.* 2005). Using genome scans, an exploratory study identified and characterised 318 single nucleotide polymorphisms (SNPs) in cod from Norway (Moen *et al.* 2008). Genome scans have identified eight

(Nielsen *et al.* 2009) and 40 (Bradbury *et al.* 2010) SNPs associated with either temperature or salinity, and suggest adaptive population divergence. A large-scale genome scan using 1,405 SNPs identified five groups with neutral markers only (eastern Atlantic, western Atlantic, Lake Tariujarusiq, Lake Ogac, and Gilbert Bay), or six groups when outliers under selection were included (separation of western Atlantic into Scotian Shelf/Georges Bank and Gulf of St. Lawrence/Newfoundland/Flemish Cap; Bradbury *et al.* 2013). When the eastern Atlantic samples were run with all SNPs, including outliers, the Davis Strait and Barents Sea separated from the more southern populations, indicating that the markers may be under temperature selection. Similar results have been seen with among the Northeast Arctic and Norwegian coastal cod populations (Berg *et al.* 2016) with significant differences among the two ecotypes caused by SNPs across three distinct genome regions. Árnason & Halldórsdóttir (2015) identified frequency differences among northern and southern populations based on the creatine muscle kinase type A locus. Again the suggestion was that the locus is under temperature selection and provides evidence of natural selection and local adaptation.

1.5 Molecular Markers

As molecular techniques are advancing, increasingly complex evolutionary questions can be addressed. Direct sequencing remains ubiquitous, whether by Sanger or "next-generation" sequencing methods (Sanger *et al.* 1977; Shendure & Ji 2008; Metzker 2010), and genome scans targeting gene-rich regions of the genome are allowing large numbers of SNPs to be identified and mapped (Beaumont & Balding

2004; Baird *et al.* 2008). A growing number of markers and methods are available, allowing previously elusive solutions to be reassessed.

The current study uses the complete mitochondrial genome to study the population genetics and phylogeography of Atlantic cod and three wolffish species in the Northwest Atlantic. MtDNA has been used extensively in both phylogeographic and phylogenetic studies due to its simple mode of inheritance and its general lack of recombination, which allows historical patterns of movement to be tracked and tested (Avise 1992, 2004). The majority of studies to date have incorporated one or a few loci (usually < 2,000 bp). Even these short fragments have allowed postglacial expansion to be studied and genetic discontinuities to be discovered in a number of vertebrate species (e.g., Avise 1992). More recent studies have moved to use the complete mitochondrial genome, again taking advantage of the benefits of mtDNA inheritance, and increasing resolution by incorporating the ~16,500 bp it contains (Avise 2004).

1.5.1 Mitochondrial DNA

Mitochondrial DNA has a number of characteristics which make it an excellent marker for studying evolutionary history both within and among species. The mitochondrial genome is a circular molecule that contains 13 protein coding regions, 22 transfer RNAs, 2 ribosomal RNAs, and a non-coding control region (Figure 1.6; Wilson *et al.* 1985; Avise 2004). With few exceptions it is uniparentally inherited, predominantly maternally, and it typically does not undergo recombination. This allows the pattern of inheritance to be followed without extraneous complications. MtDNA has a high mutation rate compared to many single-copy nuclear genes, with

an estimated mutation rate of ~2% per million years, enabling variation to accrue in a relatively short time period (Brown *et al.* 1979; Wilson *et al.* 1985). As it is uniparentally inherited and haploid, mtDNA has an effective population size four times smaller than nuclear DNA, thus allowing fixation of alleles more quickly via increased sensitivity to genetic drift (Palumbi *et al.* 2001; Brito & Edwards 2009).

MtDNA is not without limitations. Although the genome contains multiple coding and non-coding regions, the loci are physically linked (inherited as a single unit); the mitochondrial genome should therefore be treated as a single locus, albeit a high resolution marker. Homoplasy and heteroplasmy may also play a role (Wilson *et al.* 1985; Avise *et al.* 1987; Avise 2004). Homoplasy occurs when two individuals with the same sequence are assumed to be identical by descent, but instead are identical due to a reverse mutation (e.g., a nucleotide has mutated T → C → T). Heteroplasmy is when a single cell contains multiple mitochondrial sequences (or haplotypes). This can occur through mutation or biparental inheritance (or paternal leakage), and is due to the fact that unlike with nuclear DNA each cell can contain hundreds of copies of the mitochondrial genome as cells have multiple mitochondria (e.g., 10 - 100), and each mitochondrion can contain multiple copies of the genome (Wilson *et al.* 1985). Heteroplasmy can be observed through multiple chromatogram peaks, while homoplasy is difficult to detect.

Following years of debate over the use of a single locus gene tree as a species tree, studies have moved to include multiple loci such as mitochondrial DNA in combination with nuclear genes or microsatellite markers. The use of the complete mitochondrial genome (mitogenomics) has recently emerged as a high-resolution

marker in intraspecific studies (Carr & Marshall 2008b; Pope *et al.* 2011). Highly vagile species such as marine fish often show low population structure with traditional markers. Mitogenomics provides a new perspective on phylogeographic and population genetic studies (Avise 2004; Carr & Marshall 2008a).

1.5.2 Whole mitogenome studies

Complete mitochondrial genome studies were first used to study the phylogenetic relationship between humans and apes (Horai *et al.* 1995), and to expand on traditional control region studies in our understanding of human evolution (Ingman *et al.* 2000; Tanaka *et al.* 2004). Phylogenetic relationships among other higher order taxa (e.g., tetrapods, hexapods, and placental mammals) have been investigated (Curole & Kocher 1999; Nardi *et al.* 2003; Arnason *et al.* 2008), as have relationships within taxa (Cooper *et al.* 2001; Inoue *et al.* 2001; Miya *et al.* 2001; Miya *et al.* 2003; Minegishi *et al.* 2005; Coulson *et al.* 2006; Ursvik *et al.* 2007; Lin *et al.* 2012). In many cases phylogenies have been rearranged, and additional branches resolved.

Intraspecific studies have used mitogenomics to clarify ambiguous or undefined genetic structure. These studies often contradict earlier single locus studies that have found little or no variation. For example, Knaus *et al.* (2011) identified a highly endangered population of fishers *Martes pennanti* as distinct with whole mitogenome analysis that had been overlooked with traditional control region studies, while Feutry *et al.* (2014) found differences among three river drainages in speartooth sharks *Glyphis glyphis* where control region studies had shown homogeneity. Mitogenomics has been used to study domestication in goats (Doro *et al.* 2014), pigs

(Wu *et al.* 2007), and yaks (Wang *et al.* 2010), has given improved divergence date estimates in mammals (Gilbert *et al.* 2008; Stone *et al.* 2010), and has elucidated previously unknown variation in marine fish (Yanagimoto *et al.* 2004; Roques *et al.* 2006; Carr & Marshall 2008a; Carr & Marshall 2008b; Teacher *et al.* 2012; Jacobsen *et al.* 2014), reptiles (Shamblin *et al.* 2012), and mammals (Morin *et al.* 2010; Carr *et al.* 2015).

1.6 Thesis Aims

This thesis uses complete mitochondrial genome sequences to evaluate the population genetic structure in two fish taxa whose distributions span the North Atlantic Ocean. While many studies have attempted to discern the structure in Atlantic cod (Carr & Marshall 1991a; Bentzen *et al.* 1996; Ruzzante *et al.* 1998; Árnason 2004a), and to a lesser extent in the three wolffish species (McCusker & Bentzen 2010a, b, 2011), results remain inconclusive. In addition, few studies have focussed on the waters surrounding Newfoundland and Labrador, an area of special concern due to the recent population crashes and the ongoing fishery moratorium.

In order to look at the population genetic structure in these species, I applied phylogenetic analyses to complete mitochondrial DNA genome sequences. MtDNA has been used extensively in population genetic studies as it is a good marker to follow historical patterns and movement (Avise *et al.* 1987; Avise 2009). The use of the complete genome ensures that all variation is included, and minimises any

ascertainment bias from choosing only polymorphic markers. It also allows me to compare how useful different loci are for uncovering structure in different species.

The study expands upon the previous mitogenomic data in Atlantic cod (Carr & Marshall 2008a), and introduces new data for wolffish (McCusker & Bentzen 2010b). For the Atlantic wolffish, samples were collected from 14 sampling locations across the range – off Newfoundland and Labrador, the mid-Atlantic (Greenland & Iceland), and eastern Atlantic (Europe) – with the focus on Newfoundland and Labrador. In spotted and northern wolffish, samples were collected from six locations (five in each) in the Northwest Atlantic. For the Atlantic cod, samples were collected from 14 sampling locations including the waters off Atlantic Canada from Labrador to the Georges Bank, a landlocked fjord population on Baffin Island, Canada, as well as the Baltic and Barents Seas in Europe.

1.7 Predictions

The distribution of all four species is similar, extending along the continental shelves and slopes from the Barents Sea and Bay of Biscay in the east to the Grand Banks and Georges Bank in the west (Figures 1.4 and 1.5). A comparison of the genetic structure between two fish taxa that live in the same ecosystem but have different life-history characteristics will provide insights into how different species are using this environment. Despite the sympatric range, the genetic structure is expected to be different between the cod and wolffish, but similar among the three wolffish species (see Pelc *et al.* 2009). Atlantic, spotted, and northern wolffish share similar life

history characteristics: they are solitary creatures that have minimal dispersal as they are fairly sedentary and have demersal eggs and larvae. This suggests that the population genetic structure may be stronger than that seen in many marine fish species. Atlantic cod, on the other hand, live in large migratory shoals and exhibit high dispersal capabilities with long-distance migrations and millions of pelagic larvae. They are expected to show weak population genetic structure and extensive gene flow.

Previous results in Atlantic wolffish have shown no diversity in the mitochondrial ND1 and control region loci, and limited structure with microsatellites (McCusker & Bentzen 2010a, 2010b, 2011). I predict that increased variation will be identified, both among and within-populations, and that trans-Atlantic differences may be evident. Diversity patterns will provide information as to whether a single or multiple glacial refugia were present, and where the refugia may have been located. In the spotted and northern wolffish I expect to see much greater genetic variation than previously found in the two mitochondrial loci, as was eluded to in the nuclear markers in Europe (McCusker & Bentzen 2010a, 2011). The structure will likely reflect previous findings, with less structure found in the spotted wolffish and greater differentiation in the more specialist northern wolffish.

Atlantic cod have shown contrasting patterns with mtDNA and microsatellites, likely due to the nature of the markers (i.e., different inheritance modes and mutation rates). Microsatellites have suggested strong population structure within the Northwest Atlantic (Bentzen *et al.* 1996; Ruzzante *et al.* 1998), while mitochondrial DNA has consistently supported a trans-Atlantic cline in allele frequencies but little differentiation at the local scale (Árnason 2004a; Carr & Marshall 2008a). I predict

that there will be trans-Atlantic differences in Atlantic cod, with higher diversity in the eastern Atlantic, and few differences among the Northwest Atlantic populations. The exception is with the isolated Arctic population where physical isolation since the end of the Pleistocene will cause the population to show significant differences and heterogeneity. The inclusion of fish from the Barents Sea, the Baltic Sea, the Flemish Cap, and Baffin Island, all of which have previously been identified as unique populations, will further allow us to test how the cod survived the LGM, and where the current populations may have originated.

1.8 Thesis Organisation

The thesis is presented in five chapters. The first provides a general background to the research questions: what drives population genetic structure in temperate species, particularly marine species, and what patterns are commonly seen? It also looks at what molecular methods are used to study population genetic structure, and gives a thorough background on the study species in question. The second looks at the population genetic structure in Atlantic wolffish using mitogenomics, addressing whether the species shows evidence of structuring or homogeneity, and what factors may have influenced the patterns seen today. It also includes a population study where an additional 80 Atlantic wolffish samples were sequenced for a targeted 1,690 bp region. Chapter 3 begins to examine the congeneric spotted and northern wolffish. I look at whether there is evidence of any population structure when the mitochondrial genome is sequenced, and whether any regions of the genome are heavily influencing

this structure. This chapter represents preliminary work on these two species, based on samples from Newfoundland and Labrador. Chapter 4 addresses the question of population genetic structure and phylogeography in Atlantic cod. Samples from across the range are compared to identify barriers to dispersal, to test current stock designations, and to elucidate evolutionary history. The final chapter summarises the main findings of Chapters 2, 3, and 4, and discusses why the findings are different among the four species. I address possible reasons for any similarities and differences, and identify potential future research.

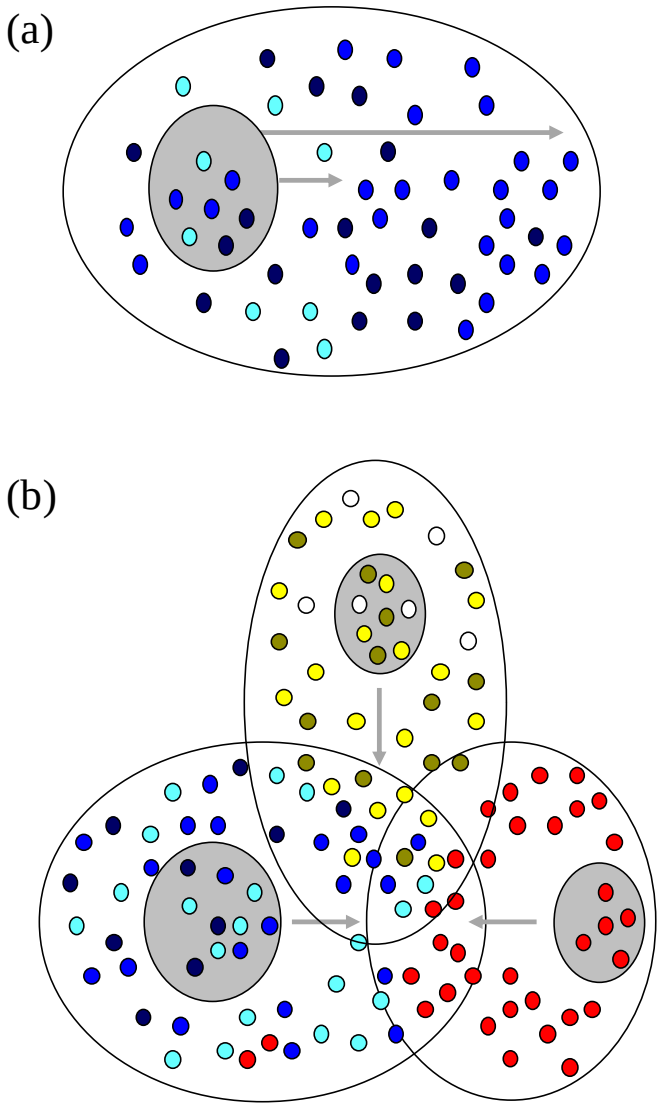


Figure 1.1. How diversity is predicted to change with increasing distance from a source population. (a) If there is a single source, diversity will be highest near the source population and will decrease with increasing distance. As species move away from the source there is a loss of diversity due to founder effect and bottlenecks. (b) With two or more sources, diversity will be high in the source populations and any secondary mixing points (where the different sources meet). The colours represent different genotypes.

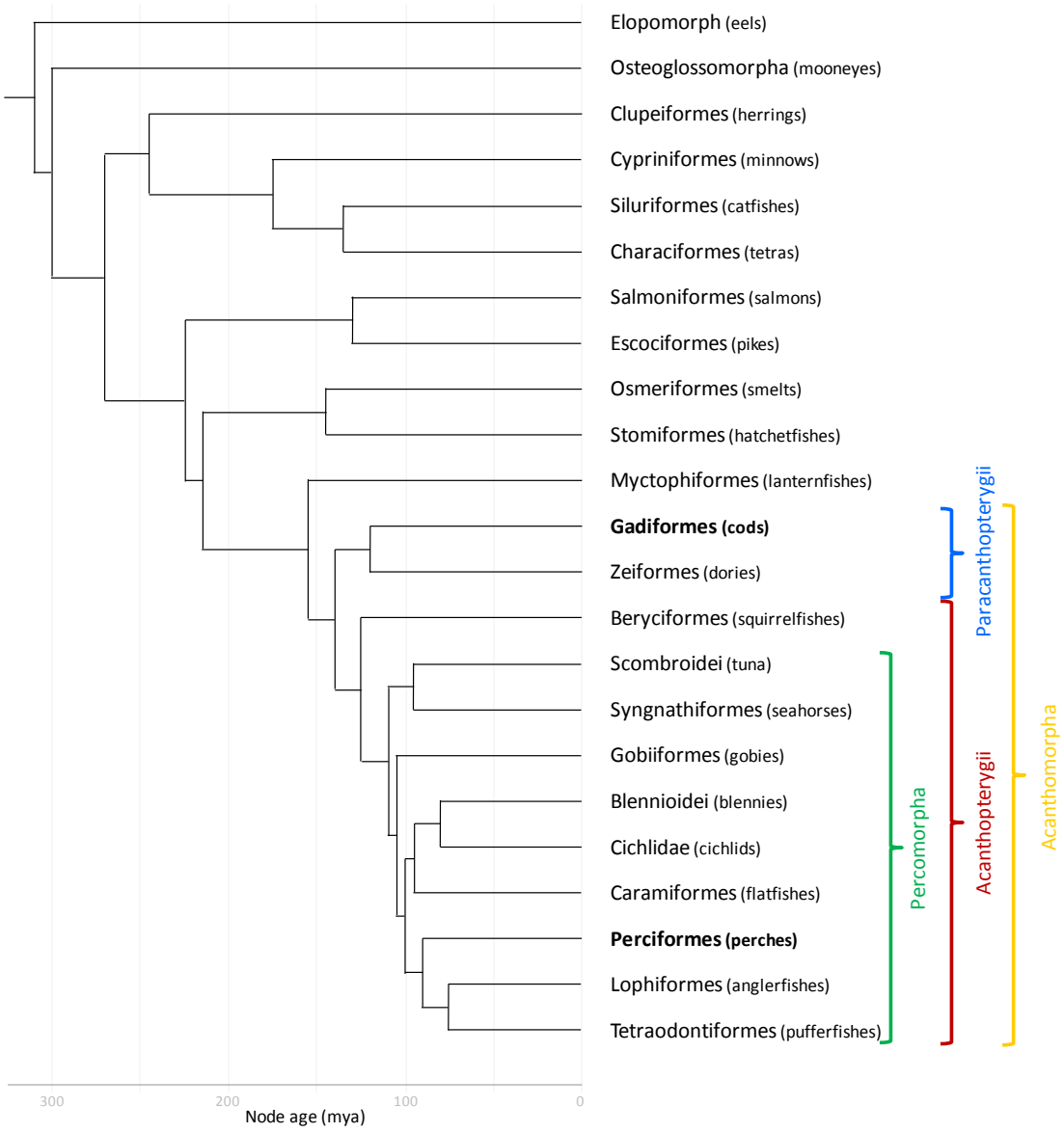


Figure 1.2. A phylogeny of select teleosts with estimated divergence times. The Gadiformes (Atlantic cod) and Perciformes (Atlantic wolffish) are highlighted in bold. Both orders belong to the Acanthomorpha. Figure modified from Miya *et al.* (2003) and Near *et al.* (2012).

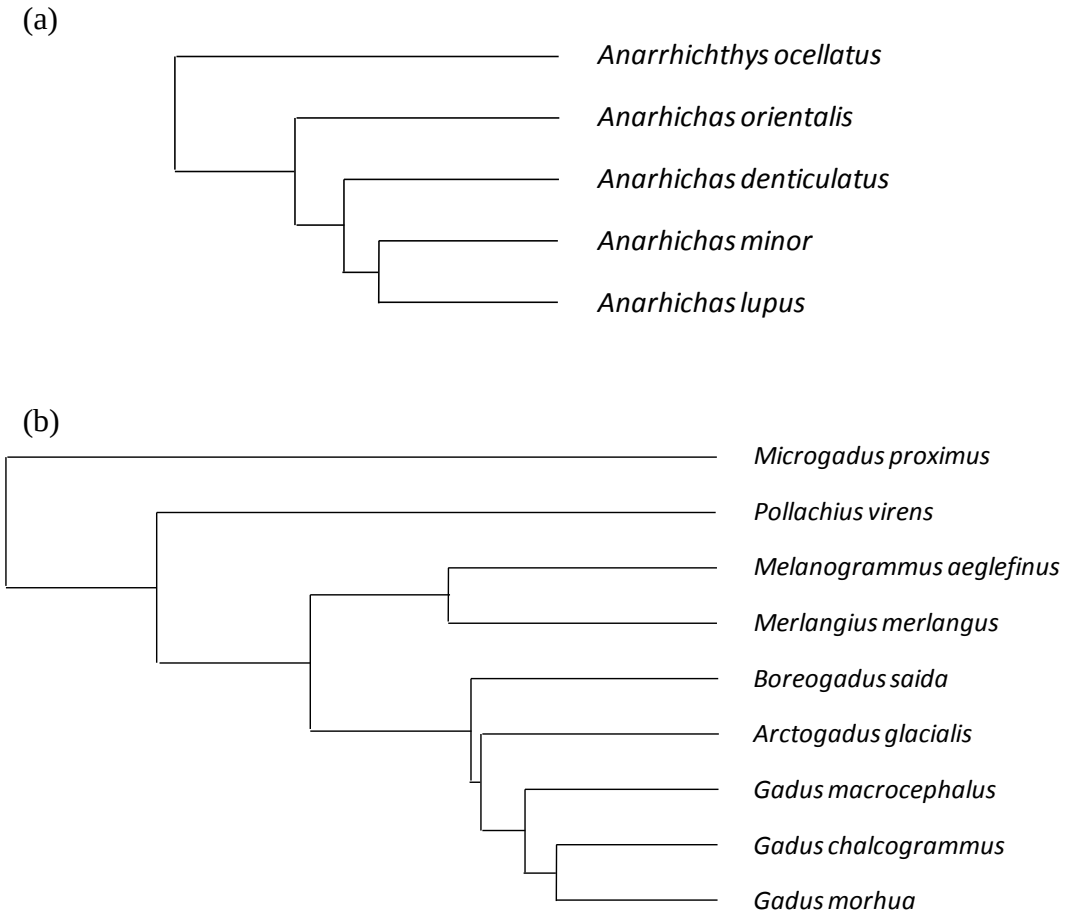


Figure 1.3. A more in-depth look at teleost phylogeny for (a) the Anarhichadidae (wolffish), and (b) Gadidae (cods and codfish). Figure modified from McCusker & Bentzen (2010b) and Coulson *et al.* (2006).

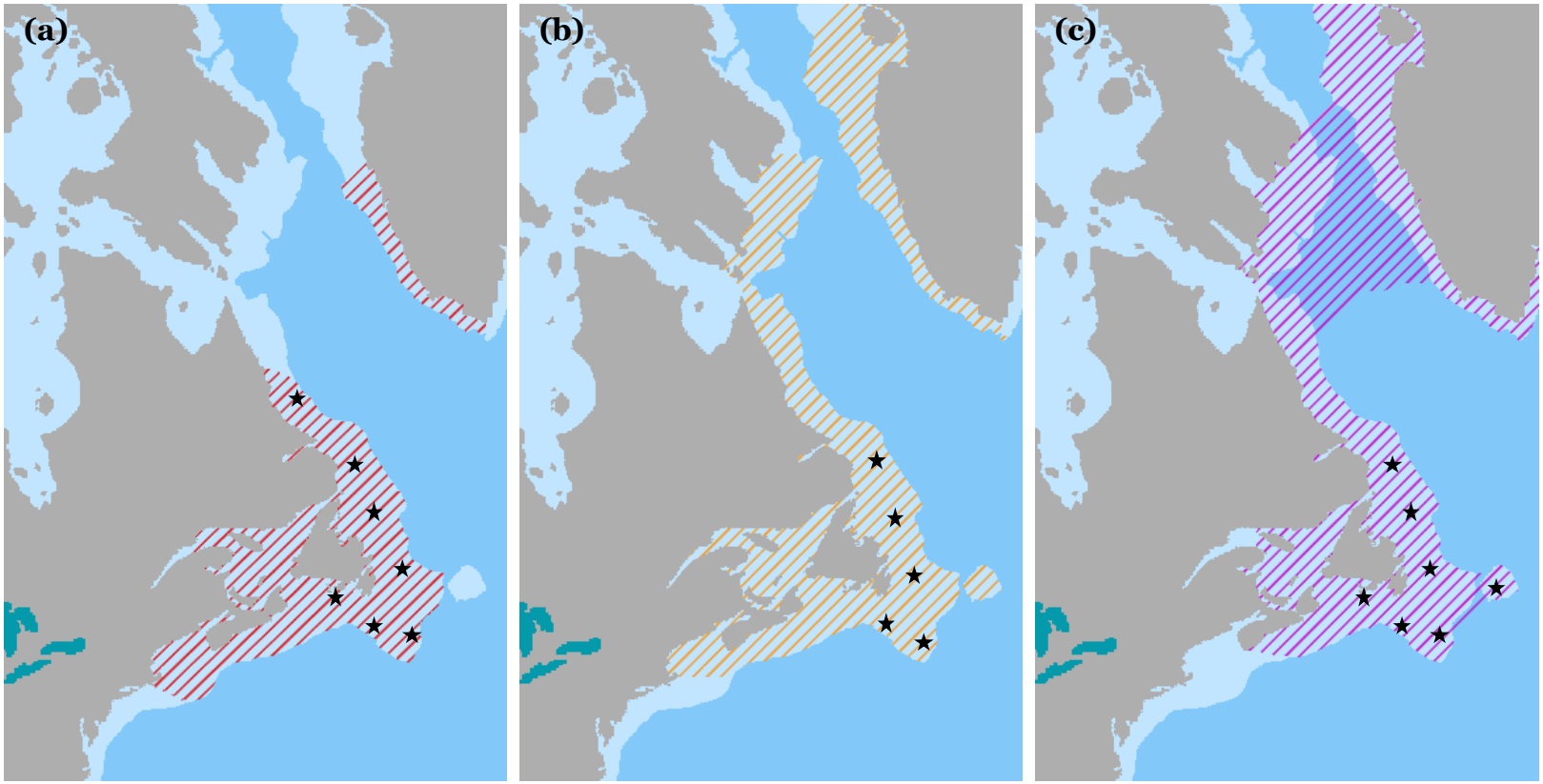


Figure 1.4. The distribution (shaded) of the (a) Atlantic wolffish (b) spotted wolffish and (c) northern wolffish in the Northwest Atlantic. Approximate sampling locations (★) are shown. Additional Atlantic wolffish samples were collected in the Northeast Atlantic (shown in Figure 2.1). Distributions are modified from FishBase (2013a, 2013b, 2013c).

Figure 1.5. The distribution (shaded) of the Atlantic cod across the North Atlantic. Approximate sampling locations (★) are shown. Distribution is modified from FishBase (2013f).

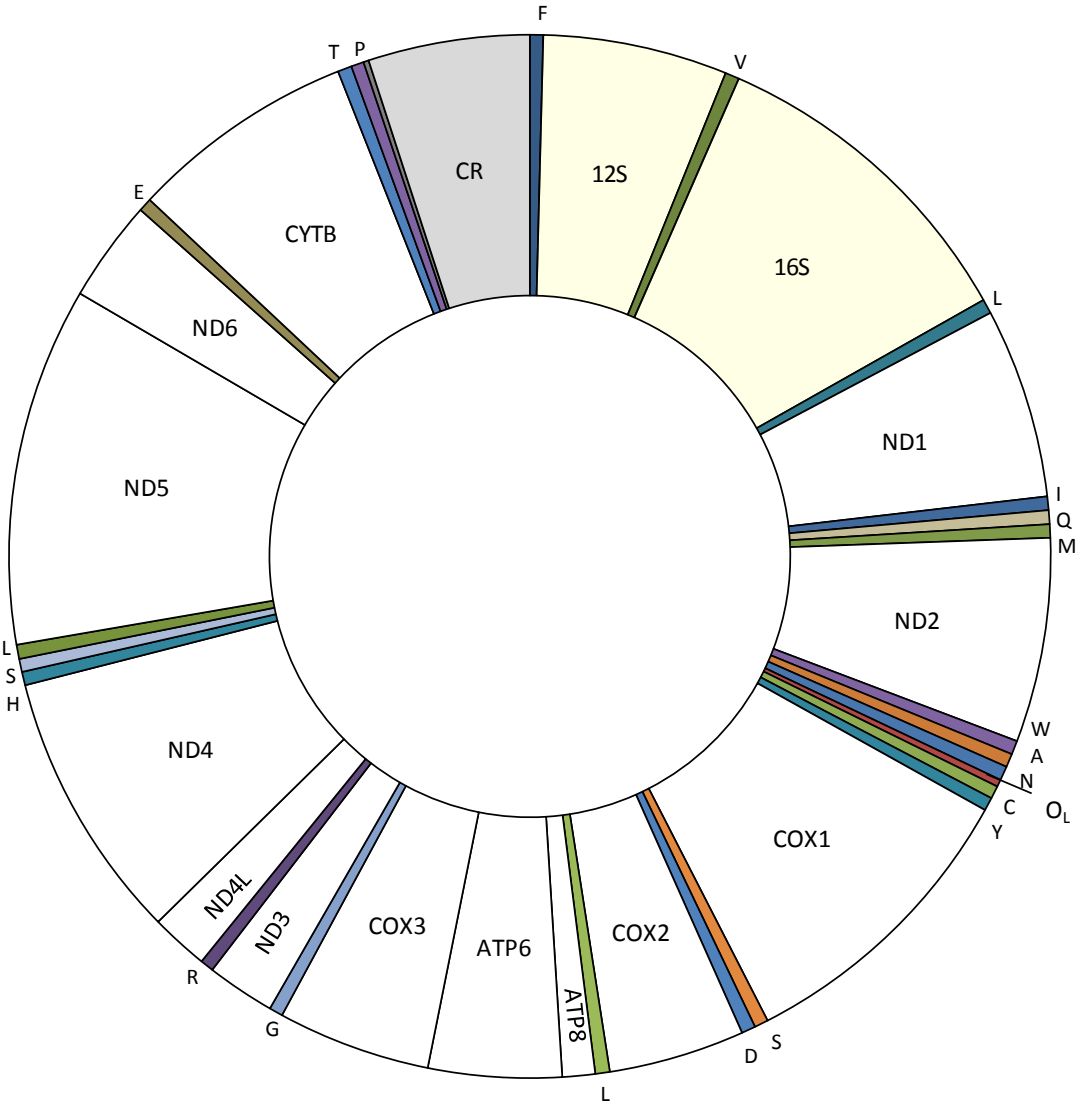


Figure 1.6. The standard structure of the teleost mitochondrial genome (mtDNA) with 13 coding regions (white), two ribosomal RNA regions (yellow), 22 transfer RNAs (coloured), and the non-coding control region (grey).

References

Albikovskaya LK (1982) Distribution and abundance of Atlantic wolffish, spotted wolffish and northern wolffish in the Newfoundland area. *NAFO Scientific Council Studies* **3**, 29-32.

Albikovskaya LK (1983) Feeding characteristics of wolffishes in the Labrador-Newfoundland region. *NAFO Scientific Council Studies* **6**, 35-38.

Allendorf FW, Leary RF (1988) Conservation and distribution of genetic variation in a polytypic species, the cutthroat trout. *Conservation Biology* **2**, 170-184.

Angel MV (1993) Biodiversity of the pelagic ocean. *Conservation Biology* **7**, 760-772.

Árnason E (2004a) Mitochondrial cytochrome b DNA variation in the high-fecundity Atlantic cod: Trans-Atlantic clines and shallow gene genealogy. *Genetics* **166**.

Árnason E (2004b) Mitochondrial cytochrome b DNA variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. *Genetics* **166**, 1871-1885.

Árnason E, Halldórsdóttir K (2015) Nucleotide variation and balancing selection at the *Ckma* gene in Atlantic cod: analysis with multiple merger coalescent models. *PeerJ* **3**, e786.

Árnason E, Pálsson S (1996) Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod *Gadus morhua*, from Norway. *Molecular Ecology* **5**, 715-724.

Árnason E, Pálsson S, Arason A (1992) Gene flow and lack of population differentiation in Atlantic cod, *Gadus morhua* L., from Iceland, and comparison of cod from Norway and Newfoundland. *Journal of Fish Biology* **40**, 751-770.

Árnason E, Petersen P, Kristinsson K, Sigurgíslason H, Pálsson S (2000) Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod from Iceland and Greenland. *Journal of Fish Biology* **56**, 409-430.

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

Arnason U, Adegoke J, Gullberg A, Harley E, Janke A, Kullberg M (2008) Mitogenomic relationships of placental mammals and molecular estimates of their divergences. *Gene* **421**, 37-51.

Avise JC (1992) Molecular population structure and the biogeographic history of a regional fauna - a case history with lessons for conservation biology. *Oikos* **63**, 62-76.

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

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

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography* **36**, 3-15.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography - the mitochondrial-DNA bridge between population-genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A (2015) Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences USA* **112**, 6110-6115.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS One* **3**, e3376.

Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2002) Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology* **11**, 659-674.

Barendregt RW, Irving E (1998) Changes in the extent of North American ice sheets during the late Cenozoic. *Canadian Journal of Earth Sciences* **35**, 504-509.

Barsukov VV (1959) *The Wolffish (Anarhichadidae)* Indian National Scientific Documentation Centre (1972 translation for the Smithsonian Institution), New Delhi, India.

Beacham TD, Brattey J, Miller KM, Le KD, Withler RE (2002) Multiple stock structure of Atlantic cod (*Gadus morhua*) off Newfoundland and Labrador determined from genetic variation. *ICES Journal of Marine Science* **59**, 650-665.

Beaugrand G, Brander KM, Alistair Lindley J, Souissi S, Reid PC (2003) Plankton effect on cod recruitment in the North Sea. *Nature* **426**, 661-664.

Beaumont M, Balding D (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**, 969 - 980.

Bentzen P, Taggart C, Ruzzante D, 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 PR, Star B, Pampoulie C, Sodeland M, Barth JMI, Knutsen H, Jakobsen KS, Jentoft S (2016) Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports* **6**, 23246.

Berggren WA, Hollister CD (1974) Paleogeography, Paleobiogeography and the History of Circulation in the Atlantic Ocean. In: *Studies in Paleo-Oceanography*, pp. 126-186. SEPM Society for Sedimentary Geology.

Bernatchez L (1997) Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt *Osmerus mordax* and their reproductive isolation in the St. Lawrence River estuary (Quebec, Canada). *Molecular Ecology* **6**, 73-83.

Bigg GR, Cunningham CW, Ottersen G, Pogson GH, Wadley MR, Williamson P (2008) Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. *Proceedings of the Royal Society B: Biological Sciences* **275**, 163-173.

Bowen BW, Karl SA (2007) Population genetics and phylogeography of sea turtles. *Molecular Ecology* **16**, 4886-4907.

Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Snelgrove PVR, Morris CJ, Gregory RS, Hardie DC, Hutchings JA, Ruzzante DE, Taggart CT, Bentzen P (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences* **277**, 3725-3734.

Bradbury IR, Hubert S, Higgins B, Bowman S, Borza T, Paterson IG, Snelgrove PVR, Morris CJ, Gregory RS, Hardie D, Hutchings JA, Ruzzante DE, Taggart CT, Bentzen P (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications* **6**, 450-461.

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

Briggs JC (1970) A faunal history of the North Atlantic Ocean. *Systematic Biology* **19**, 19-34.

Brito P, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-455.

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences USA* **76**, 1967-1971.

Burg TM, Lomax J, Almond R, Brooke MdL, Amos W (2003) Unravelling dispersal patterns in an expanding population of a highly mobile seabird, the northern fulmar (*Fulmarus glacialis*). *Proceedings of the Royal Society B: Biological Sciences* **270**, 979-984.

Carr S, Duggan A, Stenson G, Marshall H (2015) Quantitative phylogenomics of within-species mitogenome variation: Monte Carlo and non-parametric analysis of phylogeographic structure among discrete transatlantic breeding areas of harp seals (*Pagophilus groenlandicus*). *PLOS One* **10**, e0134207.

Carr S, Marshall H (2008a) Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic cod (*Gadus morhua*): origins of the "codmother", transatlantic vicariance and midglacial population expansion. *Genetics* **180**, 381-389.

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: *The Implications of Localized Fishery Stocks* (eds. Hunt von Herbing I, Kornfield I, Tupper M, Wilson J). Northeast Regional Agricultural Engineering Service, Ithaca, New York.

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, Marshall HD (1991a) Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic cod (*Gadus morhua*) by the polymerase chain reaction. *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 48-52.

Carr SM, Marshall HD (1991b) A direct approach to the measurement of genetic variation in fish populations: applications of the polymerase chain reaction to studies of Atlantic cod, *Gadus morhua* L. *Journal of Fish Biology* **39 (Supplement A)**, 101-107.

Carr SM, Marshall HD (2008b) Phylogeographic analysis of complete mtDNA genomes from walleye pollock (*Gadus chalcogrammus* Pallas, 1811) shows an ancient origin of genetic biodiversity. *Mitochondrial DNA* **19**, 490-496.

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.

Case R, Hutchinson W, Hauser L, Van Oosterhout C, Carvalho GR (2005) Macro- and micro-geographic variation in pantophysin (*PanI*) allele frequencies in NE Atlantic cod. *Marine Ecology Progress Series* **301**, 267-278.

Chen W-J, Bonillo C, Lecointre G (2003) Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular Phylogenetics and Evolution* **26**, 262-288.

Chow S, Okamoto H, Miyabe N, Hiramatsu K, Barut N (2000) Genetic divergence between Atlantic and Indo-Pacific stocks of bigeye tuna (*Thunnus obesus*) and admixture around South Africa. *Molecular Ecology* **9**, 221-227.

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM (2009) The last glacial maximum. *Science* **325**, 710-714.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.

COSEWIC (2000) COSEWIC assessment and status report on the Atlantic wolffish, *Anarhichas lupus*, in Canada, p. vi + 21 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2001a) COSEWIC assessment and status report on the northern wolffish, *Anarhichas denticulatus*, in Canada, p. vi + 21 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2001b) COSEWIC assessment and status report on the spotted wolffish, *Anarhichas minor*, in Canada, p. vi + 22 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2003) COSEWIC assessment and update status report on the Atlantic cod *Gadus morhua* in Canada p. xi + 76 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2010) COSEWIC assessment and status report on the Atlantic cod *Gadus morhua* in Canada p. xiii + 105 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2012a) COSEWIC assessment and status report on the Atlantic wolffish *Anarhichas lupus* in Canada p. x + 56 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2012b) COSEWIC assessment and status report on the northern wolffish *Anarhichas denticulatus* in Canada p. x + 41 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2012c) COSEWIC assessment and status report on the spotted wolffish *Anarhichas minor* in Canada p. x + 44 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

Coulson MW, Marshall HD, Pepin P, Carr SM (2006) Mitochondrial genomics of gadine fishes: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome* **49**, 1115-1130.

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.

Curole J, Kocher T (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology & Evolution* **14**, 394-398.

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

Dahle G, Jørstad K, Rusaas H, Otterå H (2006) Genetic characteristics of broodstock collected from four Norwegian coastal cod (*Gadus morhua*) populations. *ICES Journal of Marine Science* **63**, 209-215.

Desjardins M, Le Francois N, Fletcher G, Blier P (2007) High antifreeze protein levels in wolffish (*Anarhichas lupus*) make them an ideal candidate for culture in cold, potentially ice laden waters. *Aquaculture* **272**, 667-674.

Desjardins M, Le Francois N, Fletcher G, Blier PU (2006) Seasonal modulation of plasma antifreeze protein levels in Atlantic (*Anarhichas lupus*) and spotted wolffish (*A. minor*). *Journal of Experimental Marine Biology and Ecology* **335**, 142-150.

Doro M, Piras D, Leoni G, Casu G, Vaccargiu S, Parracciani D, Naitana S, Pirastu M, Novelletto A (2014) Phylogeny and patterns of diversity of goat mtDNA haplogroup A revealed by resequencing complete mitogenomes. *PLOS One* **9**, e95969.

Dutil J, Proulx S, Chouinard P, Borcard D, Larocque R (2014) Distribution and environmental relationships of three species of wolffish (*Anarhichas* spp.) in the Gulf of St. Lawrence. *Aquatic Conservation - Marine and Freshwater Ecosystems* **24**, 351-368.

Dyke AS, Andrews JT, Clark PU, England JH, Miller GH, Shaw J, Veillette JJ (2002) The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. *Quaternary Science Reviews* **21**, 9-31.

Eschmeyer W, Fricke R, Laan vd (2016) Catalog of Fishes: Genera, Species, References. Institute for Biodiversity Science and Sustainability, California Academy of Sciences.

Falk-Petersen I-B, Haug T, Moksness E (1990) Observations on the occurrence, size, and feeding of pelagic larvae of the common wolffish (*Anarhichas lupus*) in western Finnmark, northern Norway. *Journal du Conseil Permanent International pour l'Exploration de la Mer* **46**, 148-154.

Feutry P, Kyne P, Pillans R, Chen X, Naylor G, Grewe P (2014) Mitogenomics of the speartooth shark challenges ten years of control region sequencing. *BMC Evolutionary Biology* **14**, e232.

Fevolden S, Pogson G (1995) Differences in nuclear DNA RFLPs between the Norwegian coastal and the North-east Arctic population of Atlantic cod. In:

Ecology of Fjords and Coastal Waters (eds. Skjoldal H, Hopkins C, Erikstad K, Leinaas H), pp. 403-415. Elsevier Science B.V., Amsterdam.

Fevolden S, Pogson G (1997) Genetic divergence at the synaptophysin (Syp I) locus among Norwegian coastal and north-east Arctic populations of Atlantic cod. *Journal of Fish Biology* **51**, 895-908.

FishBase (2013a) Reviewed distribution maps for *Anarhichas denticulatus* (northern wolffish), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

FishBase (2013b) Reviewed distribution maps for *Anarhichas lupus* (Atlantic wolffish), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

FishBase (2013c) Reviewed distribution maps for *Anarhichas minor* (spotted wolffish), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

FishBase (2013d) Reviewed distribution maps for *Anarhichas orientalis* (Bering wolffish), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

FishBase (2013e) Reviewed distribution maps for *Anarrhichthys ocellatus* (wolf-eel), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

FishBase (2013f) Reviewed distribution maps for *Gadus morhua* (Atlantic cod), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

Flint RF (1947) *Glacial Geology and the Pleistocene Epoch* John Wiley & Sons, Inc, New York, NY.

Frankham R (1995) Conservation genetics. *Annual Review of Genetics* **29**, 305-327.

Frydenberg O, Møller D, Nævdal G, Sick K (1965) Haemoglobin polymorphism in Norwegian cod populations. *Hereditas* **53**, 257-271.

Galvin P, Sadusky T, McGregor D, Cross T (1995) Population genetics of Atlantic cod using amplified single locus minisatellite VNTR analysis. *Journal of Fish Biology* **47 (Supplement A)**, 200-208.

Gaudreau CM, Le François NR, Ditlecadet D, Tveiten H, Blier PU (2009) Characterization of the early-stages of the wolffish hybrid *Anarhichas minor* x *Anarhichas lupus*: conservation and aquaculture applications. *Aquatic Living Resources* **22**, 371-377.

Gibbard PL, Head MJ, Walker JC, the Subcommission on Quaternary Stratigraphy (2010) Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. *Journal of Quaternary Science* **25**, 96-102.

Gilbert M, Drautz D, Lesk A, Ho S, Qi J, Ratan A, Hsu C-H, Sher A, Dalén L, Götherström A, Tomsho L, Rendulic S, Packard M, Campos P, Kuznetsova T, Shidlovskiy F, Tikhonov A, Willerslev E, Iacumin P, Buigues B, Ericson P, Germonpré M, Kosintsev P, Nikolaev V, Nowak-Kemp M, Knight J, Irzyk G, Perbost C, Fredrikson K, Harkins T, Sheridan S, Miller W, Schuster S (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences USA* **105**, 8327-8332.

Goddard S, Wroblewski J, Taggart C, Howse K, Bailey W, Kao M, Fletcher G (1994) Overwintering of adult northern Atlantic cod (*Gadus morhua*) in cold inshore waters as evidenced by plasma antifreeze glycoprotein levels. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 2834-2842.

Gulland J, Williamson G (1962) Transatlantic journey of a tagged cod. *Nature* **4844**, 921.

Hamilton SL, Regetz J, Warner RR (2008) Postsettlement survival linked to larval life in a marine fish. *Proceedings of the National Academy of Sciences USA* **105**, 1561-1566.

Hao DM, Tashiro T, Kato M, Sohrin R, Ishibashi T, Katsuyama C, Nagaosa K, Kimura H, Thanh TD, Kato K (2010) Population dynamics of *Crenarchaeota* and *Euryarchaeota* in the mixing front of river and marine waters. *Microbes and Environments* **25**, 126-132.

Hardie D, Hutchings J (2011) The ecology of Atlantic cod (*Gadus morhua*) in Canadian Arctic lakes. *Arctic* **64**, 137-150.

Hardie DC (2003) The evolutionary ecology of Arctic fishes. *Arctic* **56**, 430-433.

Hardie DC (2004) Population genetics, life history, and ecology of arctic marine fishes. *Arctic* **57**, 444-448.

Hardie DC, Gillett RM, Hutchings JA (2006) The effects of isolation and colonization history on the genetic structure of marine-relict populations of Atlantic cod (*Gadus morhua*) in the Canadian Arctic. *Canadian Journal of Fisheries and Aquatic Sciences* **63**, 1830-1839.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences USA* **101**, 14812-14817.

Henriques R, Potts WM, Santos CV, Sauer WHH, Shaw PW (2014) Population connectivity and phylogeography of a coastal fish, *Atractoscion aequidens* (Sciaenidae), across the Benguela Current region: evident of an ancient vicariance event. *PLOS One* **9**, e87907.

Hew C, Slaughter D, Fletcher G, Joshi S (1981) Antifreeze production, freeze resistance, and overwintering of juvenile northern Atlantic cod (*Gadus morhua*). *Canadian Journal of Zoology* **59**, 2186-2192.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

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

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 183-195.

Hiddink J, MacKenzie B, Rijnsdorp A, Dulvy N, Nielsen E, Bekkevold D, Heino M, Lorance P, Ojaveer H (2008) Importance of fish biodiversity for the management of fisheries and ecosystems. *Fisheries Research* **90**, 6-8.

Hoarau G, Coyer J, Veldsink J, Stam W, Olsen J (2007) Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. *Molecular Ecology* **16**, 3606-3616.

Hoelzel AR (1999) Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. *Biological Journal of the Linnean Society* **68**, 23-39.

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proceedings of the National Academy of Sciences USA* **92**, 532-536.

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

Hutchings JA, Myers R, Lilly G (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.

Hutchinson W (2008) The dangers of ignoring stock complexity in fishery management: the case of the North Sea cod. *Biology Letters* **4**, 693-695.

Hutchinson W, Carvalho G, Rogers S (2001) Marked genetic structuring in localised spawning populations of cod *Gadus morhua* in the North Sea and adjoining waters, as revealed by microsatellites. *Marine Ecology Progress Series* **223**, 251-260.

Imsland AK, Stensland K, Johansen T, Le Francois N, Lamarre S, Nævdal G, Foss A (2008) Population genetic structure of the spotted wolffish, *Anarhichas minor*, in the North Atlantic. *The Open Marine Biology Journal* **2**, 7-12.

Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708-713.

Ingólfsson A (2009) A marine refugium in Iceland during the last glacial maximum: fact or fiction? *Zoologica Scripta* **38**, 663-665.

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

Jacobsen M, Pujolar J, Gilbert M, Moreno-Mayar J, Bernatchez L, Als T, Lobon-Cervia J, Hansen M (2014) Speciation and demographic history of Atlantic eels (*Anguilla anguilla* and *A. rostrata*) revealed by mitogenome sequencing. *Heredity* **113**, 432-442.

Jamieson A (1967) Genetic diversity on cod. *ICES Demersal Fish (Northern) Committee* **C.M.1967/F:20**, 1-5.

Jamieson A (1975) Enzyme types of Atlantic cod stocks on the North American Banks. In: *Isozymes. IV. Genetics and Evolution* (ed. Markert C). Academic Press, Inc., New York, NY.

Jamieson A, Birley A (1989) The demography of a haemoglobin polymorphism in the Atlantic cod, *Gadus morhua* L. *Journal of Fish Biology* **35 (Supplement A)**, 193-204.

Johannessen T, Gjøsæter J, Moksness E (1993) Reproduction, spawning behaviour and captive breeding of the common wolffish *Anarhichas lupus* L. *Aquaculture* **115**, 41-51.

Johannesson K, Johannesson B, Rolàn-Alvarez E (1993) Morphological differentiation and genetic cohesiveness over a microenvironmental gradient in the marine snail *Littorina saxatilis*. *Evolution* **47**, 1770-1787.

Johnstone KA, Marshall HD, Carr SM (2007) Biodiversity genomics for species at risk: patterns of DNA sequence variation within and among complete mitochondrial genomes of three species of wolffish (*Anarhichas* spp.). *Canadian Journal of Zoology* **85**, 151-158.

Jónsdóttir IG, Campana SE, Marteinsdottir G (2006) Otolith shape and temporal stability of spawning groups of Icelandic cod (*Gadus morhua* L.). *ICES Journal of Marine Science* **63**, 1501-1512.

Jónsdóttir Ó, Imsland A, Daníelsdóttir A, Marteinsdóttir G (2002) Genetic heterogeneity and prowth properties of different genotypes of Atlantic cod (*Gadus morhua* L.) at two spawning sites off south Iceland. *Fisheries Research* **55**, 37-47.

Jónsson G (1982) Contribution to the biology of catfish (*Anarhichas lupus*) at Iceland. *Rit Fiskideildar* **6**, 2-26.

Jónsson J (1996) Tagging of cod (*Gadus morhua*) in Icelandic waters 1948-1986 and tagging of haddock (*Gadus aeglefinus*) in Icelandic waters 1953-1965. *Journal of the Marine Research Institute Reykjavík* **14**, 1-108.

Keats DW, South GR, Steele DH (1985) Reproduction and egg guarding by Atlantic wolffish (*Anarhichas lupus*: Anarhichidae) and ocean pout (*Macrozoarces americanus*: Zoarcidae) in Newfoundland waters. *Canadian Journal of Zoology* **63**, 2565-2568.

King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA (2001) Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology* **10**, 807-821.

Knaus B, Cronn R, Liston A, Pilgrim K, Schwartz M (2011) Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecology* **11**, e10.

Knutsen H, Jorde P, André C, Stenseth N (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology* **12**, 385-394.

Konstantinov K (1961) Tagging bottom fish in the Barents Sea. *Voprosy Ikhtiologii* **1**, 275-280.

Köster FW, Möllmann C, Hinrichsen H-H, Wieland K, Tomkiewicz J, Kraus G, Voss R, Makarchouk A, MacKenzie BR, St. John MA, Schnack D, Rohlf N, Linkowski T, Beyer JE (2005) Baltic cod recruitment – the impact of climate variability on key processes. *ICES Journal of Marine Science* **62**, 1408-1425.

Lage C, Kuhn K, Kornfield I (2004) Genetic differentiation among Atlantic cod (*Gadus morhua*) from Browns Bank, Georges Bank, and Nantucket Shoals. *Fishery Bulletin* **102**, 289-297.

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

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

Lear W, 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 Northwest Atlantic Fisheries Science* **5**, 23-31.

Lear W, Wells R, Templeman W (1979) The discreteness of the Flemish Cap cod stock. *ICNAF Research Document* **No. 76**, Serial No. 5418.

Lee P, Luschi P, Hays G (2007) Detecting female precise natal philopatry in green turtles using assignment methods. *Molecular Ecology* **16**, 61-74.

Lessios HA (2008) The great American schism: divergence of marine organisms after the rise of the Central American Isthmus. *Annual Review of Ecology, Evolution, and Systematics* **39**, 63-91.

Liao YY, Lucas MC (2000) Diet of the common wolffish *Anarhichas lupus* in the North Sea. *Journal of the Marine Biological Association of the UK* **80**, 181-182.

Lin F, Liu Y, Sha Z, Tsang L, Chu K, Chan T, Liu R, Cui Z (2012) Evolution and phylogeny of the mud shrimps (Crustacea: Decapoda) revealed from complete mitochondrial genomes. *BMC Genomics* **13**, e631.

Love R (1970) *The Chemical Biology of Fishes* Academic Press, London.

Luhmann M (1954) Uber intermediare Formen zwischen *Anarhichas minor* Olafs. und *A. lupus* L. (Teleostei). *Ber. Dtsch. Wiss. Komm. Meeresforsch* **13**, 310-326.

Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Väinölä R, Viard F, Wares J (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* **89 (Supplement)**, S108-S122.

Magnussen E (1996) Electrophoretic studies of cod (*Gadus morhua*) from Faroe Bank and Faroe Plateau compared with results found in other distribution areas. *ICES Demersal Fish Committee* **C.M. 1996/G:10**, 1-18.

McCusker MR, Bentzen P (2010a) Historical influences dominate the population genetic structure of a sedentary marine fish, Atlantic wolffish (*Anarhichas lupus*), across the North Atlantic Ocean. *Molecular Ecology* **19**, 4228-4241.

McCusker MR, Bentzen P (2010b) Phylogeography of 3 North Atlantic wolffish species (*Anarhichas* spp.) with phylogenetic relationships within the family Anarhichadidae. *Journal of Heredity* **101**, 591-601.

McCusker MR, Bentzen P (2011) Limited population structure in northern and spotted wolffishes (*Anarhichas denticulatus* and *A. minor*) despite low apparent dispersal potential. *Marine Biology* **158**, 1869-1878.

Meier H, Kauker F (2003) Sensitivity of the Baltic Sea salinity to the freshwater supply. *Climate Research* **24**, 231-242.

Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**, 31-46.

Minegishi Y, Aoyama J, Inoue J, Miya M, Nishida M, Tsukamoto K (2005) Molecular phylogeny and evolution of the freshwater eels genus *Anguilla* based on the whole mitochondrial genome sequences. *Molecular Phylogenetics and Evolution* **34**, 134-146.

Miya M, Kawaguchi A, Nishida M (2001) Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Molecular Biology and Evolution* **18**, 1993-2009.

Miya M, Takeshima H, Endo H, Ishiguro N, Inoue J, Mukai T, Satoh T, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai S, 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.

Moen T, Hayes B, Nilsen F, Delghandi M, Fjalestad KT, Fevolden S-E, Berg PR, Lien S (2008) Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. *BMC Genetics* **9**, 1-9.

Møller D (1966) Genetic differences between cod groups in the Lofoten area. *Nature* **212**, 824.

Møller D (1968) Genetic diversity in spawning cod along the Norwegian coast. *Hereditas* **60**, 1-32.

Montresor M, Sgrosso S, Procaccini G, Kooistra W (2003) Intraspecific diversity in *Scrippsiella trochoidea* (Dinopbyceae): evidence for cryptic species. *Phycologia* **42**, 56-70.

Mooney HA, Cleland EE (2001) The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences USA* **98**, 5446-5451.

Morin P, Archer F, Foote A, Vilstrup J, Allen E, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Nielsen S, Rasmussen M, Willerslev E, Gilbert M, Harkins T (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* **20**, 908-916.

Mork J, Ryman N, Ståhl 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.

Moyle P, Cech Jr. J (2000) *Fishes: An Introduction to Ichthyology*, 4th edn. Prentice Hall, Upper Saddle River, NJ.

Myers R, Mertz G, Bishop C (1993) Cod spawning in relation to physical and biological cycles of the northern North-West Atlantic. *Fisheries Oceanography* **2**.

Nardi F, Spinsanti G, Boore J, Carapelli A, Dallai R, Frati F (2003) Hexapod origins: monophyletic or paraphyletic? *Science* **299**, 1887-1889.

Near T, Eytan R, Dornburg A, Kuhn K, Moore J, Davis M, Wainwright P, Friedmanm M, Smith W (2012) Resolution of ray-finned fish phylogeny and timing of diversification. *Proceedings of the National Academy of Sciences USA* **109**, 13698-13703.

Nelson J (2006) *Fishes of the World*, 4th edn. John Wiley & Sons, Inc., New York.

Nielsen E, Hansen MM, Ruzzante D, Meldrup D, Grønkjaer P (2003) Evidence of a hybrid-zone in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual admixture. *Molecular Ecology* **12**, 1497-1508.

Nielsen E, Hemmer-Hansen J, Poulsen N, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger G-L, Ogden R, Carvalho G (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**, 276.

Nielsen EE, Hansen MM, Meldrup D (2006) Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in nonmodel organisms. *Molecular Ecology* **15**, 3219-3229.

Nielsen EE, Hansen MM, Schmidt C, Meldrup D, Gronkjaer P (2001) Population of origin of Atlantic cod. *Nature* **413**, 272-272.

Nieto A, Ralph G, Comeros-Raynal M, Kemp J, García Criado M, Allen D, Dulvy N, Walls R, Russell B, Pollard D, García S, Craig M, Collette B, Pollom R, Biscoito M, Labbish Chao N, Abella A, Afonso P, Álvarez H, Carpenter K, Clò S, Cook R, Costa M, Delgado J, Dureuil M, Ellis J, Farrell E, Fernandes P, Florin A-B,

Fordham S, Fowler S, Gil de Sola L, Gil Herrera J, Goodpaster A, Harvey M, Heessen H, Herler J, Jung A, Karmovskaya E, Keskin C, Knudsen S, Kobyliansky S, Kovačić M, Lawson J, Lorance P, McCully Phillips S, Munroe T, Nedreaas K, Nielsen J, Papaconstantinou C, Polidoro B, Pollock C, Rijnsdorp A, Sayer C, Scott J, Serena F, Smith-Vaniz W, Soldo A, Stump E, Williams J (2015) European Red List of Marine Fishes. European Commission, IUCN Publication Services, Belgium.

Nilsson T (1983) *The Pleistocene: Geology and Life in the Quaternary Ice Age* D. Reidel Publishing Company, Dordrecht, Holland.

Nissling A, Westin L (1997) Salinity requirements for successful spawning of Baltic and Belt sea cod and the potential for cod stock interactions in the Baltic sea. *Oceanographic Literature Review* **44**, 1353.

O'Dea NR, Haedrich RL (2002) A review of the status of the Atlantic wolffish, *Anarhichas lupus*, in Canada. *Canadian Field-Naturalist* **116**, 423-432.

O'Leary DB, Coughlan J, Dillane E, McCarthy TV, Cross TF (2007) Microsatellite variation in cod *Gadus morhua* throughout its geographic range. *Journal of Fish Biology* **70 (Supplement C)**, 310-355.

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

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

Pampoulie C, Skirnisdóttir S, Daníelsdóttir AK, Gunnarsson Á (2012) Genetic structure of the Atlantic wolffish (*Anarhichas lupus* L.) at Iceland fishing grounds: another evidence of panmixia in Iceland? *ICES Journal of Marine Science* **69**, 508-515.

Patriquin D (1967) Biology of *Gadus morhua* in Ogac Lake, a landlocked fjord on Baffin Island. *Journal of the Fisheries Research Board of Canada* **24**, 2573-2594.

Pavlov DA, Novikov GG (1993) Life history and peculiarities of common wolffish (*Anarhichas lupus*) in the White Sea. *ICES Journal of Marine Science* **50**, 271-277.

Pechenik JA, Wendt DE, Jarrett JN (1998) Metamorphosis is not a new beginning. *BioScience* **48**, 901-910.

Pelc RA, Warner RR, Gaines SD (2009) Geographical patterns of genetic structure in marine species with contrasting life histories. *Journal of Biogeography* **36**, 1881-1890.

Pepin P, Carr SM (1993) Morphological, meristic, and genetic analysis of stock structure in juvenile Atlantic cod (*Gadus morhua*) from the Newfoundland shelf. *Canadian Journal of Fisheries and Aquatic Sciences* **50**, 1924-1933.

Pielou E (1991) *After the Ice Age: The Return of Life to Glaciated North America* University of Chicago Press, Chicago, IL.

Pimm SL, Raven P (2000) Biodiversity: extinction by numbers. *Nature* **403**, 843-845.

Pinhorn A (1969) Fishery and biology of Atlantic cod (*Gadus morhua*) off the southwest coast of Newfoundland. *Journal of the Fisheries Research Board of Canada* **26**, 3133-3164.

Pogson G, Mesa K (2004) Positive Darwinian selection at the pantophysin (*Pan* I) locus in marine gadoid fishes. *Molecular Biology and Evolution* **21**, 65-75.

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.

Pope AM, Carr SM, Smith KN, Marshall HD (2011) Mitogenomic and microsatellite variation in descendants of the founder population of Newfoundland: high genetic diversity in an historically isolated population. *Genome* **54**, 110-119.

Powles PM (1958) Studies of reproduction and feeding of Atlantic Cod (*Gadus callarias* L.) in the southwestern Gulf of St. Lawrence. *Journal of the Fisheries Research Board of Canada* **15**, 1383-1402.

Provan J, Bennett K (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* **23**, 564-571.

Provan J, Wattier R, Maggs C (2005) Phylogeographic analysis of the red seaweed *Palmaria palmata* reveals a Pleistocene marine glacial refugium in the English Channel. *Molecular Ecology* **14**, 793-803.

Riget F, Messtorff J (1988) Distribution, abundance and migration of Atlantic wolffish (*Anarhichas lupus*) and spotted wolffish (*Anarhichas minor*) in West Greenland waters. *NAFO Scientific Council Studies* **12**, 13-20.

Robichaud D, Rose GA (2004) Migratory behaviour and range in Atlantic cod: inference from a century of tagging. *Fish and Fisheries* **5**, 185-214.

Roques S, Fox C, Villasana M, Rico C (2006) The complete mitochondrial genome of the whiting, *Merlangius merlangus* and the haddock, *Melanogrammus aeglefinus*: a detailed genomic comparison among closely related species of the Gadidae family. *Gene* **383**, 12-23.

Rose GA (2007) *Cod: An Ecological History of the North Atlantic Fisheries* Breakwater Books Ltd, St. John's, NL.

Rose GA, Leggett WC (1988) Atmosphere-ocean coupling and Atlantic cod migrations: effect of wind-forced variations in sea temperatures and currents or nearshore

distributions and catch rates of *Gadus morhua*. *Canadian Journal of Fisheries and Aquatic Sciences* **45**, 1234-1243.

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

Ruzzante DE, Taggart CT, Cook D (1998) A nuclear DNA basis for shelf- and bank-scale population structure in northwest Atlantic cod (*Gadus morhua*)" Labrador to Georges Bank. *Molecular Ecology* **7**, 1663-1680.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**, 5463-5467.

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

Savenkoff C, Castonguay M, Chabot D, Fréchet A, Hammill M, Morissette L (2006) Main prey and predators and estimates of mortality of Atlantic cod (*Gadus morhua*) in the northern Gulf of St. Lawrence during the mid-1980s, mid-1990s, and early 2000s. *Canadian Technical Report of Fisheries and Aquatic Sciences* **2666**, vi+32 pp.

Schmidt J (1930) Racial investigations. X. The Atlantic cod (*Gadus morhua*) and local races of the same. *Comptes rendus des travaux du laboratoire Carlsberg* **18**, 1-71.

Schultz JK, Baker JD, Toonen RJ, Bowen BW (2009) Extremely low genetic diversity in the endangered Hawaiian monk seal (*Monachus schauinslandi*). *Journal of Heredity* **100**, 25-33.

Scott W, Scott M (1988) *Atlantic Fishes of Canada* University of Toronto Press, Toronto, ON.

Shamblin B, Bjorndal K, Bolten A, Hillis-Starr Z, Lundgren I, Naro-Maciel E, Nairn C (2012) Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. *Molecular Ecology* **21**, 2330-2340.

Shaw J (2006) Palaeogeography of Atlantic Canadian continental shelves from the Last Glacial Maximum to the present, with an emphasis on Flemish Cap. *Journal of Northwest Atlantic Fisheries Science* **37**, 119-126.

Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135-1145.

Sick K (1965a) Haemoglobin polymorphism of cod in the Baltic and the Danish Belt Sea. *Hereditas* **54**, 19-48.

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

Sigurgíslason H, Árnason E (2003) Extent of mitochondrial DNA sequence variation in Atlantic cod from the Faroe Islands: a resolution of gene genealogy. *Heredity* **91**, 557-564.

Smidt E (1981) The wolffish fishery at West Greenland. *NAFO Scientific Council Studies* **1**, 35-39.

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.

Species at Risk Public Registry (2016) (ed. Canada Go). COSEWIC.

Stanek E (1968) The number of rays in the second dorsal fin and the number of vertebrae in cod from Labrador, Newfoundland and Nova Scotia. *ICNAF Research Document* **No. 71**, Serial No. 2058.

Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* **16**, 608-613.

Stone A, Battistuzzi F, Kubatko L, Perry Jr G, Trudeau E, Lin H, Kumar S (2010) More reliable estimates of divergence times in *Pan* using complete mtDNA sequences and accounting for population structure. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 3277-3288.

Svetovidov A (1948) Gadiformes. In: *Fauna of the U.S.S.R. Vol. IX. No. 4. Fishes*. Zoological Institute of the Academy of Sciences of the U.S.S.R. [Translated by the Israel Program for Scientific Translations for the National Science Foundation and Smithsonian Institution as OTS 63-11071].

Swain DP, Frank K, Maillet G (2001) Delineating stocks of Atlantic cod (*Gadus morhua*) in the Gulf of St Lawrence and Cabot Strait areas using vertebral number. *ICES Journal of Marine Science* **58**, 253-269.

Tanaka M, Cabrera V, González A, Larruga J, Takeyasu T, Fuku N, Guo L-J, Hirose R, Fujita Y, Kurata M, Shinoda K-I, Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S, Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H (2004) Mitochondrial genome variation in Eastern Asia and the peopling of Japan. *Genome Research* **14**, 1832-1850.

Taylor JR, Ferrari R (2011) Ocean fronts trigger high latitude phytoplankton blooms. *Geophysical Research Letters* **38**, n/a-n/a.

Teacher A, André C, Merilä J, Wheat C (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evolutionary Biology* **12**, e248.

Templeman W (1962) Division of cod stocks in the Northwest Atlantic. *ICNAF Redbook 1962* **Part III**, 79-123.

Templeman W (1979) Migrations and intermingling of stocks of Atlantic cod, *Gadus morhua*, of the Newfoundland and adjacent areas from tagging in 1962-66. *International Commission for the Northwest Atlantic Fisheries Research Bulletin* **14**, 5-50.

Templeman W (1981) Vertebral numbers in Atlantic cod, *Gadus morhua*, of the Newfoundland and adjacent areas, 1947-71, and their use for delineating cod stocks. *Journal of Northwest Atlantic Fisheries Science* **2**, 21-45.

Templeman W (1984) Migrations of wolffishes, *Anarhichas* sp., from tagging in the Newfoundland area. *Journal of Northwest Atlantic Fisheries Science* **5**, 93-97.

Templeman W (1985) Stomach contents of Atlantic wolffish (*Anarhichas lupus*) from the Northwest Atlantic. *NAFO Scientific Council Studies* **8**, 49-51.

Templeman W (1986a) Contribution to the biology of the spotted wolffish (*Anarhichas minor*) in the Northwest Atlantic. *Journal of Northwest Atlantic Fisheries Science* **7**, 47-55.

Templeman W (1986b) Some biological aspects of Atlantic wolffish (*Anarhichas lupus*) in the Northwest Atlantic. *Journal of Northwest Atlantic Fisheries Science* **7**, 57-65.

Templeman W (1986c) Spotted forms of the northern wolffish (*Anarhichas denticulatus*). *Journal of Northwest Atlantic Fisheries Science* **7**, 77-80.

Thorson G (1950) Reproductive and larval ecology of marine bottom invertebrates. *Biological Reviews* **25**, 1-45.

Ursvik A, Breines R, Christiansen J, Fevolden S-E, DH C, Johansen S (2007) A mitogenomic approach to the taxonomy of pollocks: *Theragra chalcogramma* and *T. finnmarchica* represent one single species. *BMC Evolutionary Biology* **7**, e86.

Wang Z, Shen X, Liu B, Su J, Yonezawa T, Yu Y, Guo S, Ho S, Vilà C, Hasegawa M, Liu J (2010) Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal. *Journal of Biogeography* **37**, 2332-2344.

Ward R, Woodwark M, Skibinski D (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology* **44**, 213-232.

Wares J, Cunningham CW (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* **55**, 2455-2469.

Watling L, Norse E (1998) Disturbance of the seabed by mobile fishing gear: a comparison to forest clearcutting. *Conservation Biology* **12**, 1180-1197.

Wilson AC, Cann RL, Carr SM, George M, Gyllensten UB, Helmbychowski 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.

Wilson EO (1988) *Biodiversity* The National Academies Press, Washington D.C.

Wu G-S, Yao Y-G, Qu K-X, Ding Z-L, Li H, Palanichamy M, Duan Z-Y, Li N, Chen Y-S, Zhang Y-P (2007) Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. *Genome Biology* **8**, R245.

Yanagimoto T, Kitamura T, Kobayashi T (2004) Complete nucleotide sequence and variation of mitochondrial DNA from 10 individuals of walleye pollock, *Theragra chalcogramma*. *Fisheries Science* **70**, 885-895.

Young A, Torres C, Mack J, Cunningham C (2002) Morphological and genetic evidence for vicariance and refugium in Atlantic and Gulf of Mexico populations of the hermit crab *Pagurus longicarpus*. *Marine Biology* **140**, 1059-1066.

Yurick DB, Gaskin DE (1987) Morphometric and meristic comparisons of skulls of harbour porpoise *Phocoena Phocoena* (L.) from the North Atlantic and North Pacific. *Ophelia* **27**, 53-75.

CHAPTER TWO

A mitogenomic study of Atlantic wolffish (*Anarhichas lupus*) reveals historical separation

L.A. Lait[1] and S.M. Carr[1]

[1]Department of Biology, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, A1B 3X9

2.1 Introduction

The population genetic structure of marine species is often enigmatic. Widespread fish species, particularly those in temperate environments, have traditionally displayed limited or no population structure (Ward 1995; Vis *et al.* 1997). Those with a pelagic habit – whether larval and/or adult – tend to show extensive migration and gene flow. This has been attributed to large ranges, high dispersal capability, and few barriers to dispersal (Palumbi 1992). Species found in tropical or coastal regions, however, are more likely to be limited in their movements due to habitat and resource restrictions, resulting in greater population structuring (Shulman & Bermingham 1995; Wilson 2006). While population connectivity and presence or absence of physical barriers significantly affect population genetic structure, species found in the Northern Hemisphere were also heavily influenced by the effects of the Quaternary glaciations, particularly those of the Pleistocene and the most recent glacial events (ca. 200 - 10 kya; Hewitt 2000; Pflaumann *et al.* 2003; Hewitt 2004; Shaw 2006).

The degree of population genetic structure and its underlying causes is particularly important when considering species of conservation concern. A species' ability to deal with negative pressures, such as habitat degradation or overharvesting, is a direct result of their life history traits. Large marine fish species with long lives, slow growth, and late maturity, for example the roughhead grenadier *Macrourus berglax* or the roundnose grenadier *Coryphaenoides rupestris*, have been shown to be particularly vulnerable to overfishing and habitat destruction such as that seen in the mid-20th century (Baker *et al.* 2009). These same characteristics can affect the recovery potential of species.

The Atlantic wolffish *Anarhichas lupus* (L. 1758) is a large demersal fish found across the North Atlantic Ocean. Along with two sympatric congeners, the spotted wolffish *A. minor* and the northern wolffish *A. denticulatus*, Atlantic wolffish suffered severe population declines in the late 1970s and early 1980s due to a combination of overharvesting as bycatch in directed fisheries for other species (primarily Atlantic cod *Gadus morhua*), and to habitat destruction by the bottom trawls used in such fisheries (Watling & Norse 1998; O'Dea & Haedrich 2002). Populations decreased by 91%, 96%, and 98% for Atlantic, spotted, and northern wolffish respectively (COSEWIC 2000, 2001a, 2001b). Based on these data, the three wolffish species were the first marine fish species to be listed by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) under the Canadian Species at Risk Act (SARA). The Atlantic wolffish was assessed as Special Concern in 2000, and spotted and northern wolffish were assessed as Threatened in 2001 (COSEWIC 2000, 2001a, 2001b). The three species were reviewed in 2012 and the designations were confirmed (COSEWIC 2012a, 2012b, 2012c). The fourth species in the genus, the Bering wolffish *Anarhichas orientalis*, is designated as Data Deficient; that is, with data insufficient to assess its conservation status (COSEWIC 2002).

The distribution of Atlantic wolffish extends from New England and Newfoundland in the Northwest Atlantic, across Greenland and Iceland, to Norway and Russia in the Northeast Atlantic (Figure 2.1). They are found in relatively shallow waters along the continental shelves, typically no deeper than 450 m (Barsukov 1959; Jónsson 1982; Templeman 1984). Wolffish are solitary creatures, spending much of their time hiding in crevices and under boulders, coming out at night to feed on benthic

invertebrates such as molluscs and crustaceans (Jónsson 1982; Albikovskaya 1983; Keats *et al.* 1985). Atlantic wolffish are large marine predators that can grow up to 145 cm long, live up to 20 years, and mature anywhere from 2 - 10 years old (Jónsson 1982; Templeman 1986). Atlantic wolffish undergo internal fertilisation and lay large demersal egg masses. The eggs hatch into well-developed young which can undertake a short pelagic stage or may remain close to their hatching site (Johannessen *et al.* 1993; Pavlov & Novikov 1993). Movement in this species is limited; tagging studies show mainly short-distance movements in adults, primarily seasonal migration between feeding and spawning grounds (Barsukov 1959; Templeman 1984; Keats *et al.* 1985).

Due to the unique life history characteristics, such as sedentary lifestyle and demersal spawning, Atlantic wolffish may be expected to show strong population genetic structure, be susceptible to negative influences including fishing pressures and environmental changes, and to recover more slowly. An understanding of the species' population genetic structure is therefore imperative to devise an optimal protection and recovery plan. Previous results, however, are contradictory. A range-wide phylogeographic study using two mitochondrial DNA (mtDNA) loci found low diversity levels and a star-like phylogeny (McCusker & Bentzen 2010b). With nuclear microsatellite markers, however, McCusker & Bentzen (2010a) found evidence of divergence between populations in the Northeast and Northwest Atlantic, with higher within-group structure in the Northwest. Pampoulie *et al.*(2012), using the same microsatellite markers, found population homogeneity off Iceland.

As molecular techniques have advanced, previously unknown patterns of population divergence have been revealed in many species. For example, as we

progressed from allozymes and restriction enzymes, through sequencing of mitochondrial genes, to microsatellites and nuclear SNPs, the sensitivity and resolution of markers has identified increasing levels of structure. Given the lack of population genetic structure seen in Atlantic wolffish to date, clarification requires molecular markers capable of detecting more subtle structure. Mitochondrial DNA has been used extensively in phylogeographic and population genetic studies; however, until recently only a fraction of the genome has been analysed (usually 2 - 15%). Many single locus studies have inferred population homogeneity, when in fact it may be undersampling of the genome. The advent of mitogenomics has allowed the evolutionary history of species to be more thoroughly examined. Studies of complete mtDNA genomes have clarified evolutionary relationships in mammals (Horai *et al.* 1995; Arnason *et al.* 2008), birds (Cooper *et al.* 2001), and fish (Inoue *et al.* 2001; Miya *et al.* 2001; Miya *et al.* 2003), and have identified previously hidden diversity in humans (Tanaka *et al.* 2004; Pope *et al.* 2011), chimpanzees (Stone *et al.* 2010), fishers (Knaus *et al.* 2011), cod (Carr & Marshall 2008), and sharks (Feutry *et al.* 2014).

The present study examines the population genetic structure of Atlantic wolffish by means of complete mitogenomic sequences (16,512 bp). The degree of population genetic structure in Atlantic wolffish was tested to determine if the different populations are genetically distinct, thus forming discrete groups, or whether they are genetically homogeneous. Sampling was concentrated on the waters off Newfoundland and Labrador (Canada), with additional samples from Europe and the mid-Atlantic to allow for trans-Atlantic comparisons. Atlantic wolffish may be expected to show much greater variation in the complete genome than has been seen previously in this species. Studies on Atlantic

cod (Carr and Marshall 2008) and harp seals *Pagophilus groenlandicus* (Carr et al. 2015), on a similar geographical scale, showed that every individual had a unique genomic sequence. A comparison of single-locus, targeted multi-locus, and complete genome analyses will show how much of the genome is required to produce the maximum resolution of mitogenomic structure in Atlantic wolffish. I predict that the comparison will show that in a recent species with low levels of diversity such as the wolffish, a significant portion of the genome may be required to fully appreciate and elucidate the presence of variation and structure. The use of targeted multi-locus analyses using loci selected *a posteriori* for their variability will allow additional samples to be analysed quickly and easily. Fisheries and Oceans Canada currently have a monitoring and recovery plan for Atlantic wolffish in the Northwest Atlantic; an in-depth population genetics analysis of the sites in this area is crucial for informing management decisions.

2.2 Materials and Methods

2.2.1 Sample collection

A total of one-hundred and forty-one Atlantic wolffish samples were collected by Fisheries and Oceans Canada between August 2002 and November 2003. The fish were caught from seven sampling locations along the coast of Newfoundland and Labrador (Figure 2.1): LABC (central Labrador), LABS (southern Labrador), NENL (northeastern Newfoundland), NGB (northern Grand Banks), SEGB (southeastern Grand Banks), SWGB (southwestern Grand Banks), and SNL (southern Newfoundland). Hearts were removed and stored at -20°C. A further 23 samples were provided by Dr. Paul Bentzen

(Dalhousie University), which represented seven new sampling locations from Europe (RK (Rockall Bank), NS (North Sea), and BS (Barents Sea)), the mid-Atlantic (WG (west Greenland), EG (east Greenland), and IC (Iceland)), and Nova Scotia (SS (Scotian Shelf)), and three additional samples from NGB. These samples were provided as extracted DNA; they had previously been collected as part of research surveys by Fisheries and Oceans Canada between 2002 and 2005, with the exception of the North Sea samples which came from a fish market in Scotland (see McCusker & Bentzen 2010a).

2.2.2 DNA extraction

DNA from most samples used in this study had been extracted previously by members of the Carr lab using the methods below. Where insufficient material was available, I re-extracted DNA from the original tissue sample using the same procedure. DNA was extracted from heart tissue using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's protocol (Qiagen, Hilden, Germany). A small tissue sample (~25 mg tissue) was added to 180 µL Buffer ATL (<10% SDS) with 20 µL proteinase K (20 mg/mL) and incubated overnight at 56°C to digest the protein and other cell components. To help isolate the nucleic acids and remove any nucleases, 200 µL Buffer AL (<50% guanidinium chloride) and 200 µL 99% ethanol were added and incubated for a further 10 min at 56°C. The extracted sample was transferred to a DNeasy Mini spin column, which was centrifuged at 8,000 rpm (6,000x g) for 1 min, washed with 500 µL Buffer AW1 (> 50% guanidinium chloride) by centrifugation for 1 min at 8,000

rpm (6,000x g), and given a final wash in 500 µL Buffer AW2 with a 3 min 14,000 rpm (21,000x g) centrifugation. The DNA was eluted from the column in a final volume of 200 µL Buffer AE (10 mM Tris-Cl, 0.5 mM EDTA pH 9) and stored at -20°C.

2.2.3 DNA amplification

The complete mtDNA genomes of 50 Atlantic wolffish had been amplified by the polymerase chain reaction (PCR) prior to this study. I amplified an additional 34 Atlantic wolffish mitogenomes. Each genome was amplified in a series of 18 - 20 overlapping fragments ranging from 698 - 1,503 base pairs (bp). In some cases, two of the longest fragments were amplified as two smaller pieces. The fragments overlapped by 4 - 735 bp, with an average overlap of 245 bp. Primers were those used previously with Atlantic wolffish (Johnstone *et al.* 2007) or Atlantic cod (Coulson *et al.* 2006), or were designed specifically for this study based on the reference Atlantic wolffish genome sequence obtained by Johnstone et al. (2007; Genbank Accession Number EF427916). Primer pairs and sequences are given in Table 2.1.

PCR reactions were carried out using Qiagen PCR kits with either Taq, HotStar Taq, or TopTaq DNA polymerase. In each case the PCR reaction (25 µL) contained 10X PCR buffer, 2 mM $MgCl_2$ (Qiagen), 0.2 mM dNTP (Qiagen), 0.4 µM each of the forward and reverse primer, 1 unit DNA polymerase (Qiagen), and 2 µL of the genomic DNA extract. The PCR profile was as follows: an initial cycle of 3 min at 95°C (15 min at 95°C for HotStar and TopTaq), followed by 40 cycles of 30 s at 94°C, 45 s at a primer-pair-specific annealing temperature (T_A), 1 min at 72°C; and a final extension at 72°C for 10

minutes. T_A varied from 49 - 58°C according to the primer pair used (Table 2.1). All PCRs were performed in an Eppendorf Mastercycler epGradient S thermocycler (Eppendorf, Hamburg, Germany). PCR products were visualised under UV light on a 1% agarose gel with 0.3 μg/mL ethidium bromide.

Two shorter fragments (1,690 bp) were amplified in an additional 80 Atlantic wolffish samples, one with the primers w08f1 and g08r and one with w14f2 and str14r (Table 2.1). The amplifications were carried out with the Qiagen TopTaq DNA polymerase kit using the conditions described above.

2.2.4 Sequencing

PCR products (5 μL) were purified with an Exo-SAP clean up procedure. The PCR products were incubated at 37°C for 15 min with 0.1 U exonuclease (USB) to remove excess primers and 0.1 U shrimp alkaline phosphatase (USB) to dephosphorylate any unincorporated dNTPs. This was followed by 15 min at 80°C to denature both enzymes. Purified products were then sent to Genome Quebec (McGill University, QC) for Sanger sequencing with both forward and reverse primers. For fragments with extensive overlap, or where sequences showed no ambiguities, a single sequencing primer (either forward or reverse) was used.

2.2.5 Microarray sequencing

An additional 50 samples had been previously amplified and sequenced with Affymetrix GeneChip® CustomSeq® resequencing multispecies microarrays, the

"ArkChip" (Carr *et al.* 2009). The samples were sequenced as part of two- and four-species experiments, alongside Atlantic cod, harp seal, and Newfoundland caribou *Rangifer tarandus* (see Duggan 2007; Carr *et al.* 2008). Amplicons were pooled in equimolar quantities, fragmented, labelled, and sent to the Centre for Applied Genomics (Toronto, ON) where they were hybridised to the microarray chips and scanned with an Affymetrix GeneChip® scanner (Duggan 2007). Signal intensity data were exported to a Microsoft Excel spreadsheet program.

The microarray DNA sequences were checked using a base-calling algorithm consisting of a comparison of fluorescent signal intensity relative to the microarray reference sequence and a set of empirical rules to identify sequence polymorphisms (Carr *et al.* 2008; Carr *et al.* 2009). For each site the hybridised product was compared to the reference sequence such that the central nucleotide of a 25 bp fragment was altered between the four possible bases. The signal intensity was measured for how well the PCR product hybridised to each possible sequence. This was done in both directions, giving two calls for each base. The intensities were then compared to the reference sequence. When a signal was stronger for one base than any other by at least 13% it was considered a "strong" call, if the difference was 10 - 12.9% it was a "weaker" but still reliable call. In order for a polymorphism to be called, both strands had to agree; if not the site was excluded from further analysis.

2.2.6 Data analyses

The dideoxy DNA sequences were checked by visual inspection of the chromatograms using the program Sequencher v4.9 (Gene Codes Corporation, Ann

Arbor, MI). All sequences were aligned in Sequencher v4.9 and MEGA v6.0 (Tamura *et al.* 2013). Variable sites were double-checked against microarray calls or chromatograms and any ambiguous calls were removed. Sequences were aligned to the Atlantic wolffish reference sequence (Johnstone *et al.* 2007), coding region annotations were confirmed against the chicken mitochondrial genome (Desjardins & Morais 1990) and the Atlantic cod partial genome (Johansen *et al.* 1990), and the location of each variable site was recorded. Haplotypes (shared DNA sequences) were assigned using TCS v1.2.1 (Clement *et al.* 2000) and confirmed with Arlequin v3.5.1 (Excoffier & Lischer 2010).

2.2.7 Mitogenomic analyses

To measure genetic diversity for each population, both nucleotide (π) and haplotype (H_d) diversities were calculated in Arlequin v3.5.1 (Excoffier & Lischer 2010) and DnaSP v5.10 (Librado & Rozas 2009). Nucleotide diversity was calculated as $\pi = \sum_{i,j=1}^{s} x_i x_j \pi_{ij}$ where x_i and x_j are the frequencies of the i^{th} and j^{th} sequence, π_{ij} is the number of polymorphisms between the two, and s is the number of sequences observed (Nei & Li 1979). Haplotype diversity was calculated as $H_d = \left(\frac{n}{n-1}\right)\left(1 - \sum_i^h x_i^2\right)$ where x_i is the frequency of the i^{th} haplotype, n is the population sample size, and h is the number of haplotypes (Nei & Tajima 1981).

An analysis of molecular variance (AMOVA) was performed in Arlequin v3.5.1 (100,000 permutations; Excoffier *et al.* 1992; Excoffier & Lischer 2010) to determine whether the genetic variation was found within or among populations. The AMOVA was run on the complete dataset (n = 84) and a reduced subset containing only the North

American populations with n ≥ 9 (LABS, NENL, SENL, NGB, SGB, and SNL; n = 63). Population pairwise genetic distances (Φ_{ST} values) were calculated in Arlequin v3.5.1 (Excoffier & Lischer 2010) on the reduced dataset described above. Pairwise Φ_{ST} values, a haploid-specific measure of genetic diversity based on Wright's fixation index (Wright 1965), can range from 0 (where populations are completely homogeneous) to 1 (where populations are completely isolated from each other). A modified false discovery rate procedure (FDR; Benjamini & Yekutieli 2001) and the sequential Bonferroni method (Holm 1979) were used to correct the critical p-value (P_{crit}) for multiple tests.

The degree of isolation-by-distance was measured in Genepop v4.2 (10,000 permutations; Raymond & Rousset 1995; Rousset 2008) using a Mantel test to compare genetic and geographic distances among North American locations (Slatkin 1993). The average GPS coordinates for each location were used to calculate geographic distances in the Geographic Distance Matrix Generator v1.2.3 (Ersts 2015). Geographic distances were calculated both as straight line "as the crow flies" and shortest marine "as the fish swims" distances. Linearised Φ_{ST} values (calculated above) were used for genetic distances.

An unrooted statistical parsimony network was assembled in TCS v1.21 (Clement *et al.* 2000). The network allows visualisation of the relationship among haplotypes connected based on their DNA sequences. The network was obtained using a 95% connection limit, and all connections were evaluated by visual inspection. A principal coordinates analysis (PCoA) was performed in GenAlEx v6.5 (Peakall & Smouse 2006; Peakall & Smouse 2012). The PCoA allows visualisation of genetic differences based on individual genotypes with no *a priori* population information, and shows the distribution

of variation along the major axes. Clustering analysis was conducted using BAPS v6 (Bayesian Analysis of Population Structure; Corander *et al.* 2008). BAPS uses a Bayesian analysis to assign individuals to clusters based on genetic data with no *a priori* population information. As the data are single genomes composed of many completely linked loci, the clustering with linked loci option and the codon model of linkage were used (see Corander & Tang 2007). The number of clusters, K, was varied from 1 to 20, and the optimal K was determined based on the log marginal likelihood of the best-visited partitions. The significance of the distribution of groups was tested using both a standard chi-square (X^2) test and a modified Monte Carlo chi-square test designed for small population sizes (Roff & Bentzen 1989). The PCoA and clustering analyses were undertaken on both the complete (n = 84) and reduced (n = 63) datasets.

Phylogenetic analyses of the relationship among haplotypes was analysed using both Bayesian and distance methods. A Bayesian tree was constructed in MrBayes v3.2 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003) using the generalised time reversible model of substitution and gamma-distributed rate variation allowing for invariable sites (GTR + Γ + I). The analysis used two simultaneous runs with eight chains for 5,000,000 generations, with a 25% burn-in, final ESS > 1,000, PSRF > 0.999, and standard deviation of split frequencies < 0.005. A neighbour-joining tree (NJ) was constructed in PAUP* v4.10 (Swofford 2003) based on the absolute number of nucleotide differences (10,000 bootstrap replications). In all cases the congeneric spotted and northern wolffish were used to root the trees (Genbank accession numbers EF427917 and EF427918).

2.2.8 Divergence time estimates

Divergence times among haplogroups were estimated using two Bayesian methods: a Bayesian tree based on the strict clock model was run in MrBayes v3.2 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003), and a constant population model was run in BEAST v2.3 (Bouckaert *et al.* 2014). For MrBayes the model was run as above, with GTR + Γ+ I and uniform branch lengths. In BEAST the HKY (Hasegawa, Kishino, and Yano) + Γ + I model with a strict clock was run for 10,000,000 MCMC steps, sampled every 10,000 steps, with a 1,000,000 step burn-in. All ESS parameters were > 1,000. In both cases the trees were calibrated using a normal distribution with a mean divergence time between the congeneric spotted and northern wolffish of 2 million years and a standard deviation of 0.5 million years. A separation date of 1-2 million years ago (mya) has been suggested based on the separation of the Atlantic Ocean wolffish species from the Pacific Ocean wolffish species 3 - 4 mya with the last opening of the Bering Strait (McCusker & Bentzen 2010b). This gives an approximate mutation rate of 0.972% per million years (321 SNPs in 16,516 bp).

2.2.9 Individual loci

To test whether any of the individual loci were able to detect structure in Atlantic wolffish, an AMOVA was run separately on each of the 13 coding regions, the two rRNAs, and the control region. Population pairwise Φ_{ST} comparisons were obtained from the reduced dataset (n = 63) described above. For each locus a statistical parsimony network was constructed and NJ trees were assembled to visualise any detected variation. All analyses were carried out using the methods described above.

2.2.10 Expanded sample for two-fragment (1,690 bp) study

In order to further evaluate the population genetic structure in Atlantic wolffish, two shorter fragments were amplified in an additional 80 samples. The two fragments were a 903 bp fragment of the cytochrome oxidase 1 (COX1) gene, and a 787 bp fragment spanning a portion of the NADH dehydrogenase subunit 4 (ND4) gene (46 bp), the three contiguous tRNAs histidine (69 bp), serine (67 bp), and leucine (73 bp), and 532 bp of the NADH dehydrogenase subunit 5 (ND5). These are hereinafter referred to as the COX1 and ND5 fragments, respectively. The genetic analyses described above were conducted for the combined 1,690 bp from the 164 wolffish (84 complete and 80 partial sequences).

2.3 Results

2.3.1 Complete mitochondrial sequences

Complete mtDNA genome sequences were obtained for 84 Atlantic wolffish from 14 locations (GenBank Accession Numbers KX117921 - KX118005). The mitogenomes were 16,512 bp in length, and the sequences contained the expected 13 coding regions, 12S and 16S rRNA regions, 22 tRNAs, and the non-coding control region. There were 11 overlapping fragments consisting of 60 bp, and 12 short intergenic regions totalling 104 bp including the 37 bp origin of light strand replication (O_L). The average nucleotide composition of the heavy strand was 26.7% A, 28.1% C, 17.8% G, and 27.4% T ([G + C] = 45.9%).

Among the 84 complete sequences there were 257 variable sites including 111 parsimony informative sites (found in more than one individual) and 146 singletons (Table 2.2). There were 224 transitions, 29 transversions, and 4 insertion-deletions (indels). Coding regions contained 75% of the variable sites, and almost 15% were found in the 12S and 16S rRNA regions. The majority of the changes in the coding regions were third position transitions (137 out of 193). There were no frameshift or nonsense mutations (i.e., no unexpected stop codons). All SNP variants occur in one of two states, except for a single site in the 12S region that occurred in three variant forms (C, G, and T). The indels were all found in non-coding regions: two in the 16S rRNA, one in the tRNA tyrosine, and one in an intergenic region. The distribution of the variable sites is given in Table 2.2.

A total of 74 distinct mitogenome sequences (haplotypes or mitotypes) were identified, 71 of which were found only in one individual and three of which were shared by two or more fish (Table 2.3). Of the shared haplotypes, two were found in two individuals and one was shared by nine individuals from five North American populations (Figure 2.2a). No haplotypes were shared between North American and European individuals. Haplotype diversity (H_d) was very high in all populations (0.667 - 1.000) as most individuals had a unique DNA sequence. Nucleotide diversity (π) ranged from 0.00065 to 0.00149 (Table 2.3). The lowest diversity values were seen in SNL and RK, while the highest were in NS and BS.

2.3.2 Population structure

The overall AMOVA gave a $\Phi_{ST} = 0.05221$ ($p = 0.01135$), and allocated 94.8% of the variance within populations and 5.2% among populations. Population pairwise Φ_{ST} values ranged from 0 among the three Grand Banks populations (NGB, SEGB, and SWGB) and between NGB and SNL, to 0.236 between LABS and SNL (Table 2.4). The lowest Φ_{ST} values occurred between geographically proximal locations, and no neighbouring populations were significantly differentiated. Only three comparisons were significant after correction for multiple tests: LABS vs. NGB, LABS vs. SNL, and NENL vs. SNL. The two correction methods used gave critical p-values of 0.0151 for the modified FDR and 0.0102 for sequential Bonferroni. The method used did not affect the number of significant values. There was evidence of weak but non-significant isolation-by-distance when marine swimming distances were compared to genetic distances ($p = 0.076$, $r^2 = 0.418$; Figure 2.3). When straight-line distances were used there was no correlation between geographic and genetic distances ($p = 0.236$, $r^2 = 0.028$).

The statistical parsimony network showed two divergent groups radiating from a central group with a starburst pattern (Figure 2.2a). These three groups are also identified in the multivariate, clustering, and phylogenetic analyses, and will hereafter be referred to as haplogroups A, B, and C. There is no correlation between haplogroup membership and geographical origin of the samples, with the exception that haplogroup A was not found in LABS or the three mid-Atlantic populations (WG, EG, and IC). Haplotype diversity was high in all three groups ($H_d > 0.9$), and nucleotide diversity was high in groups B and C (Table 2.3). The pairwise Φ_{ST} comparisons among the three haplogroups were all

significant ($p < 0.0001$) with values of 0.4014 between groups B and C, 0.4610 between A and C, and 0.6084 between A and B.

The PCoA separated the samples into three groups with no apparent association to geographical origin (Figure 2.4a). Haplogroup A separated from the other two along the first coordinate (30.8% of the variation), while haplogroups B and C separated along coordinate 2 (21.3% of the variation). Coordinate 3 explained 6.3% of the variation. Results were similar in the reduced dataset (not shown), with 79.3% of the variation explained by coordinates 1 (42.6%), 2 (26.8%), and 3 (9.9%).

Bayesian clustering analysis identified the same three groups found across all populations (with the exceptions noted above for A). The highest log maximum likelihood values were -3,247.6 ($K = 3$), -3,309.1 ($K = 4$), and -3,390.4 ($K = 5$), resulting in a probability of three clusters of $p > 0.999$. The distribution of haplogroups across populations varied considerably (Table 2.3, Figure 2.1). Whether the relative frequencies of haplogroups were significantly different across populations was tested using a standard chi-squared test and a test modified for small sample sizes. The distribution was not significantly different from random by the standard method ($X^2 = 38.35$, $df = 26$, $p = 0.056$); however, the modified test (10 replicates) showed a significant departure from random ($X^2 = 30.53 - 54.87$, mean = 41.6, $p = 0.027$).

The Bayesian (Figure 2.5) and neighbour-joining (Figure 2.6) analyses both showed two distinct groups with high-levels of support and a third assemblage consisting of the remaining individuals. These correspond to the same three haplogroups found in the population analyses (i.e., statistical parsimony network, BAPS, and PCoA) with high posterior probabilities (1 for A and B) and strong bootstrap support (79 for A, 85 for B)

for haplogroups A and B. Haplogroup C is an collection of the remaining sequences, and contains at least three sub-haplogroups with high statistical confidence. These groups all converge to a central individual found in haplogroup C, as seen in the statistical parsimony network (Figure 2.2a). The Bayesian and NJ trees show identical structure among the three main groups, with the exception of a single NENL individual that is closer to haplogroups A and B in the Bayesian analysis, but falls into haplogroup C in the NJ tree. Within groups the structure is essentially conserved in haplogroups A and B, with some differences between the two methods seen in haplogroup C.

2.3.3 Divergence time estimates

On the assumption that northern and spotted wolffish diverged approximately two million years ago, the three haplogroups can be estimated to have diverged approximately 160,000 - 220,000 years ago (MrBayes; Figure 2.5) or 93,000 - 208,000 years ago (BEAST, 95% Highest Posterior Density (HPD)). Haplogroups A and B diverged around 145,000 years ago (95% HPD = 93 - 197 kya) while they both diverged from C between 151,000 and 153,000 years ago (95% HPD = 100 - 208 kya). This puts the separation time well within the Pleistocene glaciations (200 - 10 kya). Within each haplogroup, particularly in haplogroup C, there is evidence of additional divergence occurring less than 100,000 years ago, again during the last glacial event.

2.3.4 Individual loci

The analysis of molecular variance at the individual locus level gave overall Φ_{ST} values ranging from 0 to 0.21 (Supplementary Table 2.1). Only COX3, ND3, and ND5

showed significant Φ_{ST}; all of these result from an uneven distribution of haplotypes between the SNL population and the European populations. An AMOVA on the reduced dataset gave Φ_{ST} values from 0 to 0.158 (Supplementary Table 2.1). Analysis of just the six North American locations identified the same three loci as well as COX1 and ND1 with significant Φ_{ST} values. The majority of pairwise comparisons on the reduced dataset were not significant, with 12S, COX1, and ND5 showing the greatest number of significant comparisons (Supplementary Table 2.1). All significant pairwise differences involved the LABS, NENL, and SNL populations.

The statistical parsimony networks for the individual loci generally showed a single common haplotype and a starburst pattern with the exception of COX1 (Supplementary Figure 2.1). A number of the networks showed some limited diversity; for example, 12S and ND1 showed multiple common haplotypes while CYTB and ND5 showed numerous unique haplotypes radiating from several shared haplotypes. None of the networks suggested population structure.

The neighbour-joining trees for the individual loci (Supplementary Figure 2.2) varied from homogeneity in ATP8 and ND6 to highly variable with COX1. None of the trees indicated any population structure. When the samples were colour-coded as haplogroups, 11 of the 16 trees showed some degree of group structuring. Haplogroup A could be detected by 12S, COX1, ATP6, COX3, ND3, and ND5; haplogroup B by COX1, COX2, ND4, CYTB, and CR; and haplogroup C by ND1 and COX1. The only locus to identify the presence of all three groups was COX1. None of the trees showed strong bootstrap support; values range from < 50 to 79.

2.3.5 Two-fragment study

DNA sequences from two fragments (1,690 bp) were obtained for 80 additional Atlantic wolffish from seven sampling locations (GenBank Accession Numbers KX118112 - KX118271). All analyses were also performed with the original 84 samples alone for the shorter fragment; results were similar and are not shown. The 903 bp COX1 fragment contained 21 variable sites while the 787 bp ND5 fragment contained 15 variable sites; there were seven amino acid substitutions: one in COX1, one in ND4, and five in ND5 (Table 2.5). When the two fragments were concatenated there was a total of 36 variable sites: 33 transitions and three transversions. No site had more than two base variants. There were 38 haplotypes, 12 shared between two or more individuals and 26 unique to a single individual. Six of the 12 shared haplotypes were found only in the North American populations, one in North America and West Greenland, while the other five were found across the range (Table 2.6). No haplotypes were shared among the most eastern (NS and BS) and western (SS and SNL) populations. Haplotype diversity was above 0.7 in nine populations, between 0.6 and 0.7 in three (SNL, SS, and EG), and 0 in the RK population (Table 2.7). Nucleotide diversity ranged from 0 in RK to 0.00276 in NS and 0.00296 in BS (Table 2.7).

2.3.6 Population analyses with two fragments

The two-fragment AMOVA gave a $\Phi_{ST} = 0.06083$ ($p = 0.0019$), allocating 93.9% of the variance within populations and 6.1% among populations. When the AMOVA was run on the six North American populations with $n > 15$ (LABS, NENL, NGB, SEGB, SWGB, and SNL; n=142), the overall $\Phi_{ST} = 0.04916$ ($p = 0.0047$) allocated 95.1% of the

variance within populations and 4.9% among populations. Population pairwise Φ_{ST} values (Table 2.8) were similar to those obtained with the complete mitochondrial genomes, ranging from 0 among the Grand Banks populations to 0.196 between LABS and SNL. The lowest Φ_{ST} values occurred between neighbouring populations. The four significant comparisons (after correction for multiple tests with the modified FDR) all involved LABS (versus NGB, SEGB, SWGB, and SNL). After sequential Bonferroni correction only two comparisons were significant: LABS vs. NGB and LABS vs. SNL. Comparison of genetic and geographic distances among locations showed evidence of significant isolation-by-distance both with the shortest marine swimming distances ($r^2 = 0.561$, $p = 0.017$; Figure 2.3) and the straight-line distances ($r^2 = 0.246$, $p = 0.025$).

Statistical parsimony analysis of the two-fragment dataset identified three groups connected by two or three changes through a central NENL individual (Figure 2.2b). These groups correspond to the three haplogroups identified in the complete mitogenomic analyses. Both haplogroups A and B contained a common haplotype ($n = 52$ and $n = 7$, respectively) with the other group constituents differing from them by 1 - 3 substitutions. In haplogroup C there were three closely-related common haplotypes with all other haplotypes differing from them by one or two substitutions. As in the complete mitogenomic analyses there was little correspondence between sample origin and haplogroup membership. The exceptions were that only four samples from northern Newfoundland and Labrador (LABC and NENL) and Europe (BS and NS), and no samples from the mid-Atlantic (WG, EG, and IC) were found in haplogroup A. Haplotype and nucleotide diversities (Table 2.7) were highest in haplogroup B ($H_d = 0.869$, $\pi =$

0.00109) and lowest in haplogroup A (H_d = 0.360, π = 0.00035). The pairwise comparisons among the three haplotypes were all significant ($p < 0.0001$; Table 2.8).

The PCoA of the two-fragment dataset explained 81.1% of the variation along the first three axes, and separated the samples into the same three groups seen above (Figure 2.4b). There was no correlation to geographical origin except as mentioned above. Haplogroups A and C separated along the first coordinate (49.5% of the variation), while haplogroup B separated from the other two along coordinate 2 (17.6% of the variation). Coordinate 3 explained 14.0% of the variation. The PCoA was also run on the reduced dataset (n = 142); results were similar with 49.6% of the variation explained by coordinate 1, 17.4% by coordinate 2, and 15.4% by coordinate 3.

Bayesian clustering analysis identified the same three groups found in the statistical parsimony and principal coordinates analyses. These three haplogroups correspond to those found in the complete mitogenomic analyses with two exceptions: a NENL individual identified as haplogroup C in the mitogenomic analyses was assigned to haplogroup B in the two-fragment analyses, and a SEGB individual identified as haplogroup B in the mitogenomic analyses was now assigned to haplogroup A. All other samples were consistent between the two analyses, and the additional 80 samples were also assigned to one of these three groups. The highest log maximum likelihood values were -623.0 (K = 3), -625.0 (K = 3), and -632.3 (K = 4), giving a probability of three clusters of $p > 0.999$. The distribution of haplogroups across populations is given in Table 2.6. As with the mitogenomic analyses, haplogroup B had the fewest individuals (n = 27), while the other two groups had similar numbers (n = 65 for A, n = 72 for C). Both the

standard chi-squared test (X^2 = 42.06, df = 26, p = 0.024) and the modified test for small sample sizes (X^2 = 50.78, df = 26, p = 0.0025) supported the distribution of haplogroups as significantly different than random. The modified test was run for 10 replicates giving an average X^2 of 50.78; the X^2 values ranged from 39.01 to 63.73.

The Bayesian analysis identified five groups supported by posterior probabilities from 0.58 to 0.94 (Figure 2.7). These correspond to haplogroup A, and two sub-groups each within haplogroups B and C. As with the mitogenomic data, haplogroups A and B were well-defined, while haplogroup C represented the remaining samples. Only haplogroup A was supported by a posterior probability value > 0.5 (p = 0.87). In contrast, the NJ analysis showed two distinct groups: one consisting of haplogroup A and some individuals from haplogroup B, and the other consisting of haplogroup C and the remaining individuals from B (Figure 2.8). Although within the two groups haplogroup B was found clustered together, across the tree it appeared to be paraphyletic. None of the NJ groups were supported by bootstrap values; there were no values over 80%, and none of the haplogroups show bootstrap support over 30% (values not shown).

2.4 Discussion

The 74 haplotypes identified in Atlantic wolffish include three that are shared between two, two, and nine individuals. This pattern contrasts with that seen in previous mitogenomic studies of widespread marine fish species such as Atlantic cod (Carr & Marshall 2008) and Atlantic herring *Clupea harengus* (Teacher *et al.* 2012), in which every fish examined had a unique genome sequence. However, a number of marine

species have shown shared haplotypes among individuals (Morin *et al.* 2010; Shamblin *et al.* 2012; Feutry *et al.* 2014). This may be attributed to the recent separation of the species from its closest relative (Johnstone *et al.* 2007), recent recovery from loss of variation in glacial refugia, and/or a relatively low mutation rate resulting in low levels of divergence among individuals. Extremely low mutation rates have been suggested for sharks (Martin *et al.* 1992) and the giant squid *Architeuthis* sp. (Winkelmann *et al.* 2013).

2.4.1 Population genetic structure

Analysis of complete mitochondrial genomes identified three recognisable groups of haplotypes among Atlantic wolffish. Distribution of these groups over 14 trans-Atlantic locations did not support separation of the populations, nor did it support separation by region (i.e., North America, the mid-Atlantic, and Europe). There was a lack of significant pairwise Φ_{ST} values among populations, with the exception of the southern and northern extremes of Newfoundland and Labrador (Table 2.4). McCusker and Bentzen (2010b, a) previously found no population structure based on 1,830 bp from the mitochondrial ND1 and control region, but did report limited geographic structure with nuclear microsatellites and AFLPs.

The absence of population genetic structure in Atlantic wolffish could be the result of extensive gene flow among populations or incomplete lineage sorting due to a recent colonisation of the North Atlantic following the Pleistocene glaciations. Although the distribution of Atlantic wolffish is continuous along the continental shelves from Europe to Greenland to North America, it is unlikely that movement or gene flow is extensive between distant populations or regions. Tagging studies have found only

occasional long-distance movements in wolffish, with no evidence of trans-Atlantic movement (Barsukov 1959; Templeman 1984). The sedentary nature of Atlantic wolffish, coupled with the weak isolation-by-distance (Figure 2.3), are consistent with the expectation of reduced gene flow in this species.

The Bayesian clustering analysis (Figure 2.1) identified three haplogroups that also occur in the statistical parsimony network (Figure 2.2), the principal coordinates analysis (Figure 2.4), and are partially supported by Bayesian and neighbour-joining trees (haplogroups A and B; Figures 2.5 and 2.6). The distribution of clades across trans-Atlantic populations is weakly but significantly non-random ($X^2 = 41.6$, $p = 0.027$), with the two smaller haplogroups (A and B) found predominantly in the western Atlantic (27/29 and 14/15 respectively). The scarcity of European samples in haplogroups A and B may either be due to the small European sample sizes or a North American origin for these groups (see Section 2.4.3 below). McCusker and Bentzen (2010a) found three groups – Atlantic Canada, North Atlantic, and Rockall Bank – which do not correspond to those found here. The microsatellite data suggest that Atlantic Canada forms one (or more) heterogeneous groups, the mid-Atlantic and Europe a second homogeneous group, and Rockall Bank a third isolated group. The identification of Rockall Bank as a separate population is likely due to the contemporary isolation of this location. The absence of diversity in RK, with the lowest mitogenomic diversity (Table 2.3) and no diversity in the two-fragment analysis (all samples are in haplotype *j*; Table 2.7 and Figure 2.2), question Rockall Bank as a source population, although additional samples are required to confirm this. The difference in pattern seen between the markers is likely due to the different mutation rates of the markers. While nuclear markers typically evolve at a slower rate

than mtDNA, hypervariable microsatellites can allow recent patterns to emerge more quickly (Zhang & Hewitt 2003; Zink & Barrowclough 2008; Brito & Edwards 2009).

The pattern seen in Atlantic wolffish is not unprecedented. Species typically show either complete panmixia, suggesting extensive gene flow and/or recent expansion from a single source population (Pulgarín-R & Burg 2012; Dohms & Burg 2013; Winkelmann *et al.* 2013; Jacobsen *et al.* 2014), or distinct separation into clades corresponding to population membership, suggesting low levels of gene flow and/or colonisation from multiple isolated sources (Byun *et al.* 1997; Lait *et al.* 2012; Lait & Burg 2013; Feutry *et al.* 2014). The pattern observed here of multiple clades that do not correspond closely to geographic location has been seen in a number of mitogenomic studies in both marine (Carr & Marshall 2008; Teacher *et al.* 2012; Carr *et al.* 2015) and terrestrial species (Wang *et al.* 2010). For example, Teacher *et al.* (2012) identified three distinct but widespread clades in 98 Atlantic herring in the Baltic region, while Carr *et al.* (2015) found six major clades among 52 harp seals from four discrete trans-Atlantic breeding and whelping grounds. In both cases the structure is explained by historical isolation pre-dating the postglacial expansion of the species. Isolation of three (or more) groups of Atlantic wolffish prior to the end of the last glacial maximum (LGM) likely explains the separation seen in this species. This is supported by the estimated divergence dates of 93,000 - 208,000 years. If we considered an earlier opening of the Bering Strait we would see a divergence time of 3 - 4 mya among the three wolffish species, pushing the divergence among haplogroups even further into the Pleistocene glaciations.

2.4.2 Whole-genome versus single-locus or two-fragment analyses

Mitogenomics has been extremely helpful in resolving both phylogenetic and phylogeographic questions that could not previously be determined by single-locus methods. It has proved useful in resolving evolutionary relationships (Inoue *et al.* 2001; Nardi *et al.* 2003; Coulson *et al.* 2006; Arnason *et al.* 2008), narrowing estimates of divergence times (Minegishi *et al.* 2005; Stone *et al.* 2010; Knaus *et al.* 2011), and revealing hidden phylogeographic structure (Carr & Marshall 2008; Gilbert *et al.* 2008; Feutry *et al.* 2014). Variation along the mitogenome is not evenly distributed, with some regions contributing little to the uncovering of structure, and others contributing diagnostic SNPs. For example the highly variable 16S rRNA region contains no sites fixed among the haplogroups while three of the eight changes in the ND3 coding region identify haplogroup A. We can therefore ask if the lack of structure in the single-locus analyses of Atlantic wolffish and the weak structure in the two-fragment study is a consequence of marker choice (see Coulson *et al.* 2006).

The analysis of each individual gene region showed the advantage of using multiple loci or complete mitogenomes in evaluating the population genetic structure of species, particularly those with low divergence rates or recent evolutionary histories. The markers commonly used in phylogeographic and population studies, CYTB and CR, as well as those used in the previous study on Atlantic wolffish, ND1 (and CR), do not identify the structure seen with complete mitogenomes (Supplementary Figure 2.2); CYTB and CR were each able to separate haplogroup B from A and B from C, whereas ND1 separated haplogroups A and B (together) from C. The only gene region that identified the three separate mitochondrial haplogroups was COX1, although membership

was not entirely concordant with the mitogenomic structure. Other mitogenomic studies have also found that the majority of gene regions did not provide sufficient information to resolve the underlying patterns of variation. For example, Feutry *et al.* (2014) were able to identify previously undiscovered structure in the speartooth shark *Glyphis glyphis*; when individual loci were analysed eight regions were invariable, two showed no structure, four showed partial structure, and only two (12S and ND5) supported the same pattern as the whole mitogenome.

Johnstone *et al.* (2007) identified the ND2, ND4, and CYTB genes as the most variable regions among the three wolffish species found in the North Atlantic Ocean, and the COX2, ND3, and ATP8 as the least. While ND2, ND4, and CYTB do have some of the highest numbers of variable sites (though 16S, COX1, and ND5 have the same or higher; Table 2.2), none of these three markers (nor the three combined) identify the haplogroup structure in Atlantic wolffish: ND2 shows no structure, and ND4 and CYTB are able to differentiate haplogroup B but not the other two (Supplementary Figure 2.2). The most variable marker among Atlantic wolffish, ND5, is only able to differentiate one of the haplogroups (A). In a study of 10 walleye pollock *Gadus chalcogrammus*, Yanagimoto *et al.* (2004) identified ND1, ND5, and CR as all being highly variable; however, when the polymorphisms between the Japan Sea and the Bering Sea were examined only a single fixed difference was found (in ND4). This shows that the number of variable sites is not a direct measure of a locus' ability to elucidate genetic structure, and that patterns of polymorphism among closely-related species do not necessarily reflect patterns within those species.

By identification of gene regions containing fixed differences between the three identified haplogroups we were able to devise a targeted multi-locus (two-fragment) study of the population genetic structure in 164 Atlantic wolffish. Focus on the two fragments that come from two of the most variable regions (COX1 and ND5) produced the same three-group pattern as that found in the mitogenomic analysis (Figure 2.2). In contrast, however, there were only three polymorphisms separating the three haplogroups, and there was no bootstrap support for any of the clades. This lack of resolution can lead to uncertainty regarding the validity of the clades, even though Φ_{ST} comparisons of the groups show significant differences (Table 2.8). The use of complete mitogenomes provides the statistical support required to confidently identify the three clades in this species; however, this could be combined with a targeted multi-locus approach to increase sample sizes at minimal expense. With additional loci or longer fragments the resolution may be improved..

2.4.3 Consequences of Pleistocene biogeography

During the last glacial maximum (ca. 20 kya) there was a large drop in both sea surface temperatures and sea levels which reduced the available marine habitat and disrupted the spatial distribution of marine species (Pielou 1991; Rohling *et al.* 1998; Hewitt 2000). On both sides of the North Atlantic Ocean, marine species either moved to more suitable regions, primarily to the south, or died off. There is evidence that some marine fish species may have survived in small northern cryptic refugia on either side of the Atlantic (Maggs *et al.* 2008; Provan & Bennett 2008). The pattern found in Atlantic wolffish is three distinct lineages with relatively recent divergence (93 - 220 kya) and no

strong biogeographic structure (Figures 2.2 - 2.6). This may be explained by a number of *a priori* scenarios:

(1) persistence in multiple refugia on both sides of the North Atlantic Ocean;

(2) use of multiple refugia in Europe followed by recolonisation of the western Atlantic;

(3) persistence in multiple refugia in the Northwest Atlantic, followed by recolonisation of the eastern Atlantic;

(4) occupation of a single glacial refugium by three already genetically divergent groups of wolffish; or

(5) use of a single glacial refugium by a genetically panmictic population, with subsequent diversification as three lineages following deglaciation.

In each case we would see multiple distinct groups; however, the patterns of diversity and the extent of divergence would differ. I now consider each of the five scenarios in turn.

If Atlantic wolffish survived the LGM on both sides of the Atlantic Ocean we would expect to see high levels of diversity in both eastern and western Atlantic populations. While haplotype diversity is high in all populations as most individuals have a unique genome sequence, nucleotide diversity is highest in NS, BS, and SEGB, and lowest in RK and the two most westerly populations (SS and SNL; Table 2.3). While the European populations have small sample sizes that may bias the diversity levels, and increased sampling is required to confirm the diversity patterns, the two-fragment analysis also supported lower diversity in the southern populations (Table 2.7). This pattern of diversity suggests that the wolffish likely survived the LGM in the eastern Atlantic rather than on both east and west. It is possible, however, that while the majority of the wolffish survived in European refugia, a remnant population survived in a small

periglacial refugium in Atlantic Canada and upon deglaciation this small refugial population recolonised North America with little movement into the mid- and eastern Atlantic. It has been suggested that the Flemish Cap may have acted as a marine glacial refugium during this time (Pflaumann *et al.* 2003; Shaw 2006), and it has been shown that small refugia can result in low diversity levels due to founder effect and persistent small population sizes (Maggs *et al.* 2008). Evidence supporting an Atlantic Canadian refugium is the fact that haplogroup A is found predominantly in the North American populations (Table 2.3 and Figure 2.1) and is only in two European individuals and no mid-Atlantic populations. These two samples could represent long-distance migration or colonisation events. It is also possible that the lack of haplogroup A individuals in Europe is due to incomplete sampling in these populations. Evidence suggests that the marine red algae *Palmaria palmata* survived the LGM in both multiple European refugia and in a cryptic North American refugium, possibly in Atlantic Canada (Provan *et al.* 2005). The small sample size does not, however, explain the lack of haplogroup A in the LABS population (n = 9), nor the single 'A' sample in the NENL population (n = 9). A similar situation could be suggested for haplogroup B, with most of the samples being found in North America and Greenland, and a single individual in the NS population. The pattern in haplogroup B suggests a second cryptic refugium in the northern region, perhaps near southern Greenland, as there are few 'B' individuals in the southwestern populations. The presence of a Greenland refugium is unlikely, and there is currently no evidence to support this suggestion. It is more likely that the distribution of haplogroup B is the result of migration from an eastern refugium.

The second scenario, that Atlantic wolffish persisted in multiple glacial refugia in the eastern Atlantic Ocean or Mediterranean Sea, is the most likely explanation given the observed pattern of diversity (Table 2.3) and distribution of haplotypes (Figures 2.2 - 2.6). The fact that the three haplogroups are found across all regions suggests that admixture occurred following prolonged isolation but before recolonisation of glaciated regions. Genetic diversity is high in the two European populations (BS and NS) despite low sample sizes (Table 2.3). This suggests that the source of the recolonisation may have originated in Europe. Seven putative glacial refugia have been suggested for marine species in the Northeast Atlantic: the Azores (mid-Atlantic west of Portugal), the Iberian Peninsula, the Mediterranean Sea, the English Channel, southwestern Ireland, Iceland and the Faroe Islands, and the Lofoten Coast of Norway (Maggs *et al.* 2008). While there is insufficient data to determine which of the periglacial refugia may have played a role in the survival of Atlantic wolffish, it is likely that three (or more) supported wolffish populations during the LGM. This pattern has been seen in a number of marine species including the green crab *Carcinus maenas* (Roman & Palumbi 2004), the red algae *Palmaria palmata* (Provan *et al.* 2005), and the brown algae *Fucus serratus* (Hoarau *et al.* 2007). Unlike the pattern seen in the above species, however, the haplogroup membership in Atlantic wolffish does not correspond to the geographical origin of the samples. Haplogroup C is widespread and diverse, while haplogroups A and B are located primarily in the western (A and B) and mid- (B) Atlantic Ocean. This suggests at least partial admixture of the three groups before recolonisation of the rest of Europe and the Northwest Atlantic. The fact that the diversity levels are lower in haplogroups A and B may be a result of smaller glacial populations followed by a leading edge founder effect.

A similar pattern is seen in yak *Bos grunniens* where there are three differentiated lineages with no correlation to geographical origin (Wang *et al.* 2010). In this case the authors suggested that the variation had arisen in isolated Pleistocene refugia with a subsequent reunion into a single gene pool following deglaciation.

The persistence of Atlantic wolffish in multiple western Atlantic refugia is unlikely for two reasons. The genetic structure in the wolffish suggests three source populations; however, there is currently evidence supporting only two main North American refugia: one in or near Atlantic Canada, and a large refugium in the south, possibly off the southeast coast of the United States (Pflaumann *et al.* 2003; Shaw 2006). We have seen above that haplogroup A may have survived in an Atlantic Canada refugium, while haplogroup B could suggest a previously unidentified Greenland refugium. This would leave haplogroup C persisting in the large southern refuge. The pattern of diversity does not support this scenario; diversity is higher in Europe than in North America despite the higher sample size in the western Atlantic, and within North America diversity is lowest in the two southern populations. It has also been suggested that the southern refugium did not have a solid substrate which is required for the feeding habits of Atlantic wolffish, and therefore despite being unglaciated this region may not have provided suitable habitat for this species (Maggs *et al.* 2008). While there is evidence of some marine species surviving in multiple western Atlantic refugia, for example the hermit crab *Pagurus longicarpus* likely survived predominantly in the south with a small refugial population in Atlantic Canada (Young *et al.* 2002), this is not reflected in the distribution of haplotypes nor in the diversity values for Atlantic wolffish.

The final two scenarios, the use of a single glacial refugium either by three divergent groups or by a panmictic population that diverged following the LGM, can both be ruled out. If the separation of the three haplogroups had occurred before the last glaciation we would expect to see much greater divergence among the haplogroups with a higher number of fixed differences (Figure 2.2). Divergence time estimates place the separation of the three clades from 93 - 208 kya, well within the last glacial cycle. This supports the idea that separation occurred during the last glaciation – either in separate refugia or as separate pockets within a single refugium (essentially forming multiple refugia). If a panmictic population had survived in a single glacial refugium, and separated after deglaciation we would expect to see a strong correlation between haplogroup membership and population of origin.

2.4.4 Conclusions

The Atlantic wolffish is a sedentary marine species that occurs along the continental shelves across the North Atlantic Ocean. Its life history characteristics suggest *a priori* a strong population structure, with little or no movement during either larval or adult stage; however, this has been shown not to be the case. Examination of 84 complete mitogenomes from 14 sampling locations, supplemented by an additional 80 two-fragment (1,690 bp) sequences, instead shows an absence of population genetic structure across the range of Atlantic wolffish, although there is evidence of isolation-by-distance instead of uniform panmixia. We identified three distinct haplogroups supported by traditional population statistics, distance methods, and Bayesian analyses. The three haplogroups showed no significant association with geographical origin of samples,

suggesting historical isolation with subsequent admixture. This result is consistent with previous studies of nuclear gene loci over a similar range (McCusker & Bentzen 2010a). The use of the complete mitochondrial genome has allowed greater resolution of structure than seen in any single-locus or multi-locus analyses of wolffish to date.

Acknowledgements

We would like to thank Fisheries and Oceans Canada for providing the tissue samples, and everyone involved in the extraction and amplification of the samples: Dr. HD Marshall, Dr. KA Johnstone, Dr. AT Duggan, Dr. SMC Flynn, L MacDonald, E Wells, O Trela, and A Genge. We thank Dr. P Bentzen and Dr. MR McCusker for providing Atlantic wolffish samples from their study. This project was supported by a Natural Science and Engineering Research Council Discovery Grant to SMC, and an NSERC Alexander Graham Bell Canada Graduate Scholarship to LAL.

Table 2.1. Primers used for PCR amplification of Atlantic wolffish. Primer pairs are given; in some cases g03F was used with g04R (T_A = 54°C) and w17F with wcytbR (T_A = 54°C). Primers were designed for *Gadus morhua* (g) or for *Anarhichas lupus* (w or str). Annealing temperatures (T_A), location of 5'-most base of the primer, and primer sources (ref) are given.

Primer Name	Sequence (5' to 3')	T_A (°C)	loc	ref
g01F	CTGAAGATATTAGGATGGACCCTAG	49	29	1
g02R	CTATTCATTTCACAGGCAACCAGCT	49	1,498	1
g03F	ACCCCGAAACTGAGCGAGCTACTCC	55	1,357	1
g03R	TAAGCCCTCGTGATGCCATTCATAC	55	2,175	1
g04F	TTTACCAAAAACATCGCCTCTTG	49	2,018	1
g04R	TGAACCTCTGTAGAAAGGGCTTAGG	49	2,841	1
w04F	TTCAGACCGGAGTAATCCAGGTCAG	56	2,617	2
g05R	ATGTTCGGGGTATGGGCCCCAAGAGC	56	4,020	1
w06F	GTGCTTCCACTACACCACTTCCTAG	54	3,954	2
w07R	CTGGTTTGAGCGCTTAGCTGTTAAC	54	5,292	2
g07F	AAACTAGACCAAGGGCCTTCAAAGC	55	5,106	1
w08R	CAGAGGTAAAGTAAGCGCGTGTGTC	55	6,400	2
w08F1	CCCTTCACCTAGCAGGAATTTCTTCAATCC	53	5,924	2
g08R	TAACCCACAATTCTGCCTTGACAAG	53	7,158	1
w08F	ACAACGAATGTGGAGTGACTACACG	49	6,943	2
g09R	ACCCATATTAGCTTCTTAGTGAGG	49	7,893	1
w09F	GGCCATCAGTGGTACTGAAGCTATG	49	7,482	2
g10R	AGAGGGCGAATGAATAAACTAATTG	49	8,569	1

w10F1	ATGCGAAACCAACCAACCCATGCTC	58	8,457	2
w11R	TTGATCTCCTCAGGGTAGCGGGAGTAGTAG	58	9,887	2
w11F	CTACACTTGACCCATTTGAAGTGCC	56	9,153	2
w12R	TGTAGACCCTAGTTGCAAGGTCCAC	56	10,224	3
w12F	GCTAGGATTAACCGGCCTGGCCTTTCAT	54	10,101	3
wND43R	CGATGAGCGACTTCAAATCTGTGTG	54	11,214	3
wND44F	TAGACCCATTAACCCAGGAGCTAAG	53	11,104	3
wND45R	AAGACCAGCGGATGAGCTGTTATCC	53	11,908	3
w14F2	GAACATCTTCTTATGGCCCTTCACC	53	11,652	2
Str14R	GGTTGTAGACGACGGCTTGAAGAGC	53	12,488	3
w15F	TCGCCATAGTCATTCTAGTGACAGC	53	12,339	2
w16R	GCTTGTTTGTTAGGGAGGCTAGTTC	53	13,496	2
w16F	CCACAGCTTGAATGACGAGCAAGAC	54	13,006	2
w17R	TAACGCGAGGATTAAGTCGAGGAAC	54	13,823	2
w17F	TTTACCACTCCACCACTTCTCCAAC	54	13,516	2
w18R	AGCAAAGGCCGAGTAGGGAACCAAAGTTTC	54	14,499	2
w18F	CCGCTACAACAACCAACCCTAAAGC	55	14,196	2
wcytbR	GTTGTCAACTGAGAAGCCTCCTCAG	55	14,894	2
g19F	GAGGAGGTTTCTCAGTAGATAATGC	49	14,872	1
g19R	GTTTAATTTAGAATTCTAGCTTTGG	49	15,651	1
w19F	TGAATTGGCGGTATACCCGTAGAAC	54	15,387	3
g20R	GGCAGGACATTAAGGGCATTCTCAC	54	160	1

1 (Coulson *et al.* 2006)

2 (Johnstone *et al.* 2007)

3 This study

Table 2.2. Distribution of variable sites across the complete mitochondrial genome of 84 Atlantic wolffish. Total numbers of variable sites (VS), parsimony-informative sites (PI), transitions (Ts), transversions (Tv), and insertion-deletions (ID) are given. For coding regions: coding position 1, 2, or 3, and status as synonymous (S) or non-synonymous (NS) substitutions, are indicated.

Region	**VS**	**PI**	**Ts**	**Tv**	**ID**	**1**	**2**	**3**	**S**	**NS**
12S	14	8	10	4	0	n/a	n/a	n/a	n/a	n/a
16S	24	12	20	2	2	n/a	n/a	n/a	n/a	n/a
ND1	15	5	13	2	0	4	0	11	11	4
ND2	22	9	20	2	0	2	3	17	17	5
COX1	25	8	21	4	0	1	0	24	23	2
COX2	9	5	9	0	0	0	0	9	9	0
ATP8	1	0	1	0	0	0	0	1	1	0
ATP6	12	5	11	1	0	0	0	12	12	0
COX3	8	4	8	0	0	1	0	7	8	0
ND3	8	4	8	0	0	2	0	6	6	2
ND4L	4	2	3	1	0	3	0	1	2	2
ND4	24	7	21	3	0	7	5	12	12	12
ND5	41	22	37	4	0	8	2	31	32	9
ND6	5	1	4	1	0	1	0	4	4	1
CYTB	19	9	16	3	0	3	1	15	16	3
tRNA	17	6	15	1	1	n/a	n/a	n/a	n/a	n/a
CR	6	4	6	1	0	n/a	n/a	n/a	n/a	n/a
other	3	0	2	0	1	n/a	n/a	n/a	n/a	n/a
Total	**257**	**111**	**224**	**29**	**4**	**32**	**11**	**150**	**153**	**40**

Table 2.3. Sample size (n), number of haplotypes (h), haplotype diversity (H_d), nucleotide diversity (π), and assignment to haplogroup (A - C) for 84 complete mitochondrial genomes among 14 sampling locations of Atlantic wolffish. Refer to Figure 2.1 for locations.

Population	**n**	**h**	**H_d**	**π**	**A**	**B**	**C**
LABC	1	1	1.000	n/a	1	0	0
LABS	9	9	1.000	0.00104	0	3	6
NENL	9	9	1.000	0.00108	1	1	7
NGB	12	11	0.985	0.00095	8	2	2
SEGB	13	13	1.000	0.00122	4	2	7
SWGB	10	10	1.000	0.00101	4	2	4
SNL	10	7	0.867	0.00069	7	1	2
SS	3	3	1.000	0.00069	2	0	1
WG	3	3	1.000	0.00101	0	2	1
EG	3	3	1.000	0.00085	0	1	2
IC	3	3	1.000	0.00077	0	0	3
NS	3	3	1.000	0.00149	1	1	1
RK	3	2	0.667	0.00065	0	0	3
BS	2	2	1.000	0.00121	1	0	1
A	29	21	0.911	0.00044			
B	15	14	0.990	0.00072			
C	40	39	0.999	0.00082			
Total	**84**	**74**	**0.989**	**0.00106**	**29**	**15**	**40**

Table 2.4. Population pairwise Φ_{ST} values (below diagonal) and corresponding p-values (above diagonal) based on 100,172 permutations for complete mitogenomes from (a) six North American populations of Atlantic wolffish, and (b) three identified haplogroups (A - C). Significant p-values after modified FDR and sequential Bonferroni corrections (P_{crit} = 0.0151, P_{crit} = 0.0102, respectively) are shown in bold. Refer to Figure 2.1 for locations.

(a)	**LABS**	**NENL**	**NGB**	**SEGB**	**SWGB**	**SNL**
LABS	*	0.1174	**0.0036**	0.4275	0.1646	**0.0027**
NENL	0.0366	*	0.0306	0.3133	0.3693	**0.0102**
NGB	0.1530	0.0966	*	0.1448	0.4377	0.4346
SEGB	0.0013	0.0058	0.0273	*	0.8678	0.0936
SWGB	0.0400	0.0062	0.0000	0.0000	*	0.1533
SNL	0.2355	0.1891	0.0000	0.0527	0.0425	*

(b)	**A**	**B**	**C**
A	*	**<0.0001**	**<0.0001**
B	0.6084	*	**<0.0001**
C	0.4610	0.4014	*

Table 2.5. Distribution of variable sites in the two-fragment study of Atlantic wolffish. Total numbers of variable sites (VS), parsimony-informative sites (PI), transitions (Ts), and transversions (Tv) are given. For coding regions: coding position 1, 2, or 3, and amino acid change for any non-synonymous substitutions (Sub), are given.

Region	**Length (bp)**	**VS**	**PI**	**Ts**	**Tv**	**1**	**2**	**3**	**Sub**
COX1									
- COX1	903	21	9	19	2	1	0	20	Pro ↔ Ser
ND5									
- ND4	46*	1	1	1	0	1	0	0	Val ↔ Ile
- tRNA His	69*	1	0	1	0				
- tRNA Ser	67	2	0	2	0				
- other	4	1	0	1	0				
- tRNA Leu	73	0	0	0	0				
- ND5	532	10	5	9	1	4	1	5	Val ↔ Leu Val ↔ Ala Val ↔ Ile His ↔ Tyr
Total	**1,690**	**36**	**15**	**32**	**3**	**6**	**1**	**32**	

* ND4 and His overlap by 5 bp

Table 2.6. Distribution of shared and unique haplotypes among 14 Atlantic wolffish populations for the two-fragment study, with sample size (n) and number of haplotypes (h). The populations are grouped as North America (NA), mid-Atlantic (mid), and eastern Atlantic (Eur). The haplotypes are grouped by haplogroup membership, with ***a*** **-** ***e*** falling in haplogroup A, ***f*** **-** ***h*** in haplogroup B, and ***i*** **-** ***l*** in haplogroup C. Refer to Figure 2.1 for locations.

Haplogroup	**Haplotype**	**NA**								**mid**			**Eur**			
		LABC	**LABS**	**NENL**	**NGB**	**SEGB**	**SWGB**	**SNL**	**SS**	**WG**	**EG**	**IC**	**NS**	**RK**	**BS**	**TOTAL**
A	***a***			1											1	**2**
	b				2								1			**3**
	c	1	2	6	9	9	13	10	2							**52**
	d				1		1									**2**
	e		1			1										**2**
B	***f***		1		1	2		1								**5**
	g		2	1	3					1						**7**
	h		3	1	1		1									**6**
C	***i***		5	2		1	1						1		1	**11**
	j		6	2	2	3	1	2	1		2	1		3		**23**
	k		2	6	3	4	5	2				1				**23**
	l			1			1									**2**
# unique		1	4	5	1	5	4	1	0	2	1	1	1	0	0	**26**
h		2	12	13	9	11	11	5	2	3	2	3	3	1	2	**38**
n		**2**	**26**	**25**	**23**	**25**	**27**	**16**	**3**	**3**	**3**	**3**	**3**	**3**	**2**	**164**

Table 2.7. Sample size (n), number of haplotypes (h), haplotype diversity (H_d), nucleotide diversity (π), and distribution among haplogroups (A - C) for the two-fragment dataset across 14 populations of Atlantic wolffish. Refer to Figure 2.1 for locations.

Population	**n**	**h**	**H_d**	**π**	**A**	**B**	**C**
LABC	2	2	1.000	0.00118	2	0	0
LABS	26	12	0.905	0.00150	4	8	14
NENL	25	13	0.893	0.00184	7	4	14
NGB	23	9	0.826	0.00155	13	5	5
SEGB	25	11	0.847	0.00153	11	2	12
SWGB	27	11	0.749	0.00169	14	2	11
SNL	16	5	0.608	0.00126	10	2	4
SS	3	2	0.667	0.00118	2	0	1
WG	3	3	1.000	0.00158	0	2	1
EG	3	2	0.667	0.00118	0	1	2
IC	3	3	1.000	0.00118	0	0	3
NS	3	3	1.000	0.00276	1	1	1
RK	3	1	0.000	0.00000	0	0	3
BS	2	2	1.000	0.00296	1	0	1
A	65	9	0.360	0.00035			
B	27	12	0.869	0.00109			
C	72	17	0.780	0.00082			
Total	**164**	**38**	**0.855**	**0.0017**	**65**	**27**	**72**

Table 2.8. Population pairwise Φ_{ST} values (below diagonal) and corresponding p-values (above diagonal) based on 100,172 permutations for the two-fragment study from (a) six North American populations, and (b) three identified haplogroups (A - C). Significant values after modified FDR correction (P_{crit} = 0.0151) are shown in bold (for sequential Bonferroni correction P_{crit} = 0.0038). Refer to Figure 2.1 for locations.

(a)	**LABS**	**NENL**	**NGB**	**SEGB**	**SWGB**	**SNL**
LABS	*	0.1022	**0.0013**	**0.0145**	**0.0041**	**0.0008**
NENL	0.0290	*	0.0364	0.4193	0.2412	0.0327
NGB	0.1443	0.0651	*	0.1683	0.4033	0.8278
SEGB	0.0781	0.0000	0.0210	*	0.5610	0.2146
SWGB	0.1107	0.0100	0.0000	0.0000	*	0.4012
SNL	0.1961	0.0896	0.0000	0.0188	0.0000	*

(b)	**A**	**B**	**C**
A	*	**<0.0001**	**<0.0001**
B	0.5462	*	**<0.0001**
C	0.7470	0.7283	*

Supplementary Table 2.1. Overall Φ_{ST} and p-values (ns = not significant) for the AMOVAs on individual loci for the complete (n = 84) and reduced (n = 63) datasets. The number of significant population pairwise comparisons (SP), after correction for multiple tests, and which populations are involved (pop), are given for the reduced dataset.

	n = 84		**n = 63**			
Region	**Φ_{ST}**	**p**	**Φ_{ST}**	**p**	**SP**	**pop**
12S	0.074	ns	0.060	ns	3	SNL
16S	0.036	ns	0.000	ns	0	n/a
ND1	0.073	ns	0.072	< 0.05	1	LABS
ND2	0.064	ns	0.000	ns	0	n/a
COX1	0.071	ns	0.095	< 0.01	5	LABS NENL
COX2	0.008	ns	0.000	ns	0	n/a
ATP8	0.216	ns	0.000	ns	0	n/a
ATP6	0.076	ns	0.064	ns	1	SNL
COX3	0.210	< 0.001	0.158	< 0.01	2	LABS
ND3	0.160	< 0.01	0.126	< 0.05	1	LABS SNL
ND4L	0.054	ns	0.000	ns	0	n/a
ND4	0.024	ns	0.000	ns	0	n/a
ND5	0.117	< 0.001	0.103	< 0.01	4	LABS NENL
ND6	0.000	ns	0.001	ns	0	n/a
CYTB	0.022	ns	0.000	ns	0	n/a
CR	0.000	ns	0.000	ns	0	n/a

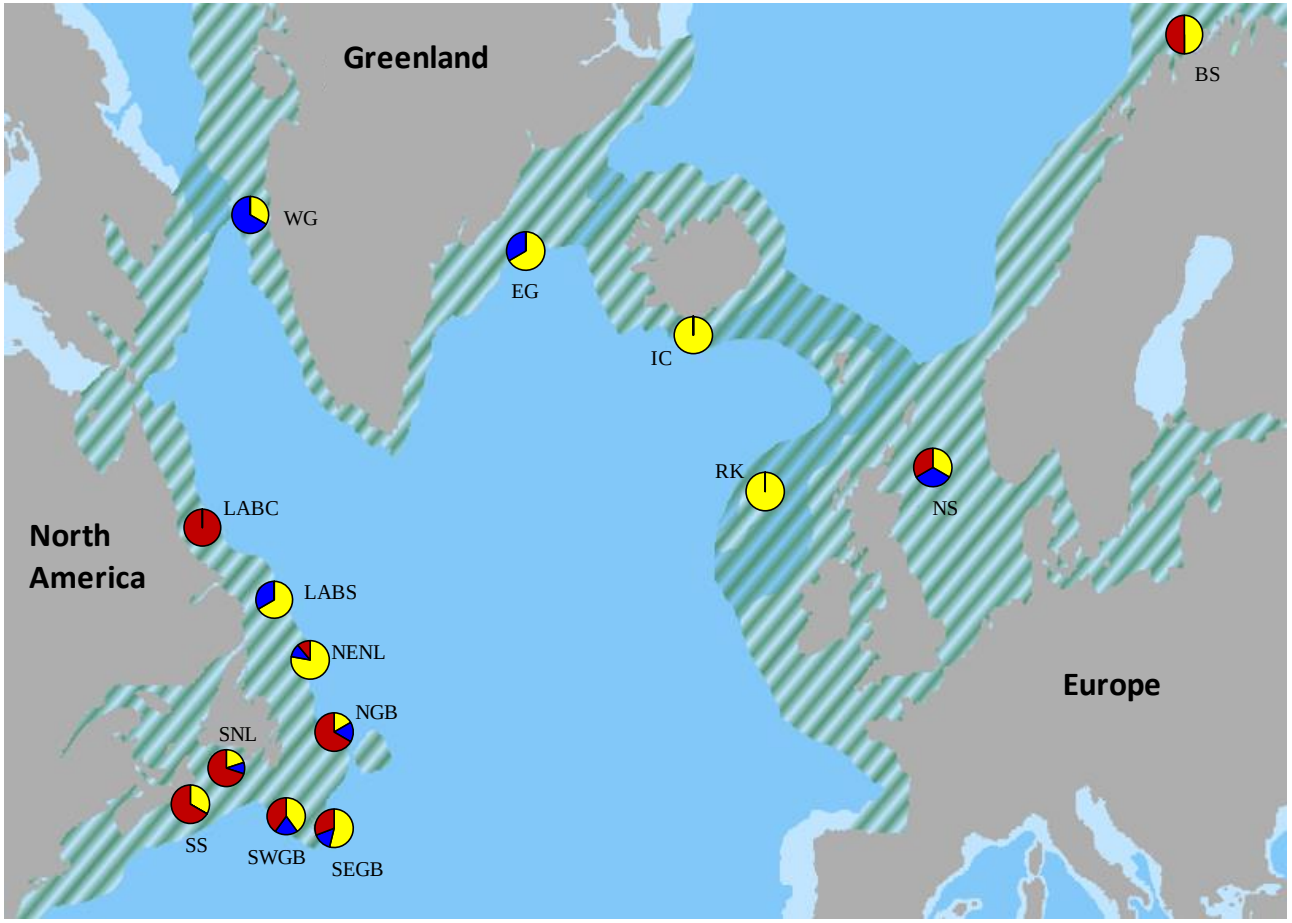
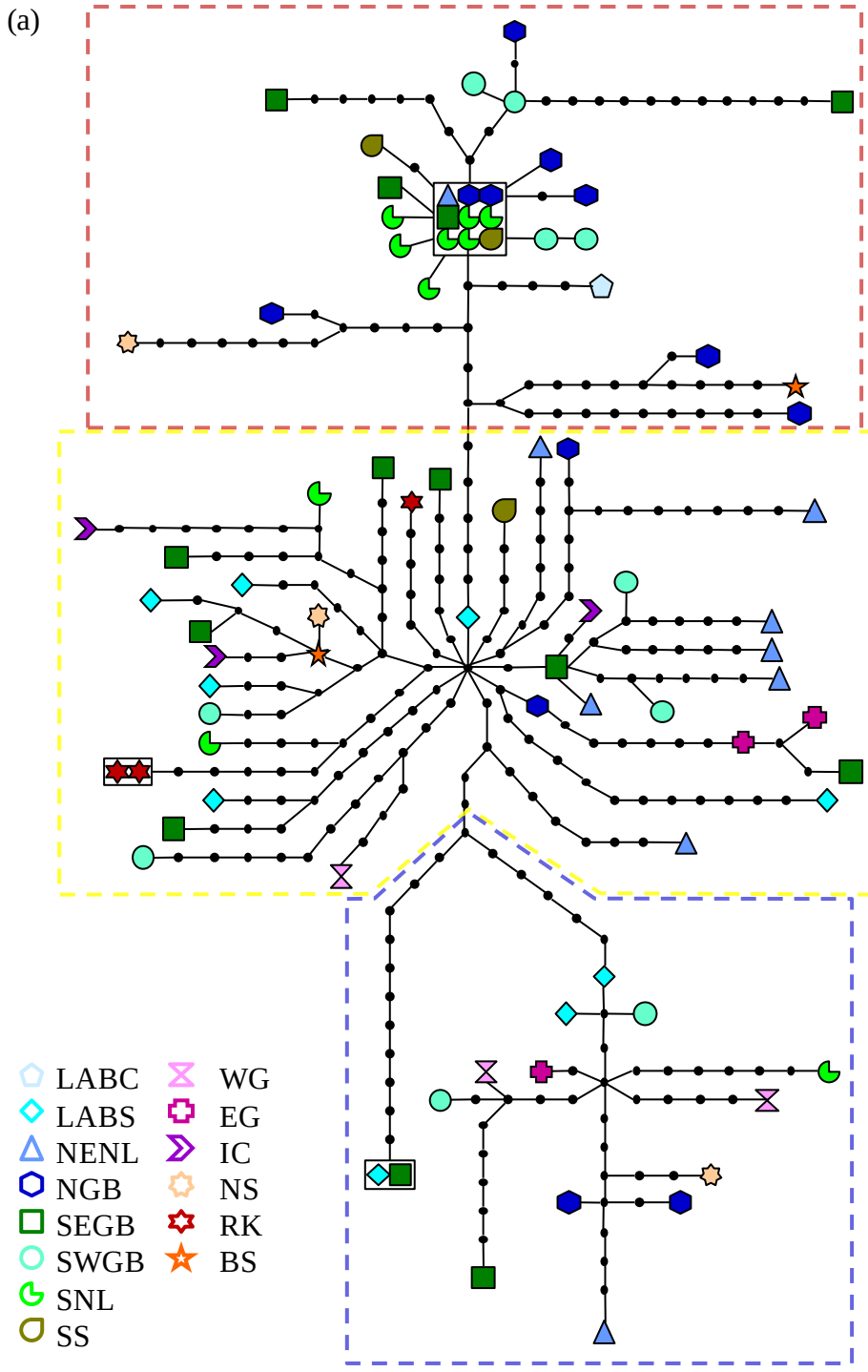


Figure 2.1. Distribution of Atlantic wolffish in the North Atlantic (shaded) and the 14 sampling locations. Pie charts show the geographic distribution of the three haplogroups identified in BAPS v6: A (red), B (blue), and C (yellow). Figure modified from FishBase (2013).


(a)
LABC
LABS
NENL
NGB
SEGB
SWGB
SNL
SS
WG
EG
IC
NS
RK
BS

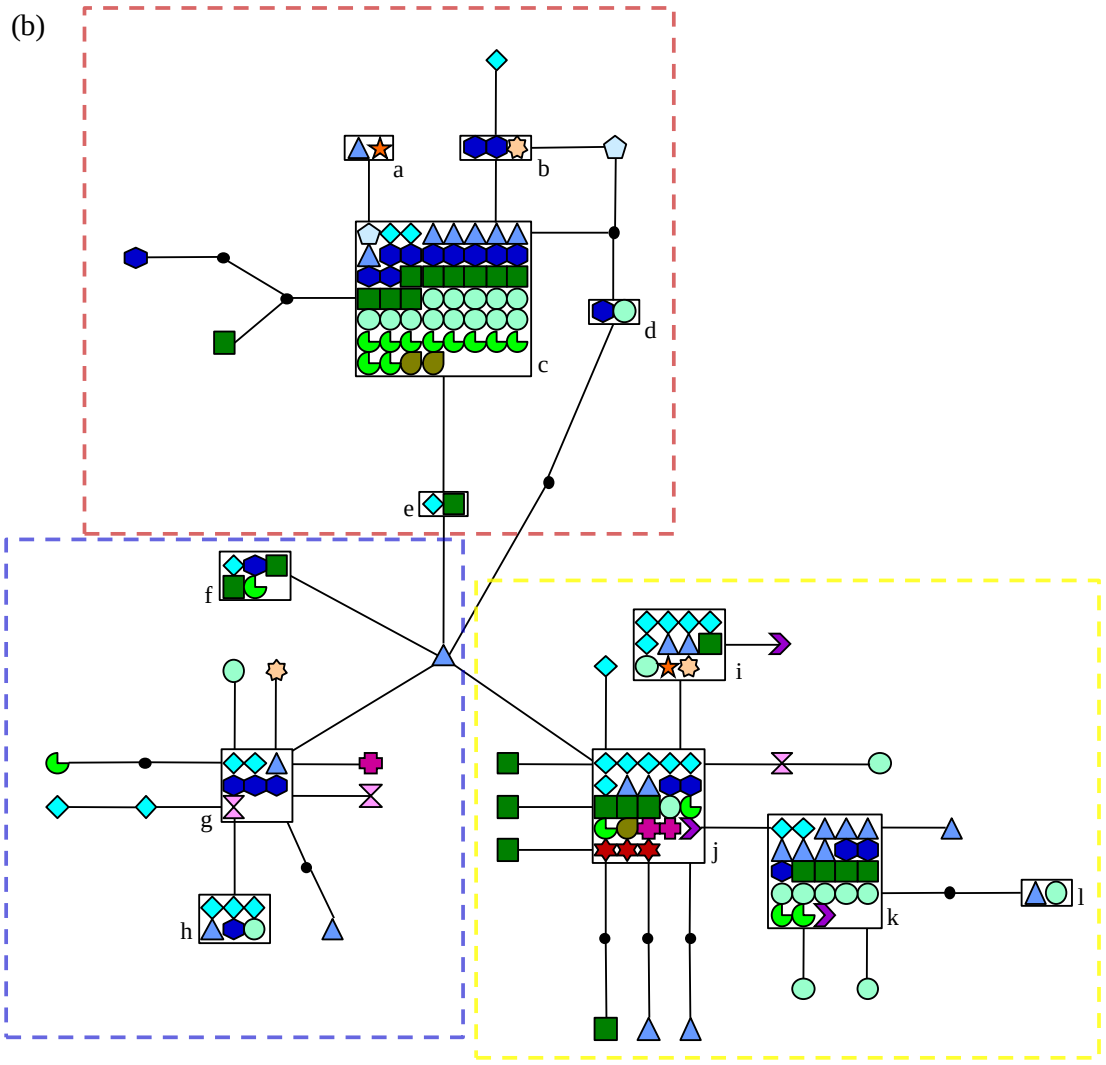


Figure 2.2. Statistical parsimony network of (a) complete mitogenomes and (b) the two-fragment dataset for Atlantic wolffish. Each symbol represents an individual, samples are colour-coded by sampling location, and shared haplotypes are encased by black boxes. The black dots are inferred or unsampled haplotypes, and each connection represents one nucleotide change. The dashed boxes correspond to the haplogroups (A - C) found in BAPS v6. Refer to Figure 2.1 for locations.

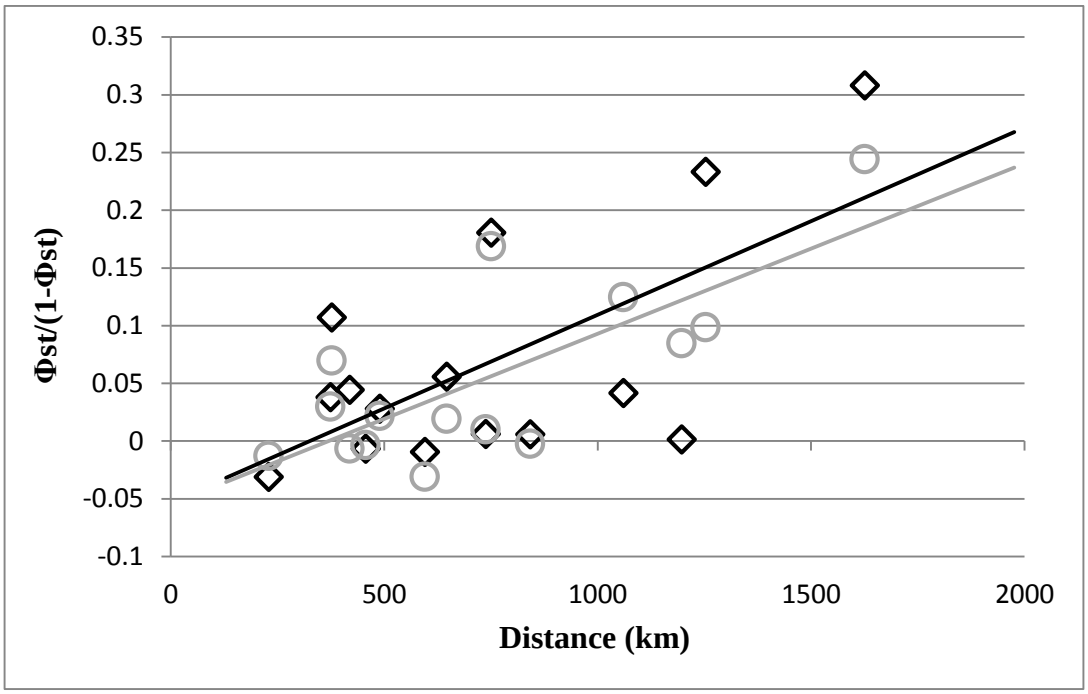


Figure 2.3. Mantel test of isolation-by-distance showing the association between genetic distance ($\Phi_{ST}/ (1-\Phi_{ST})$) and shortest marine swimming distance (km), for the mitogenome (◊) and the two-fragment analysis (○). The mitogenome analysis shows a weak correlation ($r^2 = 0.4179$) as compared with the two-fragment analysis ($r^2 = 0.561$).

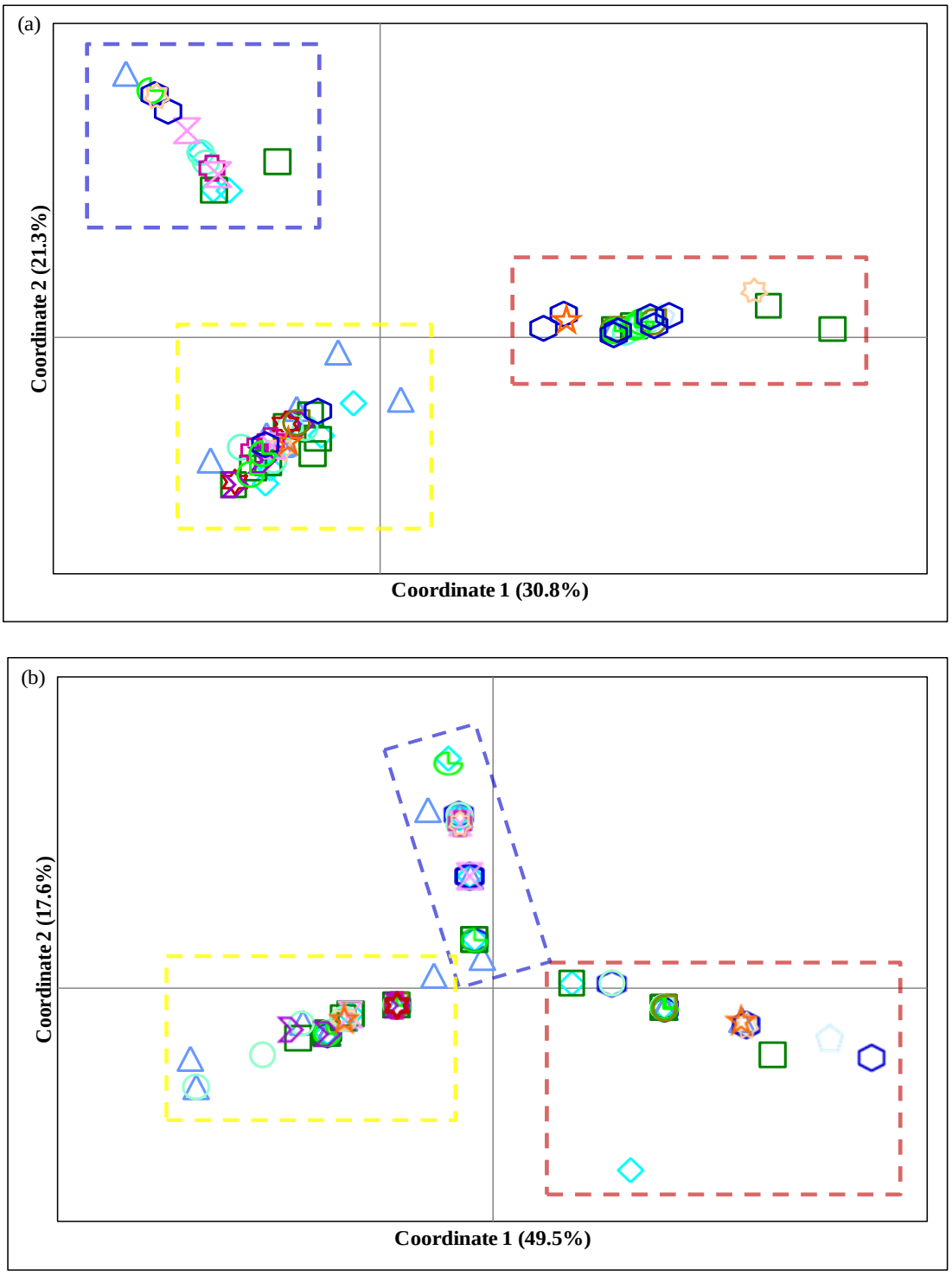


Figure 2.4. Principal coordinates analysis (PCoA) of (a) complete mitogenomes and (b) the two-fragment dataset for Atlantic wolffish. Sampling locations are coded as per Figure 2.2, and dashed boxes outline haplogroups (A - C) identified by BAPS v6.

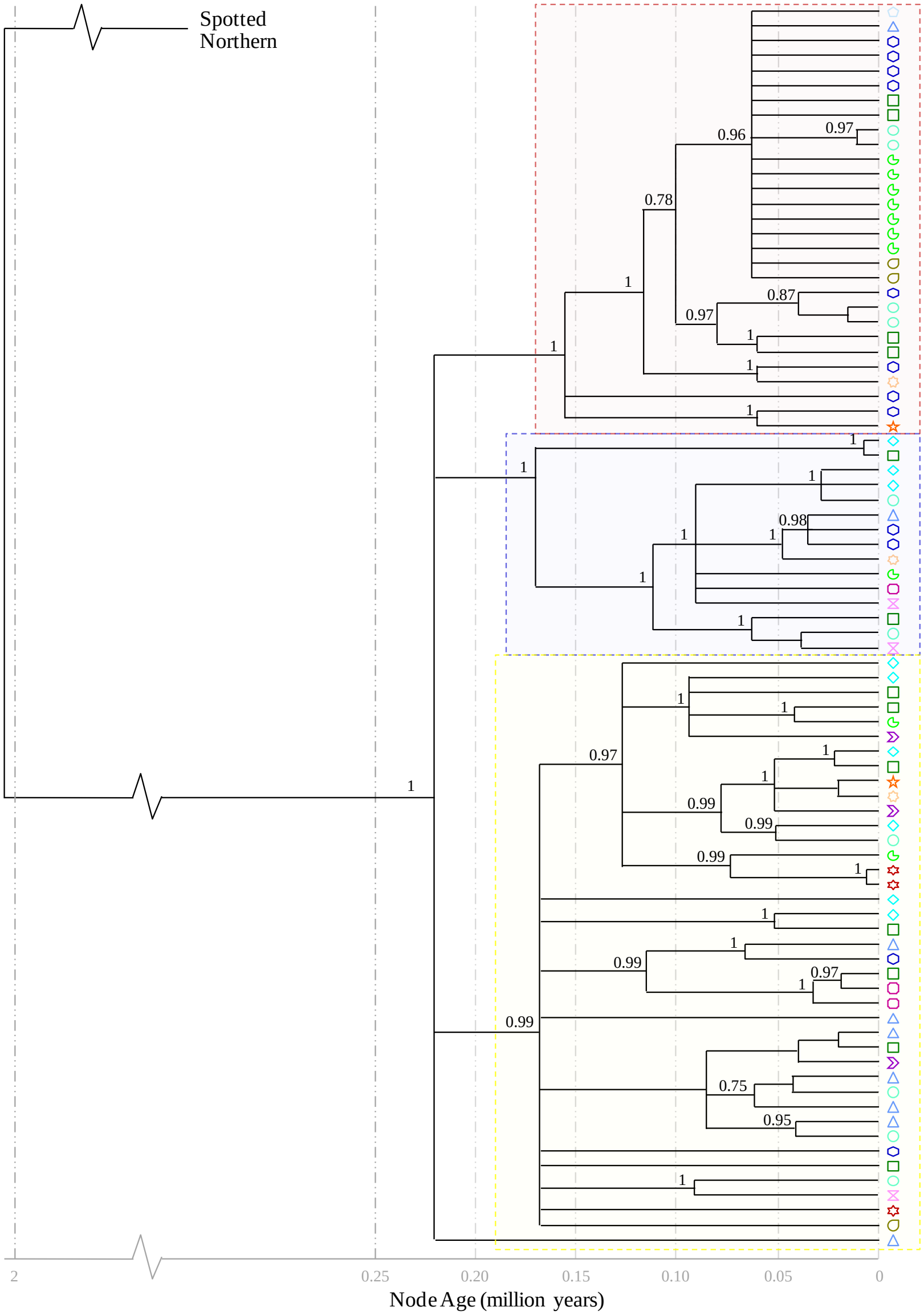


Figure 2.5. Clock-calibrated Bayesian analysis of complete mitogenomes from Atlantic wolffish. Posterior probabilities above 0.75 are given, all others fell between 0.5 and 0.749. The x-axis shows estimated separation time in millions of years. Sampling locations are coded as in Figure 2.2, and dashed boxes outline haplogroups identified by BAPS v6. The rooting is the same when either of the other two Northwest Atlantic species is used.

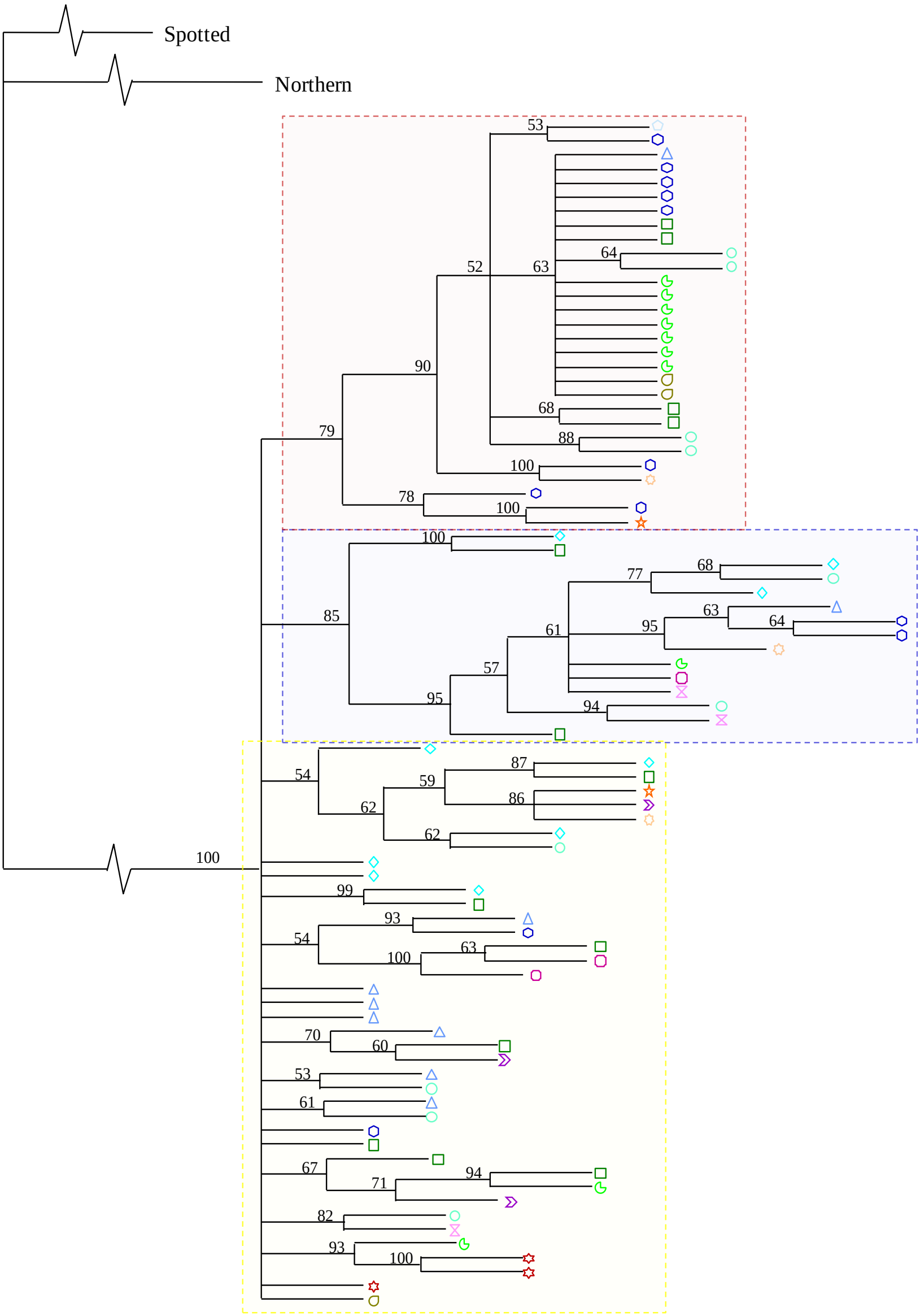


Figure 2.6. Neighbour-joining analysis of complete mitogenomes from Atlantic wolffish. Bootstrap values > 50% are given (10,000 replicates). Sampling locations are coded as in Figure 2.2; dashed boxes outline haplogroups identified by BAPS v6. The rooting is the same when either of the other two Northwest Atlantic species is used.

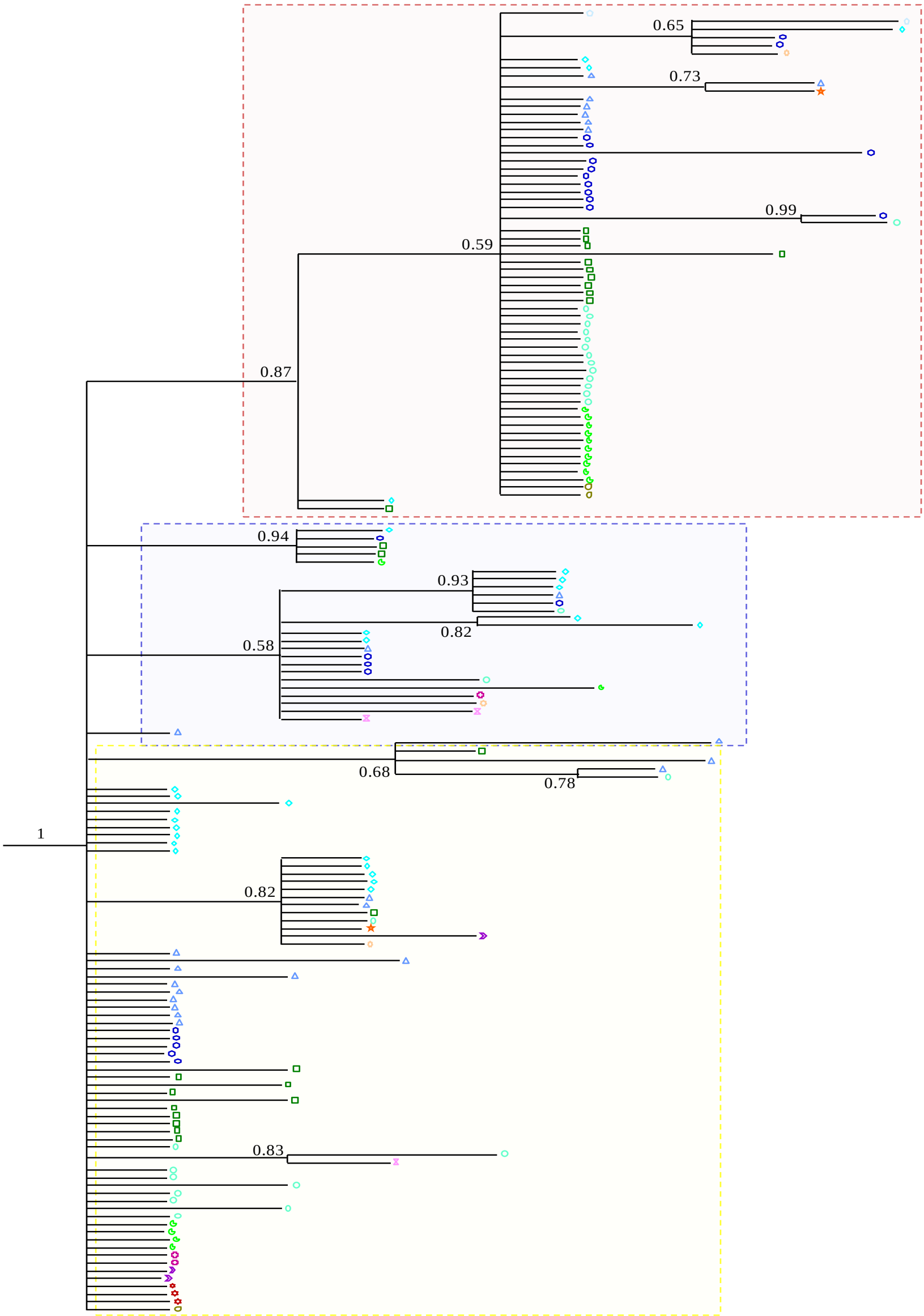


Figure 2.7. Bayesian analysis of the two-fragment dataset of Atlantic wolffish. Posterior probabilities above 0.5 are given. Sampling locations are coded as in Figure 2.2; dashed boxes outline haplogroups identified by BAPS v6.

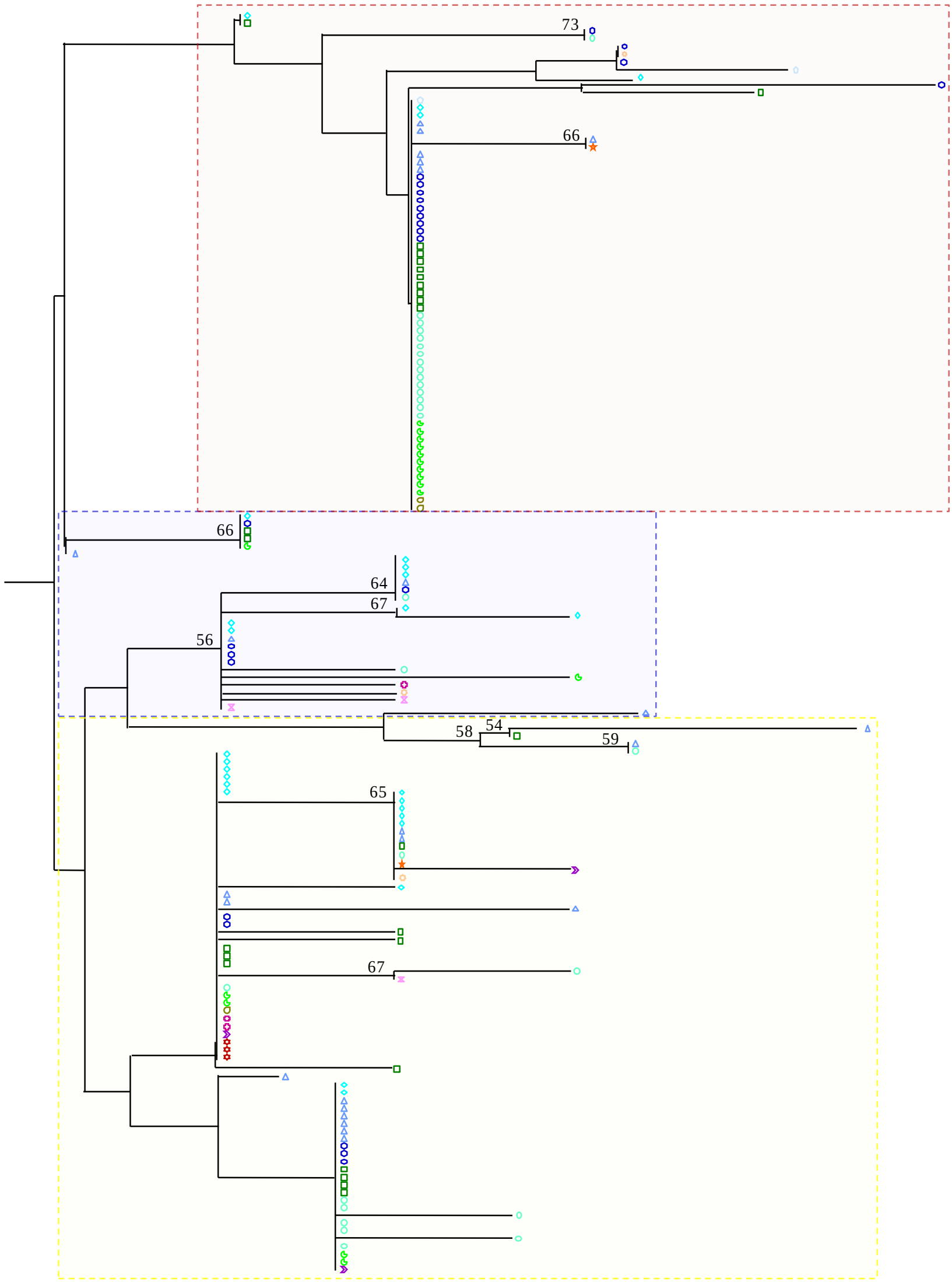
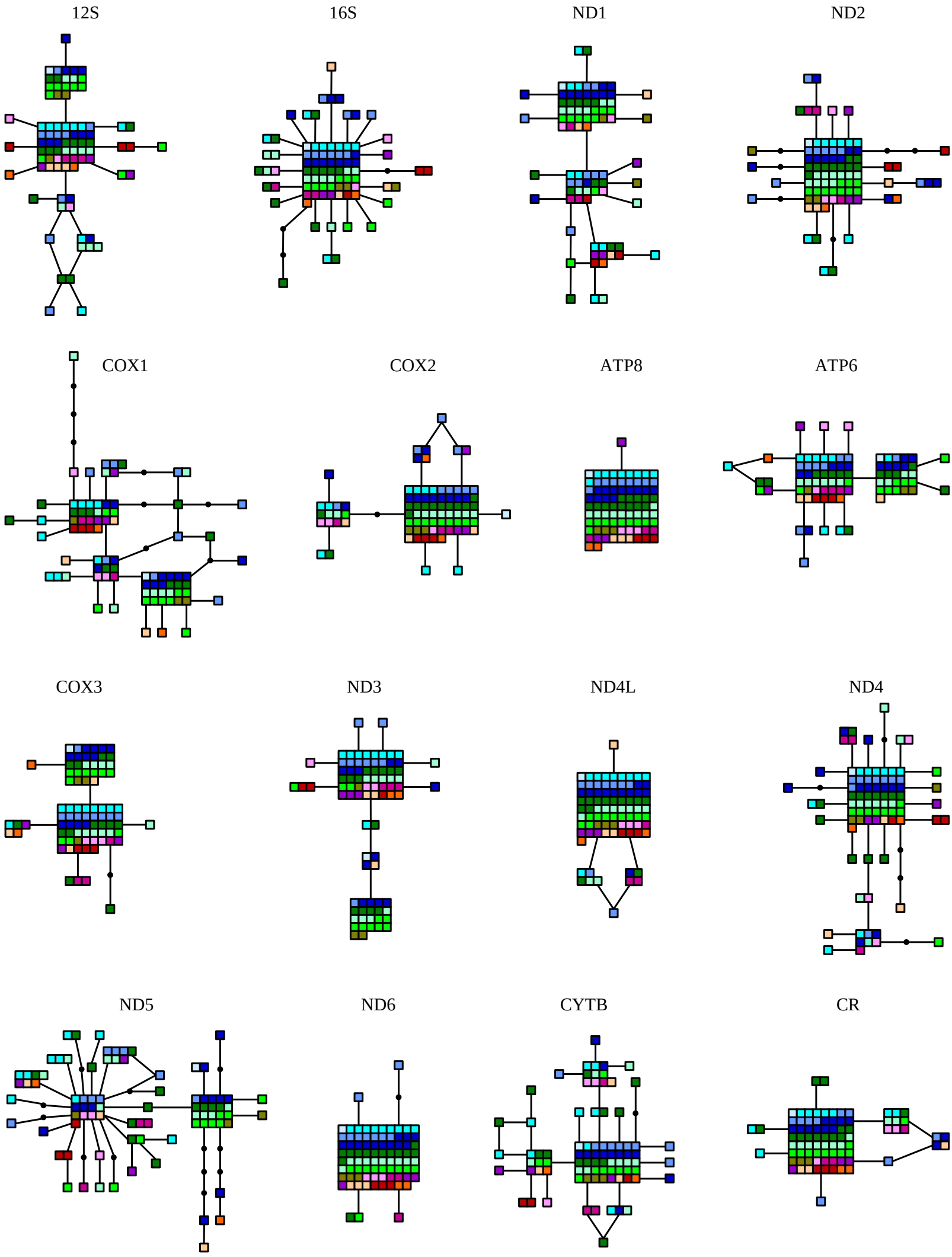
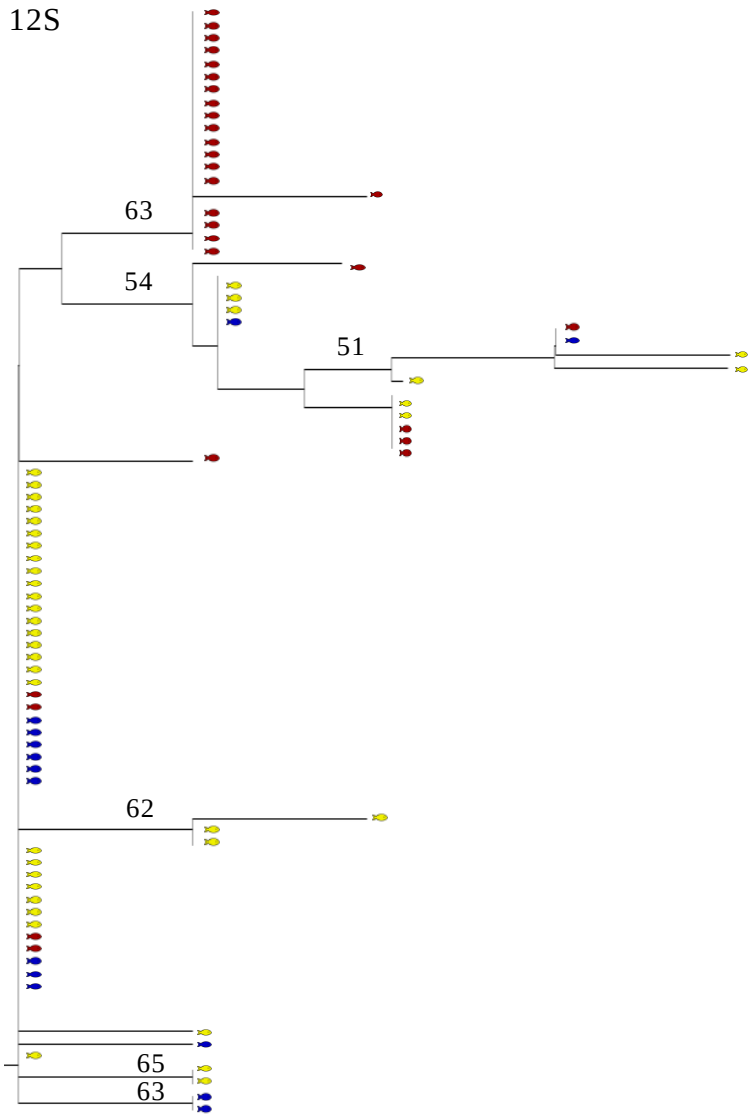
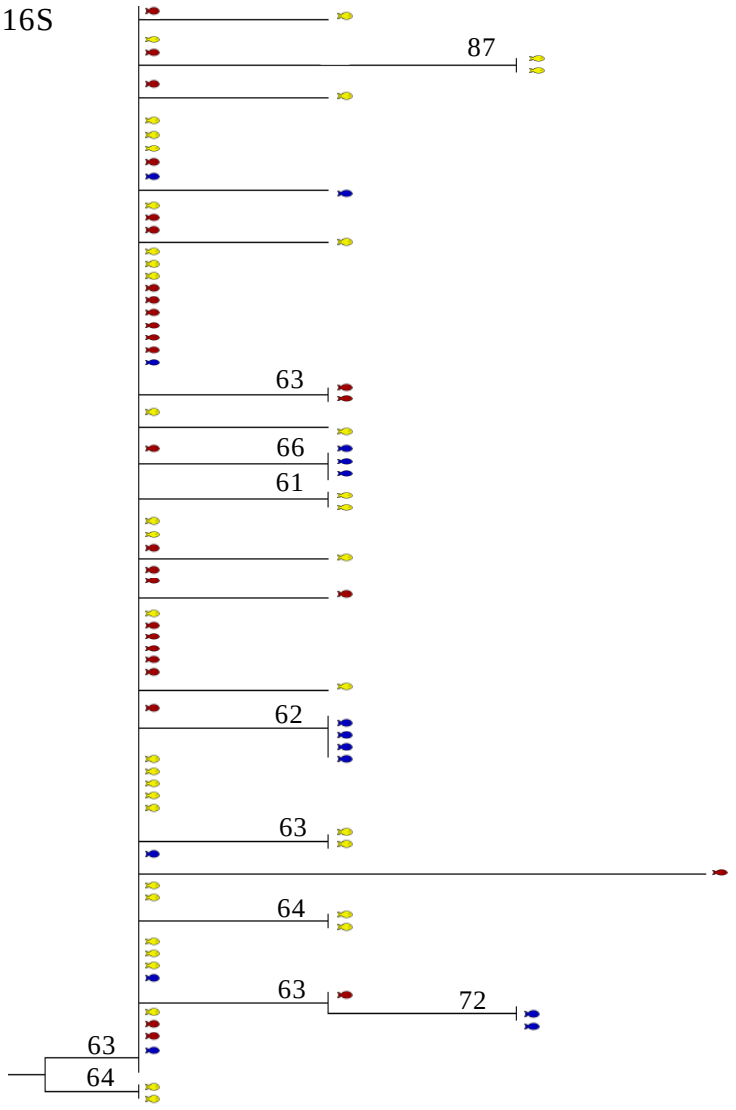


Figure 2.8. Neighbour-joining analysis of the two-fragment dataset of Atlantic wolffish. Bootstrap values > 50% are given (10,000 replicates). Sampling locations are coded as in Figure 2.2; dashed boxes outline haplogroups identified by BAPS v6. The rooting is the same when either of the other two Northwest Atlantic species is used.

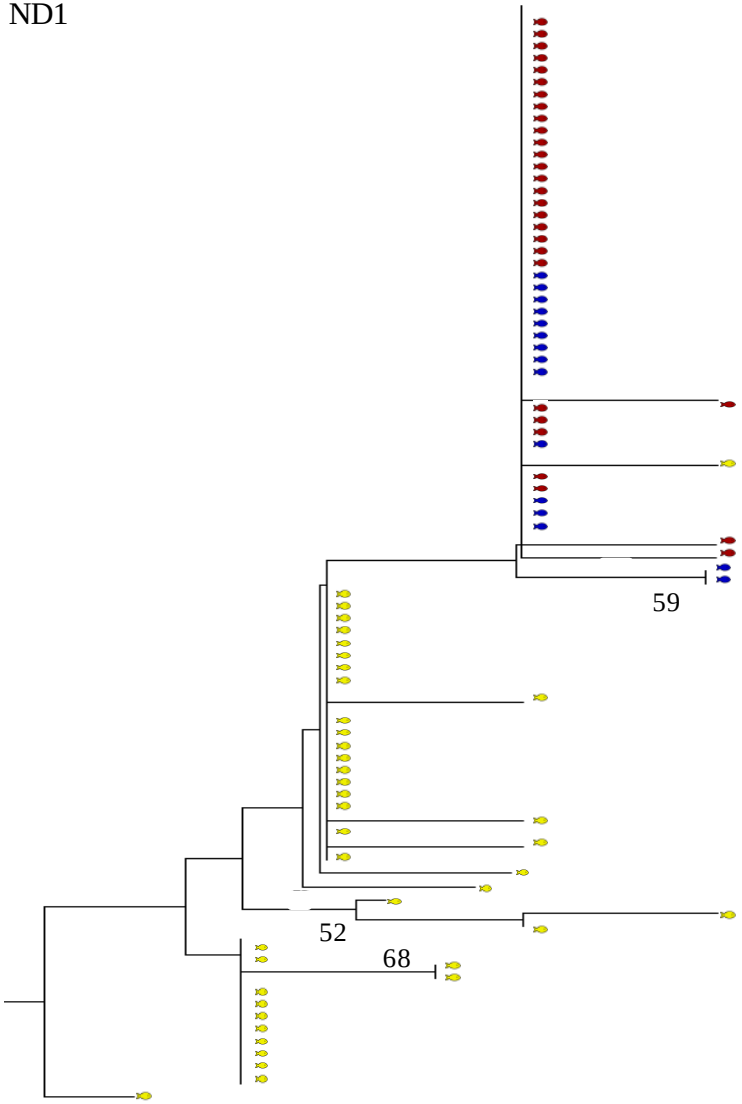


Supplementary Figure 2.1. Statistical parsimony networks for each individual locus for 84 Atlantic wolffish samples. Shared haplotypes are represented by clusters of squares. Sampling locations are colour-coded as in Figure 2.2.

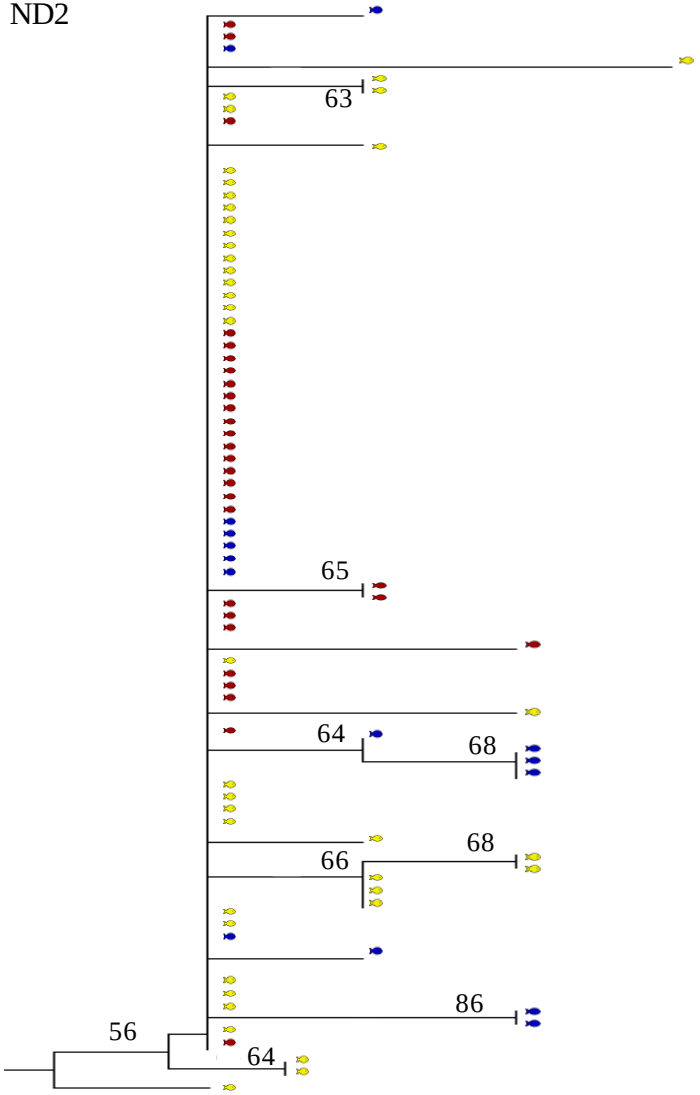

12S
63
54
51
62
65
63


16S
87
63
66
61
62
63
64
63
72
63
64

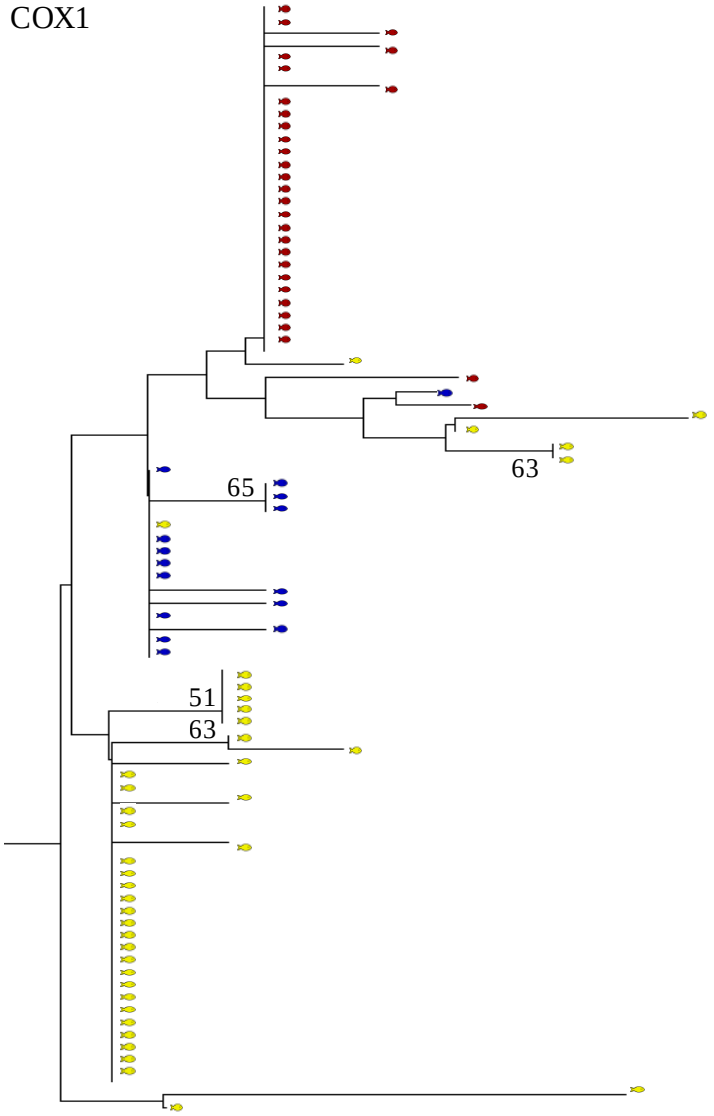
ND1



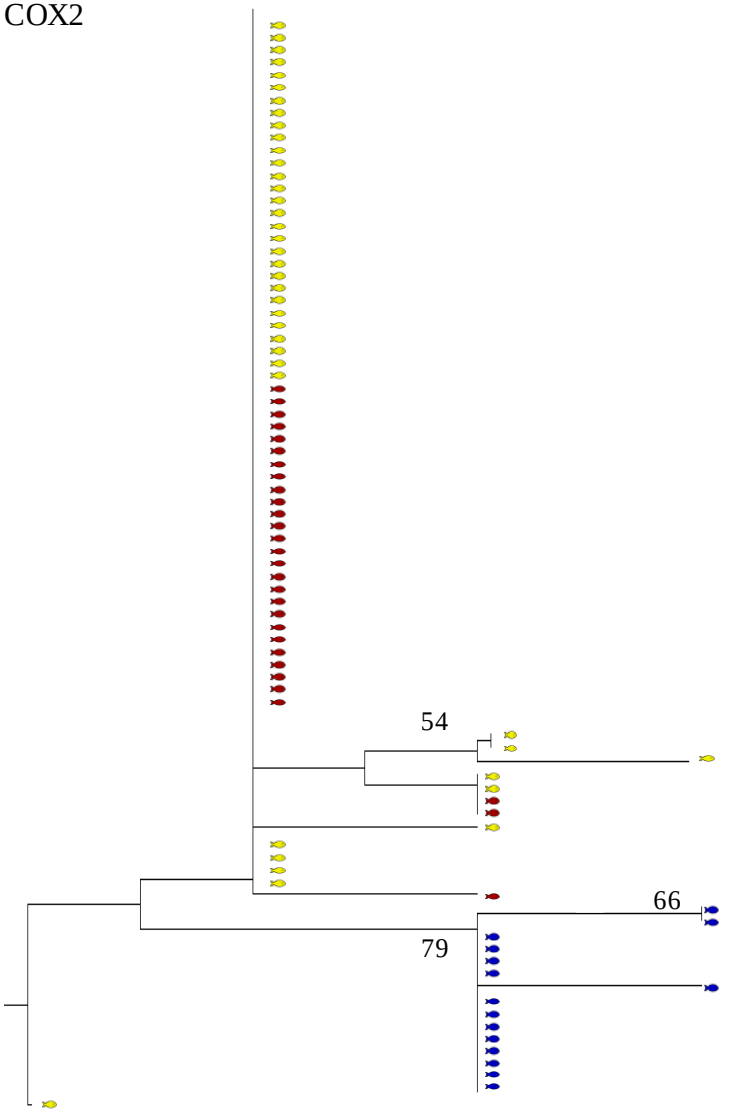
ND2

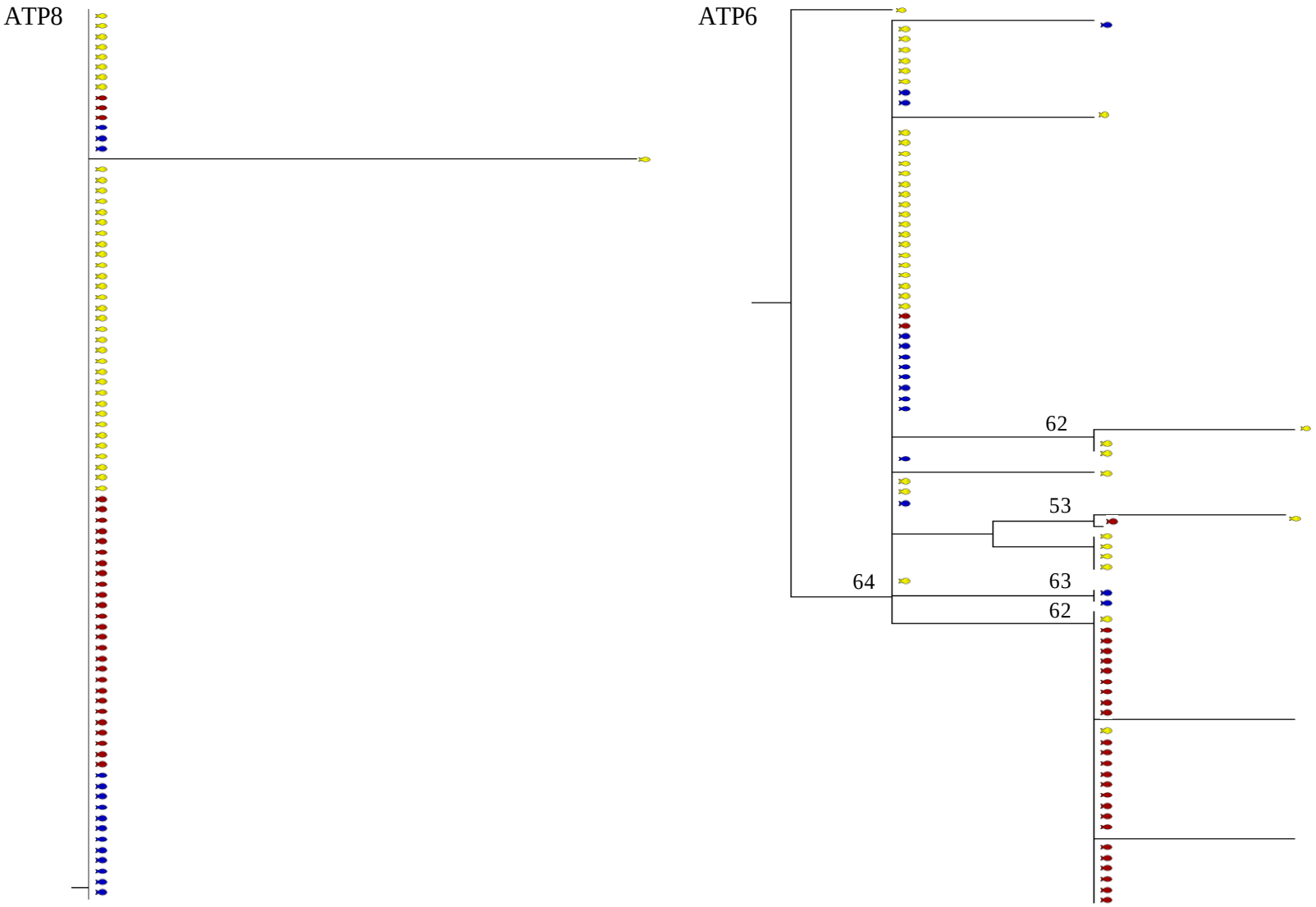


COX1

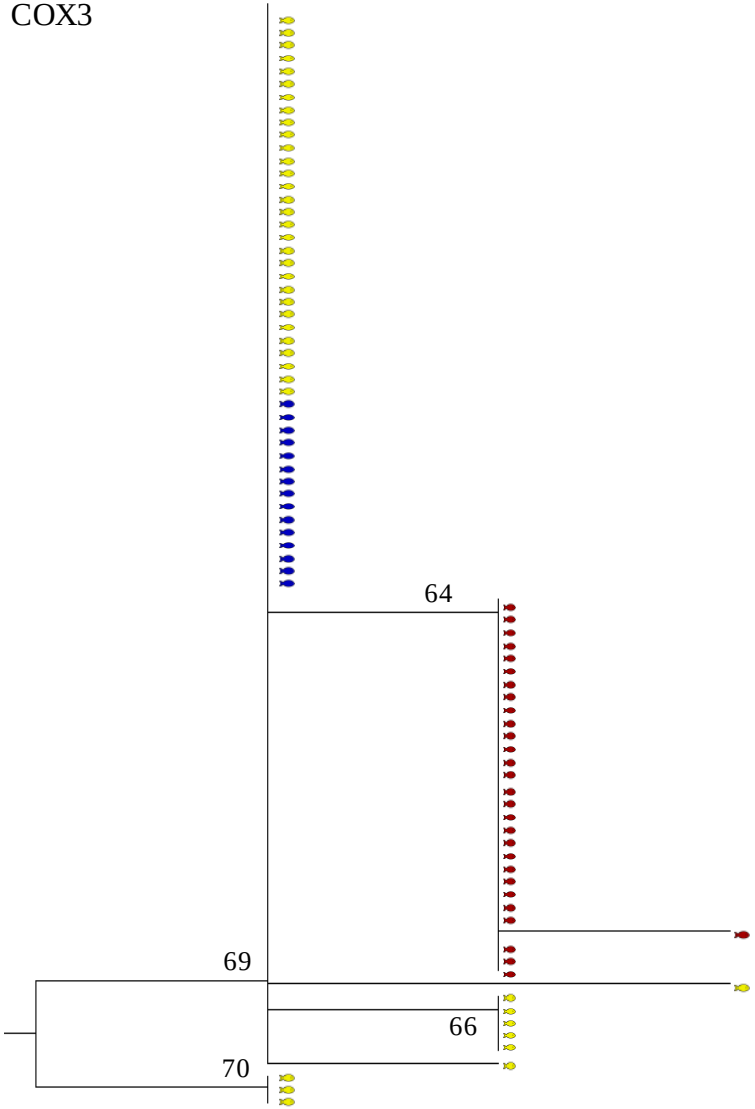


COX2

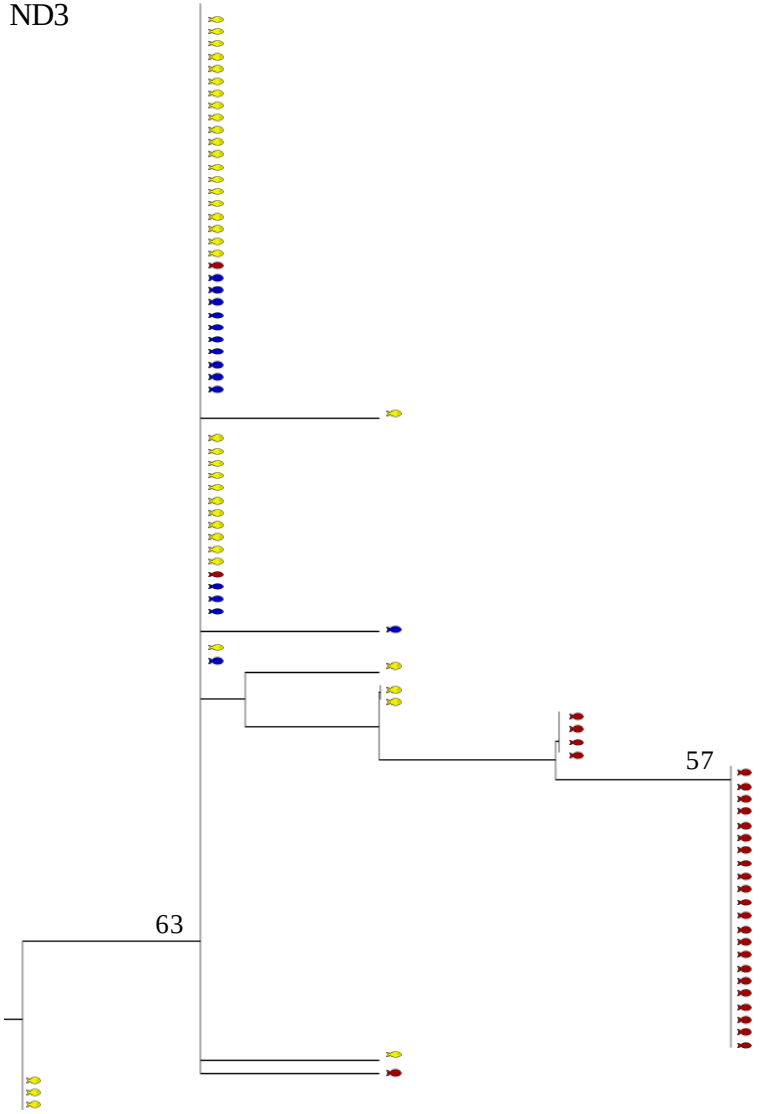



ATP8
ATP6
62
53
64
63
62

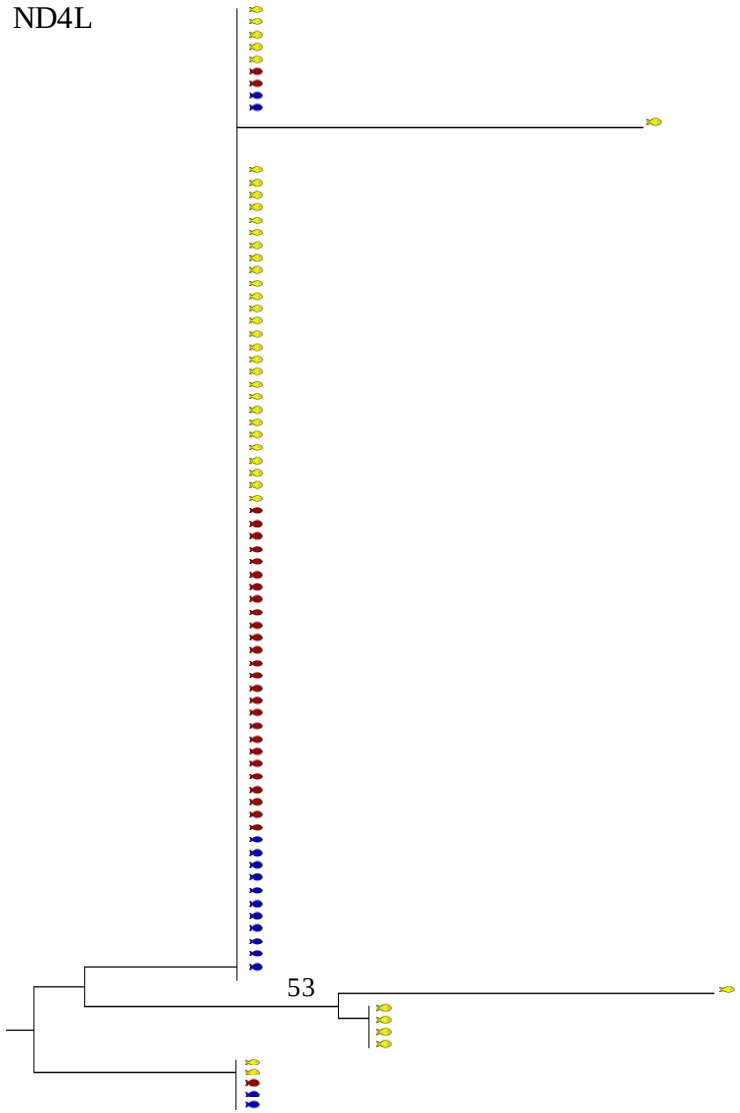
COX3



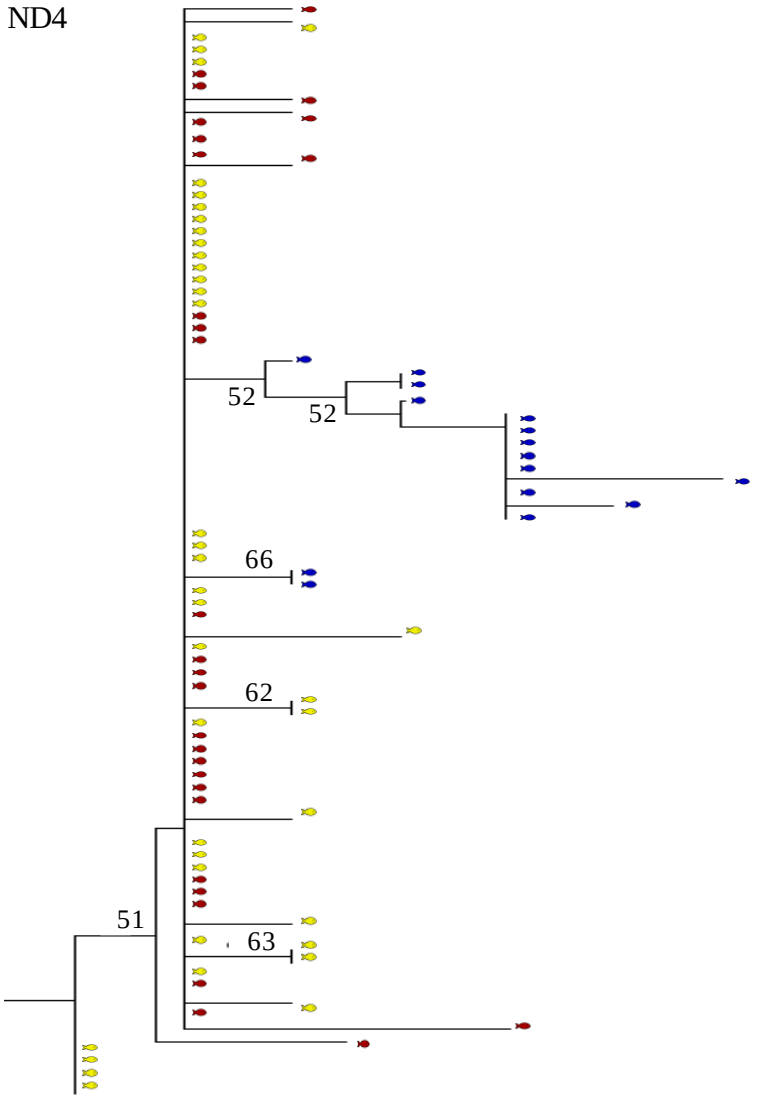
ND3

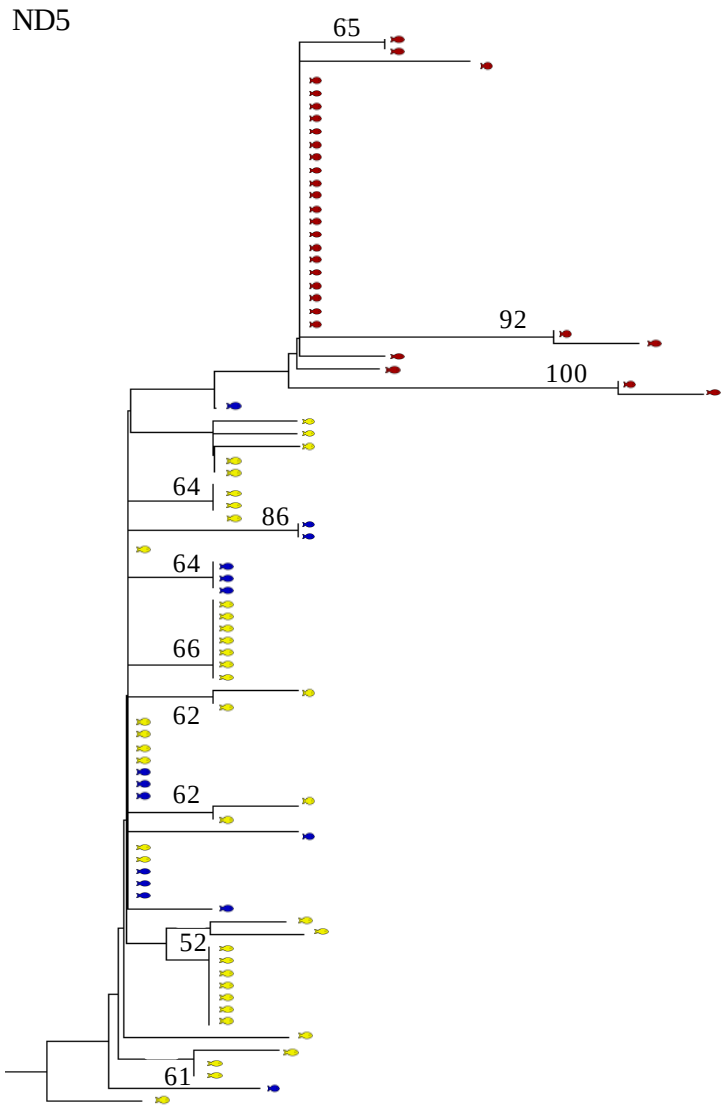


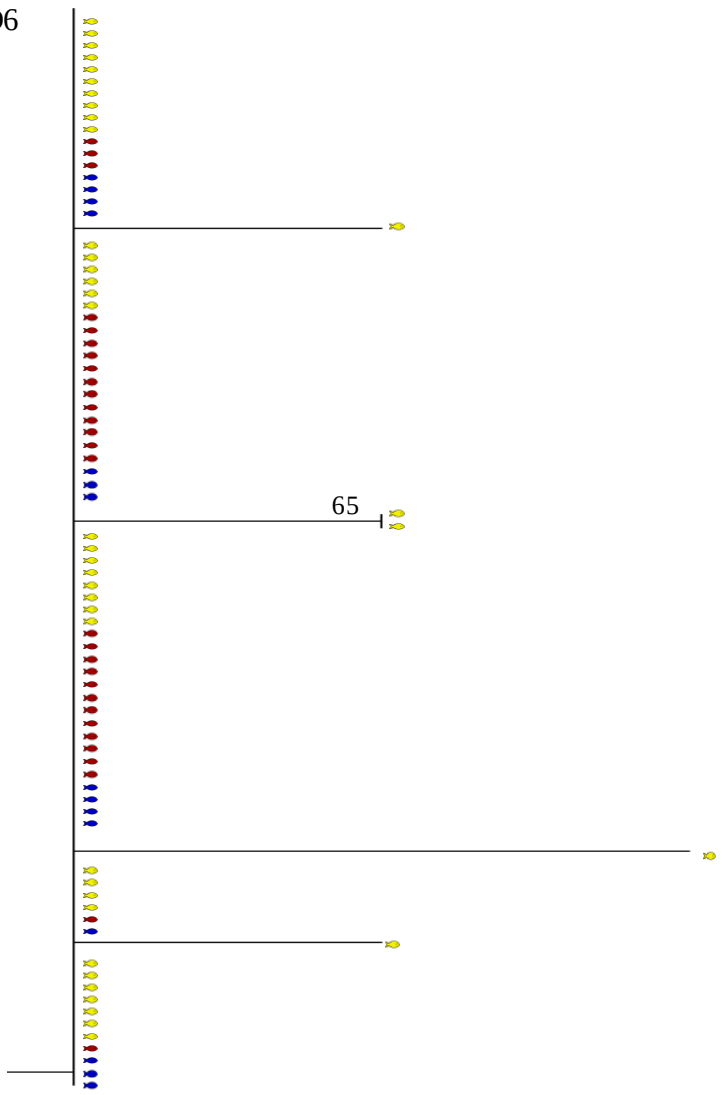
ND4L



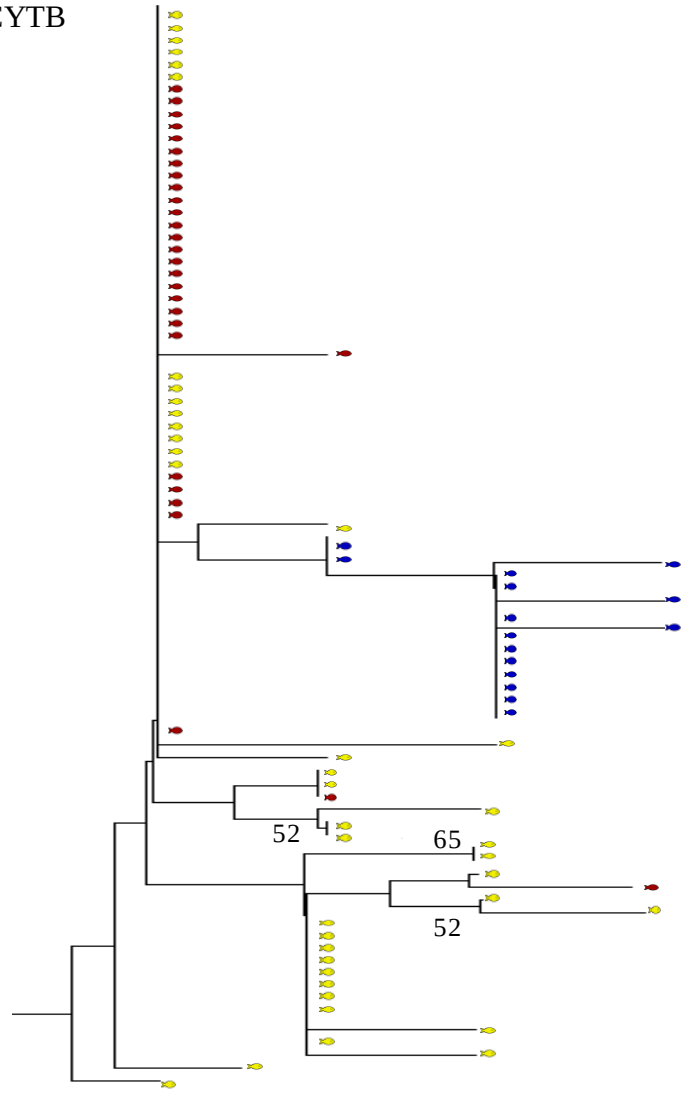
ND4



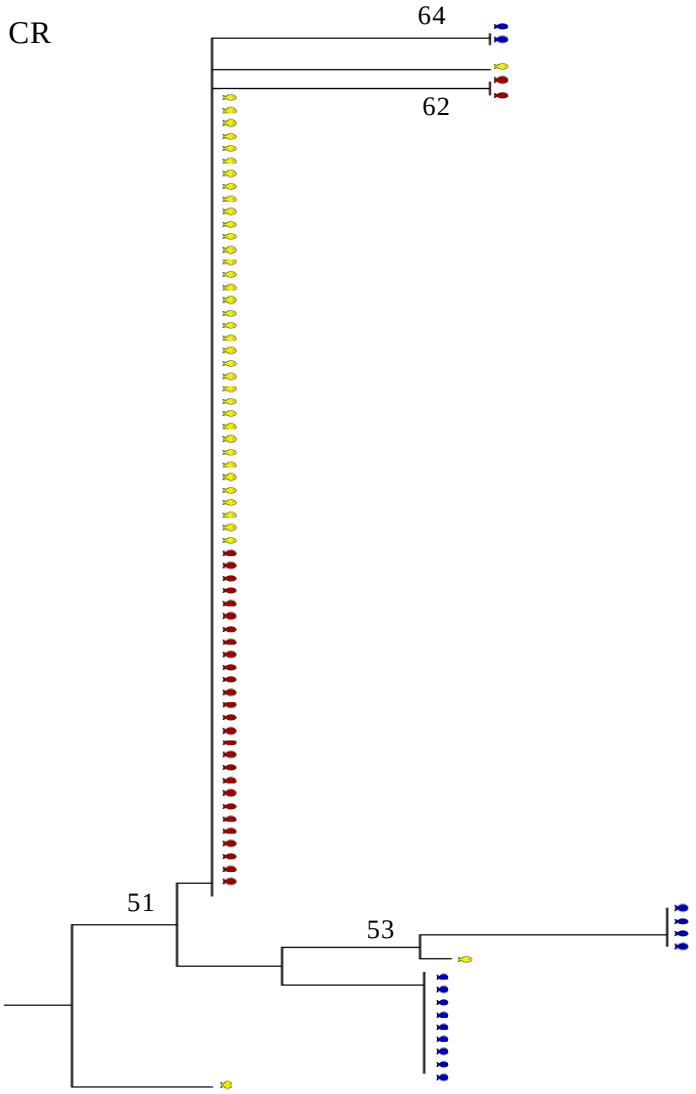

ND5
65
92
100
64
86
64
66
62
62
52
61


ND6
65

CYTB



CR



Supplementary Figure 2.2. NJ analyses for each individual locus for 84 Atlantic wolffish samples. Bootstrap values > 50% are given (10,000 replicates). Samples are colour-coded by haplogroups identified in BAPS v6 (A red, B blue, C yellow). The rooting is the same when either of the other two Northwest Atlantic species is used.

References

Albikovskaya LK (1983) Feeding characteristics of wolffishes in the Labrador-Newfoundland region. *NAFO Scientific Council Studies* **6**, 35-38.

Arnason U, Adegoke J, Gullberg A, Harley E, Janke A, Kullberg M (2008) Mitogenomic relationships of placental mammals and molecular estimates of their divergences. *Gene* **421**, 37-51.

Baker K, Devine J, Haedrich RL (2009) Deep-sea fishes in Canada's Atlantic: population declines and predicted recovery times. *Environmental Biology of Fishes* **85**, 79-88.

Barsukov VV (1959) *The Wolffish (Anarhichadidae)* Indian National Scientific Documentation Centre (1972 translation for the Smithsonian Institution), New Delhi, India.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165-1188.

Bouckaert R, J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard M, Rambaut A, Drummond A (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537.

Brito P, Edwards SV (2009) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-455.

Byun S, Koop B, Reimchen T (1997) North American black bear mtDNA phylogeography: implications for morphology and the Haida Gwaii glacial refugium controversy. *Evolution* **51**, 1647-1653.

Carr S, Duggan A, Stenson G, Marshall H (2015) Quantitative phylogenomics of within-species mitogenome variation: Monte Carlo and non-parametric analysis of

phylogeographic structure among discrete transatlantic breeding areas of harp seals (*Pagophilus groenlandicus*). *PLOS One* **10**, e0134207.

Carr S, Marshall H (2008) Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic cod (*Gadus morhua*): origins of the "codmother", transatlantic vicariance and midglacial population expansion. *Genetics* **180**, 381-389.

Carr SM, Duggan AT, Marshall HD (2009) Iterative DNA sequencing on microarrays: a high-throughput NextGen technology for ecological and evolutionary mitogenomics. *Laboratory Focus* **13**, 8-12.

Carr SM, Marshall HD, Duggan AT, Flynn SMC, Johnstone KA, Pope AM, Wilkerson CD (2008) Phylogeographic genomics of mitochondrial DNA: highly-resolved patterns of intraspecific evolution and a multi-species, microarray-based DNA sequencing strategy for biodiversity studies. *Comparative Biochemistry and Physiology D-Genomics & Proteomics* **3**, 1-11.

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

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.

Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539.

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* **205**, 19-31.

COSEWIC (2000) COSEWIC assessment and status report on the Atlantic wolffish, *Anarhichas lupus*, in Canada, p. vi + 21 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2001a) COSEWIC assessment and status report on the northern wolffish, *Anarhichas denticulatus*, in Canada, p. vi + 21 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2001b) COSEWIC assessment and status report on the spotted wolffish, *Anarhichas minor*, in Canada, p. vi + 22 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2002) COSEWIC assessment and update status report on the Bering wolffish *Anarhichas orientalis* in Canada p. x + 13 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2012a) COSEWIC assessment and status report on the Atlantic wolffish *Anarhichas lupus* in Canada p. x + 56 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2012b) COSEWIC assessment and status report on the northern wolffish *Anarhichas denticulatus* in Canada p. x + 41 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2012c) COSEWIC assessment and status report on the spotted wolffish *Anarhichas minor* in Canada p. x + 44 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

Coulson MW, Marshall HD, Pepin P, Carr SM (2006) Mitochondrial genomics of gadine fishes: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome* **49**, 1115-1130.

Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *Journal of Molecular Biology* **212**, 599-634.

Dohms K, Burg T (2013) Molecular markers reveal limited population genetic structure in a North American corvid, Clark's nutcracker (*Nucifraga columbiana*). *PLOS One* **8**, e79621.

Duggan A (2007) *Efficiency and accuracy of a multi-species iterative DNA sequencing microarray*, BSc Honours Dissertation, Memorial University of Newfoundland.

Ersts P (2015) Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation.

Excoffier L, Lischer H (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

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

Feutry P, Kyne P, Pillans R, Chen X, Naylor G, Grewe P (2014) Mitogenomics of the speartooth shark challenges ten years of control region sequencing. *BMC Evolutionary Biology* **14**, e232.

FishBase (2013) Reviewed distribution maps for *Anarhichas lupus* (Atlantic wolffish), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

Gilbert M, Drautz D, Lesk A, Ho S, Qi J, Ratan A, Hsu C-H, Sher A, Dalén L, Götherström A, Tomsho L, Rendulic S, Packard M, Campos P, Kuznetsova T, Shidlovskiy F, Tikhonov A, Willerslev E, Iacumin P, Buigues B, Ericson P,

Germonpré M, Kosintsev P, Nikolaev V, Nowak-Kemp M, Knight J, Irzyk G, Perbost C, Fredrikson K, Harkins T, Sheridan S, Miller W, Schuster S (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences USA* **105**, 8327-8332.

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

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 183-195.

Hoarau G, Coyer J, Veldsink J, Stam W, Olsen J (2007) Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. *Molecular Ecology* **16**, 3606-3616.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70.

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proceedings of the National Academy of Sciences USA* **92**, 532-536.

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

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

Jacobsen M, Pujolar J, Gilbert M, Moreno-Mayar J, Bernatchez L, Als T, Lobon-Cervia J, Hansen M (2014) Speciation and demographic history of Atlantic eels (*Anguilla*

anguilla and *A. rostrata*) revealed by mitogenome sequencing. *Heredity* **113**, 432-442.

Johannessen T, Gjøsæter J, Moksness E (1993) Reproduction, spawning behaviour and captive breeding of the common wolffish *Anarhichas lupus* L. *Aquaculture* **115**, 41-51.

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

Johnstone KA, Marshall HD, Carr SM (2007) Biodiversity genomics for species at risk: patterns of DNA sequence variation within and among complete mitochondrial genomes of three species of wolffish (*Anarhichas* spp.). *Canadian Journal of Zoology* **85**, 151-158.

Jónsson G (1982) Contribution to the biology of catfish (*Anarhichas lupus*) at Iceland. *Rit Fiskideildar* **6**, 2-26.

Keats DW, South GR, Steele DH (1985) Reproduction and egg guarding by Atlantic wolffish (*Anarhichas lupus*: Anarhichidae) and ocean pout (*Macrozoarces americanus*: Zoarcidae) in Newfoundland waters. *Canadian Journal of Zoology* **63**, 2565-2568.

Knaus B, Cronn R, Liston A, Pilgrim K, Schwartz M (2011) Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecology* **11**, e10.

Lait L, Burg T (2013) When east meets west: population structure of a high-latitude resident species, the boreal chickadee (*Poecile hudsonicus*). *Heredity* **111**, 321-329.

Lait L, Friesen V, Gaston A, Burg T (2012) The post-Pleistocene population genetic structure of a western North American passerine: the chestnut-backed chickadee (*Poecile rufescens*). *Journal of Avian Biology* **43**, 541-552.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Väinölä R, Viard F, Wares J (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* **89 (Sup)**, S108-S122.

Martin A, Naylor G, Palumbi S (1992) Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. *Nature* **357**, 153-155.

McCusker MR, Bentzen P (2010a) Historical influences dominate the population genetic structure of a sedentary marine fish, Atlantic wolffish (*Anarhichas lupus*), across the North Atlantic Ocean. *Molecular Ecology* **19**, 4228-4241.

McCusker MR, Bentzen P (2010b) Phylogeography of 3 North Atlantic wolffish species (*Anarhichas* spp.) with phylogenetic relationships within the family Anarhichadidae. *Journal of Heredity* **101**, 591-601.

Minegishi Y, Aoyama J, Inoue J, Miya M, Nishida M, Tsukamoto K (2005) Molecular phylogeny and evolution of the freshwater eels genus *Anguilla* based on the whole mitochondrial genome sequences. *Molecular Phylogenetics and Evolution* **34**, 134-146.

Miya M, Kawaguchi A, Nishida M (2001) Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Molecular Biology and Evolution* **18**, 1993-2009.

Miya M, Takeshima H, Endo H, Ishiguro N, Inoue J, Mukai T, Satoh T, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai S, 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.

Morin P, Archer F, Foote A, Vilstrup J, Allen E, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Nielsen S, Rasmussen M, Willerslev E, Gilbert M, Harkins T (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* **20**, 908-916.

Nardi F, Spinsanti G, Boore J, Carapelli A, Dallai R, Frati F (2003) Hexapod origins: monophyletic or paraphyletic? *Science* **299**, 1887-1889.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 5269-5273.

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145-164.

O'Dea NR, Haedrich RL (2002) A review of the status of the Atlantic wolffish, *Anarhichas lupus*, in Canada. *Canadian Field-Naturalist* **116**, 423-432.

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

Pampoulie C, Skirnisdóttir S, Daníelsdóttir AK, Gunnarsson Á (2012) Genetic structure of the Atlantic wolffish (*Anarhichas lupus* L.) at Iceland fishing grounds: another evidence of panmixia in Iceland? *ICES Journal of Marine Science* **69**, 508-515.

Pavlov DA, Novikov GG (1993) Life history and peculiarities of common wolffish (*Anarhichas lupus*) in the White Sea. *ICES Journal of Marine Science* **50**, 271-277.

Peakall R, Smouse P (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* **28**, 2537-2539.

Peakall ROD, Smouse PE (2006) GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Pflaumann U, Sarnthein M, Chapman M, d'Abreu L, Funnell B, Huels M (2003) Glacial North Atlantic: sea-surface conditions reconstructed by GLAMAP 2000. *Paleoceanography* **18**, 1-28.

Pielou E (1991) *After the Ice Age: The Return of Life to Glaciated North America* University of Chicago Press, Chicago, IL.

Pope AM, Carr SM, Smith KN, Marshall HD (2011) Mitogenomic and microsatellite variation in descendants of the founder population of Newfoundland: high genetic diversity in an historically isolated population. *Genome* **54**, 110-119.

Provan J, Bennett K (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* **23**, 564-571.

Provan J, Wattier R, Maggs C (2005) Phylogeographic analysis of the red seaweed *Palmaria palmata* reveals a Pleistocene marine glacial refugium in the English Channel. *Molecular Ecology* **14**, 793-803.

Pulgarín-R P, Burg T (2012) Genetic signals of demographic expansion in downy woodpecker (*Picoides pubescens*) after the last North American glacial maximum. *PLOS One* **7**, e40412.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Roff D, Bentzen P (1989) The statistical analysis of mitochondrial DNA polymorphisms: X^2 and the problem of small samples. *Molecular Biology and Evolution* **6**, 539-545.

Rohling E, Fenton M, Jorissen F, Bertrand P, Ganssen G, Caulet J (1998) Magnitudes of sea-level lowstands of the past 500,000 years. *Nature* **394**, 162-165.

Roman J, Palumbi S (2004) A global invader at home: population structure of the green crab, *Carcinus maenas*, in Europe. *Molecular Ecology* **13**, 2891-2898.

Ronquist F, Huelsenbeck J (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**.

Rousset F (2008) Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

Shamblin B, Bjorndal K, Bolten A, Hillis-Starr Z, Lundgren I, Naro-Maciel E, Nairn C (2012) Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. *Molecular Ecology* **21**, 2330-2340.

Shaw J (2006) Palaeogeography of Atlantic Canadian continental shelves from the Last Glacial Maximum to the present, with an emphasis on Flemish Cap. *Journal of Northwest Atlantic Fisheries Science* **37**, 119-126.

Shulman M, Bermingham E (1995) Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* **49**, 897-910.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**, 264-279.

Stone A, Battistuzzi F, Kubatko L, Perry Jr G, Trudeau E, Lin H, Kumar S (2010) More reliable estimates of divergence times in *Pan* using complete mtDNA sequences and accounting for population structure. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 3277-3288.

Swofford D (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Tanaka M, Cabrera V, González A, Larruga J, Takeyasu T, Fuku N, Guo L-J, Hirose R, Fujita Y, Kurata M, Shinoda K-I, Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S, Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H (2004) Mitochondrial genome variation in Eastern Asia and the peopling of Japan. *Genome Research* **14**, 1832-1850.

Teacher A, André C, Merilä J, Wheat C (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evolutionary Biology* **12**, e248.

Templeman W (1984) Migrations of wolffishes, *Anarhichas* sp., from tagging in the Newfoundland area. *Journal of Northwest Atlantic Fisheries Science* **5**, 93-97.

Templeman W (1986) Some biological aspects of Atlantic wolffish (*Anarhichas lupus*) in the Northwest Atlantic. *Journal of Northwest Atlantic Fisheries Science* **7**, 57-65.

Vis M, Carr S, Bowering W, Davidson W (1997) Greenland halibut (*Reinhardtius hippoglossoides*) in the North Atlantic are genetically homogeneous. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 1813-1821.

Wang Z, Shen X, Liu B, Su J, Yonezawa T, Yu Y, Guo S, Ho S, Vilà C, Hasegawa M, Liu J (2010) Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal. *Journal of Biogeography* **37**, 2332-2344.

Ward R (1995) Population genetics of tunas. *Journal of Fish Biology* **47 (Supplement A)**, 259-280.

Watling L, Norse E (1998) Disturbance of the seabed by mobile fishing gear: a comparison to forest clearcutting. *Conservation Biology* **12**, 1180-1197.

Wilson A (2006) Genetic signature of recent glaciation on populations of a near-shore marine fish species (*Syngnathus leptorhynchus*). *Molecular Ecology* **15**, 1857-1871.

Winkelmann I, Campos P, Strugnell J, Cherel Y, Smith P, Kubodera T, Allcock L, Kampmann M-L, Schroeder H, Guerra A (2013) Mitochondrial genome diversity and population structure of the giant squid *Architeuthis*: genetics sheds new light on one of the most enigmatic marine species. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130273.

Wright S (1965) The interpretation of population structure by F-Statistics with special regard to systems of mating. *Evolution* **19**, 395-420.

Yanagimoto T, Kitamura T, Kobayashi T (2004) Complete nucleotide sequence and variation of mitochondrial DNA from 10 individuals of walleye pollock, *Theragra chalcogramma*. *Fisheries Science* **70**, 885-895.

Young A, Torres C, Mack J, Cunningham C (2002) Morphological and genetic evidence for vicariance and refugium in Atlantic and Gulf of Mexico populations of the hermit crab *Pagurus longicarpus*. *Marine Biology* **140**, 1059-1066.

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

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107-2121.

CHAPTER THREE

Variation among the mitochondrial genome sequences of two threatened marine fish, the spotted wolffish (*Anarhichas minor*) and the northern wolffish (*A. denticulatus*), detects patterns of isolation and reintegration

L.A. Lait[1] and S.M. Carr[1]

[1]Department of Biology, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, A1B 3X9

3.1 Introduction

Understanding population structure and genetic diversity is necessary for the conservation and recovery of endangered species. While it has long been accepted that maintenance of biodiversity is crucial for sustaining ecosystem health and function (Chapin *et al.* 2000; Hooper *et al.* 2005), it is now becoming clear that preserving intraspecific diversity among populations is also an important means of conservation (Schindler *et al.* 2010). This is particularly true for species subject to overexploitation and habitat degradation. Lack of genetic diversity in a species limits their ability to effectively respond to and recover from environmental threats such as disease, pollution, and parasites (Frankham 1995; Amos & Harwood 1998), and maintenance or enhancement of variation can help mitigate the effects of climatic extremes and environmental changes (Hilborn *et al.* 2003).

The marine environment has been heavily impacted by overexploitation, habitat destruction, and climatic changes. Since the advent of improved fishery technologies in the late 20th century, marine fish stocks worldwide have declined by as much as 99% (Hutchings 2000). Marine fish species were historically regarded as resilient species that would recover quickly and easily due to their large population sizes, high fecundity, and extensive dispersal capabilities (Palumbi 1992; Hutchings 2000). However, in most fish species this has not been the case. In the North Atlantic Ocean, species subject to intense commercial exploitation, such as Atlantic cod *Gadus morhua*, haddock *Melanogrammus aeglefinus*, and pollock *Pollachius virens*, have all undergone significant population declines since the 1960s (Frank *et al.* 2005). Other non-commercial species, such as the

roundnose grenadier *Coryphaenoides rupestris*, blue hake *Atimora rostrata*, and Kaup's arrowtooth eel *Synaphobranchus kaupi*, have been affected by habitat destruction and bycatch mortality (Devine *et al.* 2006; Jørgensen *et al.* 2014). Few fish species in the North Atlantic have shown appreciable population increases since the 1992 groundfish moratorium (Hutchings 2000).

Two species heavily impacted by the fisheries of the mid-20th century are the spotted wolffish *Anarhichas minor* (Ólafsson, 1772) and the northern wolffish *Anarhichas denticulatus* (Krøyer, 1845). These species suffered dramatic population declines of 96% and 98%, respectively, and have since been listed as Threatened by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC; COSEWIC 2001a, 2001b) under the Canadian Species at Risk Act (SARA). The population declines were caused by a combination of overharvesting as bycatch and habitat destruction by bottom trawls in the large commercial fisheries (Watling & Norse 1998; O'Dea & Haedrich 2002). Recent status reports suggest that populations are beginning to stabilise and have even seen slight improvements over the past decade, although population declines over the assessment period remain at 80% and 90%, respectively (COSEWIC 2012a, 2012b). Along with the congeneric Atlantic wolffish *Anarhichas lupus*, which is listed as Special Concern (COSEWIC 2000), the three wolffish species were the first marine fish species to be listed by the Minister under SARA.

Wolffish are large demersal fish in the family Anarhichadidae. The family comprises wolfeel *Anarrhichthys ocellatus* and Bering wolffish *Anarhichas orientalis* in the North Pacific Ocean and Atlantic, spotted, and northern wolffish in the North Atlantic Ocean (Barsukov 1959). Spotted and northern wolffish are sympatric along the

continental shelves of North America and Europe (Figure 3.1), although they are generally found in somewhat different habitats. Spotted wolffish are found at moderate depths (to 550 m) and prefer a sandy bottom, while northern wolffish occur at much greater depths (around 900 m) and prefer a softer, muddy bottom (Barsukov 1959). The two species also have slightly different feeding strategies – spotted wolffish eat primarily larger molluscs with harder shells, whereas northern wolffish prey more on echinoderms and mobile organisms such as jellyfish (Barsukov 1959; Albikovskaya 1983). Despite differences in habitat preference and feeding behaviour, spotted and northern wolffish have been known to hybridise (Templeman 1986).

Unlike many marine fish species with high gene flow and fecundity, wolffish are sedentary, slow-growing species with late maturity and slow population growth which may subject them to a greater risk of local extirpation and decreased recovery potential (Hiddink *et al.* 2008; Baker *et al.* 2009). The sedentary nature of wolffish, coupled with the fact that eggs and larvae are generally demersal rather than pelagic (Barsukov 1959), suggest that they may show greater population structure than is seen in many marine fish species. Because of this, it is critical that the population structure be determined and used to inform management decisions (Schindler *et al.* 2010). Studies to date have been ambiguous; previous work with nuclear microsatellites and AFLPs identified the Barents Sea population of northern wolffish as genetically isolated; however, in spotted wolffish few population comparisons were significant (McCusker & Bentzen 2011). In both species, diversity levels were lower than that seen in the sympatric Atlantic wolffish. In contrast, a study using two mitochondrial markers, ND1 and the control region, found no phylogeographic pattern in either species (McCusker & Bentzen 2010).

Here we aim to measure genetic variability in the two threatened marine species by comparing multiple complete mitogenomic sequences. While the control region has been useful in many population studies (Vigilant *et al.* 1991; Bernatchez *et al.* 1992; Marshall & Baker 1997), it has recently been shown that some slowly evolving marine fish species, including wolffish, show little or no variation in this marker (Johnstone *et al.* 2007; Feutry *et al.* 2014). The objective of this study is to determine whether the complete mitochondrial genome can identify previously undetected structure in spotted and northern wolffish as it has in a number of other species, and to identify where the variation lies in this genome. This will allow more in-depth analyses of these two species to be devised. The results are compared with those of the more extensively sampled Atlantic wolffish (Chapter 2).

3.2 Materials and Methods

3.2.1 Sample collection

A total of 17 spotted wolffish and 18 northern wolffish samples were collected by Fisheries and Oceans Canada between October 2002 and November 2003. The fish were caught from seven sampling locations along the coast of Newfoundland and Labrador (Figure 3.1). Hearts were removed and stored at -20°C. An additional northern wolffish sample from Europe was provided as extracted DNA by Dr. Paul Bentzen (Dalhousie University).

3.2.2 DNA extraction, amplification, and sequencing

Samples were extracted from heart tissue using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's protocol (see Section 2.2.2 for details; Qiagen, Hilden, Germany). The complete mtDNA genome was amplified in a series of 18 - 20 overlapping standard PCR fragments (698 - 1503 bp), or five long-range PCR fragments (2,058 - 4,969 bp). The average overlap was 245 bp (4 - 735 bp) for standard PCR fragments, and 224 bp (156 - 303 bp) for long-range fragments. Primers were originally designed for use in either the congeneric Atlantic wolffish (Chapter 2; Johnstone *et al.* 2007) or in Atlantic cod (Coulson *et al.* 2006).

For the standard PCRs, primer pairs were the same as those used in Atlantic wolffish (see Table 3.1 for combinations, sequences, and annealing temperatures). PCR reactions were carried out using Qiagen Taq DNA polymerase or Qiagen TopTaq DNA polymerase kits. The reactions (25 μL) contained 10X PCR buffer, 2 mM $MgCl_2$ (Qiagen), 0.2 mM dNTP (Qiagen), 0.4 μM each of forward and reverse primer, 1 U DNA polymerase (Qiagen), and 2 μL genomic DNA. The following thermal cycling program was used: an initial cycle of 3 min at 95°C (15 min at 95°C for TopTaq), followed by 40 cycles of 30 s at 94°C, 45 s at a primer-pair-specific T_A, 60 s at 72°C; and a final extension at 72°C for 10 min.

For long-range PCRs, the primer pairs used were w18F/g03R, g04F/w07R, g07F/g08R, w08F/wND45R, and w14F2/w18R (see Table 3.1 for sequences). PCR reactions were carried out using the Qiagen LongRange PCR kit. The reactions (50 μL) contained 10X LR PCR buffer (Qiagen), 1.5 mM $MgCl_2$ (Qiagen), 0.5 mM dNTP

(Qiagen), 0.4 μM forward and reverse primer, 3 units enzyme (Qiagen), and 4 μL genomic DNA. The PCRs were run on the following thermal cycling program: an initial cycle of 2 min at 92°C; 10 cycles of 10 s at 92°C, 15 s at 58°C, 8 min at 68°C; 20 cycles of 15 s at 92°C, 30 s at 58°C, 8 to 17 min at 68°C (time increased by 20 s per cycle), and a final extension at 68°C for 7 min.

All PCR reactions were performed in an Eppendorf Mastercycler epGradient S thermocycler (Eppendorf, Hamburg, Germany). Products were visualised on a 1% agarose gel with 0.3 μg/mL ethidium bromide added, and compared to either a 100 bp ladder (for standard PCRs) or a 1 kbp ladder (for long-range PCRs). From 10 μL (for standard PCR) to 50 μL (for long-range PCR) of PCR product was purified with an Exo-SAP clean up: products were incubated with 0.1 U exonuclease (USB) and 0.1 U shrimp alkaline phosphatase (USB) for 15 min at 37°C, followed by denaturation at 80°C for 15 min. The purified products were sent to Genome Quebec (McGill University, QC) for Sanger sequencing. Standard PCR products were sequenced in both directions, and long-range PCR products were sequenced using a series of forward and reverse internal primers.

3.2.3 Genetic analyses

The chromatograms were visually inspected and aligned in Sequencher v4.9 (Gene Codes Corporation, Ann Arbor, MI). All variable sites were double-checked and any ambiguous calls were removed. Sequences were annotated using the Atlantic wolffish mitochondrial genome sequence. Genetic diversity was measured using an analysis of

molecular variance (AMOVA; 100,000 permutations; Excoffier *et al.* 1992) in Arlequin v3.5.1 (Excoffier & Lischer 2010), and nucleotide (π) and haplotype (H_d) diversities were run in DNAsp (Librado & Rozas 2009).

Clustering analysis was performed in BAPS v6 (Bayesian Analysis of Population Structure; Corander *et al.* 2008) with the linked loci option and the codon model of linkage (see Corander & Tang 2007). We varied K from 1 to 15, and determined the best number of groups based on the lowest log marginal likelihood of the best visited partitions. The relationship among sequences was visualised using two methods: an unrooted statistical parsimony network constructed in TCS v1.21 (95% connection limit; Clement *et al.* 2000), and a principal coordinates analysis (PCoA) in GenAlEx v6.5 (Peakall & Smouse 2006; Peakall & Smouse 2012).

Both Bayesian analyses and distance methods were used to evaluate the phylogenetic relationship among samples. Bayesian trees were constructed in both BEAST (Bayesian Evolutionary Analysis Sampling Trees; Bouckaert *et al.* 2014) and MrBayes v3.2 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003). The analyses were run (separately) with the generalised-time-reversible model with gamma-distributed rate variation allowing for invariable sites (GTR + Γ + I) and the Hasegawa, Kishino, and Yano (HKY) + Γ + I model; the two models were compared to ensure that model selection did not affect the results. In BEAST, the analysis was run for 10,000,000 generations, with a 25% burn-in, and final ESS > 1000. In MrBayes, the analysis was run for two simultaneous runs with four chains for 10,000,000 generations, with a 25% burn-in, final ESS > 1000, PSRF > 0.999, and standard deviation of split frequencies < 0.005.

A neighbour-joining tree (NJ) was constructed in PAUP* v4.10 (Swofford 2003) based on the absolute number of nucleotide differences (10,000 bootstrap replications). Trees were rooted with the congeneric northern/spotted wolffish and Atlantic wolffish.

Divergence times for the most recent common ancestors in spotted and northern wolffish were estimated in BEAST v2.3 (Bouckaert *et al.* 2014). The trees were run using the HKY + Γ + I model with a strict clock and were run for 10,000,000 MCMC steps, sampled every 10,000, with a 1,000,000 step burn-in. All ESS parameters were > 1000. In both cases the trees were calibrated by setting the separation time between spotted and northern wolffish at 2 million years (see Section 2.2.8 for details).

3.2.4 Individual loci

The effectiveness of an individual locus to detect structure in either spotted or northern wolffish was tested. Comparisons were made among the 13 coding regions, the two rRNAs, and the control region using the above analyses.

3.3 Results

3.3.1 Complete mitochondrial sequences

Complete mtDNA sequences were obtained for 17 spotted and 19 northern wolffish (GenBank Accession Numbers KX118006 - KX118041). The sequences contained the standard 13 coding regions, two rRNA regions, 22 tRNAs, and the non-coding control region in the standard vertebrate order. The length of the mitogenomes varied from 16,510 - 16,515 bp in spotted wolffish and from 16,510 - 16,516 bp in

northern wolffish. In all three North Atlantic wolffish species there is a homopolymer run of cytosine residues between the tRNA aspartic acid (Asp) and the cytochrome oxidase II (COX2) gene; in spotted and northern wolffish the number of residues is variable (7 - 12 and 7 - 13, respectively), whereas Atlantic wolffish have an invariable five residues. The average nucleotide composition of the mtDNA was 26.7% A, 28.3% C, 17.8% G, and 27.2% T in both spotted and northern wolffish.

A total of 61 variable sites were identified in the 17 spotted wolffish sequences, 17 of which were found in more than one individual (i.e., parsimony informative *sensu* Nei 1987), and 44 singletons. Approximately 80% (48/61) of the changes were transitions (22 A/G and 26 C/T), with seven transversions and six insertion-deletion events. Of the 61 variable sites, 39 were in the coding regions, mostly third position (82%) leading to synonymous (79%) changes. There were seven amino acid changes among the samples, with no unexpected stop codons. Patterns of variation are given in Table 3.2. No site had more than two base variants, and the insertion-deletion events all occurred in non-coding regions (five in the intergenic region between Asp and COX2, and one in the 16S rRNA).

There were 81 variable sites identified in the 19 northern wolffish sequences, 37 of which were parsimony informative. Of these sites, 70 were transitions (41 A/G and 29 C/T), four transversions (one of each kind), and seven insertion-deletions. Fifty-seven of the 81 changes were in the coding regions, with the majority (87.7%) in the third position. Only six amino acid changes were found (Table 3.2), there were no unexpected stop codons, and no sites had more than two base variants. As with spotted wolffish, all insertion-deletion events occurred in non-coding regions.

3.3.2 Population structure

There were no shared haplotypes in either species; each individual had a unique DNA sequence (H_d = 1.000). Pairwise differences ranged from 2 to 28 in the spotted wolffish and 1 to 28 in the northern wolffish. Nucleotide diversity (π) was lower in spotted (π = 0.00045) than northern wolffish (π = 0.00086). AMOVA analyses showed no evidence of phylogeographic structure in either species, with 99.7% and 91.9% of the variance seen within rather than among sampling location (Φ_{ST} = 0.0028; p = 0.46 and Φ_{ST} = 0.0814; p = 0.14 for spotted and northern wolffish, respectively). In both cases sample sizes are small which will affect overall Φ_{ST} values.

Bayesian clustering analysis identifies groups based on DNA sequence with no *a priori* location information. In spotted wolffish, BAPS identified three groups found across all sampling sites (D - F, shown in grey, cyan, and orange, respectively; Figure 3.1). The highest log maximum likelihood values were -449.6 (K = 3), -472.2 (K = 4), and -488.5 (K = 5); the probability of three clusters was 1. In northern wolffish, BAPS identified two groups (G and H), one found primarily in northern Newfoundland and Europe (G, shown in green), and the other found on the Grand Banks, Flemish Cap, and in Labrador (H, shown in purple). The three highest log maximum likelihood values were -752.7 (K = 2), -760.8 (K = 3), and -770.0 (K = 4); the probability of two clusters was 0.999. While the three haplogroups in spotted wolffish did not correspond to geographical location, there appears to be a trend in haplogroup distribution in northern wolffish; additional sampling will be required to test this further.

The statistical parsimony networks identified the same groups as the clustering analyses. Neither species showed phylogeographic structure. In spotted wolffish (Figure 3.2a) there were at least three groups evident, separated by three to five substitutions. Two of the groups, D and E, were of similar size and pattern of variation, arranged as star phylogenies around a non-extant haplotype, while the third group (F) was divergent and may represent multiple additional groups. In northern wolffish two main groups were observed (Figure 3.2b); a large group found across the southern locations as well as in Labrador, and a smaller group (or two) present in the northern locations and in Europe. In both species there were a number of divergent individuals, all of which clustered with the species in question.

PCoA analysis identified the same groups as the clustering and statistical parsimony analyses. For spotted wolffish (Figure 3.3a), haplogroup F separated from the other two along coordinate 1 (35.1% of the variation), and haplogroups D and E separated slightly along coordinate 2 (20.8% of the variation). Coordinate 3 explained 17.1% of the variation. In northern wolffish (Figure 3.3b), the separation of the two groups was very pronounced along coordinate 1 (52.8% of the variation), while haplogroup H showed separation within the group along coordinate 2 (17.0% of the variation). Coordinate 3 explained 9.4% of the variation.

3.3.3 Phylogenetic analyses

As in Atlantic wolffish, all three methods of analysis identified the same groups in both spotted and northern wolffish; the BEAST trees are similar and are not shown. The NJ analysis supported the relationship among the three species seen in Johnstone *et al.*

(2007; Figure 3.4a). In spotted wolffish (Figures 3.4b and 3.5a) there was strong statistical support for haplogroup F with a posterior distribution of 0.9 and bootstrap support of 84, haplogroup E had strong support in the Bayesian tree although lower support in the NJ tree (posterior distribution = 1, bootstrap = 62), and haplogroup D consisted of the remaining individuals and potentially contained at least one additional group (Figure 3.5a, posterior distribution = 1). In northern wolffish (Figures 3.4c and 3.5b), both trees showed strong statistical support for the two haplogroups (posterior distribution = 0.9 for G and 1 for H, bootstrap = 97 and 90). In both haplogroups additional substructure was evident, with an additional two groups seen in haplogroup G, and potentially two or more groups in haplogroup H. The statistical support is lower for this substructure.

In spotted wolffish, the three identified haplogroups were estimated to have diverged approximately 81,000 years ago (95% Highest Posterior Density (HPD) = 33 - 132 kya), with the divergence time between haplogroups D and E at 63,000 years ago (95% HPD = 28 - 104 kya). This puts the separation time towards the end of the Pleistocene glaciations, although still within them, and suggests that the separation is not postglacial. In northern wolffish there is clear separation of the two identified haplogroups, with an estimated divergence time of 126,000 years ago (95% HPD = 57 - 204 kya), well within the Pleistocene glaciations. Within each haplogroup there was evidence of additional divergence occurring prior to the last glacial maximum (LGM, ca. 21 kya).

3.3.4 Individual loci

The AMOVA analyses on the individual loci gave overall Φ_{ST} values of 0 to 1 for spotted wolffish and 0 to 0.31 for northern wolffish (Table 3.3). In spotted wolffish none of the values were significant; the locus with Φ_{ST} = 1 (ND1) was due to the two populations with n = 1 being different from all other individuals. Without this marker the Φ_{ST} values ranged from 0 to 0.32. In northern wolffish, two loci gave significant values: ND2 and the control region.

The statistical parsimony networks for the individual loci showed limited structure in spotted wolffish (Supplementary Figure 3.1). While only one of the loci was invariable (ATP8), six loci consisted of only two haplotypes. Of these, 12S was still able to identify one of the three haplogroups (F), as was ND6 with three haplotypes. Two other loci were able to identify haplogroup E (COX2 and CYTB). A single locus, ND2, was able to differentiate the three haplogroups (D - F). A much higher degree of variation was seen in northern wolffish (Supplementary Figure 3.2). Again only a single locus was invariable (ND3), and three loci had only two haplotypes. Eight of the 16 loci were able to identify the two haplogroups (G and H), from ND6 with only two haplotypes to the much more variable 16S, ND5, and CYTB with nine, nine, and 13 haplotypes respectively.

The NJ trees showed a similar pattern to the statistical parsimony networks in both species. In spotted wolffish (Supplementary Figure 3.3), 12S and ND6 were able to separate haplogroup F, COX2 separated haplogroup B (CYTB identified it but not as a monophyletic group), and ND2 identified the three haplogroups, although all with low bootstrap support. Northern wolffish (Supplementary Figure 3.4) showed clear separation of haplogroups G and H in six of the 16 loci. 16S and ND6 which had both identified the

haplogroups in the statistical parsimony networks (Supplementary Figure 3.2) no longer separated them – 16S grouped some of the constituents of haplogroup G but not in a monophyletic group and with no bootstrap support, while ND6 did not identify the groups at all. The other six loci separated the groups with bootstrap support between 60 and 89.

3.4 Discussion

The application of mitogenomics has allowed us to identify previously undetected genetic structure in spotted and northern wolffish. In both species there was evidence of multiple divergent lineages (haplogroups) not seen in previous studies. The distribution of variation within the genome shows that the markers used in the previous mitochondrial DNA study (McCusker & Bentzen 2010), ND1 and the control region, were not highly variable in either species. The lack of phylogeographic structure despite low migration rates is likely due to low mutation rates in the wolffish as seen in the Atlantic wolffish (Chapter 2), the speartooth shark *Glyphis glyphis* (Feutry *et al.* 2014), and the giant squid *Architeuthis dux* (Winkelmann *et al.* 2013).

3.4.1 Population genetic structure and haplogroup dating

Analysis of complete mitochondrial genomes of 17 spotted and 19 northern wolffish identified a lack of geographical structure in either species. This supports the findings of McCusker & Bentzen (2010) who found no apparent phylogeographical pattern in any of the three North Atlantic wolffish species despite low dispersal capabilities and trans-Atlantic sampling. However, both species showed evidence of

multiple divergent haplogroups in all analyses – three in the spotted wolffish and two in the northern wolffish.

In spotted wolffish the three groups were found across the sampling locations, although at different frequencies; the differences are unlikely to be significant though additional sampling is required to confirm this. Although the three groups did not have strong statistical support, with bootstrap values ranging from 62 for haplogroup E to 84 for haplogroup F, estimated divergence dates of 63 - 81 kya place the separation of these groups within the last glaciation. This is the same pattern seen in the Atlantic wolffish, the sister-species to the spotted wolffish with very similar life history and habitat requirements (Chapter 2), and suggests that spotted wolffish separated into multiple glacial refugia, although likely all in the same region (e.g., in three nearby refuges in Europe). As the ice sheets melted and additional habitat became suitable for the wolffish the refugial populations may have converged before recolonising previously glaciated or emerged regions. This scenario has been suggested in the Atlantic wolffish (Chapter 2) and the Eurasian yak *Bos grunniens* (Wang *et al.* 2010).

In northern wolffish two haplogroups were identified, again with no significant geographical structure, although haplogroup G appears to be more common in the northern and eastern locations, and haplogroup H in the south and west. Additional sampling, particularly in the mid-Atlantic and Europe, is needed to test if there is a significant difference in haplogroup distribution. Statistical support was high for the two groups in northern wolffish (bootstrap > 90), and all analyses showed clear separation of the lineages into monophyletic groups. An estimated divergence date of 126 kya indicates that the two haplogroups are not of postglacial origin, as was suggested by McCusker&

Bentzen (2010) based on the ND1 gene and the control region. The separation may have been in multiple isolated refugia on one or both sides of the Atlantic Ocean. In the eastern Atlantic there have been as many as seven putative marine refugia suggested, while in the western Atlantic there were likely two main refugia (Pflaumann *et al.* 2003; Shaw 2006; Maggs *et al.* 2008). Unlike Atlantic and spotted wolffish, northern wolffish do not require a hard, stony bottom to feed; they can instead catch mobile prey such as comb jellies and jellyfish (Albikovskaya 1982, 1983). This lack of dependence on a solid substrate may have facilitated their survival in a large southern refugium in the western Atlantic which was likely not suitable for the other two wolffish species. This may help to explain the different pattern seen in northern wolffish.

3.4.2 Diversity levels

In contrast to Atlantic wolffish (Chapter 2) and other slowly evolving fishes (Feutry *et al.* 2014), we identified unique sequences in each of 17 spotted and 19 northern wolffish from the North Atlantic. This is analogous to the patterns previously found in Atlantic cod (Carr & Marshall 2008) and Atlantic herring *Clupea harengus* (Teacher *et al.* 2012). Despite an absence of shared haplotypes, diversity levels were low, particularly in spotted wolffish (π = 0.00045). This is higher than that seen in the endangered speartooth shark (π = 0.00019; Feutry *et al.* 2014), but lower than most other species including the giant squid (π = 0.00066; Winkelmann *et al.* 2013), Atlantic wolffish (π = 0.00106; Chapter 2), and American eel *Anguilla rostrata* (π = 0.00291; Jacobsen *et al.* 2014). The low diversity levels may be in part due to the low sample size; however, the

Atlantic wolffish study had twice as many individuals as the American eel and less than half the diversity, while we had only two more northern wolffish than spotted wolffish yet they had almost twice the diversity, suggesting that sample size plays only a small role in diversity differences.

3.4.3 Distribution of variation

The distribution of variable sites varied across the genome, and differed between the three wolffish species. In spotted wolffish the greatest number of SNPs were seen in ND2 (8 SNPs; 0.0076 substitutions per site) and ND4 (9 SNPs; 0.0065 substitutions per site), while no variation was found in ATP8. The ribosomal RNA loci were much less variable with only 0.0011 and 0.0035 substitutions per site for 12S and 16S respectively, the transfer RNAs were moderately variable with 0.0054 substitutions per site, and the control region showed among the lowest variation with 2 SNPs (0.0024 substitutions per site). In contrast, northern wolffish saw the highest percent of polymorphisms in CYTB (13 SNPs; 0.0112 substitutions per site), followed by ND4L (3 SNPs; 0.0101 substitutions per site) and ND2 (9 SNPs; 0.0086 substitutions per site); there was no variation in ND3. The ribosomal RNA markers were slightly more variable than in the spotted wolffish (0.0021 and 0.0041 substitutions per site), the transfer RNAs less variable (0.0033 substitutions per site), and the control region twice as variable (4 SNPs; 0.0048 substitutions per site). The distribution is again different in the congeneric Atlantic wolffish where the greatest variation was seen in ND5 (41 SNPs; 0.0223 substitutions per site), ND3 (8 SNPs; 0.0228 substitutions per site), and ND2 (22 SNPs; 0.0210 substitutions per site), and the lowest in ATP8 (1 SNP; 0.0060 substitutions per site), CR

(6 SNPs; 0.0072 substitutions per site), and ND6 (5 SNPs; 0.0096 substitutions per site; Chapter 2). Despite the fact that these three species are closely related with few polymorphisms among them, none of them share the same locus as their most variable marker, although ND2 is among the top three in each species.

The distribution of variation corresponded to the pattern seen in the statistical parsimony networks and the neighbour-joining trees (Supplementary Figures 3.1 to 3.4, Chapter 2). In spotted wolffish, the greatest variation was seen in ND2 and ND4, with very little structure in 12S, ATP8, and CR. Northern wolffish had the greatest variation in CYTB, ND5, and 16S, and little or no variation in ATP8 and ND3. In Atlantic wolffish, the greatest variation was in COX1, ND4, ND5, and CYTB, while ATP8 had only a single substitution in a single sample. The loci with the greatest variation were not necessarily the markers that identified the haplogroup structure. In spotted wolffish, ND2, the locus with the highest variation, was the only marker to differentiate the three haplogroups D, E, and F, while the second most variable locus, ND4, was unable to identify any of the groups. At the same time some of the loci with little variation, 12S, COX2, and ND6, were able to identify one of the three haplogroups (F, E, and F, respectively). In northern wolffish the three most variable markers above were all able to identify the two haplogroups (G and H), as were less variable markers such as COX2 and ND6. In Atlantic wolffish only one locus, COX1, was able to identify the three haplogroups A, B, and C (Chapter 2).

3.4.4 Conclusions

Although no phylogeographic structure was detected in either spotted or northern wolffish, there are distinct haplogroups in both species that likely originated before or during the latest Pleistocene glacial period. In spotted wolffish the three haplogroups found were reminiscent of those seen in the sister-species Atlantic wolffish, as the three groups showed strong statistical support but did not correspond to population of origin. The two northern wolffish haplogroups were much more distinct and likely originated earlier. In each case population admixture following glacial isolation is likely responsible for the lack of geographical structure, although incomplete lineage sorting may also have played a role. For both species sample size was small so a more thorough sampling of both the North American and European populations is needed to confirm whether the current group pattern is upheld, if the distribution of haplogroups is random across regions, and to try and identify approximate locations for the putative glacial refugia (i.e., east or west Atlantic). The fact that the pattern of variation in the threatened northern wolffish differs from that seen in spotted and Atlantic wolffish, with two more divergent haplogroups of older origin, suggests that these species may need to be managed differently.

Use of complete mitogenomes together with nuclear markers is the best way to order to ensure that population structure is fully revealed. However, we realise that in many cases this is not feasible due to quality of material, time constraints, and financial limitations. Typically the marker of study has been chosen based on what has been previously used in the study species, what markers have been useful in similar species, or what loci are variable among closely related species. As we have shown here, given the

pattern of variation seen in these two wolffish species, it is important to decide on a species-by-species basis which markers, if any, should be pursued further. For a more extensive population study of spotted wolffish with mtDNA the use of ND2 is necessary to identify the presence of haplogroups, perhaps in combination with CYTB. In northern wolffish a combination of ND5 and CYTB should elucidate much of the variation in this species. Applying as many additional markers as possible will improve resolution.

Acknowledgements

We would like to thank Fisheries and Oceans Canada for providing the tissue samples, and everyone involved in the extraction and amplification of the samples: Dr. HD Marshall, Dr. KA Johnstone, Dr. AT Duggan, Dr. SMC Flynn, L MacDonald, E Wells, O Trela, and A Genge. We thank Dr. P Bentzen and Dr. MR McCusker for providing a northern wolffish sample from their study. This project was supported by a Natural Science and Engineering Research Council Discovery Grant to SMC, and an NSERC Alexander Graham Bell Canada Graduate Scholarship to LAL.

Table 3.1. Primers used for PCR amplification of spotted and northern wolffish. Primer pairs are given – each reaction used one forward (F) and one reverse (R) primer. Primers were designed either in *Gadus morhua* (g) or *Anarhichas lupus* (w or str). Annealing temperatures (T_A) and primer sources (ref) are given.

Primer Name	**Sequence (5' to 3')**	**T_A (°C)**	**ref**
g01F	CTGAAGATATTAGGATGGACCCTAG	49	1
g02R	CTATTCATTTCACAGGCAACCAGCT		1
g03F	ACCCCGAAACTGAGCGAGCTACTCC	55	1
g03R	TAAGCCCTCGTGATGCCATTCATAC		1
g04F	TTTACCAAAAACATCGCCTCTTG	49	1
g04R	TGAACCTCTGTAGAAAGGGCTTAGG		1
w04F	TTCAGACCGGAGTAATCCAGGTCAG	56	2
g05R	ATGTTCGGGGTATGGGCCCCAAGAGC		1
w06F	GTGCTTCCACTACACCACTTCCTAG	54	2
w07R	CTGGTTTGAGCGCTTAGCTGTTAAC		2
g07F	AAACTAGACCAAGGGCCTTCAAAGC	55	1
w08R	CAGAGGTAAAGTAAGCGCGTGTGTC		2
w08F1	CCCTTCACCTAGCAGGAATTTCTTCAATCC	53	2
g08R	TAACCCACAATTCTGCCTTGACAAG		1
w08F	ACAACGAATGTGGAGTGACTACACG	49	2
g09R	ACCCATATTAGCTTCTTAGTGAGG		1
w09F	GGCCATCAGTGGTACTGAAGCTATG	49	2
g10R	AGAGGGCGAATGAATAAACTAATTG		1
w10F1	ATGCGAAACCAACCAACCCATGCTC	58	2
w11R	TTGATCTCCTCAGGGTAGCGGGAGTAGTAG		2

w11F	CTACACTTGACCCATTTGAAGTGCC	56	2
w12R	TGTAGACCCTAGTTGCAAGGTCCAC		3
w12F	GCTAGGATTAACCGGCCTGGCCTTTCAT	54	3
wND43R	CGATGAGCGACTTCAAATCTGTGTG		3
wND44F	TAGACCCATTAACCCAGGAGCTAAG	53	3
wND45R	AAGACCAGCGGATGAGCTGTTATCC		3
w14F2	GAACATCTTCTTATGGCCCTTCACC	53	2
Str14R	GGTTGTAGACGACGGCTTGAAGAGC		3
w15F	TCGCCATAGTCATTCTAGTGACAGC	53	2
w16R	GCTTGTTTGTTAGGGAGGCTAGTTC		2
w16F	CCACAGCTTGAATGACGAGCAAGAC	54	2
w17R	TAACGCGAGGATTAAGTCGAGGAAC		2
w17F	TTTACCACTCCACCACTTCTCCAAC	54	2
w18R	AGCAAAGGCCGAGTAGGGAACCAAAGTTTC		2
w18F	CCGCTACAACAACCAACCCTAAAGC	55	2
wcytbR	GTTGTCAACTGAGAAGCCTCCTCAG		2
g19F	GAGGAGGTTTCTCAGTAGATAATGC	49	1
g19R	GTTTAATTTAGAATTCTAGCTTTGG		1
w19F	TGAATTGGCGGTATACCCGTAGAAC	54	3
g20R	GGCAGGACATTAAGGGCATTCTCAC		1

1 (Coulson *et al.* 2006)

2 (Johnstone *et al.* 2007)

3 Chapter 2

Table 3.2. Distribution of variable sites across the mitochondrial genome in spotted and northern wolffish. The length of each region and the total number of variable sites (VS) are given. The position of the change is given for coding regions (1,2, or 3), and the amino acid change for any non-synonymous substitution (Sub).

Region	**Length**	**Spotted**					**Northern**				
	(bp)	**VS**	**1**	**2**	**3**	**Sub**	**VS**	**1**	**2**	**3**	**Sub**
12S	947	1	n/a	n/a	n/a		2	n/a	n/a	n/a	
16S	1,696	6	n/a	n/a	n/a		7	n/a	n/a	n/a	
ND1	975	3	0	0	3		4	0	1	3	Val/Ala
ND2	1,047	8	2	0	6	Thr/Ala Ile/Val	9	3	0	6	Val/Ile Ala/Thr
COX1	1,551	3	1	0	2	Phe/Val	6	0	0	6	
COX2	699	2	0	0	2		1	0	0	1	
ATP8	168	0	0	0	0		1	0	0	1	
ATP6	684	1	0	0	1		2	0	0	2	
COX3	786	1	1	0	0	Ala/Thr	3	0	0	3	
ND3	351	1	0	0	1		0	0	0	0	
ND4L	297	1	0	0	1	Gln/His	3	0	0	3	
ND4	1,386	9	2	0	7	Val/Ile	4	0	1	3	Ser/Thr
ND5	1,839	3	0	0	3		10	1	0	9	Ile/Leu
ND6	522	2	0	0	2		1	0	0	1	
CYTB	1,161	5	0	1	4	Arg/Gln	13	1	0	12	Asp/Asn
tRNA	1,504	7	n/a	n/a	n/a		5	n/a	n/a	n/a	
CR	829	2	n/a	n/a	n/a		4	n/a	n/a	n/a	
Other	68 - 74	6	n/a	n/a	n/a		6	n/a	n/a	n/a	
Total		**61**	**6**	**1**	**32**		**81**	**5**	**2**	**50**	

Table 3.3. Overall Φ_{ST} values and corresponding p-values for AMOVAs (100,172 permutations) run on the individual loci for 17 spotted wolffish and 19 northern wolffish. Significant p-values are given in bold.

	Spotted		Northern	
Region	**Φ_{ST}**	**p**	**Φ_{ST}**	**p**
12S	0.000	1.000	0.000	1.000
16S	0.090	0.216	0.135	0.155
ND1	1.000	0.118	0.262	0.272
ND2	0.041	0.334	**0.306**	**0.039**
COX1	0.322	0.098	0.000	0.645
COX2	0.026	0.401	0.310	0.084
ATP8	0.000	1.000	0.000	0.420
ATP6	0.000	1.000	0.243	0.122
COX3	0.000	0.648	0.000	0.824
ND3	0.000	1.000	0.000	1.000
ND4L	0.000	1.000	0.266	0.085
ND4	0.000	0.453	0.000	0.757
ND5	0.295	0.200	0.147	0.129
ND6	0.000	1.000	0.000	0.735
CYTB	0.000	1.000	0.109	0.153
CR	0.000	0.646	**0.269**	**0.037**

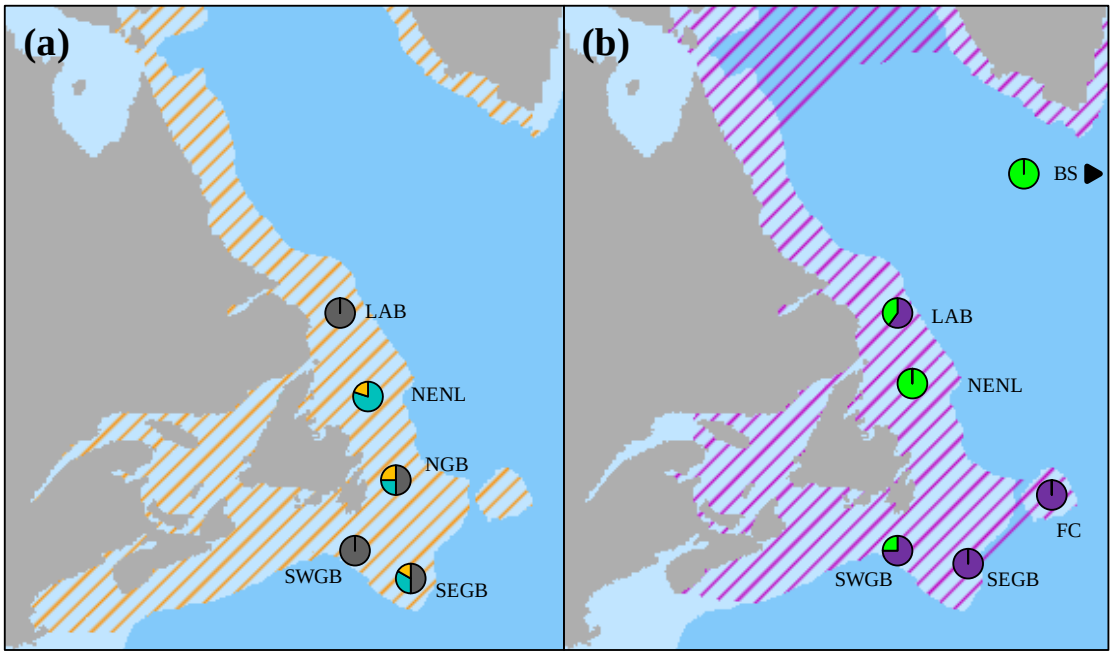


Figure 3.1. Distribution (shaded) of spotted and northern wolffish and the seven sampling locations in the North Atlantic. The pie charts represent the haplogroups identified in BAPS v6. (a) Northwest Atlantic sampling locations for spotted wolffish: LAB (Labrador, n=1), NENL (northeastern Newfoundland, n=5), NGB (northern Grand Banks, n=4), SEGB (southeastern Grand Banks, n=6), and SWGB (southwestern Grand Banks, n=1). (b) Sampling locations for northern wolffish: LAB (n=5), NENL (n=1), FC (Flemish Cap, n=2), SEGB (n=6), SWGB (n=4), and BS (Barents Sea, n=1). Figure modified from FishBase (2013a, 2013b).

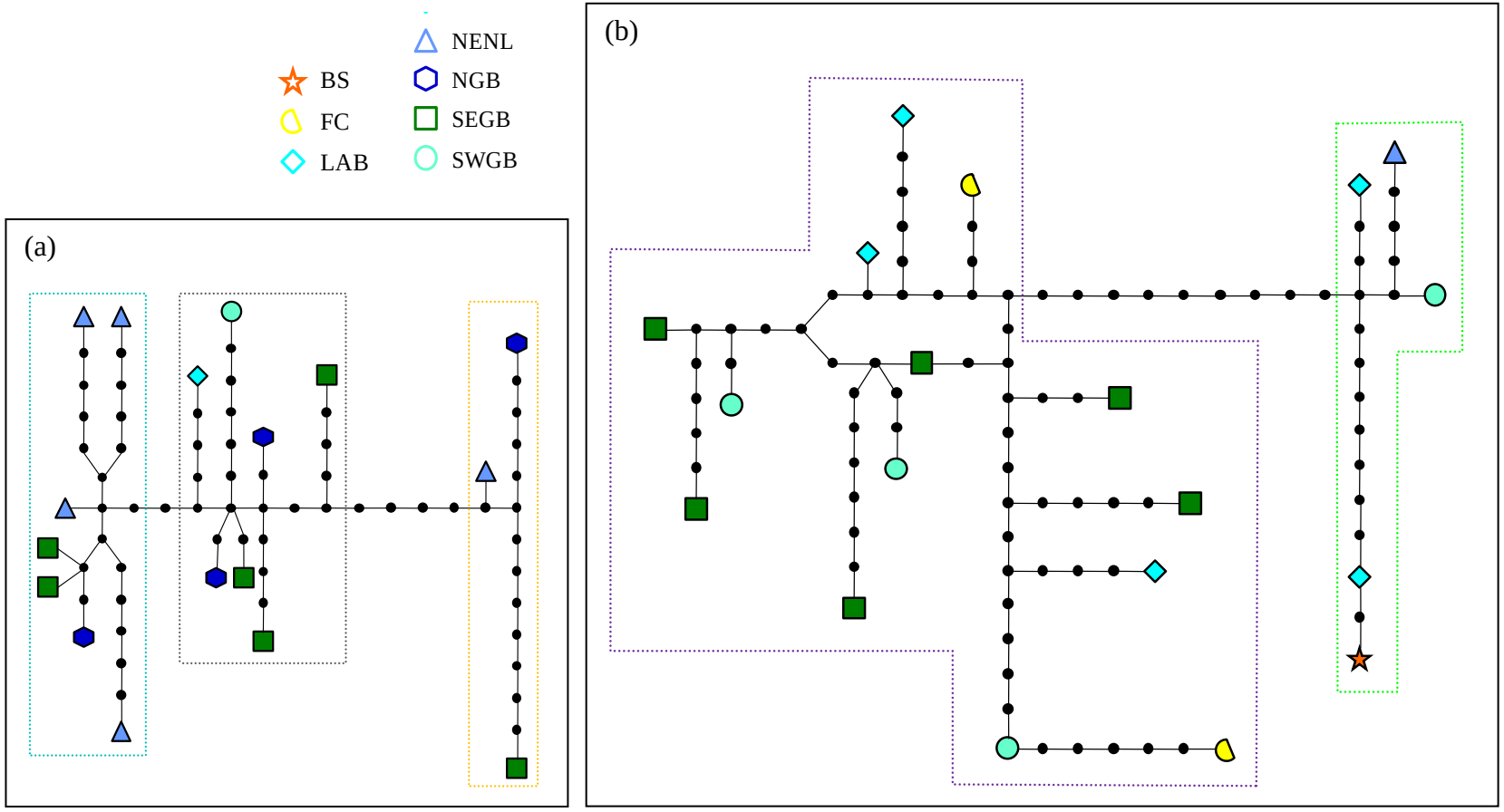


Figure 3.2. Statistical parsimony network of complete mtDNA sequences for (a) spotted and (b) northern wolffish. Each symbol represents an individual, colour-coded by sampling location. The black dots are inferred or intermediate haplotypes, separated by one nucleotide change each. The dashed boxes correspond to the haplogroups found in BAPS v6.

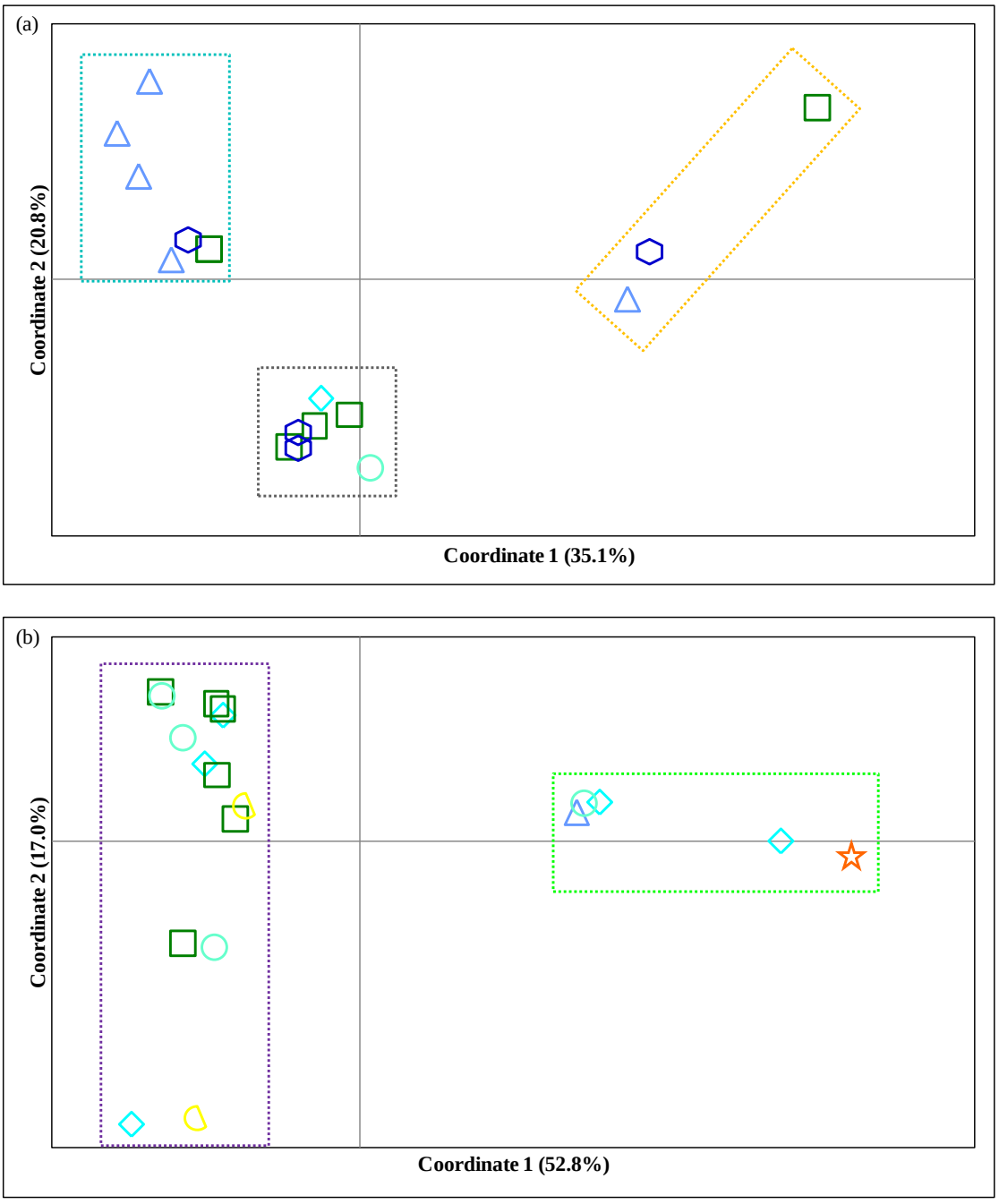
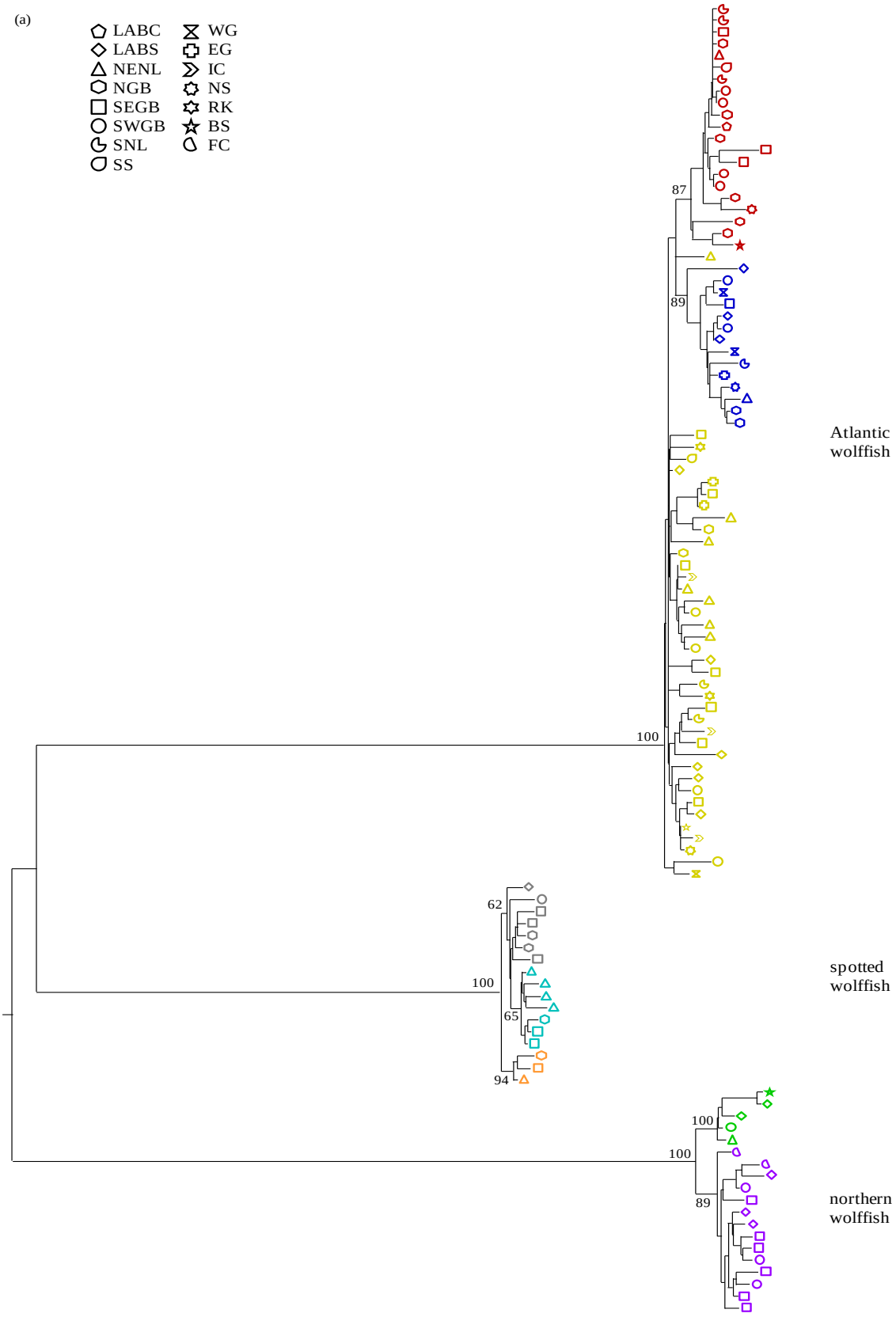


Figure 3.3. Principal coordinates analysis of complete mtDNA genomes from (a) spotted and (b) northern wolffish. Sampling locations are coded as in Figure 3.2; dashed boxes outline haplogroups identified by BAPS v6.


(a)
LABC
LABS
NENL
NGB
SEGB
SWGB
SNL
SS
WG
EG
IC
NS
RK
BS
FC
87
89
100
62
100
65
94
100
100
89
Atlantic wolffish
spotted wolffish
northern wolffish

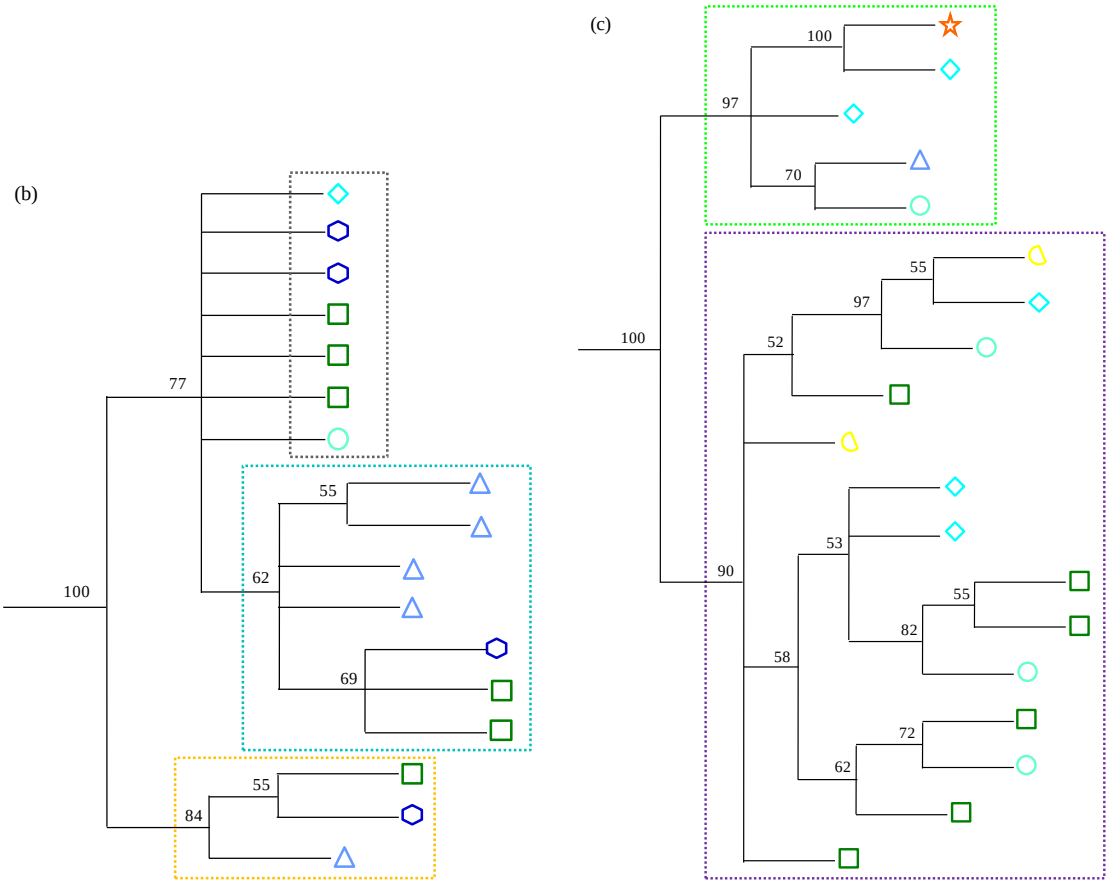


Figure 3.4. (a) Neighbour-joining analysis of complete mtDNA genomes from the three North Atlantic wolffish species. Bootstrap values (10,000 replications) are given for the major haplogroups. Samples are colour-coded by haplogroups. NJ analysis for (b) spotted and (c) northern wolffish. Bootstrap values based on 10,000 replications are given. Sampling locations are colour-coded by sampling location as in Figure 3.2; dashed boxes outline haplogroups identified by BAPS v6. The rooting of each species is the same when either of the other two Northwest Atlantic species is used.

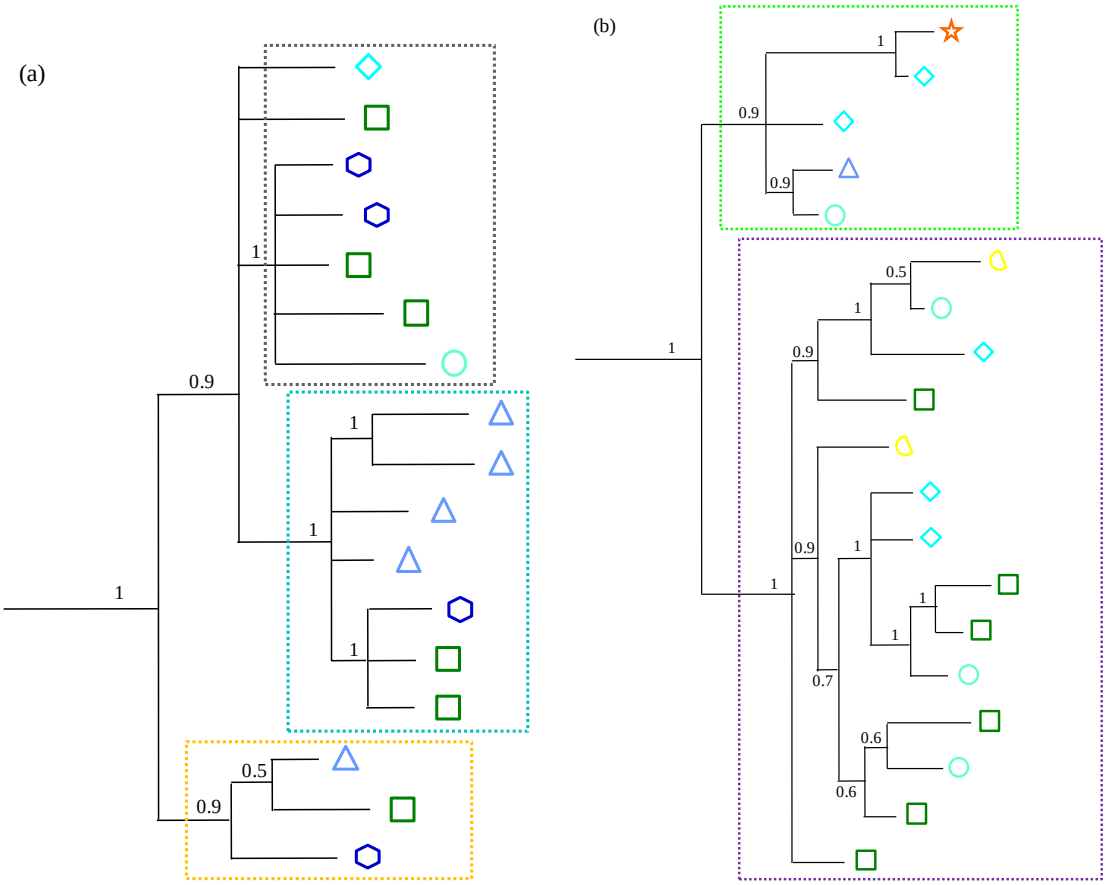
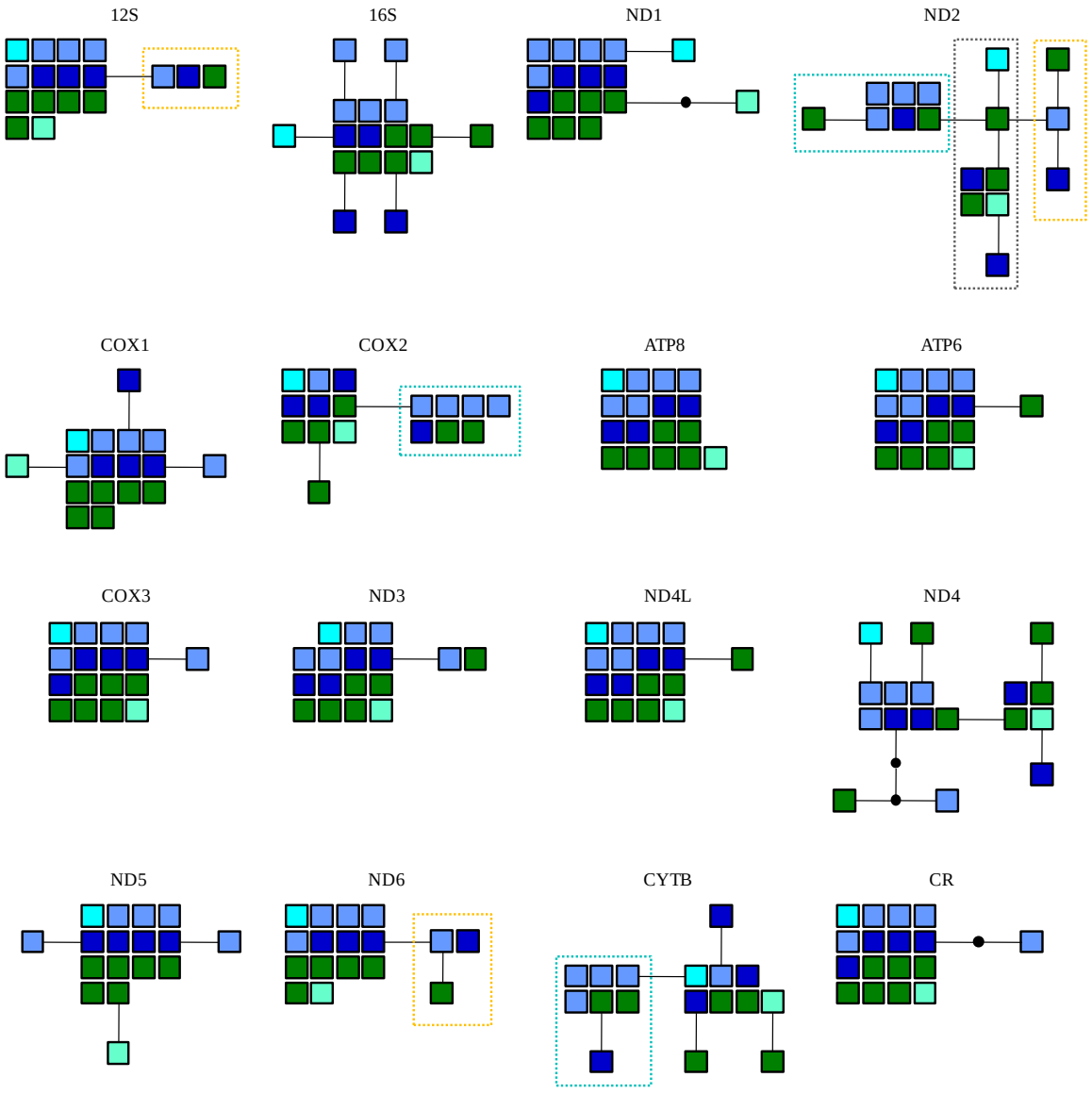
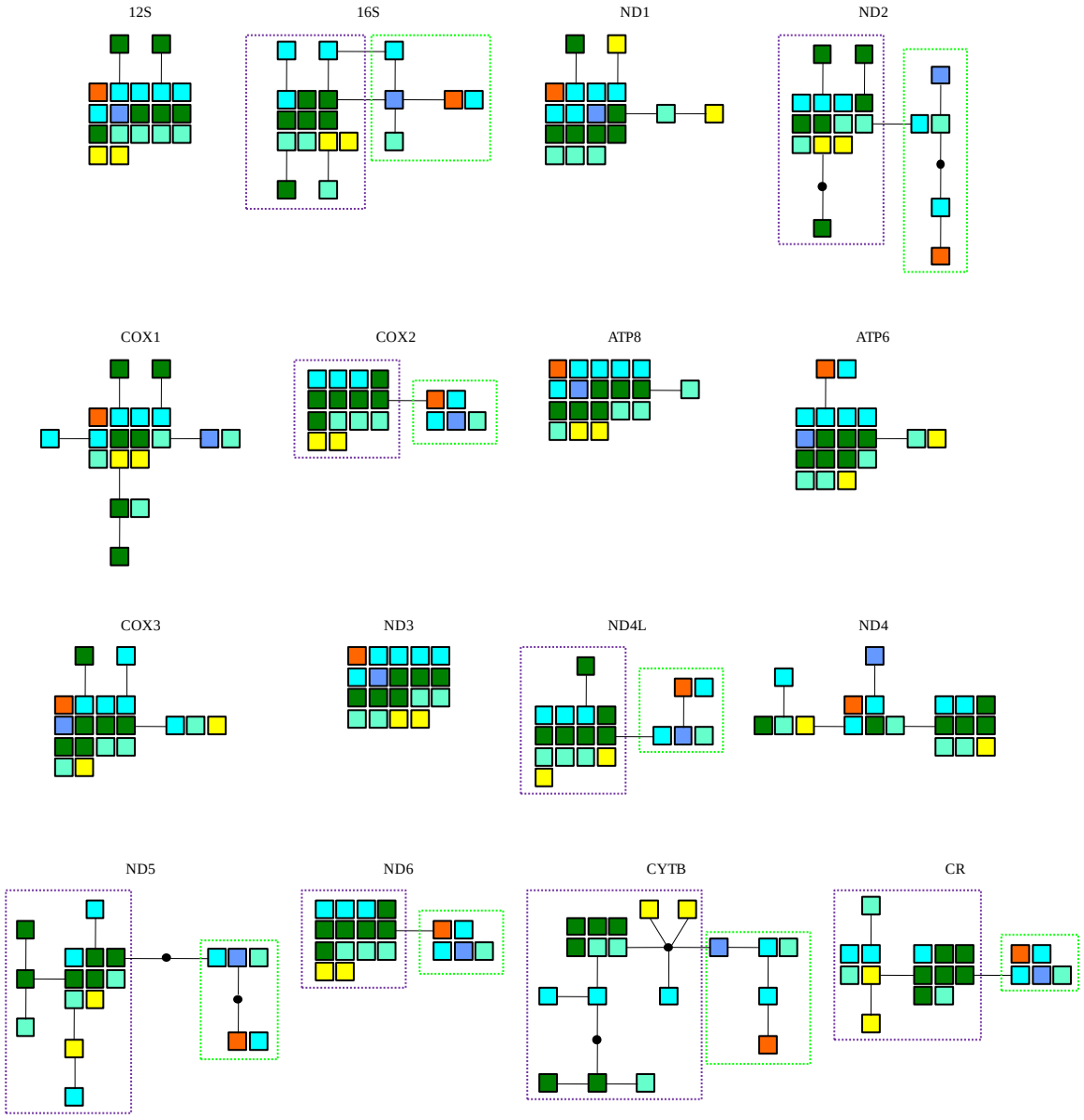


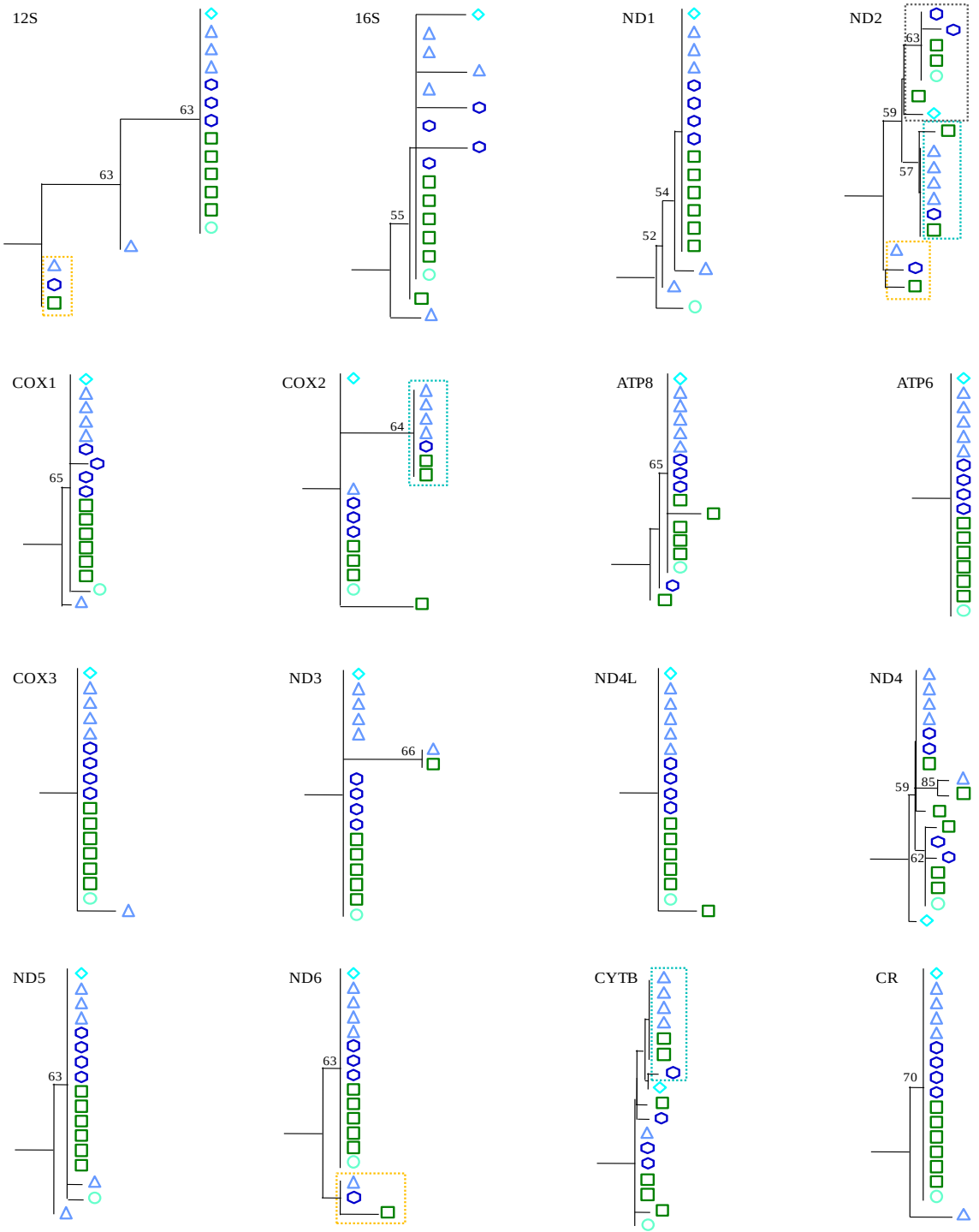
Figure 3.5. Bayesian analysis of complete mtDNA genomes from (a) spotted and (b) northern wolffish. Posterior probabilities are given. Sampling locations are coded as in Figure 3.2; dashed boxes outline haplogroups identified by BAPS v6. The rooting of each species is the same when either of the other two Northwest Atlantic species is used.



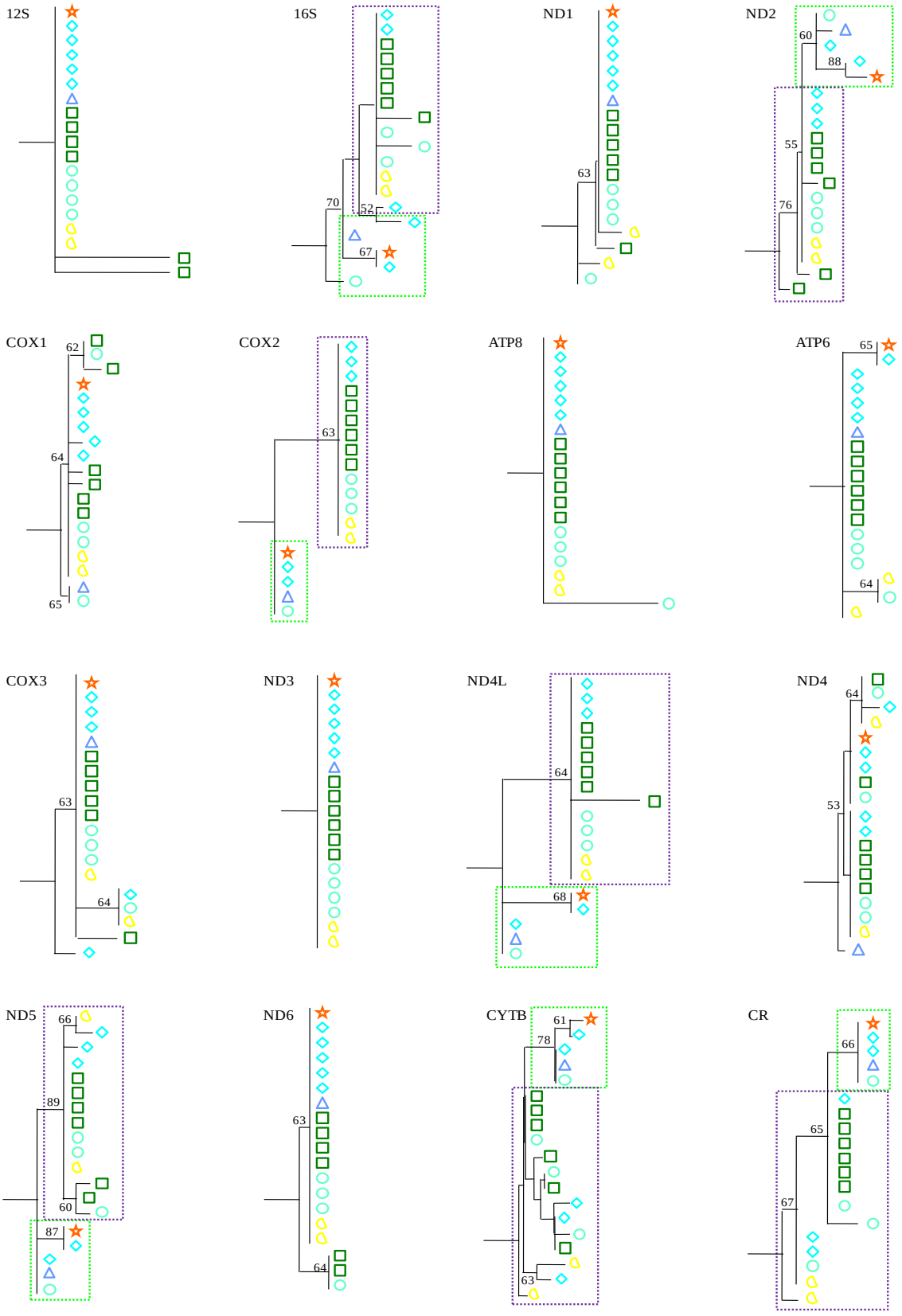
Supplementary Figure 3.1. Statistical parsimony networks for each individual locus for 17 spotted wolffish samples. Shared haplotypes are represented by clusters of squares. Sampling locations are colour-coded as in Figure 3.2; dashed boxes outline haplogroups identified by BAPS v6.



Supplementary Figure 3.2. Statistical parsimony networks for each individual locus for 19 northern wolffish samples. Shared haplotypes are represented by clusters of squares. Sampling locations are colour-coded as in Figure 3.2; dashed boxes outline haplogroups identified by BAPS v6.



Supplementary Figure 3.3. NJ analysis for each individual locus for 17 spotted wolffish samples rooted with the congeneric Atlantic and northern wolffish. Bootstrap support > 50% (10,000 replicates) is given. Sampling locations are coded as in Figure 3.2, and dashed boxes are given where possible to show the haplogroups identified in BAPS v6.


12S
16S
70
52
67
ND1
63
ND2
60
88
55
76
COX1
62
64
65
COX2
63
ATP8
ATP6
65
64
COX3
63
64
ND3
ND4L
64
68
ND4
64
53
ND5
66
89
60
87
ND6
63
64
CYTB
61
78
63
CR
66
65
67

Supplementary Figure 3.4. NJ analysis for each individual locus for 19 northern wolffish samples rooted with the congeneric Atlantic and spotted wolffish. Bootstrap support > 50% (10,000 replicates) is given. Sampling locations are coded as in Figure 3.2, and dashed boxes are given where possible to show the haplogroups identified in BAPS v6.

References

Albikovskaya LK (1982) Distribution and abundance of Atlantic wolffish, spotted wolffish and northern wolffish in the Newfoundland area. *NAFO Scientific Council Studies* **3**, 29-32.

Albikovskaya LK (1983) Feeding characteristics of wolffishes in the Labrador-Newfoundland region. *NAFO Scientific Council Studies* **6**, 35-38.

Amos W, Harwood J (1998) Factors affecting levels of genetic diversity in natural populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **353**, 177-186.

Baker K, Devine J, Haedrich RL (2009) Deep-sea fishes in Canada's Atlantic: population declines and predicted recovery times. *Environmental Biology of Fishes* **85**, 79-88.

Barsukov VV (1959) *The Wolffish (Anarhichadidae)* Indian National Scientific Documentation Centre (1972 translation for the Smithsonian Institution), New Delhi, India.

Bernatchez L, Guyomard R, Bonhomme F (1992) DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. *Molecular Ecology* **1**, 161-173.

Bouckaert R, J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard M, Rambaut A, Drummond A (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537.

Carr S, Marshall H (2008) Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic cod (*Gadus morhua*): origins of the "codmother", transatlantic vicariance and midglacial population expansion. *Genetics* **180**, 381-389.

Chapin FS, Zavaleta ES, Eviner VT, Naylor RL, Vitousek PM, Reynolds HL, Hooper DU, Lavorel S, Sala OE, Hobbie SE, Mack MC, Diaz S (2000) Consequences of changing biodiversity. *Nature* **405**, 234-242.

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

Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539.

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* **205**, 19-31.

COSEWIC (2000) COSEWIC assessment and status report on the Atlantic wolffish, *Anarhichas lupus*, in Canada, p. vi + 21 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2001a) COSEWIC assessment and status report on the northern wolffish, *Anarhichas denticulatus*, in Canada, p. vi + 21 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2001b) COSEWIC assessment and status report on the spotted wolffish, *Anarhichas minor*, in Canada, p. vi + 22 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2012a) COSEWIC assessment and status report on the northern wolffish *Anarhichas denticulatus* in Canada p. x + 41 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2012b) COSEWIC assessment and status report on the spotted wolffish *Anarhichas minor* in Canada p. x + 44 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

Coulson MW, Marshall HD, Pepin P, Carr SM (2006) Mitochondrial genomics of gadine fishes: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome* **49**, 1115-1130.

Devine J, Baker K, Haedrich R (2006) Fisheries: deep-sea fishes qualify as endangered. *Nature* **439**, 29.

Excoffier L, Lischer H (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

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

Feutry P, Kyne P, Pillans R, Chen X, Naylor G, Grewe P (2014) Mitogenomics of the speartooth shark challenges ten years of control region sequencing. *BMC Evolutionary Biology* **14**, e232.

FishBase (2013a) Reviewed distribution maps for *Anarhichas denticulatus* (northern wolffish), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

FishBase (2013b) Reviewed distribution maps for *Anarhichas minor* (spotted wolffish), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

Frank K, Petrie B, Choi J, Leggett W (2005) Trophic cascades in a formerly cod-dominated ecosystem. *Science* **308**, 1621-1623.

Frankham R (1995) Conservation genetics. *Annual Review of Genetics* **29**, 305-327.

Hiddink J, MacKenzie B, Rijnsdorp A, Dulvy N, Nielsen E, Bekkevold D, Heino M, Lorance P, Ojaveer H (2008) Importance of fish biodiversity for the management of fisheries and ecosystems. *Fisheries Research* **90**, 6-8.

Hilborn R, Quinn T, Schindler D, Rogers D (2003) Biocomplexity and fisheries sustainability. *Proceedings of the National Academy of Sciences USA* **100**, 6564-6568.

Hooper DU, Chapin FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, Schmid B, Setälä H, Symstad AJ, Vandermeer J, Wardle DA (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs* **75**, 3-35.

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

Hutchings JA (2000) Collapse and recovery of marine fishes. *Nature* **406**, 882-885.

Jacobsen M, Pujolar J, Gilbert M, Moreno-Mayar J, Bernatchez L, Als T, Lobon-Cervia J, Hansen M (2014) Speciation and demographic history of Atlantic eels (*Anguilla anguilla* and *A. rostrata*) revealed by mitogenome sequencing. *Heredity* **113**, 432-442.

Johnstone KA, Marshall HD, Carr SM (2007) Biodiversity genomics for species at risk: patterns of DNA sequence variation within and among complete mitochondrial genomes of three species of wolffish (*Anarhichas* spp.). *Canadian Journal of Zoology* **85**, 151-158.

Jørgensen OA, Bastardie F, Eigaard OR (2014) Impact of deep-sea fishery for Greenland halibut (*Reinhardtius hippoglossoides*) on non-commercial fish species off West Greenland. *ICES Journal of Marine Science* **71**, 845-852.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Väinölä R, Viard F, Wares J (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* **89 (Sup)**, S108-S122.

Marshall HD, Baker AJ (1997) Structural conservation and variation in the mitochondrial control region of fringilline finches (Fringilla spp.) and the greenfinch (Carduelis chloris). *Molecular Biology and Evolution* **14**, 173-184.

McCusker MR, Bentzen P (2010) Phylogeography of 3 North Atlantic wolffish species (*Anarhichas* spp.) with phylogenetic relationships within the family Anarhichadidae. *Journal of Heredity* **101**, 591-601.

McCusker MR, Bentzen P (2011) Limited population structure in northern and spotted wolffishes (*Anarhichas denticulatus* and *A. minor*) despite low apparent dispersal potential. *Marine Biology* **158**, 1869-1878.

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

O'Dea NR, Haedrich RL (2002) A review of the status of the Atlantic wolffish, *Anarhichas lupus*, in Canada. *Canadian Field-Naturalist* **116**, 423-432.

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

Peakall R, Smouse P (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* **28**, 2537-2539.

Peakall ROD, Smouse PE (2006) GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Pflaumann U, Sarnthein M, Chapman M, d'Abreu L, Funnell B, Huels M (2003) Glacial North Atlantic: sea-surface conditions reconstructed by GLAMAP 2000. *Paleoceanography* **18**, 1-28.

Ronquist F, Huelsenbeck J (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**.

Schindler D, Hilborn R, Chasco B, Boatright C, Quinn T, Rogers L, Webster M (2010) Population diversity and the portfolio effect in an exploited species. *Nature* **465**, 609-613.

Shaw J (2006) Palaeogeography of Atlantic Canadian continental shelves from the Last Glacial Maximum to the present, with an emphasis on Flemish Cap. *Journal of Northwest Atlantic Fisheries Science* **37**, 119-126.

Swofford D (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Teacher A, André C, Merilä J, Wheat C (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evolutionary Biology* **12**, e248.

Templeman W (1986) Spotted forms of the northern wolffish (*Anarhichas denticulatus*). *Journal of Northwest Atlantic Fisheries Science* **7**, 77-80.

Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson A (1991) African populations and the evolution of human mitochondrial DNA. *Science* **253**, 1503-1507.

Wang Z, Shen X, Liu B, Su J, Yonezawa T, Yu Y, Guo S, Ho S, Vilà C, Hasegawa M, Liu J (2010) Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal. *Journal of Biogeography* **37**, 2332-2344.

Watling L, Norse E (1998) Disturbance of the seabed by mobile fishing gear: a comparison to forest clearcutting. *Conservation Biology* **12**, 1180-1197.

Winkelmann I, Campos P, Strugnell J, Cherel Y, Smith P, Kubodera T, Allcock L, Kampmann M-L, Schroeder H, Guerra A (2013) Mitochondrial genome diversity and population structure of the giant squid *Architeuthis*: genetics sheds new light on one of the most enigmatic marine species. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20130273.

CHAPTER FOUR

Mitogenomics of Atlantic cod (*Gadus morhua*): variation in trans-Atlantic, Baffin Island, and Baltic Sea populations

L.A. Lait[1], H.D. Marshall[1], and S.M. Carr[1]

[1]Department of Biology, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, A1B 3X9

4.1 Introduction

An understanding of population genetic structure, particularly in those species threatened with extinction, is critical to conservation and preservation of biodiversity. Genetic structure can be influenced by a number of factors including habitat loss, changes to environmental conditions, overexploitation, and the introduction of invasive species (Wilson 1988; Frankham 1995). While all of these factors have contributed to population declines and extinctions in aquatic and terrestrial environments, the marine environment has been particularly impacted by overexploitation and environmental change (Lluch-Belda *et al.* 1992; Alheit & Hagen 1997; Hoelzel 1999; Carpenter *et al.* 2008; Schultz *et al.* 2009). For example, the most recent Pleistocene glaciations (ca. 110 - 12 kya) disrupted ocean currents, decreased sea temperatures, and caused considerable sea level drops in the North Atlantic Ocean (Pielou 1991; Shaw 2006). More recently, heavy fishing pressures have caused severe population declines in many marine fish species (O'Dea & Haedrich 2002; Rose 2007).

Unlike terrestrial and freshwater species that frequently exhibit strong population genetic structure from barriers to gene flow such as rivers, waterfalls, mountains, and deserts, marine species typically experience fewer and less formidable physical obstacles. They often have wide distributions, large population sizes, high fecundity, and extensive gene flow (Palumbi 1992). Some species are restricted due to habitat requirements or feeding habits (e.g., tropical reef fishes), and exhibit clear population genetic structure (Shulman & Bermingham 1995). In contrast, pelagic species tend to display weaker

population structure in the absence of natal philopatry (Ward 1995; Vis *et al.* 1997; Castro *et al.* 2007; Carr *et al.* 2015).

Atlantic cod *Gadus morhua* (L.1758) are found along the continental shelves of the North Atlantic Ocean. The distribution extends as far south as Cape Hatteras (North Carolina), through New England and Atlantic Canada in the west, across the continental shelves of Greenland and Iceland, through the North and Baltic Seas to the Barents Sea in the east (Figure 4.1). Cod in the main part of their range are demersal omnivores that commonly migrate in large aggregations between offshore spawning grounds and inshore feeding grounds to find their main food source, capelin *Mallostus villosus*. Populations at the southern extreme tend to be more sedentary, feeding on herring, sand eels and invertebrates (Powles 1958; Rose 2007). There are also three landlocked Arctic populations in North America found in coastal fjords on Baffin Island, Nunavut, Canada: Lake Ogac off Frobisher Bay, and Lakes Qasigialiminiq and Tariujarusiq off Cumberland Sound (Hardie 2003). All three lakes consist of a freshwater layer at the surface and saltwater of varying concentrations underneath, a condition called meromictic (McLaren 1967; Hardie 2003). Their populations are estimated at no more than 500 - 1,000 cod each. The fish exhibit significantly different life history characteristics than their marine counterparts: they survive in extremely cold and seasonally frozen waters, they mature at a much greater size and age, and they feed on echinoderms, molluscs, and polychaetes, as well as cannibalising other cod (Patriquin 1967; Hardie 2004; Hardie & Hutchings 2011).

Atlantic cod previously supported one of the most important commercial fisheries in the world (Halliday & Pinhorn 1996). Heavy fishing in the latter half of the 20th century led to severe population declines first in Europe and then in North America. The

northeast Newfoundland and Labrador stock (NAFO 2J3KL) declined by > 99%, and other areas by > 90% (COSEWIC 2003; Rose 2007). As a result a moratorium on commercial cod fishing was introduced in Canadian waters in 1992, although a food fishery and some limited local commercial fishing has since been reopened (Rose 2007). Populations in the Northwest Atlantic are categorised as Vulnerable on the IUCN Red List (International Union for Conservation of Nature 2011), and the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) made a recommendation of Special Concern in 1998 for the species as a whole, which was then upgraded in 2003 to Endangered for the Newfoundland and Labrador populations, Threatened for the North Laurentian populations, and Special Concern for the Maritimes and Arctic populations (COSEWIC 2003). In 2010 Atlantic cod were reassessed and the North Laurentian, South Laurentian, and Southern populations were upgraded (or reassigned) to Endangered (COSEWIC 2010).

Atlantic cod in the Northwest Atlantic Ocean have traditionally been managed as a single population (COSEWIC 2003), although following recommendations based on meristic analysis of vertebral averages (Templeman 1979, 1981), cod around Newfoundland and Labrador were managed as five stocks separated along North Atlantic Fisheries Organization (NAFO) divisions (Figure 4.2). In other regions cod are delineated by marine area (e.g., Baltic Sea, Barents Sea, North Sea, and Iceland) with numerous regions subdivided further (e.g., Northeast Arctic vs. Norwegian coast). Considerable effort has been made to determine whether these stocks represent biological populations; the question remains unanswered.

Extensive work has been done on Atlantic cod to uncover population genetic structure. Tagging studies found evidence of local stock structure with either natal homing or limited movement (Gulland & Williamson 1962; Lear 1984; Hovgård & Christensen 1990), whereas long-distance migration has been recorded in Iceland (Jónsson 1996) and the North Sea (Gulland & Williamson 1962). Early studies found significant frequency differences among geographic locations with haemoglobin (Frydenberg *et al.* 1965; Sick 1965b, a; Møller 1968; Cross & Payne 1978; Gjøsæter *et al.* 1992), although selection at the haemoglobin marker and temperature differences have since explained many of these results (Mork & Sundnes 1985; Jamieson & Birley 1989). Trans-Atlantic clines are evident in protein loci (Jamieson 1967; Mork *et al.* 1985), mitochondrial cytochrome *b* (Carr & Marshall 1991a; Sigurgíslason & Árnason 2003), and nuclear markers (Pogson *et al.* 1995), although results at the local scale are equivocal. Extensive mtDNA analyses have found little or no structure within North America (Carr *et al.* 1995), Iceland (Árnason *et al.* 1992; Árnason *et al.* 2000), the Faroe Islands (Sigurgíslason & Árnason 2003), Norway (Árnason & Pálsson 1996), or the Baltic Sea (Árnason *et al.* 1998). Microsatellite studies, however, suggest range-wide isolation-by-distance (O'Leary *et al.* 2007b), and significant structure in North America (e.g., Flemish cap, Bentzen *et al.* 1996; inshore versus offshore, Ruzzante *et al.* 1996; Beacham *et al.* 2002; Arctic populations, Hardie *et al.* 2006), Norway (Knutsen *et al.* 2003), and the North Sea (Hutchinson *et al.* 2001). Genome scans of nuclear SNP markers have suggested the presence of local adaptation and temperature gradients at both large and small scales (Nielsen *et al.* 2009; Bradbury *et al.* 2010; Bradbury *et al.* 2013; Árnason & Halldórsdóttir 2015; Berg *et al.* 2016). A recent study based on sequence variation in

complete mitochondrial genomes (Carr & Marshall 2008a) identified a much deeper time scale of population divergence than previously suspected, as well as paraphyly of short cytochrome *b* genotypes within a pattern of six major lineages. The study was limited to a small number of fish (n = 32) from four populations (three Northwest Atlantic, one Europe); genomic patterns suggesting older lineages in the Northwest Atlantic may have been an artefact of undersampling of the Northeast Atlantic Ocean.

The different results question the utility of different molecular markers in fully elucidating population genetic structure and evolutionary history in species. Genetic markers vary in resolution, inheritance, and power to identify genetic structure. Mitochondrial DNA is commonly employed as it is (usually) uniparentally inherited, free from the complications of recombination, mutates at a predictable and measurable rate, and is in high copy number thus easy to isolate and replicate even from poor quality material. Nuclear markers are biparentally inherited, undergo recombination, and have variable mutation rates and mechanisms. In particular, hypervariable markers such as microsatellites can provide higher resolution due to high mutations rates, and are not affected by the sex-biased dispersal evident in many organisms. Both mtDNA and microsatellites, therefore, are excellent vehicles for population genetic studies. Recently, complete mitochondrial sequences (usually 16 - 17 kbp) have proven useful in clarifying the phylogenetic relationship among species (Horai *et al.* 1995; Cooper *et al.* 2001; Inoue *et al.* 2001; Miya *et al.* 2001; Miya *et al.* 2003; Coulson *et al.* 2006; Arnason *et al.* 2008), and in intraspecific studies of fish (Carr & Marshall 2008a; Teacher *et al.* 2012; Feutry *et al.* 2014), humans (Tanaka *et al.* 2004; Pope *et al.* 2011), and other mammals (Stone *et al.* 2010; Knaus *et al.* 2011). The use of the complete mitochondrial genome provides a

potential new direction for identifying previously undetected variation (Avise 2004; Carr & Marshall 2008a).

The present study uses sequence variation in complete mitogenomes (16,576 bp) to measure the extent of population genetic structure in Atlantic cod. A previous mitogenomic study showed a much greater degree and more highly structured level of inter-individual variation than had been seen with short cytochrome *b* sequences (Carr & Marshall 2008a). A deficiency of that study was its limited sampling, especially outside North American waters. This study extends the previous study by inclusion of additional samples from the Northwest Atlantic, as well as the Baltic and Barents Seas, and a landlocked Arctic fjord population on Baffin Island (Lake Qasigialiminiq). Previous intraspecific mitogenomic studies have found patterns in which every individual has a unique mitogenomic sequence (Carr & Marshall 2008a; Carr & Marshall 2008b; Pope *et al.* 2011; Teacher *et al.* 2012) or where multiple individuals share haplotypes (Morin *et al.* 2010; Knaus *et al.* 2011; Shamblin *et al.* 2012; Feutry *et al.* 2014). I predict that the cod will continue to show unique genome sequences with a small number of changes between individuals. The application of mitogenomics to the question of population genetic structure, in combination with previous nuclear marker results, will allow a more complete picture of the current and past structure of Atlantic cod in the North Atlantic Ocean. This will provide crucial data for informing management and conservation decisions.

4.2 Materials and Methods

4.2.1 Sample collection

A total of 153 Atlantic cod from 14 sampling locations are included in this study (Table 4.1). Between January 1992 and July 1993, samples were collected by Fisheries and Oceans Canada from eight sampling locations in the Northwest Atlantic and by commercial trawlers from the Norwegian coast (Figure 4.1). The Baltic Sea samples were collected by individuals at the Sopot Institute for Oceanology (Sopot, Poland). The three Labrador and Baffin Island locations were collected by local fishermen between September 2013 and November 2014 (see Acknowledgements). Thirty-two samples from four locations (Norwegian coast, Hawke Channel, northern Grand Banks, and Flemish Cap) were previously analysed and are included here (Genbank Accession Numbers EU877710 – EU877741; Carr & Marshall 2008a).

4.2.2 DNA extraction and amplification

DNA from 72 samples had previously been extracted and amplified by members of the Carr lab using the methods below (the “ArkChip” and "C&M" samples; see Table 4.1). DNA was extracted from heart tissue using the Qiagen DNeasy Blood and Tissue Kit according to the manufacturer's protocol as described in Section 2.2.2 (Qiagen, Hilden, Germany). The complete mitochondrial genome was amplified in series of 23 - 24 standard PCR fragments ranging from 534 bp to 1,345 bp, with an average overlap of 169 bp (88 bp to 273 bp). Primers were originally designed for use in the Atlantic cod (Coulson *et al.* 2006) or were designed for this study based previous cod sequences. The

control region for the 32 C&M mitogenomes was amplified using g20F with g20-2R1 and g20-2F1 with g20R. Primer details are given in Table 4.2.

PCR reactions were carried out using Qiagen PCR kits with either Taq, HotStar Taq, or TopTaq DNA polymerase. The reactions (25 μL) contained 10X PCR buffer, 2 mM $MgCl_2$ (Qiagen), 0.2 mM dNTP (Qiagen), 0.4 μM forward and reverse primer, 1 unit DNA polymerase (Qiagen), and 2 μL of the genomic DNA extract. The thermal cycling program was as follows: an initial cycle of 3 min at 95°C (15 min at 95°C for HotStar and TopTaq), followed by 40 cycles of 30 s at 94°C, 45 s at a primer-specific annealing temperature (T_A), 60 s at 72°C; and a final extension at 72°C for 10 min. The annealing temperature varied from 47 - 57°C depending on the primer-pair used (Table 4.2).

Five of the samples were also amplified in a series of five overlapping long-range (LR) PCR fragments ranging from 2,876 bp to 5,741 bp, with an average overlap of 406 bp (77 bp to 846 bp). For LR PCRs the primer pairs used were g01F/g05R1, g05aF/g07bR, g07bF/g10R, g11F/g14R, and g14F/g20R (g05R1: CGAGTAAACGGCG AGACTTGAAAGG, g05aF: CCTATGCCCTTTCCTGTAGC TGATC, see Table 4.2 for all other sequences). PCR reactions (50 μL) were carried out using the Qiagen LongRange PCR kit; each reaction contained 10X LR PCR buffer (Qiagen), 1.5 mM $MgCl_2$ (Qiagen), 0.5 mM dNTP (Qiagen),), 0.4 μM forward and reverse primer, 1.5 μL DMSO, 3 units enzyme (Qiagen), and 4 μL of the genomic DNA extract. The thermal cycling program was as follows: an initial cycle of 2 min at 92°C; 10 cycles of 10 s at 92°C, 15 s at 58°C, 8 min at 68°C; 20 cycles of 15 s at 92°C, 30 s at 58°C, 8 to 17 min at 68°C (time increased by 20 s per cycle); and a final extension at 68°C for 7 min. All PCR

reactions were performed in an Eppendorf Mastercycler epGradient S thermocycler (Eppendorf, Hamburg, Germany) and then visualised under UV light on a 1% agarose gel with 0.3 μg/mL ethidium bromide.

A total of 38 samples were tested for control region repeat copy number (see Section 4.3.1 for details). The control region fragment containing the repeat was amplified with Qiagen HotStar DNA polymerase following the standard HotStar thermal cycling program given above (T_A = 47°C). The primer pair used was codrep1 (AATTCTAAATTTAAACTACC) and codrep2 (GGACATATTATGAGGCAGGG; Árnason & Rand 1992). The PCR products were visualised under UV light on a 3% agarose gel and a picture was taken. The number of repeats was estimated by measuring the relative intensity of the bands in Image-J (Schindelin *et al.* 2012; Schneider *et al.* 2012).

4.2.3 Sequencing

PCR products were purified with an Exo-SAP clean-up. From 10 μL (for standard PCR) to 50 μL (for long-range PCR) of PCR product was incubated with 0.1 U exonuclease (USB) and 0.1 U shrimp alkaline phosphatase (USB) at 37°C for 15 min, followed by denaturation at 80°C for 15 min. The purified products were then sent to Genome Quebec (McGill University, QC) for Sanger sequencing. Long-range PCR products were sequenced using a series of forward and reverse internal primers. Standard PCR products were sequenced in one or both directions; for fragments with extensive

overlap, or where sequences showed no ambiguities, a single sequencing primer (either forward or reverse) was used

An additional 36 samples had previously been sequenced with the Affymetrix GeneChip® CustomSeq® resequencing multispecies microarray, the "ArkChip" (Carr *et al.* 2009). The samples were sequenced as part of two- and four-species experiments, alongside Atlantic wolffish *Anarhichas lupus*, harp seals *Pagophilus groenlandicus*, and Newfoundland caribou *Rangifer tarandus* (Duggan 2007; Carr *et al.* 2008). For each sample PCR products were pooled in equimolar quantities, fragmented, labelled, and sent to the Centre for Applied Genomics (Toronto, ON). They were then hybridised to the microarray chips and scanned with an Affymetrix GeneChip® scanner (Duggan 2007). Signal intensity data were exported in a Microsoft Excel spreadsheet.

4.2.4 Data analyses

All sequences were aligned in Sequencher v4.9 (Gene Codes Corporation, Ann Arbor, MI) and MEGA v6.0 (Tamura *et al.* 2013). Variable sites were double-checked across all genomes (microarray and/or chromatogram data); positions at which ambiguities could not be reliably resolved were excluded from analyses. Sequences were aligned to the Atlantic cod reference sequence (Coulson *et al.* 2006), and the coding region annotations were confirmed against the partial Atlantic cod mitochondrial genome (Johansen & Bakke 1996) and the chicken mitochondrial genome (Desjardins & Morais 1990). Haplotypes were confirmed using TCS v1.2.1 (Clement *et al.* 2000) and Arlequin v3.5.1 (Excoffier & Lischer 2010).

The microarray DNA sequences were determined with a base-calling algorithm based on a decision-tree from empirical rules (Carr *et al.* 2008; Carr *et al.* 2009). For each site the hybridised product was compared to a reference sequence such that the central nucleotide of a 25 bp fragment was altered between the four possible bases. For each individual the data matrix consisted of a 4 x 16,552 array in which each of the four columns represent one of the four possible sequence variants. Experimental products bind preferentially to the variant that is the perfect match. This was done in both directions, resulting in two calls for each base. The highest fluorescent signal intensity is the presumptive call; the quality of the call is determined by calculation of $\Delta S/N$ as the difference between the two highest signal intensities divided by the sum of all intensities (i.e., the signal-to-noise ratio). Empirically, calls made on both strands with $\Delta S/N \geq 0.13$ are considered "strong", whether in agreement with the reference sequence or for a SNP variant, and calls with $0.10 \leq \Delta S/N < 0.13$ are considered "weaker" but reliable (Pope *et al.* 2011). Positions at which the two strands disagreed, or where $\Delta S/N < 0.10$, were flagged with IUPAC ambiguity codes for further analysis.

4.2.5 Genetic analyses

Genetic diversity was measured with both haplotype (H_d) and nucleotide (π) indices (Nei & Li 1979; Nei & Tajima 1981) calculated in DnaSP v5.10 (Librado & Rozas 2009). To test whether the genetic variation was allocated within or among populations, an analysis of molecular variance (AMOVA) was run in Arlequin v3.5.1 (100,000 permutations; Excoffier *et al.* 1992; Excoffier & Lischer 2010). Spatial analysis of molecular variance (SAMOVA) uses both geographical and genetic data to partition

populations into groups so as to maximise the among-group variance (Φ_{CT}), thus identifying underlying structure and/or genetic barriers. The SAMOVA was calculated for K = 2 to 13 (100 iterations; Dupanloup *et al.* 2002).

Population pairwise genetic distances (Φ_{ST} values) for haploid mtDNA genomes, based on Wright's fixation index for diploid loci (Wright 1965), were calculated in Arlequin v3.5.1 (Excoffier & Lischer 2010) on the sampling locations, both including and excluding GIL (due to small sample size). A modified false discovery rate (Benjamini & Yekutieli 2001) and sequential Bonferroni correction (Holm 1979) were applied to correct for multiple tests. Isolation-by-distance was tested using a Mantel test in Genepop v4.2 (10,000 permutations; Raymond & Rousset 1995; Rousset 2008). The straight-line geographic distances were calculated from average GPS coordinates with the Geographic Distance Matrix Generator v1.2.3 (Ersts 2015) using great circle distances. Swimming distances were calculated as the shortest marine distance between locations. These geographic distances were compared to linearised Φ_{ST} values (genetic distances).

Phylogenetic relationships among haplotypes were analysed with both neighbour-joining (NJ) and Bayesian networks. The NJ analysis was run in PAUP* v4.10 (Swofford 2003) based on the absolute number of nucleotide differences (10,000 bootstrap replications). The Bayesian network was assembled in MrBayes v3.2 (Huelsenbeck & Ronquist 2001) using the generalised time reversible model with gamma-distributed rate variation and allowing for invariable sites (GTR + Γ + I). The analysis was run over two simultaneous runs with eight chains for 5,000,000 generations, a 25% burn-in, final PSRF > 0.999, ESS > 1000, and standard deviation of split frequencies < 0.01. The congeneric

Gadus chalcogrammus (GenBank Accession Number AB094061) was used to root both analyses.

Clustering analysis was run using Bayesian Analysis of Population Structure (BAPS) v6 (Corander *et al.* 2008). BAPS allows haploid sequence data to be clustered using Bayesian analysis in order to assign individuals to K clusters. The samples are assigned based on genetic data with no *a priori* population information. The analysis was run with the linked loci option, the codon model of linkage, and variable K ($K \leq 13$; see Corander & Tang 2007). The optimal K was determined based on the log marginal likelihood of the best-visited partitions. An unrooted statistical parsimony network was constructed in TCS v1.21 (Clement *et al.* 2000) to visualise the relationship between haplotypes. All connections were confirmed by visual examination. A principal coordinates analysis (PCoA) was run in GenAlEx v6.5 (Peakall & Smouse 2006; Peakall & Smouse 2012) on individual samples and population pairwise differences for both sampling locations and putative stock classifications (see Section 4.2.7 below). The PCoA allows visualisation of patterns of variation based on DNA sequence with no *a priori* population information. Although the PCoA on the individuals has far more statistical power, the summary statistics were also used as the resulting separation is much easier to distinguish.

4.2.6 Divergence time estimates

Divergence time estimates were calculated with the Bayesian tree based on the strict clock model in MrBayes v3.2 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003), and with the constant population model in BEAST v2.3 (Bouckaert

et al. 2014). The MrBayes analysis was run as described above with uniform branch lengths. The BEAST analysis was run using a strict clock and the Hasegawa, Kishino, and Yano (HKY) model with gamma-distributed rate variation and allowing for invariable sites (HKY + Γ + I). The models were run for 5,000,000 steps with a 1,000,000 step burn-in and sampled every 10,000 steps. All ESS parameters were > 1,000. The trees were calibrated with a normal distribution on the assumption that the Atlantic cod separated from the Alaska pollock *G. chalcogrammus* ca. 3.8 million years ago during the last opening of the Bering Strait (standard deviation = 0.5 mya; Grant & Ståhl 1988; Vermeij 1991; Coulson *et al.* 2006). This resulted in a mean genetic distance of 0.040 substitutions per site, a divergence rate of 1.14×10^{-8} substitutions / site / year, and a temporal interval of 5,592 years / substitution. These values are similar to those calculated by Carr & Marshall (2008a) from the 15,655 bp excluding the control region, and only slightly higher than that seen in the Alaska pollock (Carr & Marshall 2008b) and in *Homo* (Achilli *et al.* 2004). Estimates of time since expansion were calculated in DnaSP v5.10 (Librado & Rozas 2009) using pairwise mismatches and $\tau = 2ut$ (Rogers & Harpending 1992).

4.2.7 Stock analyses

The Northwest Atlantic Fisheries Organization divides the Northwest Atlantic into seven divisions (NAFO 0 - 6) comprising 25 subdivisions, various combinations of which are treated as management stocks (Figure 4.2). There are currently ten cod management stocks (including the five original Newfoundland stocks): west Greenland (NAFO divisions 1A - 1F), northern Labrador (2G, 2H), southern Labrador - eastern

Newfoundland (2J, 3K, 3L), southern Grand Bank (3N, 3O), Flemish Cap (3M), St. Pierre Bank and the northern Gulf of St. Lawrence (3Ps, 3Pn, 4R, 4S), southern Gulf of St. Lawrence (4T, 4Vn), northern Scotian Shelf (4Vs, 4W), southern Scotian Shelf (4X), and the Gulf of Maine-Georges Bank (5Y, 5Ze, 5Zw; Figure 4.2; Fisheries and Oceans Canada 2016).

The stock classifications were tested using an AMOVA and population pairwise Φ_{ST} values as described above. The three populations outside of the NAFO zones – Barents Sea at Tromsö (NC), Baltic Sea at Sopot (BAL), and the landlocked fjord Lake Qasigialiminiq on Baffin Island, Nunavut (QAS) – were treated as separate entities, while samples within NAFO were assigned to stocks as follows: Hopedale, Makkovik, and Postville (LBH, LBM, and LBP) were grouped as northern Labrador; samples from Hawke Channel, Gilbert Bay, inshore fish at Random Island, and the North Cape of the Grand Banks (HC, GIL, RAN, and NGB) were grouped as southern Labrador - eastern Newfoundland, often referred to as "Northern Cod" or "2J3KL"; FC is on the Flemish Cap, an offshore seamount; ESS is on the northern Scotian Shelf; FSS is on the southern Scotian Shelf; and GEO is from the Gulf of Maine - Georges Bank. The analyses were repeated with ESS, FSS, and GEO grouped together as "South of the St. Lawrence", and with LBM and LBP grouped as "Northern Cod" as the sampling locations are on the boundary of NAFO divisions 2H and 2J.

4.2.8 Analysis of individual loci

To determine how much of the genetic variation in Atlantic cod was captured by any one loci, an AMOVA, pairwise Φ_{ST} comparisons, and NJ trees were run on each of

the 13 individual protein-coding regions, the two rRNAs, and the control region. All analyses were performed with the methods described above.

4.3 Results

4.3.1 Complete mitochondrial sequences

Complete mitogenomic sequences were obtained for 121 Atlantic cod from 14 locations (GenBank Accession Numbers KX266969 - KX267089). These were combined with the 32 known mitogenome sequences (with added 921 bp control region sequences; GenBank Accession Numbers KX432219 - KX432250) for analyses. The mitogenomes were 16,576 bp in length (not including repeat units, see below), and contained the standard 13 coding regions, 12S and 16S rRNA regions, 22 tRNAs, and the non-coding control region in the standard order expected for vertebrates. There were 11 overlapping regions totalling 63 bp, and 14 non-coding regions including 12 short intergenic regions totalling 106 bp, the 33 bp origin of light strand replication (O_L), and the 881 bp control region (CR). The average nucleotide composition was 28.0% A, 25.6% C, 16.8% G, and 29.6% T ([G + C] = 42.4%).

A total of 887 variable sites were identified in the 153 complete sequences: 372 parsimony informative (*sensu* Nei 1987) and 515 singletons (Table 4.3). There were 779 transitions, 102 transversions, and six single-base insertion-deletion events (five in the control region and one in the intergenic region between tRNA Thr and tRNA Pro). At 36 positions SNP variants occurred in three (n = 34) or four (n = 2) of the possible variant forms. Almost 80% of the variable sites occurred in coding regions, 75.2% of which were

third position transitions (531/706; Table 4.3). There were no frameshift mutations or unexpected stop codons. Eighty-seven mutations led to amino acid substitutions; in three cases there were three variant amino acids (Val ↔ Met or Leu, Met ↔ Val or Leu, and Ala ↔ Val or Gly).

There were two tandem repeats found in the non-coding regions: the previously described 40 bp repeat in the control region (sites 15,696 - 15,735) and a perfect 29 bp repeat in the intergenic region between tRNA Thr and tRNA Pro (15,577 - 15,605). The 40 bp fragment contained 13 variable sites, and some individuals showed changes from one repeat to the next (e.g., a 'C' in the first two repeats and a 'G' in the last two repeats). In thirty-eight samples we tested how many repeats were present; from one to six repeat units were identified with the most common state being three or four (Supplementary Table 4.1). The 29 bp repeat was found in three individuals; two samples from NC and one from ESS had two copies of the repeat, while all other individuals had a single copy of the sequence. In the single copy individuals five variable sites were found. The three samples were re-sequenced to confirm the insertion.

4.3.2 Population structure

A total of 142 distinct mitogenome sequences (haplotypes or mitotypes) were identified in 153 Atlantic cod. Of these, 139 were unique to a single individual, and the other three were found in two, five, or seven of the 18 fish from Lake Qasigialiminiq (QAS). As a result, haplotype diversity (H_d) was 0.791 in QAS and 1.0 in all other locations. Nucleotide diversity (π) ranged from 0.0016 in LBM to 0.0028 in the adjacent LBH (Table 4.4), reflecting the basal position of two of the Hopedale (LBH) samples.

Due to the small sample size (n = 2) the Gilbert Bay population was excluded from all population comparisons.

The AMOVA identified the majority of the variation as within (89.5%) rather than among (10.5%) populations (Φ_{ST} = 0.1051, p < 0.001). When run excluding GIL, slightly less of the variation was allocated among populations (9.5%; Φ_{ST} = 0.0978, p < 0.001). The SAMOVA detected the presence of two groups (QAS vs. all others; Φ_{CT} = 0.241, p = 0.065). When K was increased to K = 3, BAL was pulled out (Φ_{CT} = 0.217, p = 0.016), and at K = 4 NC separated (Φ_{CT} = 0.199, p = 0.003). Pairwise Φ_{ST} values (Table 4.5a) detected QAS as significantly different from all populations except BAL (p = 0.024, non-significant following correction for multiple tests). The two European locations, NC and BAL, were significantly different from many of the North American populations (NC 5/11 and BAL 6/11). None of the other locations showed significant values following correction for multiple tests. All negative Φ_{ST} values were treated as zero. The QAS samples separated into two clusters of divergent individuals (see NJ and statistical parsimony analysis below); to ensure that the frequency difference of the two groups was not the cause of the significant results, the pairwise comparisons were run with QAS separated into two groups. Results were similar to Table 4.5a with QAS1 and QAS2 significantly different than all other populations including each other (p < 0.001). The correlation between genetic and geographic distance was not significant (p = 0.088, r^2 = 0.005; Figure 4.3) although the largest genetic distances not including QAS were in trans-Atlantic comparisons. When NC, BAL, and QAS samples were excluded the correlation reduced further (p=0.783, r^2=0.020).

The neighbour-joining analysis identified 49 branches with > 70% bootstrap support (Figure 4.4). These represent four of the six inclusive clades previously identified (B, D, E, and I; Carr & Marshall 2008a), an additional two inclusive clades (H and K), and numerous subclades within the larger groups. The remaining two clades, F and G, were supported at 58% and 64%, respectively. Clade H was basal to all other groups. The Bayesian analysis identified 62 branches with posterior probabilities > 0.95 (Figure 4.5, Supplementary Figure 4.1), and included the same eight inclusive clades found in the NJ analysis with high posterior probabilities (0.9 - 1.0). Clade membership was identical between the two methods. The distribution of samples among clades was significant (X^2 = 214.5, df = 117, p = 0.004); NC, BAL, and QAS were found almost exclusively together (in B, E, and G), and no European samples and only a single QAS sample were found in the two most common clades (F and I). If NC, BAL, and QAS were excluded, the distribution of samples among clades was no longer significantly different than random (X^2 = 106.7, df = 90, p = 0.908).

Bayesian clustering analysis (Table 4.4) identified two groups found across most populations (group 1 was not present in LBM or GIL). The three highest log maximum likelihood values were -14,637 (K = 2), -15,080 (K = 3), and -15,831 (K = 4). The probability of two clusters was one. There was no association between group membership and sample origin; the distribution of populations between the two groups was not significantly heterogeneous (X^2 = 13.13, df = 13, p = 0.438). When compared to the NJ diagram (Figure 4.4), group 1 included the four basal clades (H, B, D, and K) and group 2 the four more recent clades (E, F, G, and I).

The statistical parsimony network (Figure 4.6) showed extensive variation in Atlantic cod and identified the same groups found in the NJ and Bayesian analyses. The QAS samples occurred essentially as monomorphic haplotypes, although in two disjunct groups (clades B and E), and one individual most closely related to a geographically distant individual in clade F. Half of the BAL samples clustered together in a non-exclusive subclade within E, and the others in B, D, and basally within the FGI superclade. The BAL and QAS samples in clade E are sister clades. Norwegian coast (NC) samples were not closely grouped with any others (including each other, minimum distance = 16 changes). The two most abundant clades (F and I) both showed a starburst pattern of variation, did not contain any European (NC or BAL) samples, and only a single QAS sample in F.

The PCoA on the individual samples (Figure 4.7a) showed a slight separation of samples into groups 1 and 2, but not into sampling location, and there was no defined structure. The QAS samples stood out from the others in two clusters as the most different along both coordinates 1 and 2. The first three coordinates explained 37.1 % of the variation (15.8% coordinate 1, 12.5% coordinate 2, and 8.8% coordinate 3). When viewed by clade membership (Figure 4.7b), the samples clustered with the older clades (H, B, D, and K) separating from EFGI along coordinate 1. The more recent European clade (E) segregated along coordinate 2. The PCoA on the population Φ_{ST} values (Figure 4.7c) again showed the limited structure in the North American populations with the exception of QAS. Most of the variation was explained by the first coordinate (62.0% coordinate 1, 16.2% coordinate 2, and 10.1% coordinate 3). QAS was isolated from all other

populations along coordinate 1, NC separated along coordinate 2, and BAL along coordinates 1 and 2.

4.3.3 Divergence time estimates

On the assumption that Atlantic cod separated from the congeneric Alaska pollock at the most recent opening of the Bering Strait, the most recent common ancestor of Atlantic cod dates to 328 kya (95% Highest Posterior Density (HPD) = 212 - 413 kya), and the superclade containing the more recent expansions in North America (FGI) dates to 150 kya (Figure 4.5; HPD = 87 - 172 kya). The separation of the clades all date to before the last glacial maximum (LGM), with estimates ranging from 45 - 139 kya (Figure 4.5). The pairwise mismatch distribution suggests a sudden expansion rather than a constant population size, with $\tau = 23.39$ corresponding to an expansion time of 124 kya. The expansion of the North American FGI superclade dates to 99 kya, and coalescence times in the eight major clades ranged from 43 to 109 kya. As would be expected, population expansion estimates all postdate the clade separation estimates.

4.3.4 Stock structure

When the AMOVA was run with the stock classifications as "populations" the results were similar to above, with 89.7% of the variation found within populations and 10.3% of the variation among populations ($\Phi_{ST} = 0.1031$, $p < 0.001$). Pairwise Φ_{ST} values were also similar with 12 of 36 comparisons significant, all involving NC, BAL, or QAS, and QAS was again significantly different than all populations except BAL (Table 4.5b). Within the Northwest Atlantic the Φ_{ST} values ranged from 0.0 to 0.023, with the only

value above 0.01 between northern Labrador and the southern Scotian Shelf (p = 0.099). Results were similar when LBM and LBP were grouped as part of the Northern Cod, and when the three populations south of the St. Lawrence were combined (ESS, FSS, and GEO).

4.3.5 Individual loci

The AMOVAs on the individual loci gave Φ_{ST} values ranging from 0.02 to 0.15, with the greatest among-population variance seen with ND6 followed by 16S. (Supplementary Table 4.2). All Φ_{ST} values were significant. The pairwise Φ_{ST} values consistently supported QAS as significantly different from the other populations, although not all loci had significant values (12S, ATP8, and ND4L regions did not show any significant comparisons, COX2 values were not significant following correction for multiple tests, and the CR comparisons only showed some QAS comparisons as different; Supplementary Table 4.2). The other populations to show at least some separation at the single locus level were BAL (with ATP6, CYTB, ND2, ND4, and ND6), NC (with ND2), and FSS (with 16S). ND2 was the only locus able to identify NC, BAL, and QAS as divergent from the Northwest Atlantic populations.

The neighbour-joining trees on the individual loci generally suggested clustering of the eight major clades, but these were not well-supported. Supplementary Table 4.2 outlines which clades were supported by each locus, and the corresponding bootstrap support. Note that in few cases were these clades identical to those found with the complete genome (e.g., there may have been an extra individual or a few missing). The ND1 and ND2 markers showed the best clustering of groups with no support in ND1 and

support for H, D, and I in ND2. Surprisingly, the control region identified three of the eight clades, one with bootstrap support of 85%. None of the individual loci identified all of the clades.

4.4 Discussion

A total of 887 variable sites were identified in the complete mitogenomes of 153 Atlantic cod from North America and Europe. Each fish (except for 14 samples from QAS) had a unique DNA sequence, as has been seen in the highly variable Atlantic herring *Clupea harengus* (Teacher *et al.* 2012), Alaska pollock *Gadus chalcogrammus* (Carr & Marshall 2008b), and harp seal *Pagophilus groenlandicus* (Carr *et al.* 2015), but in contrast to the slowly-evolving and more recent Atlantic wolffish *Anarhichas lupus* (Chapter 2), green turtle *Chelonia mydas* (Shamblin *et al.* 2012) and speartooth shark *Glyphis glyphis* (Feutry *et al.* 2014). The Atlantic cod shows a relatively high mutation rate of 5,592 years/substitution, similar to that calculated for Alaska pollock (5,255 years/substitution; Carr & Marshall 2008b), and somewhat higher than that seen in the sympatric Atlantic wolffish (6,078 years/substitution, 0.020 substitutions/site; Chapter 2).

4.4.1 Population genetic structure

The population genetic structure of the Atlantic cod shows trans-Atlantic differences, high diversity in Norway and the Baltic Sea, an isolated Arctic population, and limited differentiation within the Northwest Atlantic. The overall Φ_{ST} value (Φ_{ST} = 0.1051) was similar to that seen with the cytochrome *b* marker in Atlantic cod (Φ_{ST} =

0.081; Árnason 2004). This is attributable to differences among the Lake Qasigialiminiq, Baltic Sea, and Norway populations versus the Northwest Atlantic populations. Differentiation among the Northwest Atlantic samples alone was not significant (Φ_{ST} = 0.0061, p = 0.2163). This pattern is similar to that seen in the pan-global albacore tuna *Thunnus alalunga* (Φ_{ST} = 0.092; Chow & Ushiama 1995) and bigeye tuna *Thunnus obesus* (Φ_{ST} = 0.260; Alvarado Bremer *et al.* 1998); in those cases, the large significant differences could be attributed to Atlantic-Pacific divergence (Ward 2000). Much lower geographic structuring has been seen in yellowfin tuna *Thunnus albacares* (Φ_{ST} = 0.01; Ward *et al.* 1994a) and Atlantic wolffish (Φ_{ST} = 0.052; Chapter 2). The average F_{ST} value for marine fish is estimated at 0.062, comparable to that of highly mobile birds (average F_{ST} = 0.076), and much lower than that seen in freshwater fish (average F_{ST} = 0.222; Ward *et al.* 1992; Ward *et al.* 1994b). That Atlantic cod exhibit a higher Φ_{ST} value suggests either contemporary barriers to dispersal and/or historical separation. Within the Northwest Atlantic there was no support for the current management stocks; when the locations were grouped together based on the current management delineations there were no significant pairwise differences among North American 'stocks'.

Extensive gene flow and admixture is suggested by the large non-significant Φ_{ST} values among all but the Qasigialiminiq and European sampling locations (Table 4.5a), and the absence of isolation-by-distance among the Northwest Atlantic samples. The separation of trans-Atlantic populations is incomplete – in many comparisons (e.g., with LBH, FC, HC, and GEO) there were no differences between populations from either side of the Atlantic, whereas in others there was strong isolation (e.g., LBP, LBM, NC, and FSS; Table 4.5a). This pattern is similar to that seen in the congeneric Pacific cod *Gadus*

macrocephalus, where a combination of mtDNA (ND2 and CYTB) and microsatellites showed significant differences between trans-Pacific populations and no distinction among geographically proximal locations (Canino *et al.* 2010). In this case the authors suggested multiple glacial refugia on either side of the North Pacific Ocean with significant admixture following deglaciation.

Bayesian analysis (Figure 4.5) supports the presence of eight major clades (and a further 55 nodes with Bayesian posterior probabilities ≥ 0.95) that are supported by NJ, PCoA, and statistical parsimony analyses (Figures 4.4, 4.6a, and 4.7a). There was little apparent correlation between location of sample origin and location on the networks. Some exceptions to this were that the Lake Qasigialiminiq samples grouped together and the Norwegian and Baltic samples were generally found in the older clades. Distribution of clade membership was significantly heterogeneous for Norway, the Baltic Sea, and Lake Qasigialiminiq, but not for the Northwest Atlantic samples. This suggests either panmixia among the Northwest Atlantic populations, incomplete lineage sorting in the mitochondrial DNA with insufficient time for genetic isolation to occur, or a combination of the two.

These results can be compared to earlier work on Atlantic cod. Initial analyses with partial mtDNA cytochrome *b* sequences (307 - 401 bp) have consistently shown a lack of genetic differentiation on the small scale while trans-Atlantic differences in allele frequencies have been found (Carr & Marshall 1991a, b; Árnason *et al.* 1992; Carr *et al.* 1995; Árnason & Pálsson 1996; Árnason *et al.* 1998; Carr & Crutcher 1998; Árnason *et al.* 2000; Árnason 2004). Using a set of five (or six) hypervariable microsatellite markers (*Gmo 2*, *Gmo4*, *Gmo120*, *Gmo132*, (*Gmo141*),and *Gmo145*), however, both Ruzzante *et*

al. (1996) and Bentzen *et al.* (1996) found significant differences among the Northwest Atlantic cod stocks – Ruzzante *et al.* identified differences between offshore and onshore spawning aggregations, and Bentzen *et al.* found heterogeneity between Labrador (North) and Grand Banks (South) samples, both from the 2J3KL Northern Cod stock complex. Bentzen *et al.* also suggested that the Flemish Cap and Scotian Shelf populations were significantly divergent. A different suite of microsatellite markers produced different results. Hardie *et al.* (2006) used *Gmo 3*, *Gmo8*, *Gmo19*, *Gmo34*, *Gmo35*, *Tch5*, and *Mae9* on samples from three landlocked Baffin Island populations (see below), the St. Pierre Bank, southern Gulf of St. Lawrence, northern Scotian Shelf, and southern Scotian Shelf; he found significant differences among Baffin Island populations but no differences among the Northwest Atlantic samples. Similarly, Beacham *et al.* (2002), using the same five *Gmo* markers as Hardie *et al.* and two new markers (*Gmo36* and *Gmo37*), identified the Flemish Cap and Gilbert Bay populations as unique and suggested limited inshore vs. offshore differences, but found no separation among samples from the 2J3KL offshore locations. A range-wide study using a combination of the *Gmo* markers used above suggested strong divergence and isolation-by-distance among populations (O'Leary *et al.* 2007a), although only a single North American sample was used (Scotian Shelf) with large geographic distances to the closest neighbouring populations (Greenland). These findings suggest that marker choice is crucial for interpreting results at all levels, not only between marker types.

In the present study, the two inshore locations, Random Island and Gilbert Bay (NAFO 3L and 2J, respectively), are not differentiated from the nearby offshore Grand Banks or Hawke Channel cod. This result is consistent with that found with the 307 bp

cytochrome *b* sequences from 80 cod at Random Island and the nearby Heart's Ease Ledge which found no significant differentiation among these and cod from the northern Grand Banks (Carr *et al.* 1995). Results are also consistent with tagging data which shows that most fish move from inshore feeding grounds to offshore spawning grounds with some inconsistencies (Templeman 1979; Lear 1984); and Trinity Bay tagging data which shows that some fish stay in Trinity Bay, some move to other nearby inshore bays, and some move offshore (Goddard *et al.* 1994; Wroblewski *et al.* 1994).

In both cases there were limitations. For the Random Island samples, the fish could be offshore cod that have migrated to the coast to feed. In order to ensure that the sample is representative of overwintering fish, either antifreeze levels must be measured or sampling must be done before migration begins (i.e., late fall, early winter). The Gilbert Bay population had a very small sample size (n = 2) so conclusions here are limited. The two fish were each other's closest relative and differed by seven variable sites, and Φ_{ST} analysis showed only three pairwise comparisons were significant after correcting for multiple tests (GIL with QAS, LBP, and LBM; data not shown).

The two most divergent populations, Lake Qasigialiminiq and the Baltic Sea, show dissimilar patterns despite similar population age. Both Baffin Island and the Baltic Sea were glaciated until ca. 5,000 years ago. However, the Baltic Sea has a very large population size and the samples exhibit high diversity and basal lineages, while Lake Qasigialiminiq has a very small population size and shows lower diversity and clustering of individuals. This reflects the isolated nature of the landlocked population and the influence of genetic drift on a small population.

4.4.2 Landlocked Arctic fjord population

The shared haplotypes seen in the landlocked fjord population at Lake Qasigialiminiq, at the southwest end of Cumberland Sound, Baffin Island, appear to reflect a long period of isolation at small population size (500 - 1000 individuals). Genomic haplotypes shared between individuals, much less multiple individuals from the same sample, has not previously been observed in Atlantic cod, and is unusual in mitogenomic studies. Given the small population size, however, it is feasible that samples in such a restricted population may represent family group sampling. Hardie *et al.* (2006) examined short cytochrome *b* fragments and found that all cod sampled in the three Baffin Island lakes were of haplotypes '*A'*, '*E'*, or '*G'* (I denote these in italics to distinguish from the mitogenomic clades). Of the 18 Qasigialiminiq samples in the present study, seven of them had CYTB haplotype '*A'* – the most common western Atlantic haplotype – and 11 had CYTB haplotype '*E'*. In the whole-genome analyses the 11 '*E*' samples were assigned to clade E, while of the seven '*A*' samples, six were assigned to the basal clade B and one to the more recent clade F (most closely related to an individual from northern Grand Banks, NAFO 3L; Figures 4.4 - 4.7). The single Qasigialiminiq sample in haplogroup F may be a very recent introduction from a haplogroup otherwise restricted to the Northwest Atlantic.

As would be expected in an isolated population with little or no gene flow to adjacent marine populations, diversity levels are lower in QAS than in many of the marine populations (Table 4.3). Hardie *et al.* (2006) examined seven microsatellite loci and found the lowest levels of genetic diversity in the landlocked populations. Pairwise F_{ST} values for the nuclear markers were large among Arctic populations (F_{ST} = 0.195 -

0.238), and much greater between Arctic and marine populations (F_{ST} = 0.081 - 0.214) than they were among marine populations alone (F_{ST} = 0.000 - 0.058; Hardie *et al.* 2006). The among-marine F_{ST} values were lower still when the inshore Gilbert Bay sample was excluded (F_{ST} < 0.003). A similar pattern was seen among mitogenomes, where Φ_{ST} values ranged from 0.153 to 0.340 with QAS, as compared with 0.00 to 0.06 among the rest of the Northwest Atlantic samples ($\Phi_{ST} \leq 0.03$ excluding northern Labrador comparisons; Table 4.4). Pairwise comparisons involving QAS were significant with all populations except BAL. The low diversity could be due to a small inbred population (Hardie 2003), genetic drift amplified by the small population size, or may be caused by the physical separation of the population.

The facts that 17 of the 18 QAS samples are most closely related with either specific trans-Atlantic Norway or Baltic Sea samples, and that the only non-significant pairwise Φ_{ST} comparison was with BAL (p = 0.024), suggest that the QAS population may have originated from a Northeast Atlantic source population, one that then continued down into Atlantic Canadian waters. This contrasts with the suggestion that the three Arctic saline lakes were colonised by Northwest Atlantic cod 5,000 - 8,000 years ago (Hardie *et al.* 2006). The 18th individual is most closely related to an individual from the Northwest Atlantic (although not one that is geographically proximal), which indicates a third, more recent, invasion. That the lake at present includes descendents of at least three distinct lineages suggests either that the source population was variable and/or that there were multiple invasion events. Atlantic cod are currently absent from adjacent marine waters, not being found farther north than northern Labrador, making it unlikely that a second (or third) colonisation event happened in recent history. However, water

temperatures in the Arctic Ocean and the Davis Strait were several degrees higher from 3,000 - 6,000 years ago (Aitken & Gilbert 1989), suggesting that cod may have persisted in the northern marine environment until being forced to retreat southwards.

The presence of three distinct lineages in Lake Qasigialiminiq had previously been overlooked by single locus and microsatellite studies. The cytochrome *b* data found a starburst pattern within each lake population, as well as among lakes, with all samples having haplotypes '*A*', '*E*', or '*G*', a group of haplotypes differing by a single mutation from a common haplotype ('*A*'). However, upon closer inspection, what registered as '*A*' was really quite distinct from '*E*'. Carr & Marshall (2008a) discussed the consequences of unrecognised paraphyly within the cytochrome *b* haplotype '*A*' versus the complete mtDNA haplogroups. Despite the additional resolution provided by the microsatellites, they also missed the pattern seen in QAS. The microsatellites identified strong differences among populations, but nothing significant within each lake. In fact, allelic diversity and heterozygosity were lowest in the three landlocked fjord populations (as they were here due to multiple identical samples; Hardie *et al.* 2006). It is likely that the microsatellites detected contemporary isolation, but could not detect the historical pattern.

4.4.3 Glacial refugia

The Pleistocene consisted of long cold spells (ice ages) interspersed by shorter warm periods (Barendregt & Irving 1998; Hewitt 2000, 2004). During the most recent glaciations, the Wisconsinan (Nearctic) or Würm (Palaearctic) periods (ca. 110 - 12 kya), ice sheets covered much of the Northern Hemisphere. Sea levels were as much as 120 m lower than today, exposing continental shelves that were then covered with ice (Pielou

1991; Mitrovica 2003). Clade expansion in Atlantic cod coincides with the interglacial and warming periods experienced during the Wisconsinan and Illinoian glaciations (Gibbard & van Kolfschoten 2004). Origins of the most common predominantly Northwest Atlantic clades occurred ca. 136 kya (F and G) and 150 kya (F, G, and I), both of which correspond to the warmer Sangamonian interglacial. Although many of the other separations are estimated to occur during the Illinoian glaciation, there were at least two warming periods ca. 250 kya and 300 kya (Gibbard & van Kolfschoten 2004). It seems likely that much of the early divergence occurred during these warmer periods or even prior to the Illinoian.

Patterns of variation seen in Atlantic cod suggest distinctiveness of Northeast and Northwest Atlantic cod populations. Pairwise comparisons show that the Baltic Sea and Norway populations are significantly differentiated from many of the Northwest Atlantic populations (Table 4.4). There were also significant differences in clade membership with respect to NC and BAL, they separated from the Northwest Atlantic populations in the PCoA (Figure 2.6), and the AMOVA was heavily influenced by these two populations. As with the cytochrome *b* and nuclear studies, there appears to be significant trans-Atlantic differences, although there are no fixed differences. A recent study using nuclear SNP markers identified consistent differences between Northwest and Northeast Atlantic samples with both neutral and non-neutral markers (Bradbury *et al.* 2013). The magnitude of the separation suggests that it predates the most recent glacial cycle, and may be a relic of the Illinoian glaciation or earlier. The lack of fixed differences and geographical structuring among cod populations (except QAS) suggests that historical admixture may be confounding older patterns by increased gene flow and reduction of local genetic drift.

During the last glacial maximum (ca. 20 kya) suitable habitat may have occurred along the coasts of Europe, Iceland, Atlantic Canada (off the Flemish Cap and Grand Banks), and in the Davis Strait between Baffin Island and Greenland (Bigg *et al.* 2008). It is likely that more than one of these areas acted as a refugium for Atlantic cod. Nucleotide diversity (Table 4.4) was highest in the Norway and Baltic Sea samples, which may represent older source populations. Although the Baltic Sea was covered in ice during the last glaciation, the source may have persisted in a nearby refuge and subsequently recolonised as a whole. In the present data the northernmost Labrador sample, LBH, contains the most basal lineage (H), which may reflect a trans-Atlantic connection. Europe may have provided glacial refugia for Atlantic cod, either as a single large refuge or multiple smaller regions. The trans-Atlantic differences seen in Atlantic cod, and the lack of NC and BAL samples in the younger F and I clades, suggest that there may have been an additional glacial refugium in the Northwest Atlantic; however, whether this was the putative Flemish Cap or a more southerly trans-Laurentian refugium is uncertain. Cross & Payne (1978), on the basis of allozyme genotypes, and Carr & Marshall (2008a), from mitogenomes, suggested that shared genotypes between Flemish Cap and Barents Sea samples indicated trans-Atlantic dispersal by way of middle latitudes. However, the Flemish Cap population does not differ from cod in Newfoundland and Labrador (including the 2J3KL Northern Cod), the Scotian Shelf, or Georges Bank (Φ_{ST} = 0.0 to 0.023, $p > 0.15$; Table 4.4a). This suggests that the Flemish Cap likely did not act as a refugium for Atlantic cod, and that the Northwest Atlantic cod may be derived from a southern refugium, with secondary mixing.

4.4.4 Tandem repeats

Atlantic cod from the Northeast Atlantic and adjacent waters have been shown to have one to six copies of a 40 bp tandem repeat in the control region (sites 15,696 - 15,735), consistent with the results found here (Johansen *et al.* 1990; Árnason & Rand 1992; Johansen & Bakke 1996; Kijewska *et al.* 2009). In contrast to what has been reported to date, however, the control region repeats found in this study were imperfect (i.e., the repeat could vary between copies) and showed 13 variable sites among individuals and zero to three variable sites within individuals. The novel 29 bp repeat found in the non-coding region between tRNA Thr and tRNA Pro was a perfect copy (i.e., the repeat unit was identical between copies within individuals, and among the three samples showing multiple copies). Given that only three samples (two from NC and one from ESS) had the 29 bp repeat, and that there was no variation in the repeated units, it seems likely that this is a relatively recent mutation. There were, however, five variable sites in the corresponding 29 bp region of single-copy individuals.

4.4.5 Individual loci

Complete mitochondrial genomes are proving useful in identification of population structure, where previously on the basis of short sequences there was thought to be none (Wang *et al.* 2010; Shamblin *et al.* 2012; Feutry *et al.* 2014). Single locus studies have often failed to identify structure due to poor marker choice – loci are chosen based on what has been done before, what is available, and what is variable in closely (or not so closely) related species. However, variation is not evenly distributed across the mitochondrial genome, and even within coding regions there is heterogeneity of variation

levels (Table 4.3 and Supplementary Table 4.2). The most variable regions in Atlantic cod by number of SNPs were ND2, ND4, and ND5, and by percent variation ND2, ND6, and the non-coding regions (not including the control region). Even where variation was high, this did not necessarily reveal population genetic structure (as seen in Chapter 2).

Before mitogenomics was introduced to study the population genetics of Atlantic cod, a short fragment of cytochrome *b* was used as the sequencing marker of choice. In the current study the 307 bp fragment contained 17 variable sites representing 5.5% variation. This is roughly the same as the mitogenome as a whole (5.4%). Another commonly used marker for interspecific studies is the "barcoding" gene (COX1) which showed only 3.8% variation (59 variable sites). Despite the smaller number of variable sites, both CYTB and COX1 were able to identify clustering of clades and the distinctiveness of QAS, as it is not the variability so much as the ability to discern true structure which is key (Supplementary Table 4.2). All but two loci (12S and ATP8) identified QAS as distinctive, and most of the loci could identify some of the groups found in the Bayesian and NJ analyses (Figures 4.4 and 4.5). The locus with the highest resolution was ND2 which allowed pairwise distinction of NC, BAL, and QAS, and was also able to generally cluster all eight major clades, three with bootstrap support > 60. Other species have suggested COX1 (Chapter 2), 12 S and ND5 (Feutry *et al.* 2014), or COX2 (Teacher *et al.* 2012) as showing the greatest population structure. The ability of a specific gene to identify the true clade structure is likely species-specific.

4.4.6 Conclusions

Atlantic cod population genetics have been studied extensively since the 1960s, by a gamut of molecular markers including blood transferrins, allozymes, restriction mapping, sequencing of mtDNA, microsatellites, and now genome scans. The current picture is based on sequence analysis of mitochondrial and nuclear sequences (Carr & Marshall 2008a; Johansen *et al.* 2009) as well as nuclear SNP markers (Bradbury *et al.* 2010; Bradbury *et al.* 2013). Due to its wide-scale distribution and pelagic lifestyle, Atlantic cod have always been expected to have limited population structure over small geographical scales, and even in trans-Atlantic comparisons. This has consistently been the case in mtDNA single locus studies. However, a greater degree of population structure has been suggested based on some nuclear microsatellites at local scales, and with most markers at large scales.

Examination of 153 complete mitogenome sequences provides evidence for trans-Atlantic differences and Arctic isolation, but finds no evidence of stock differentiation within the commercial range of the Northwest Atlantic. The statistical parsimony network shows two clade expansions (starburst patterns) from the most common haplotypes found in North America. This supports a western Atlantic refugium; however, diversity measures and older clade membership strongly support Europe as a refugial and source population. As the variation does not correspond to geographical location of origin, it is likely that historically separated populations have undergone secondary admixture.

The contrast between mitochondrial DNA and microsatellite markers is likely due to the nature of the loci – mtDNA is more suitable for studying historical patterns of divergence and evolution while microsatellites are good at detecting contemporary

patterns. Additional samples from across the range, especially intensive sampling of European and mid-Atlantic populations, may help to resolve such questions as:

(1) Given that multiple glacial refugia is likely, where were these refugia located?

(2) Did Iceland and/or Greenland play a historical or contemporary role in trans-Atlantic movement and the colonisation of Baffin Island?

(3) Did the Flemish Cap act as a glacial refuge?

(4) What is the origin and age of the Baltic Sea cod?

(5) Is a similar pattern evident in the other two North American Arctic lake populations?

The high level of variation seen in the cod mitogenomic sequences, including the trans-Atlantic differences and isolation of Arctic fjord populations on Baffin Island, show that the complete mitogenome has sufficient resolution to answer these questions. Used in combination with nuclear markers it will allow a complete picture of Atlantic cod evolutionary history to be formed.

Acknowledgements

We would like to thank Fisheries and Oceans Canada for providing the tissue samples, and both the Nunatsiavut Government and Pangnirtung Hunters and Trappers Organisation for helping to organise field work. A special thank you to Ian Winters, Errol Andersen, George Gear, Noah, and Jackie for their invaluable help arranging field seasons in the north, and William Hunter, Errol Andersen, Leon Jacque, and Norman Mike for catching the cod. We thank everyone involved in the extraction and amplification of the samples: Dr. HD Marshall, Dr. KA Johnstone, Dr. AT Duggan, Dr.

SMC Flynn, and L MacDonald. This project was supported by a Natural Science and Engineering Research Council Discovery Grant to SMC, an NSERC Alexander Graham Bell Canada Graduate Scholarship to LAL, a W. Garfield Weston Award for Northern Research to LAL, and a Northern Scientific Training Program Award to LAL.

Table 4.1. Distribution of sample origins and method of sequencing for 153 Atlantic cod samples from 14 sampling locations. The "C&M" samples are those from Carr & Marshall (2008a), the "Arkchip" samples were previously sequenced and were reanalysed as part of the present study, and the "New" samples were amplified, sequenced, and analysed for this study.

	C&M	**Arkchip**	**New**	**Total**
NC	6		1	**7**
BAL		10		**10**
QAS			18	**18**
LBH			12	**12**
LBM			14	**14**
LBP			14	**14**
HC	8		2	**10**
GIL			2	**2**
RAN			11[2]	**11**
NGB	9	4		**13**
FC	9[1]	2		**11**
ESS		10	1	**11**
FSS			10[2]	**10**
GEO		10		**10**
Total	**32**	**36**	**85**	**153**

[1] Four of the FC "C&M" samples were amplified using the Arkchip

[2] Two RI samples and three FSS samples were previously amplified and sequenced by Sanger sequencing, but not previously analysed

Table 4.2. Primers used for PCR amplification of Atlantic cod. Primer pairs are given; in some cases g01F was used with g02R (T_A = 52°C) and g09aF with g09bR (T_A = 52°C). Annealing temperatures (T_A; °C), location of 5'-most base of the primer (loc), and primer sources (ref) are given.

Name	Sequence (5' to 3')	T_A	loc	ref
g01F	CTGAAGATATTAGGATGGACCCTAG	49	29	1
g01R	CTAGTCCCTACTTACTGCTAAATCC	49	871	1
g02F	CCAAAAACGTCAGGTCGAGGTGTAG	55	742	1
g02R	CTATTCATTTCACAGGCAACCAGCT	55	1,490	1
g03F	ACCCCGAAACTGAGCGAGCTACTCC	52	1,351	1
g03R	TAAGCCCTCGTGATGCCATTCATAC	52	2,150	1
g04F	TTTACCAAAAACATCGCCTCTTG	52	1,995	1
g04R	TGAACCTCTGTAGAAAGGGCTTAGG	52	2,815	1
g05F	GGAGTAATCCAGGTCAGTTTCTATCTATG	52	2,600	1
g05aR	AAAGTGGTGTAGTGGAAGCAGTAGG	52	3,944	2
g06F	GGTTAAAGTCCCTTCAACTCCTTAG	52	3,856	1
g06aR	GGTAGTGTCCTGCAGATCTTATCAG	52	5,153	2
g07aF	CTACCTTACCTCTCGCTATTTCAGC	55	4,978	2
g07aR	CTACACCAGAGGATGCTAAAAGGAG	55	5,798	2
g07bF	ATAATTGGAGGCTTTGGGAACTGAC	55	5,663	2
g07bR	TGTTAAGCCCCCGACTGTAAAGAGG	55	6,505	2
g08F	ATGGGTATAGTCTGAGCTATGATGG	48	6,254	1
g08R	TAACCCACAATTCTGCCTTGACAAG	48	7,121	1
g09aF	ACATTCGAGGAACCCGCATTCGTTC	52	6,953	2
g09aR	GAGGTCTTCGTAATCGGTSTSCTCG	52	7,499	2
g09bF	GCCCTTCCATCATTACGAATTCTTTATC	52	7,380	2
g09bR	ATTAAGGGTGGTTGGGAGTCACCTGCTT	52	7,914	2

g10F	TCCCGGAGTTTTCTACGGACAATG	47	7,715	1
g10R	AGAGGGCGAATGAATAAACTAATTG	47	8,539	1
g11F	TAGCAACTGTCCTTATCGGCATACG	47	8,407	1
g11R	TAATACTGTGGTGAGCTCAGGTTAC	47	9,212	1
g12aF	ATTATCCGAGAGGGGACCTTTCAGG	55	8,939	2
g12aR	CGTAGGGAGATAGCTTTTCGTAGTC	55	9,728	2
g12bF	CAACAGTTATCCTTATTGCCTCAGC	55	9,627	2
g12bR	GTTAAGAGCAGTCCATCCGGTATCC	55	10,483	2
g13F	CTTTCTCCGCTTGTGAAGCAAG	47	10,225	1
g13R	CAATTAGAGATTTCAGGTCAGTTTG	47	11,183	1
g14F	CTGTTGCAGGCTCAATAGTTCTTGC	52	10,995	1
g14R	TTCGAGGGAGCCTTGGGGTCTAACC	52	11,841	1
g15F	TAACCAAGACATTAGATTGTGATTC	48	11,715	1
g15R	TGGTAGTCATGGGTGAAGTCCAAAC	48	12,634	1
g16F	GGTGATGACACGGCCGAGCAGATG	57	12,396	1
g16R	AATAATTGCATCTTTGGAGAAGAAGC	57	13,120	1
g17F	ATTCATAGCCTAAACGATGAACAAG	52	12,971	1
g17R	GTCGTTTTTCATATCATTAGTCCTG	52	14,316	1
g18F	GCTACTAAGACCAGTCCTAAAGCAG	48	14,164	1
g18R	CTGTGGGATTATTTGAGCCTGTTTC	48	14,971	1
g19F	GAGGAGGTTTCTCAGTAGATAATGC	48	14,837	1
g19R	GTTTAATTTAGAATTCTAGCTTTGG	48	15,690	1
g20F	GAATGAAACTGCCCTAGTAGCTCAG	47	15,475	1
g20-2R1	TGGACCTGAAGCTAGGCAGAATAGC	47	16,098	2
g20-2F1	GTAAACATAACCGGACTTTCCTTGC	49	15,981	2
g20R	GGCAGGACATTAAGGGCATTCTCAC	49	160	1

1 (Coulson *et al.* 2006)

2 This study

Table 4.3. Distribution of variable sites across the complete mitochondrial genome of 153 Atlantic cod. Size of region in bp (BP), percent variation (%V), total numbers of variable sites (VS), parsimony-informative sites (PI), transitions (Ts), transversions (Tv), and insertion-deletions (ID) are given. For coding regions: coding positions 1, 2, or 3 and status as synonymous (S) or non-synonymous (NS) substitutions are indicated.

Region	**BP**	**%V**	**VS**	**PI**	**Ts**	**Tv**	**ID**	**1**	**2**	**3**	**S**	**NS**
12S	947	2.7	26	6	25	1	0	n/a	n/a	n/a	n/a	n/a
16S	1,696	2.4	40	11	36	4	0	n/a	n/a	n/a	n/a	n/a
ND1	975	6.6	64	31	57	7	0	11	1	52	56	8
ND2	1,047	8.3	87	35	74	13	0	14	1	72	75	12
COX1	1,551	3.8	59	21	52	7	0	2	1	56	56	3
COX2	699	2.7	19	4	17	2	0	0	1	18	18	1
ATP8	168	1.2	2	1	2	0	0	0	1	1	1	1
ATP6	684	6.3	43	10	41	2	0	7	1	35	37	6
COX3	786	4.2	33	9	29	4	0	2	1	30	31	2
ND3	351	6.3	22	12	20	2	0	3	1	18	20	2
ND4L	297	2.4	7	3	6	1	0	0	1	6	6	1
ND4	1,386	7.8	108	52	95	13	0	15	4	89	92	16
ND5	1,839	7.2	133	54	120	13	0	26	5	102	107	26
ND6	522	9.8	51	27	48	3	0	8	5	38	44	7
CYTB	1,161	6.7	78	35	73	5	0	6	0	72	76	2
tRNA	1,504	3.3	49	18	41	8	0	n/a	n/a	n/a	n/a	n/a
CR	829	6.0	50	35	31	14	5	n/a	n/a	n/a	n/a	n/a
Other	70	22.9	16	8	12	3	1	n/a	n/a	n/a	n/a	n/a
Total	**16,512**	**5.4**	**887**	**372**	**779**	**102**	**6**	**94**	**23**	**589**	**619**	**87**

Table 4.4. Sample size (n), number of haplotypes (h), haplotype diversity (H_d), nucleotide diversity (π), and assignment to BAPS group (Gp1 or Gp2) for 153 complete mitochondrial genomes among 14 sampling locations of Atlantic cod. Refer to Figure 4.1 for locations.

Population	**n**	**h**	**H_d**	**π**	**Gp1**	**Gp2**
NC	7	7	1.000	0.0026	1	6
BAL	10	10	1.000	0.0027	3	7
QAS	18	7	0.791	0.0020	6	12
LBH	12	12	1.000	0.0028	4	8
LBM	14	14	1.000	0.0016	1	13
LBP	14	14	1.000	0.0018	0	14
HC	10	10	1.000	0.0024	3	7
GIL	2	2	1.000	0.0004	0	2
RAN	11	11	1.000	0.0021	3	8
NGB	13	13	1.000	0.0017	1	12
FC	11	11	1.000	0.0018	1	10
ESS	11	11	1.000	0.0017	3	8
FSS	10	10	1.000	0.0023	2	8
GEO	10	10	1.000	0.0022	2	8
Total	**153**	**142**	**0.997**	**0.0023**	**30**	**123**

Table 4.5. Population pairwise Φ_{ST} values (below diagonal) and corresponding p-values (above diagonal) based on 100,172 permutations for complete mitogenomes from (a) 13 populations of Atlantic cod, and (b) putative stock classifications. Significant p-values after modified FDR (P_{crit} = 0.010 and P_{crit} = 0.012 respectively) are shown in bold (after sequential Bonferroni correction P_{crit} = 0.001 and 0.002). Refer to Figure 4.1 for locations.

(a)	**NC**	**BAL**	**QAS**	**LBH**	**LBM**	**LBP**	**HC**	**RAN**	**NGB**	**FC**	**ESS**	**FSS**	**GEO**
NC	*	0.057	**0.009**	0.123	**0.003**	**0.010**	0.169	0.091	**<0.001**	0.012	0.030	**0.010**	0.151
BAL	0.082	*	0.024	0.018	**<0.001**	**<0.001**	0.039	**0.004**	**<0.001**	0.052	**0.009**	**0.010**	0.044
QAS	0.253	0.153	*	**<0.001**	**<0.001**	**<0.001**	**0.002**	**<0.001**	**<0.001**	**0.003**	**<0.001**	**0.001**	**0.003**
LBH	0.029	0.082	0.236	*	0.042	0.3946	0.756	0.751	0.234	0.281	0.812	0.409	0.583
LBM	0.085	0.153	0.315	0.042	*	0.646	0.073	0.142	0.049	0.154	0.039	0.018	0.078
LBP	0.062	0.140	0.301	0.001	0.000	*	0.367	0.780	0.247	0.488	0.738	0.175	0.536
HC	0.029	0.086	0.263	0.000	0.040	0.004	*	0.557	0.192	0.200	0.764	0.297	0.453
RAN	0.033	0.121	0.276	0.000	0.020	0.000	0.000	*	0.234	0.342	0.622	0.306	0.641
NGB	0.120	0.180	0.340	0.014	0.046	0.008	0.030	0.013	*	0.230	0.187	0.328	0.299
FC	0.064	0.075	0.245	0.009	0.023	0.000	0.020	0.005	0.013	*	0.484	0.581	0.587
ESS	0.047	0.114	0.292	0.000	0.044	0.000	0.000	0.000	0.019	0.000	*	0.580	0.702
FSS	0.065	0.105	0.271	0.000	0.060	0.015	0.012	0.006	0.005	0.000	0.000	*	0.568
GEO	0.033	0.093	0.258	0.000	0.036	0.000	0.000	0.000	0.009	0.000	0.000	0.000	*

(b)	**NC**	**BAL**	**QAS**	**2GH**	**2J3KL**	**3M**	**4VsW**	**4X**	**5YZ**
NC	*	0.057	**0.009**	0.025	0.017	0.013	0.030	**0.009**	0.153
BAL	0.082	*	0.026	**<0.001**	**<0.001**	0.052	**0.008**	**0.009**	0.044
QAS	0.253	0.153	*	**<0.001**	**<0.001**	**0.003**	**<0.001**	**0.001**	**0.002**
2GH	0.061	0.133	0.279	*	0.253	0.381	0.621	0.099	0.322
2J3KL	0.071	0.148	0.295	0.003	*	0.300	0.690	0.269	0.447
3M	0.064	0.075	0.245	0.000	0.005	*	0.481	0.583	0.588
4VsW	0.047	0.114	0.292	0.000	0.000	0.000	*	0.582	0.699
4X	0.065	0.105	0.271	0.023	0.007	0.000	0.000	*	0.568
5YZ	0.033	0.093	0.258	0.003	0.000	0.000	0.000	0.000	*

Supplementary Table 4.1. Frequency of 40 bp control region tandem repeat units in 38 Atlantic cod samples.

Population	Copy number (frequency)					
	1	**2**	**3**	**4**	**5**	**6**
NC	0.00	0.10	0.23	0.42	0.25	0.00
NC	0.00	0.00	0.00	0.46	0.54	0.00
BAL	0.00	0.16	0.65	0.19	0.00	0.00
BAL	0.00	0.09	0.52	0.39	0.00	0.00
HC	0.00	0.10	0.59	0.30	0.00	0.00
HC	0.00	0.00	0.57	0.43	0.00	0.00
RAN	0.00	0.40	0.34	0.26	0.00	0.00
RAN	0.00	0.09	0.25	0.42	0.23	0.00
RAN	0.00	0.32	0.35	0.33	0.00	0.00
RAN	0.00	0.23	0.55	0.23	0.00	0.00
RAN	0.00	0.23	0.78	0.00	0.00	0.00
RAN	0.00	0.15	0.33	0.52	0.00	0.00
NGB	0.00	0.15	0.31	0.54	0.00	0.00
NGB	0.00	0.14	0.41	0.27	0.00	0.18
NGB	0.00	0.15	0.37	0.47	0.00	0.00
NGB	0.00	0.33	0.41	0.26	0.00	0.00
NGB	0.06	0.27	0.39	0.16	0.11	0.00
NGB	0.00	0.15	0.58	0.27	0.00	0.00
NGB	0.07	0.48	0.29	0.16	0.00	0.00

NGB	0.00	0.33	0.67	0.00	0.00	0.00
NGB	0.00	0.35	0.65	0.00	0.00	0.00
FC	0.00	0.15	0.33	0.36	0.15	0.00
FC	0.00	0.13	0.49	0.25	0.13	0.00
ESS	0.00	0.42	0.36	0.22	0.00	0.00
ESS	0.00	0.57	0.43	0.00	0.00	0.00
FSS	0.00	0.24	0.76	0.00	0.00	0.00
FSS	0.00	0.09	0.37	0.54	0.00	0.00
GEO	0.00	0.25	0.75	0.00	0.00	0.00
GEO	0.00	0.14	0.50	0.36	0.00	0.00
GEO	0.00	0.18	0.38	0.44	0.00	0.00
GEO	0.00	0.19	0.81	0.00	0.00	0.00
GEO	0.00	0.13	0.41	0.46	0.00	0.00
GEO	0.00	0.18	0.23	0.37	0.22	0.00
GEO	0.00	0.17	0.33	0.31	0.19	0.00
GEO	0.00	0.19	0.62	0.19	0.00	0.00
GEO	0.00	0.13	0.31	0.27	0.30	0.00
GEO	0.00	0.24	0.39	0.36	0.00	0.00
GEO	0.00	0.19	0.25	0.31	0.25	0.00

Supplementary Table 4.2. Overall Φ_{ST} and p-values for AMOVAs of individual loci. The populations with significant pairwise comparisons are given, those not significant after correction for multiple tests are shown in grey. The clades supported in the individual neighbour-joining trees (NJ) and their supporting bootstrap values are given.

	AMOVA		**Pairwise**	**Clades**
Region	**Φ_{ST}**	**p**	**Φ_{ST}**	**NJ**
12S	0.021	0.014	ns	D(54)
16S	0.144	< 0.001	QAS, FSS	none
ND1	0.101	< 0.001	QAS	generally cluster
ND2	0.107	< 0.001	NC, BAL, QAS	generally cluster H(65) D(65) I(64)
COX1	0.069	< 0.001	QAS	some clustering
COX2	0.072	< 0.001	some NC, QAS	H(65) G(66)
ATP8	0.057	0.038	ns	none
ATP6	0.091	< 0.001	BAL, QAS	D(66) E(65)
COX3	0.122	< 0.001	QAS	I(63)
ND3	0.081	0.002	QAS	H(77) I(53)
ND4L	0.024	0.040	Ns	none
ND4	0.108	< 0.001	some BAL, QAS	generally cluster B(54) D(42) E(42)
ND5	0.111	< 0.001	QAS	H(54) B(49) D(47)
ND6	0.148	< 0.001	BAL, QAS	B(41) E(45) F(11)
CYTB	0.092	< 0.001	some BAL, QAS	D(60) E(10) G(23) K(46)
CR	0.079	< 0.001	some QAS	H(85) B(25) D(46)

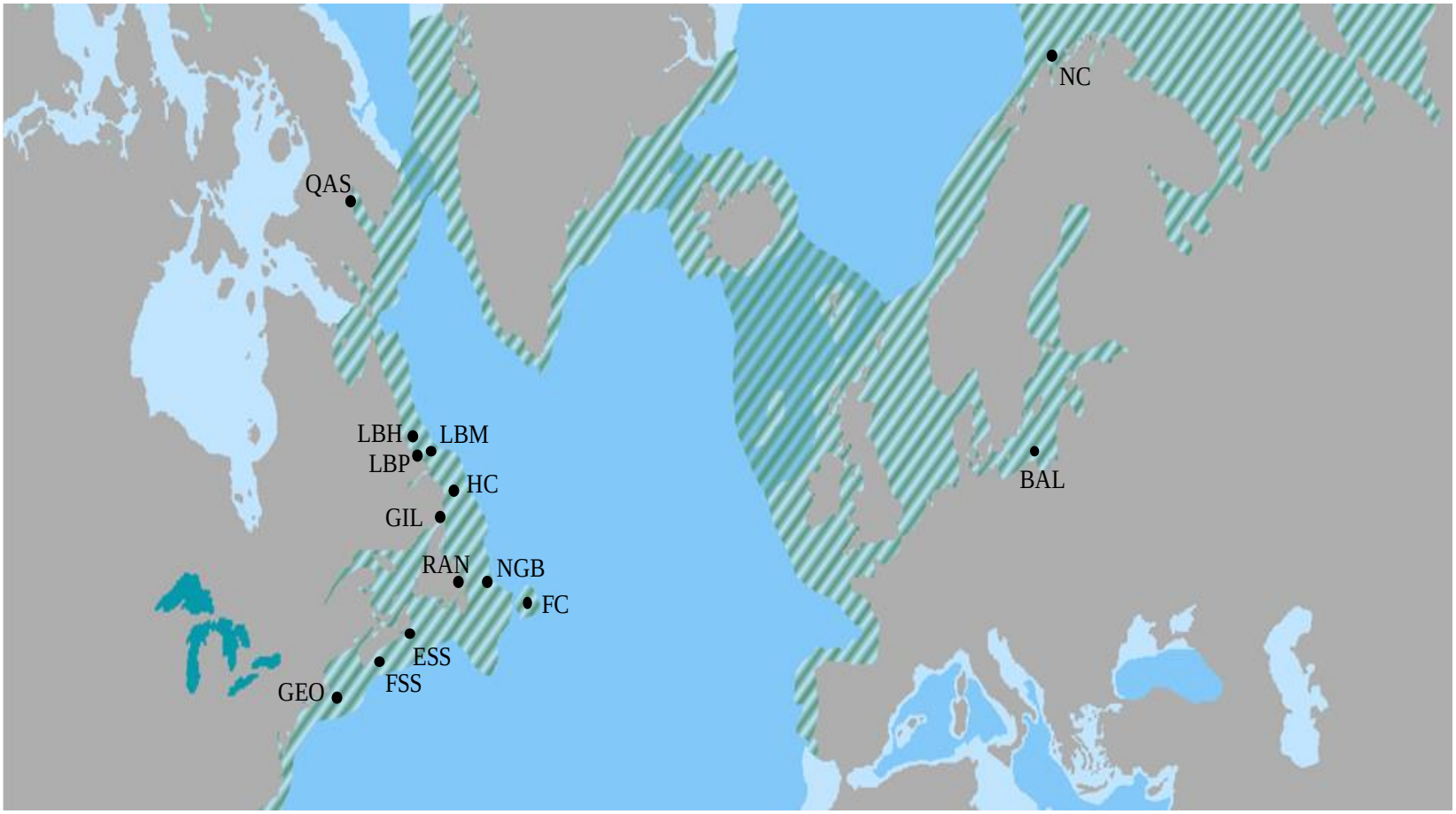

NC
QAS
LBH
LBM
LBP
HC
GIL
BAL
RAN
NGB
FC
ESS
FSS
GEO

Figure 4.1. Distribution of Atlantic cod across the North Atlantic Ocean (shaded). Sampling locations are shown by the circles: NC (Norwegian Coast at Tromsö), BAL (Baltic Sea at Sopot, Poland), QAS (Lake Qasigialiminiq, Baffin Island, Nunavut), LBH, LBM, and LBP (Labrador – Hopedale, Makkovik, and Postville, NAFO 2H), HC (Hawke Channel, NAFO 2J offshore), GIL (Gilbert Bay, 2J inshore), RAN (Random Island, NAFO 3L inshore), NGB (North Cape of the Grand Banks, NAFO 3L offshore), FC (Flemish Cap, NAFO 3M), ESS (eastern Scotian Shelf, NAFO 4W), FSS (Fundy-Scotian Shelf, NAFO 4X), and GEO (Georges Bank, NAFO 5Ze). Samples from NC, HC, NGB, and FC are from Carr & Marshall (2008a) with additional sequences added. Figure modified from FishBase (2013).

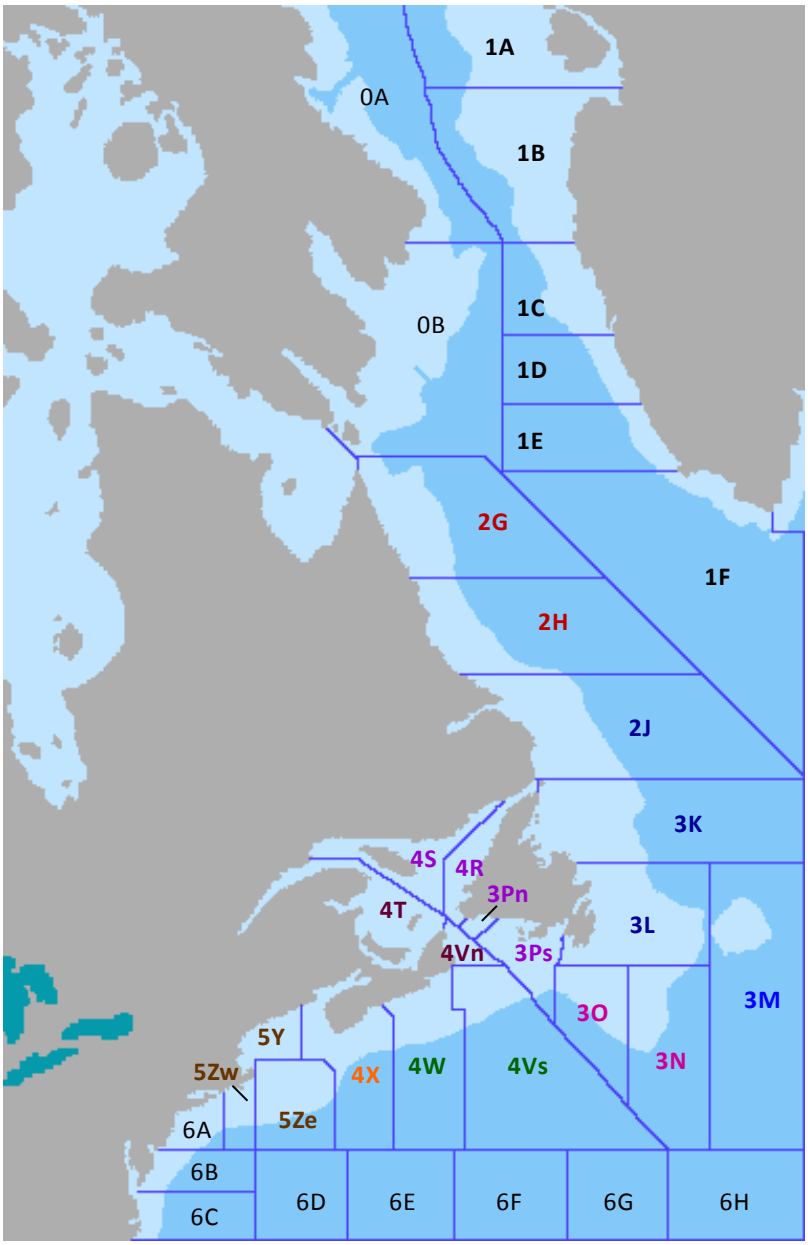


Figure 4.2. Delineation of the seven NAFO divisions (0 - 6) and 25 subdivisions in the Northwest Atlantic Ocean. The 10 Atlantic cod management stocks (in divisions 1 - 5) are shown in bold and grouped by colour. Divisions 0 and 6 are not included in management stocks due to low cod densities.

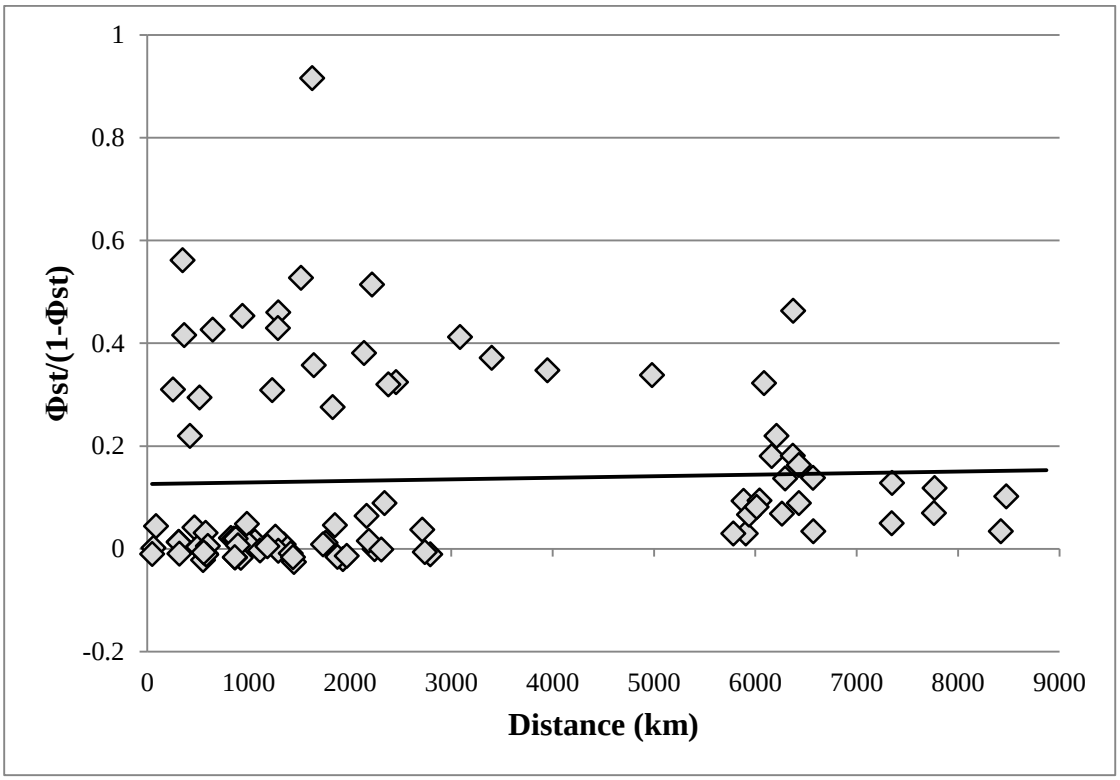


Figure 4.3. Mantel's test for isolation-by-distance showing the relationship between genetic (Φ_{ST}/ (1-Φ_{ST})) and geographic distances for the shortest marine swimming distance (p=0.083, r^2 = 0.001). Similar results were obtained when straight-line distances were used (p=0.087, r^2 = 0.005).

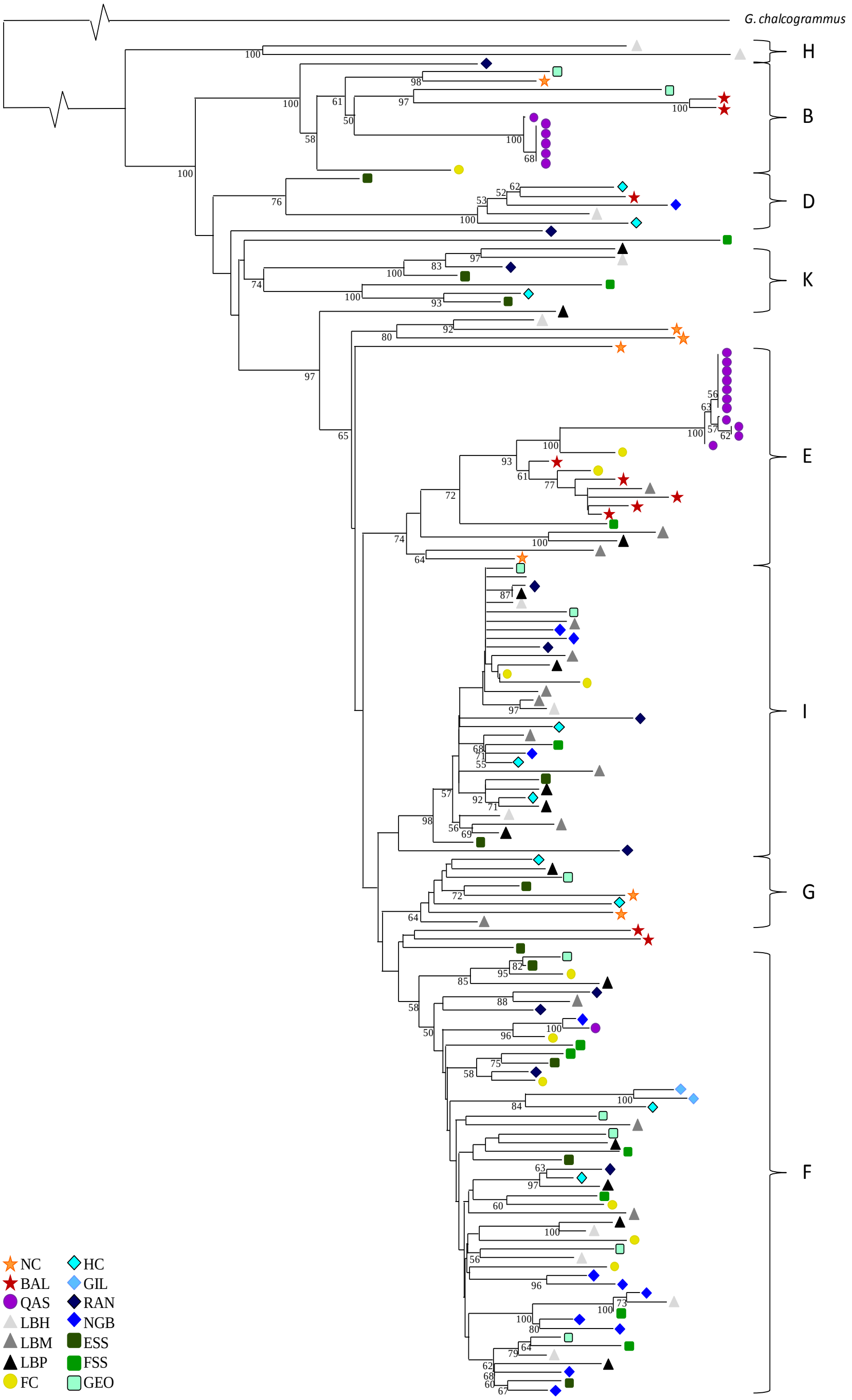

G. chalcogrammus
H
B
D
K
E
I
G
F
NC
BAL
QAS
LBH
LBM
LBP
FC
HC
GIL
RAN
NGB
ESS
FSS
GEO

Figure 4.4. Neighbour-joining analysis of complete mitogenomes from 153 Atlantic cod. Bootstrap values > 50% are given (10,000 replicates). Each symbol represents an individual colour-coded by sampling location. Refer to Figure 4.1 for locations.


G. chalcogrammus
H
B
τ = 12.6
D
τ = 19.5
K
τ = 20.6
E
τ = 8.09
I
τ = 10.0
G
τ = 15.1
F
τ = 20.3
NC
BAL
QAS
LBH
LBM
LBP
FC
HC
GIL
RAN
NGB
ESS
FSS
GEO
3.5
0.30
Node age (MY)
0.20
0.10
0.0

Figure 4.5. Clock-calibrated Bayesian analysis of complete mitogenomes from 153 Atlantic cod. Grey lines represent time since separation in million years. Posterior probabilities above 0.5 are given. Clade membership and τ are given on the right. Refer to Figure 4.1 for locations. A more detailed image is given in Supplementary Figure 1.

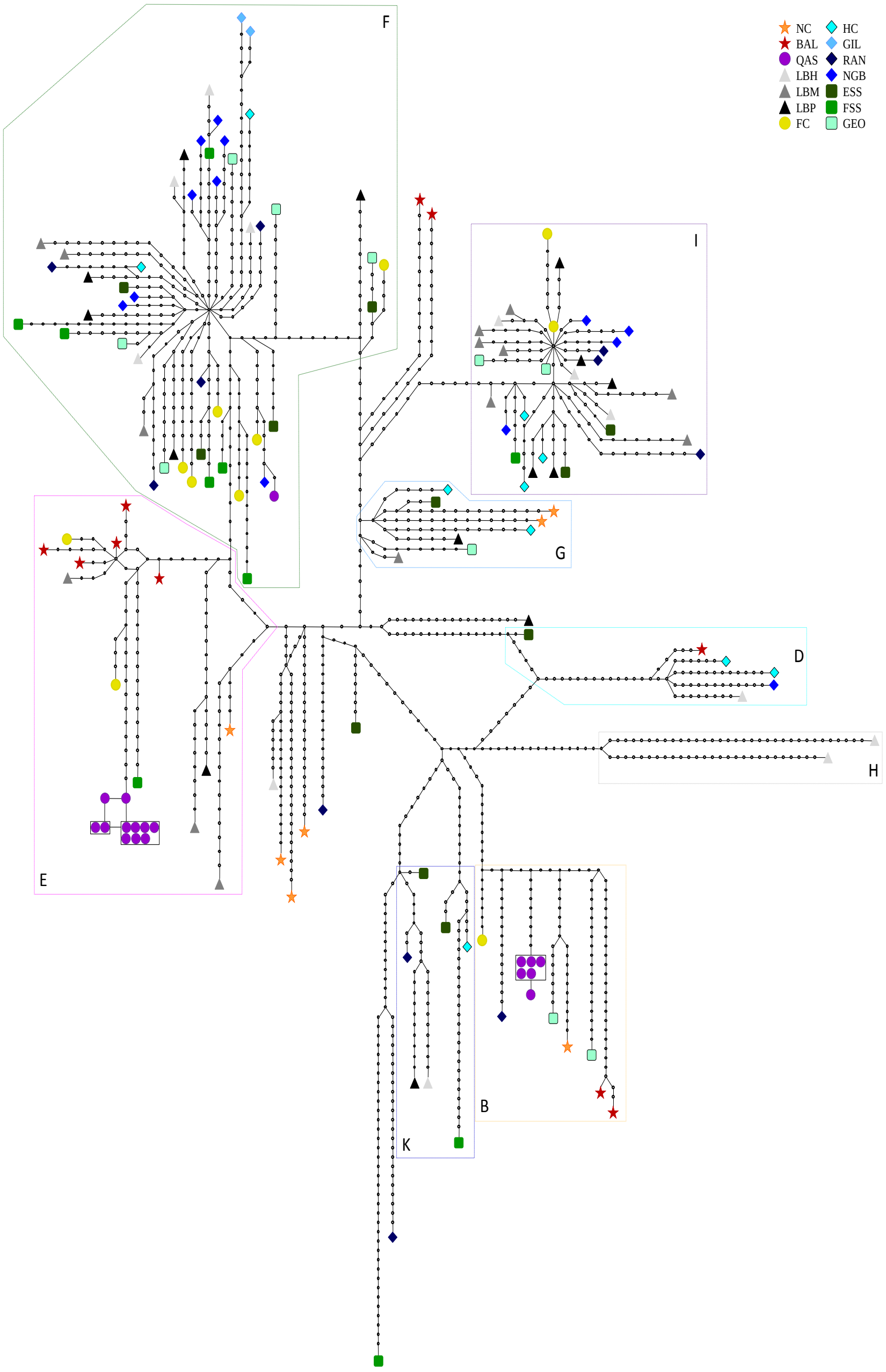
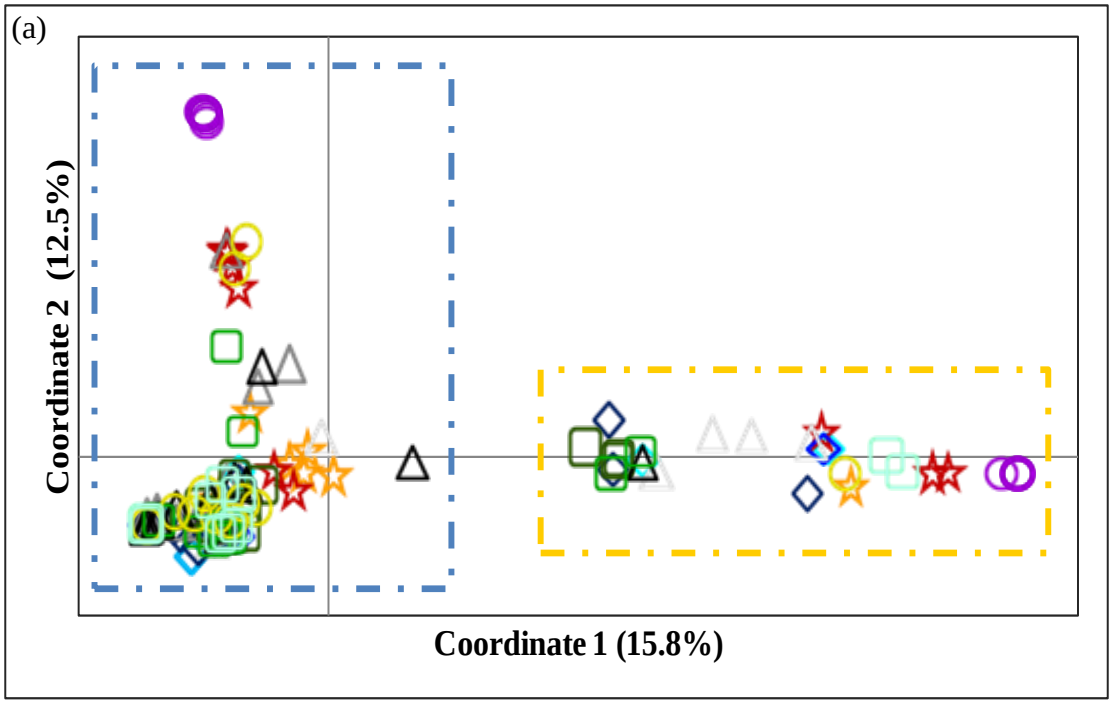
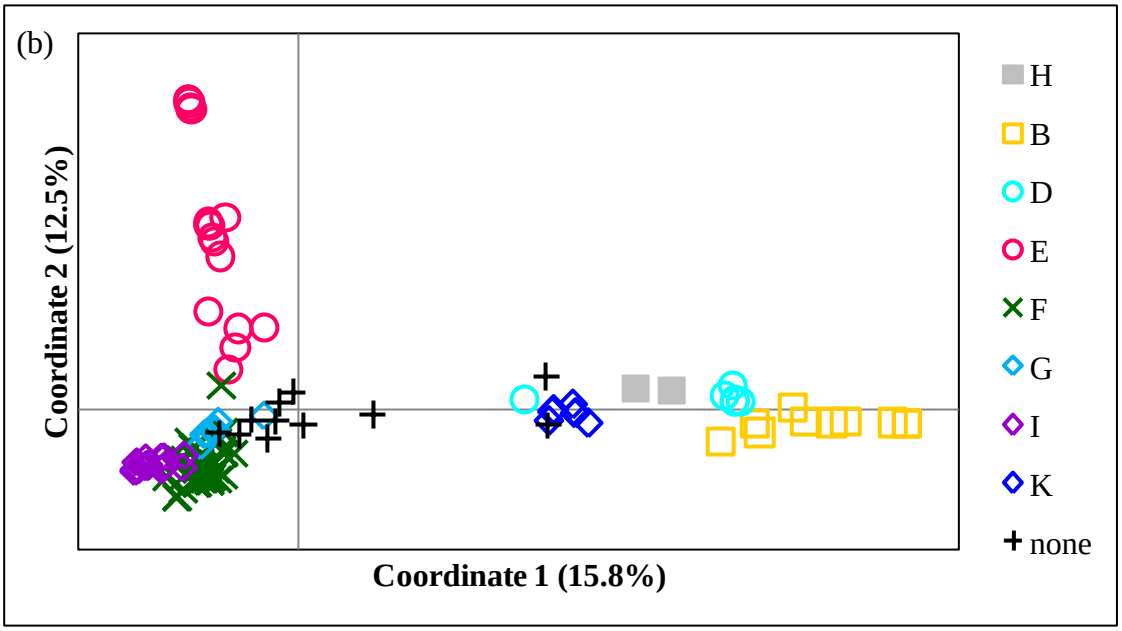


Figure 4.6. Statistical parsimony network of complete mitogenomes for 153 Atlantic cod. Each symbol represents an individual, the black dots are inferred or intermediate haplotypes, and each connection represents one nucleotide change. Shared haplotypes are encased by a black box. The dashed boxes correspond to the clades identified in Figure 4.4. Refer to Figure 4.1 for locations.


(a)
Coordinate 2 (12.5%)
Coordinate 1 (15.8%)


(b)
Coordinate 2 (12.5%)
Coordinate 1 (15.8%)
H
B
D
E
F
G
I
K
none

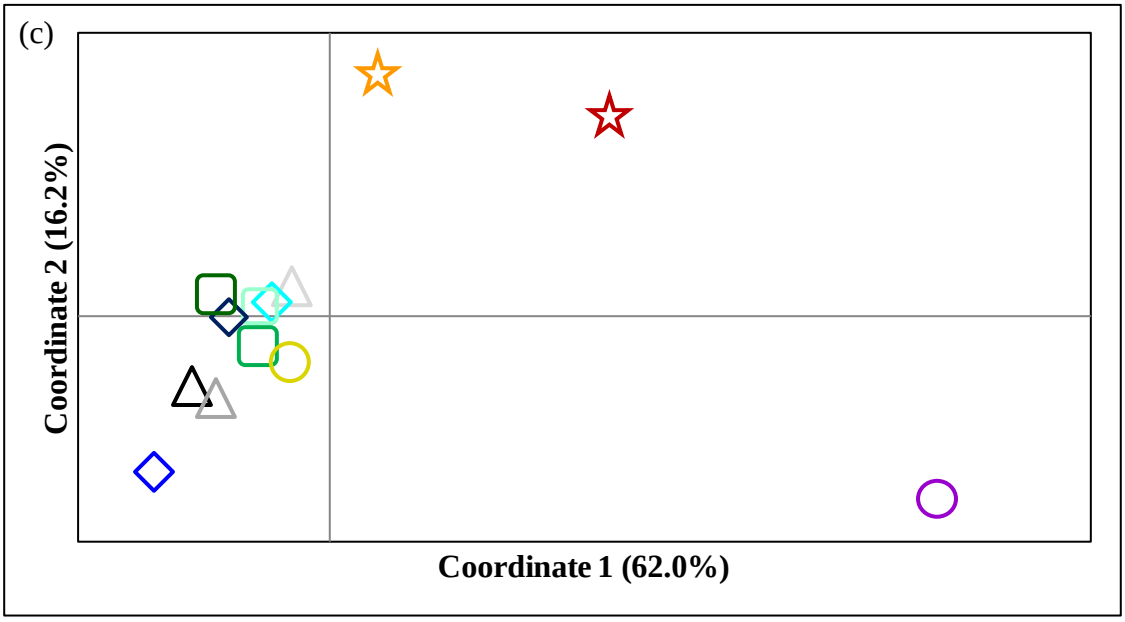
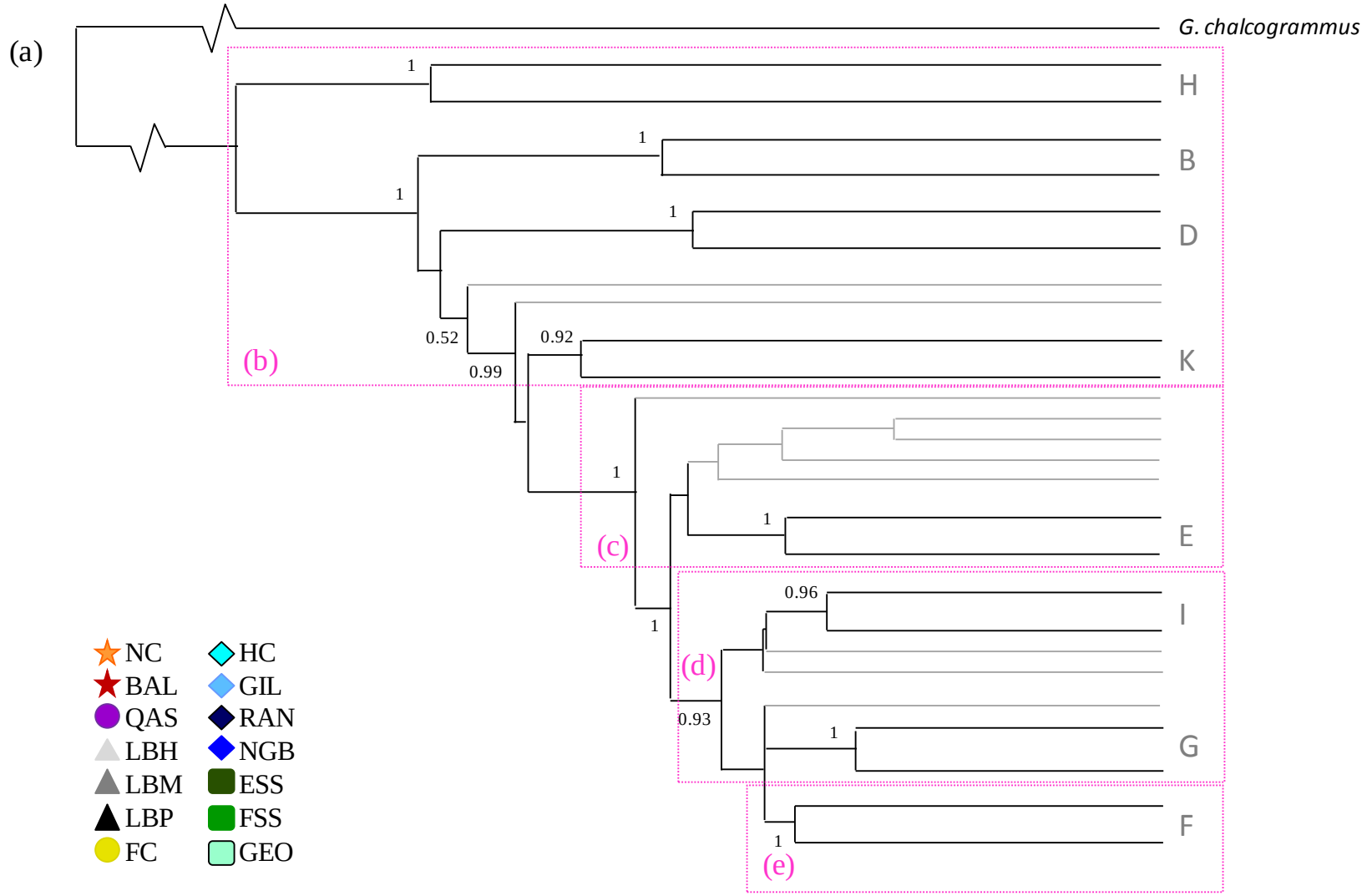
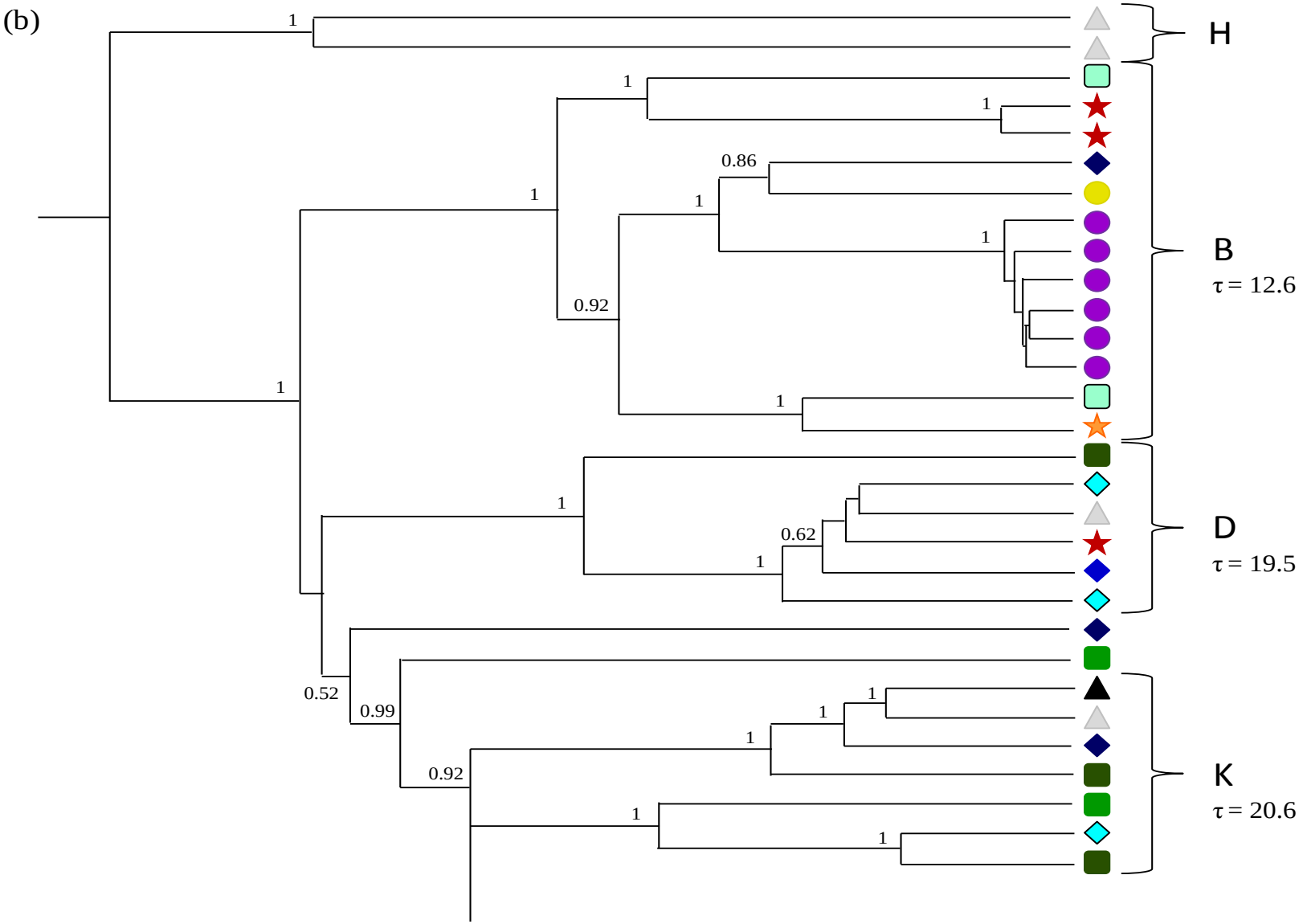
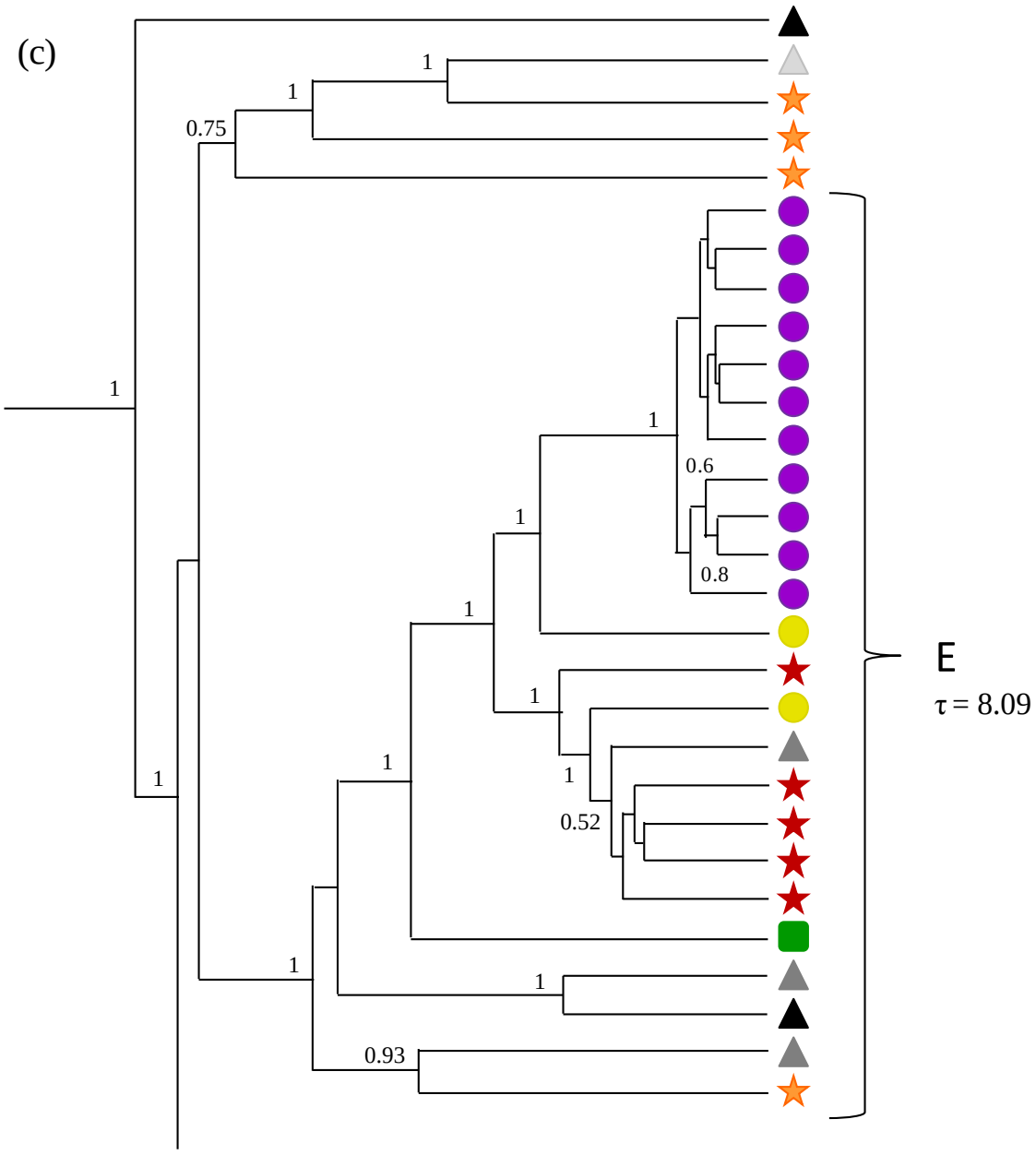


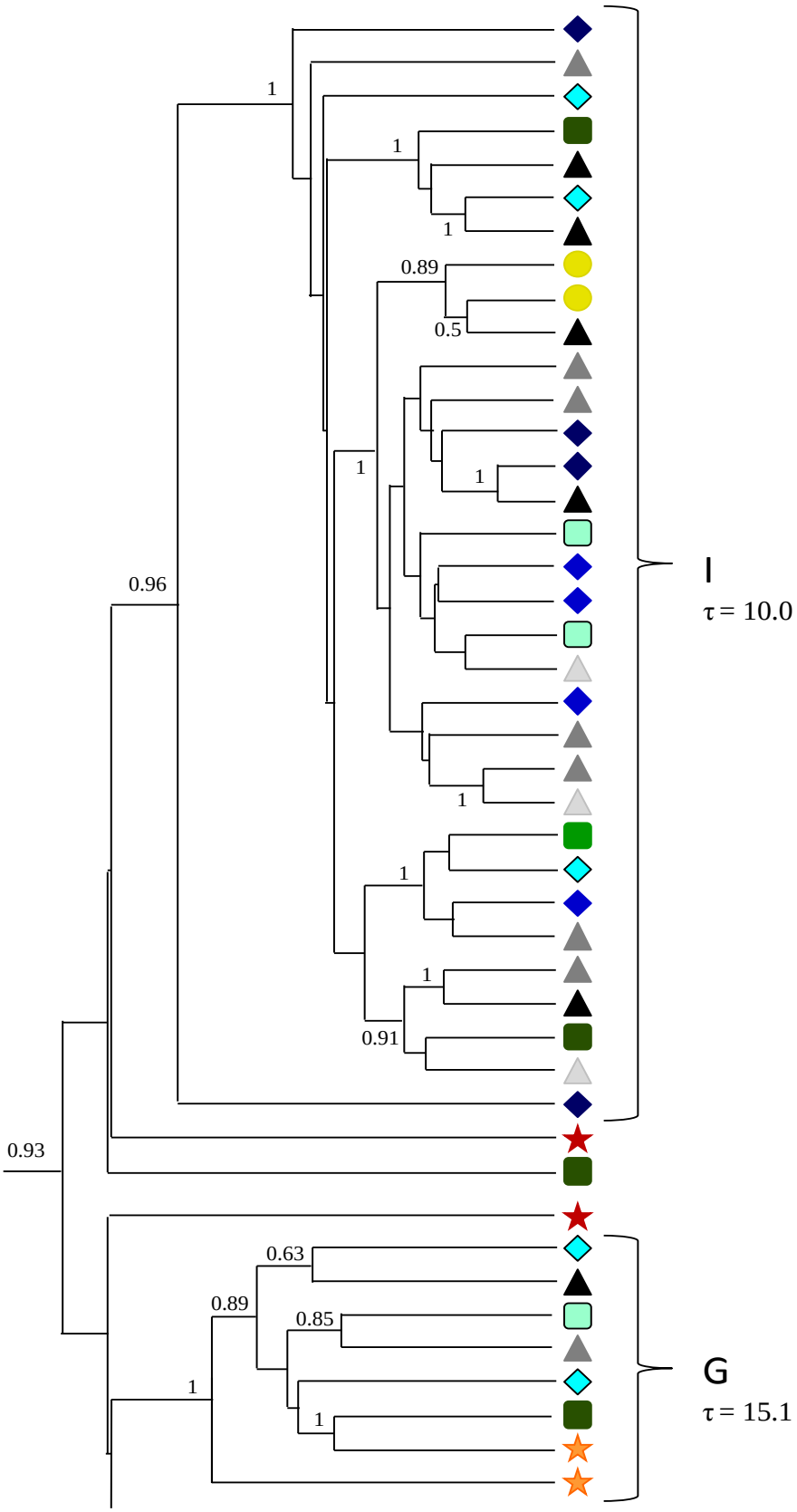
Figure 4.7. Principal coordinates analysis (PCoA) based on (a) 153 individual mtDNA sequences coloured by population, (b) coloured by NJ clade and (c) the population Φ_{ST} values for 13 populations (excluding GIL). Populations are colour-coded as per Figure 4.4. The dashed boxes in (a) represent the two groups identified in BAPS v6. Refer to Figure 4.1 for locations.


(a)
G. chalcogrammus
H
B
D
K
E
I
G
F
(b)
(c)
(d)
(e)
1
1
1
1
0.52
0.92
0.99
1
1
0.96
1
0.93
1
NC
BAL
QAS
LBH
LBM
LBP
FC
HC
GIL
RAN
NGB
ESS
FSS
GEO

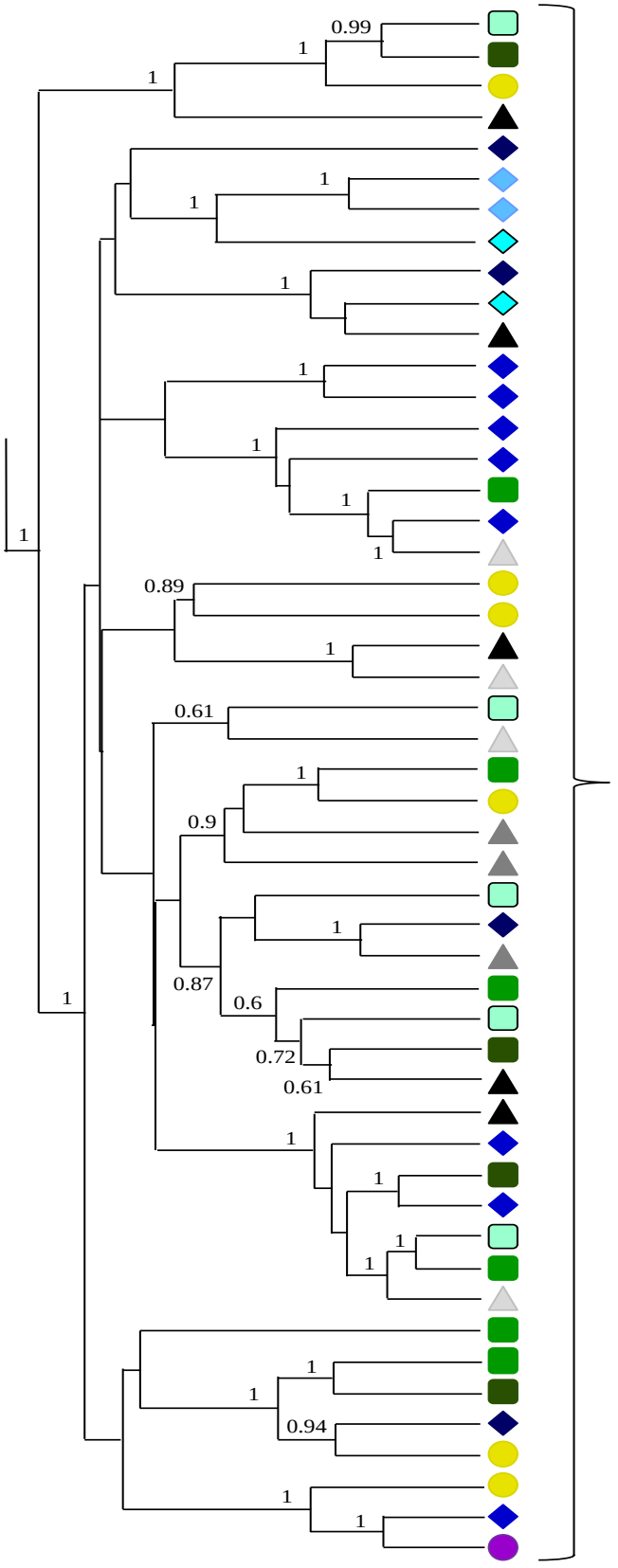

(b)
H
B
τ = 12.6
D
τ = 19.5
K
τ = 20.6


(c)
1
1
0.75
1
1
0.6
1
0.8
1
1
1
1
0.52
1
1
0.93
E
τ = 8.09

(d)



(e)


0.99
1
1
1
1
1
1
1
1
1
1
1
0.89
1
0.61
1
0.9
1
0.87
0.6
0.72
0.61
1
1
1
1
1
1
0.94
1
1
F
τ = 20.3

Supplementary Figure 1. (a) Simplified Bayesian analysis of complete mitogenomes from 153 Atlantic cod. Grey lines represent samples that do not fall into one of the eight major clades (labelled), and (b-e) show a close-up of the samples found in each of the eight major clades. Bayesian probabilities ≥ 0.5 are given. Refer to Figure 4.1 for locations.

References

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon J-M, Roostalu U, Loogväli E-L, Kivisild T, Bandelt H-J, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A (2004) The molecular dissection of mtDNA Haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *The American Journal of Human Genetics* **75**, 910-918.

Aitken A, Gilbert R (1989) Holocene nearshore environments and sea-level history in Pangnirtung Fiord, Baffin Island, N.W.T., Canada. *Arctic and Alpine Research* **21**, 34-44.

Alheit J, Hagen E (1997) Long-term climate forcing of European herring and sardine populations. *Fisheries Oceanography* **6**, 130-139.

Alvarado Bremer J, Stequert B, Robertson N, Ely B (1998) Genetic evidence for inter-oceanic subdivision of bigeye tuna (*Thunnus obesus* Lowe) populations. *Marine Biology* **132**, 547-557.

Árnason E (2004) Mitochondrial cytochrome b DNA variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. *Genetics* **166**, 1871-1885.

Árnason E, Halldórsdóttir K (2015) Nucleotide variation and balancing selection at the *Ckma* gene in Atlantic cod: analysis with multiple merger coalescent models. *PeerJ* **3**, e786.

Árnason E, Pálsson S (1996) Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod *Gadus morhua*, from Norway. *Molecular Ecology* **5**, 715-724.

Árnason E, Pálsson S, Arason A (1992) Gene flow and lack of population differentiation in Atlantic cod, *Gadus morhua* L., from Iceland, and comparison of cod from Norway and Newfoundland. *Journal of Fish Biology* **40**, 751-770.

Árnason E, Petersen P, Kristinsson K, Sigurgíslason H, Pálsson S (2000) Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod from Iceland and Greenland. *Journal of Fish Biology* **56**, 409-430.

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

Árnason E, Rand DM (1992) Heteroplasmy of short tandem repeats in mitochondrial DNA of Atlantic cod, *Gadus morhua*. *Genetics* **132**, 211-220.

Arnason U, Adegoke J, Gullberg A, Harley E, Janke A, Kullberg M (2008) Mitogenomic relationships of placental mammals and molecular estimates of their divergences. *Gene* **421**, 37-51.

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

Barendregt RW, Irving E (1998) Changes in the extent of North American ice sheets during the late Cenozoic. *Canadian Journal of Earth Sciences* **35**, 504-509.

Beacham TD, Brattey J, Miller KM, Le KD, Withler RE (2002) Multiple stock structure of Atlantic cod (*Gadus morhua*) off Newfoundland and Labrador determined from genetic variation. *ICES Journal of Marine Science* **59**, 650-665.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165-1188.

Bentzen P, Taggart C, Ruzzante D, 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 PR, Star B, Pampoulie C, Sodeland M, Barth JMI, Knutsen H, Jakobsen KS, Jentoft S (2016) Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports* **6**, 23246.

Bigg GR, Cunningham CW, Ottersen G, Pogson GH, Wadley MR, Williamson P (2008) Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. *Proceedings of the Royal Society B: Biological Sciences* **275**, 163-173.

Bouckaert R, J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard M, Rambaut A, Drummond A (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537.

Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Snelgrove PVR, Morris CJ, Gregory RS, Hardie DC, Hutchings JA, Ruzzante DE, Taggart CT, Bentzen P (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences* **277**, 3725-3734.

Bradbury IR, Hubert S, Higgins B, Bowman S, Borza T, Paterson IG, Snelgrove PVR, Morris CJ, Gregory RS, Hardie D, Hutchings JA, Ruzzante DE, Taggart CT, Bentzen P (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications* **6**, 450-461.

Canino MF, Spies IB, Cunningham KM, Hauser L, Grant WS (2010) Multiple ice-age refugia in Pacific cod, *Gadus macrocephalus*. *Molecular Ecology* **19**, 4339-4351.

Carpenter KE, Abrar M, Aeby G, Aronson RB, Banks S, Bruckner A, Chiriboga A, Cortés J, Delbeek JC, DeVantier L, Edgar GJ, Edwards AJ, Fenner D, Guzmán HM, Hoeksema BW, Hodgson G, Johan O, Licuanan WY, Livingstone SR, Lovell ER, Moore JA, Obura DO, Ochavillo D, Polidoro BA, Precht WF, Quibilan MC, Reboton C, Richards ZT, Rogers AD, Sanciangco J, Sheppard A, Sheppard C, Smith J, Stuart S, Turak E, Veron JEN, Wallace C, Weil E, Wood E (2008) One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science* **321**, 560-563.

Carr S, Duggan A, Stenson G, Marshall H (2015) Quantitative phylogenomics of within-species mitogenome variation: Monte Carlo and non-parametric analysis of phylogeographic structure among discrete transatlantic breeding areas of harp seals (*Pagophilus groenlandicus*). *PLOS One* **10**, e0134207.

Carr S, Marshall H (2008a) Intraspecific phylogeographic genomics from multiple complete mtDNA genomes in Atlantic cod (*Gadus morhua*): origins of the "codmother", transatlantic vicariance and midglacial population expansion. *Genetics* **180**, 381-389.

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: *The Implications of Localized Fishery Stocks* (eds. Hunt von Herbing I, Kornfield I, Tupper M, Wilson J). Northeast Regional Agricultural Engineering Service, Ithaca, New York.

Carr SM, Duggan AT, Marshall HD (2009) Iterative DNA sequencing on microarrays: a high-throughput NextGen technology for ecological and evolutionary mitogenomics. *Laboratory Focus* **13**, 8-12.

Carr SM, Marshall HD (1991a) Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic cod (*Gadus morhua*) by the

polymerase chain reaction. *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 48-52.

Carr SM, Marshall HD (1991b) A direct approach to the measurement of genetic variation in fish populations: applications of the polymerase chain reaction to studies of Atlantic cod, *Gadus morhua* L. *Journal of Fish Biology* **39 (Supplement A)**, 101-107.

Carr SM, Marshall HD (2008b) Phylogeographic analysis of complete mtDNA genomes from walleye pollock (*Gadus chalcogrammus* Pallas, 1811) shows an ancient origin of genetic biodiversity. *Mitochondrial DNA* **19**, 490-496.

Carr SM, Marshall HD, Duggan AT, Flynn SMC, Johnstone KA, Pope AM, Wilkerson CD (2008) Phylogeographic genomics of mitochondrial DNA: highly-resolved patterns of intraspecific evolution and a multi-species, microarray-based DNA sequencing strategy for biodiversity studies. *Comparative Biochemistry and Physiology D-Genomics & Proteomics* **3**, 1-11.

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.

Castro ALF, Stewart BS, Wilson SG, Hueter RE, Meekan MG, Motta PJ, Bowen BW, Karl SA (2007) Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). *Molecular Ecology* **16**, 5183-5192.

Chow S, Ushiama H (1995) Global population structure of albacore (*Thunnus alalunga*) inferred by RFLP analysis of the mitochondrial ATPase gene. *Marine Biology* **123**, 39-45.

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

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**, 704-707.

Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* **9**, 539.

Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* **205**, 19-31.

COSEWIC (2003) COSEWIC assessment and update status report on the Atlantic cod *Gadus morhua* in Canada p. xi + 76 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

COSEWIC (2010) COSEWIC assessment and status report on the Atlantic cod *Gadus morhua* in Canada p. xiii + 105 pp., Committee on the Status of Endangered Wildlife in Canada. Ottawa.

Coulson MW, Marshall HD, Pepin P, Carr SM (2006) Mitochondrial genomics of gadine fishes: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome* **49**, 1115-1130.

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.

Desjardins P, Morais R (1990) Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. *Journal of Molecular Biology* **212**, 599-634.

Duggan A (2007) *Efficiency and accuracy of a multi-species iterative DNA sequencing microarray*, BSc Honours Dissertation, Memorial University of Newfoundland.

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**, 2571-2581.

Ersts P (2015) Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation.

Excoffier L, Lischer H (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

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

Feutry P, Kyne P, Pillans R, Chen X, Naylor G, Grewe P (2014) Mitogenomics of the speartooth shark challenges ten years of control region sequencing. *BMC Evolutionary Biology* **14**, e232.

FishBase (2013) Reviewed distribution maps for *Gadus morhua* (Atlantic cod), with modeled year 2100 native range map based on IPCC A2 emissions scenario. In: *www.aquamaps.org*.

Fisheries and Oceans Canada (2016) Atlantic cod. In: *Underwater World*.

Frankham R (1995) Conservation genetics. *Annual Review of Genetics* **29**, 305-327.

Frydenberg O, Møller D, Nævdal G, Sick K (1965) Haemoglobin polymorphism in Norwegian cod populations. *Hereditas* **53**, 257-271.

Gibbard P, van Kolfschoten T (2004) The Pleistocene and Holocene Epochs. In: *A Geologic Time Scale 2004* (eds. Gradstein F, Ogg J, Smith A). Cambridge University Press, Cambridge.

Gjøsæter J, Jørstad K, Nævdal G, Thorkildsen S (1992) Genotype distributions of cod from the Norwegian Skagerrak coast. *Sarsia* **76**, 255-259.

Goddard S, Wroblewski J, Taggart C, Howse K, Bailey W, Kao M, Fletcher G (1994) Overwintering of adult northern Atlantic cod (*Gadus morhua*) in cold inshore waters as evidenced by plasma antifreeze glycoprotein levels. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 2834-2842.

Grant W, Ståhl G (1988) Evolution of Atlantic and Pacific cod: loss of genetic variation and gene expression in Pacific cod. *Evolution* **42**, 138-146.

Gulland J, Williamson G (1962) Transatlantic journey of a tagged cod. *Nature* **4844**, 921.

Halliday R, Pinhorn A (1996) North Atlantic fishery management systems: a comparison of management methods and resource trends. *Journal of Northwest Atlantic Fisheries Science* **20**, 1-143.

Hardie D, Hutchings J (2011) The ecology of Atlantic cod (*Gadus morhua*) in Canadian Arctic lakes. *Arctic* **64**, 137-150.

Hardie DC (2003) The evolutionary ecology of Arctic fishes. *Arctic* **56**, 430-433.

Hardie DC (2004) Population genetics, life history, and ecology of arctic marine fishes. *Arctic* **57**, 444-448.

Hardie DC, Gillett RM, Hutchings JA (2006) The effects of isolation and colonization history on the genetic structure of marine-relict populations of Atlantic cod (*Gadus morhua*) in the Canadian Arctic. *Canadian Journal of Fisheries and Aquatic Sciences* **63**, 1830-1839.

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

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**, 183-195.

Hoelzel AR (1999) Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. *Biological Journal of the Linnean Society* **68**, 23-39.

Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70.

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proceedings of the National Academy of Sciences USA* **92**, 532-536.

Hovgård H, Christensen S (1990) Population structure and migration patterns of Atlantic cod (*Gadus morhua*) in West Greenland waters based on tagging experiments from 1946 to 1964. *NAFO Scientific Council Studies* **14**, 45-50.

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

Hutchinson W, Carvalho G, Rogers S (2001) Marked genetic structuring in localised spawning populations of cod *Gadus morhua* in the North Sea and adjoining waters, as revealed by microsatellites. *Marine Ecology Progress Series* **223**, 251-260.

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

Jamieson A (1967) Genetic diversity on cod. *ICES Demersal Fish (Northern) Committee* **C.M.1967/F:20**, 1-5.

Jamieson A, Birley A (1989) The demography of a haemoglobin polymorphism in the Atlantic cod, *Gadus morhua* L. *Journal of Fish Biology* **35 (Supplement A)**, 193-204.

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

Johansen SD, Coucheron DH, Andreassen M, Karlsen BO, Furmanek T, Jørgensen TE, Emblem Å, Breines R, Nordeide JT, Moum T, Nederbragt AJ, Stenseth NC, Jakobsen KS (2009) Large-scale sequence analyses of Atlantic cod. *New Biotechnology* **25**, 263-271.

Jónsson J (1996) Tagging of cod (*Gadus morhua*) in Icelandic waters 1948-1986 and tagging of haddock (*Gadus aeglefinus*) in Icelandic waters 1953-1965. *Journal of the Marine Research Institute Reykjavík* **14**, 1-108.

Kijewska A, Burzyński A, Wenne R (2009) Variation in the copy number of tandem repeats of mitochondrial DNA in the North-East Atlantic cod populations. *Marine Biology Research* **5**, 186-192.

Knaus B, Cronn R, Liston A, Pilgrim K, Schwartz M (2011) Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecology* **11**, e10.

Knutsen H, Jorde P, André C, Stenseth N (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Molecular Ecology* **12**, 385-394.

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

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Lluch-Belda D, Schwartzlose RA, Serra R, Parrish R, Kawasaki T, Hedgecock D, Crawford RJM (1992) Sardine and anchovy regime fluctuations of abundance in four regions of the world oceans: a workshop report. *Fisheries Oceanography* **1**, 339-347.

McLaren I (1967) Physical and chemical characteristics of Ogac Lake, a landlocked fjord on Baffin Island. *Journal of the Fisheries Research Board of Canada* **24**, 981-1015.

Mitrovica JX (2003) Recent controversies in predicting post-glacial sea-level change. *Quaternary Science Reviews* **22**, 127-133.

Miya M, Kawaguchi A, Nishida M (2001) Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Molecular Biology and Evolution* **18**, 1993-2009.

Miya M, Takeshima H, Endo H, Ishiguro N, Inoue J, Mukai T, Satoh T, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai S, 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.

Møller D (1968) Genetic diversity in spawning cod along the Norwegian coast. *Hereditas* **60**, 1-32.

Morin P, Archer F, Foote A, Vilstrup J, Allen E, Wade P, Durban J, Parsons K, Pitman R, Li L, Bouffard P, Nielsen S, Rasmussen M, Willerslev E, Gilbert M, Harkins T (2010) Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* **20**, 908-916.

Mork J, Ryman N, Ståhl 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.

Mork J, Sundnes G (1985) Haemoglobin polymorphism in Atlantic cod (*Gadus morhua*): allele frequency variation between yearclasses in a Norwegian fjord stock. *Helgolander Meeresuntersuchungen* **39**, 55-62.

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

Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences USA* **76**, 5269-5273.

Nei M, Tajima F (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145-164.

Nielsen E, Hemmer-Hansen J, Poulsen N, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger G-L, Ogden R, Carvalho G (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**, 276.

O'Dea NR, Haedrich RL (2002) A review of the status of the Atlantic wolffish, *Anarhichas lupus*, in Canada. *Canadian Field-Naturalist* **116**, 423-432.

O'Leary D, Coughlan J, Dillane E, McCarthy T, Cross T (2007a) Microsatellite variation in cod *Gadus morhua* throughout its geographic range. *Journal of Fish Biology* **70**, 310 - 335.

O'Leary DB, Coughlan J, Dillane E, McCarthy TV, Cross TF (2007b) Microsatellite variation in cod *Gadus morhua* throughout its geographic range. *Journal of Fish Biology* **70 (Supplement C)**, 310-355.

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

Patriquin D (1967) Biology of *Gadus morhua* in Ogac Lake, a landlocked fjord on Baffin Island. *Journal of the Fisheries Research Board of Canada* **24**, 2573-2594.

Peakall R, Smouse P (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* **28**, 2537-2539.

Peakall ROD, Smouse PE (2006) GenAlEx 6: Genetic Analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Pielou E (1991) *After the Ice Age: The Return of Life to Glaciated North America* University of Chicago Press, Chicago, IL.

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.

Pope AM, Carr SM, Smith KN, Marshall HD (2011) Mitogenomic and microsatellite variation in descendants of the founder population of Newfoundland: high genetic diversity in an historically isolated population. *Genome* **54**, 110-119.

Powles PM (1958) Studies of reproduction and feeding of Atlantic Cod (*Gadus callarias* L.) in the southwestern Gulf of St. Lawrence. *Journal of the Fisheries Research Board of Canada* **15**, 1383-1402.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248-249.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-569.

Ronquist F, Huelsenbeck J (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**.

Rose GA (2007) *Cod: An Ecological History of the North Atlantic Fisheries* Breakwater Books Ltd, St. John's, NL.

Rousset F (2008) Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.

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

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White D, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open source platform for biological-image analysis. *Nature Methods* **9**, 676-682.

Schneider C, Rasband W, Eliceiri K (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675.

Schultz JK, Baker JD, Toonen RJ, Bowen BW (2009) Extremely low genetic diversity in the endangered Hawaiian monk seal (*Monachus schauinslandi*). *Journal of Heredity* **100**, 25-33.

Shamblin B, Bjorndal K, Bolten A, Hillis-Starr Z, Lundgren I, Naro-Maciel E, Nairn C (2012) Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. *Molecular Ecology* **21**, 2330-2340.

Shaw J (2006) Palaeogeography of Atlantic Canadian continental shelves from the Last Glacial Maximum to the present, with an emphasis on Flemish Cap. *Journal of Northwest Atlantic Fisheries Science* **37**, 119-126.

Shulman M, Bermingham E (1995) Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* **49**, 897-910.

Sick K (1965a) Haemoglobin polymorphism of cod in the Baltic and the Danish Belt Sea. *Hereditas* **54**, 19-48.

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

Sigurgíslason H, Árnason E (2003) Extent of mitochondrial DNA sequence variation in Atlantic cod from the Faroe Islands: a resolution of gene genealogy. *Heredity* **91**, 557-564.

Stone A, Battistuzzi F, Kubatko L, Perry Jr G, Trudeau E, Lin H, Kumar S (2010) More reliable estimates of divergence times in *Pan* using complete mtDNA sequences and accounting for population structure. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 3277-3288.

Swofford D (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Tanaka M, Cabrera V, González A, Larruga J, Takeyasu T, Fuku N, Guo L-J, Hirose R, Fujita Y, Kurata M, Shinoda K-I, Umetsu K, Yamada Y, Oshida Y, Sato Y, Hattori N, Mizuno Y, Arai Y, Hirose N, Ohta S, Ogawa O, Tanaka Y, Kawamori R, Shamoto-Nagai M, Maruyama W, Shimokata H, Suzuki R, Shimodaira H (2004) Mitochondrial genome variation in Eastern Asia and the peopling of Japan. *Genome Research* **14**, 1832-1850.

Teacher A, André C, Merilä J, Wheat C (2012) Whole mitochondrial genome scan for population structure and selection in the Atlantic herring. *BMC Evolutionary Biology* **12**, e248.

Templeman W (1979) Migrations and intermingling of stocks of Atlantic cod, *Gadus morhua*, of the Newfoundland and adjacent areas from tagging in 1962-66. *International Commission for the Northwest Atlantic Fisheries Research Bulletin* **14**, 5-50.

Templeman W (1981) Vertebral numbers in Atlantic cod, *Gadus morhua*, of the Newfoundland and adjacent areas, 1947-71, and their use for delineating cod stocks. *Journal of Northwest Atlantic Fisheries Science* **2**, 21-45.

Vermeij G (1991) Anatomy of invasion: the trans-Arctic interchange. *Paleobiology* **17**, 281-307.

Vis M, Carr S, Bowering W, Davidson W (1997) Greenland halibut (*Reinhardtius hippoglossoides*) in the North Atlantic are genetically homogeneous. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 1813-1821.

Wang Z, Shen X, Liu B, Su J, Yonezawa T, Yu Y, Guo S, Ho S, Vilà C, Hasegawa M, Liu J (2010) Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal. *Journal of Biogeography* **37**, 2332-2344.

Ward R (1995) Population genetics of tunas. *Journal of Fish Biology* **47 (Supplement A)**, 259-280.

Ward R (2000) Genetics in fisheries management. *Hydrobiologia* **420**, 191-201.

Ward R, Elliot N, Grewe P, Smolenski A (1994a) Allozyme and mitochondrial DNA variation in yellowfin tuna (*Thunnus albacares*) from the Pacific Ocean. *Marine Biology* **118**, 531-539.

Ward R, Skibinski D, Woodwark M (1992) Protein heterozygosity, protein structure and taxonomic differentiation. *Evolutionary Biology* **26**, 73-159.

Ward R, Woodwark M, Skibinski D (1994b) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology* **44**, 213-232.

Wilson EO (1988) *Biodiversity* The National Academies Press, Washington D.C.

Wright S (1965) The interpretation of population structure by F-Statistics with special regard to systems of mating. *Evolution* **19**, 395-420.

Wroblewski J, Bailey W, JHowse K (1994) Observations of adult northern cod (*Gadus morhua*) overwintering in nearshore waters of Trinity Bay, Newfoundland. *Canadian Journal of Fisheries and Aquatic Sciences* **51**, 142-150.

CHAPTER FIVE - SUMMARY

Comparison of multiple complete mitochondrial genomes in two genera of At-Risk marine fish reveals contrasting patterns that reflect their distinctive life histories and responses to Pleistocene glaciations. Extensive polymorphism and trans-Atlantic differences were seen in Atlantic cod, with high diversity in Baltic and Barents Sea populations, increased evidence for the absence of a mitogenomic break in the Northwest Atlantic, an unusual pattern in the landlocked Arctic population, and the discovery of a new basal clade in northern Labrador. Previously unknown diversity was uncovered within three sympatric species of wolffish, with similar patterns in the sister-species Atlantic and spotted wolffish, and a somewhat different pattern in the more divergent northern wolffish.

5.1 Pleistocene Glacial Refugia

The genetic patterns seen in the complete mitochondrial genome support the existence of multiple glacial refugia by Atlantic, spotted, and northern wolffish, and Atlantic cod. Atlantic cod exhibited significant trans-Atlantic differences and a lack of European samples in clades F and I, suggesting two or more isolated refugia. Atlantic and spotted wolffish showed very similar patterns to each other, with three groups separating in the late Pleistocene (ca. 145 - 153 kya and 63 - 81 kya, respectively). Northern wolffish showed a slightly different pattern, with two distinct groups separating ca. 126 kya. In all cases the groups were geographically unstructured. This suggests the occupation of

multiple glacial refugia through the LGM, or historical separation as separate refugia within a single refugium (Gómez & Lunt 2007; Shafer *et al.* 2010), followed by secondary admixture upon recolonisation. This could represent multiple proximal refugial populations meeting as the ice sheets melted (in Atlantic and spotted wolffish), or by more distant locations (as suggested in northern wolffish). A number of other northern pelagic fish species show evidence of multiple distinct lineages: Atlantic herring exhibit strong trans-Atlantic clines in microsatellite studies (McPherson *et al.* 2004), whereas capelin *Mallotus villosus* form four groups, two in the North Atlantic Ocean, with mitochondrial DNA (Dodson *et al.* 2007). These species likely persisted in two or more glacial refugia, often with extensive admixture following recolonisation. Other species show population homogeneity, suggestive of a large single refugium (Bremer *et al.* 1996; Vis *et al.* 1997; Côté *et al.* 2013).

Diversity levels in both Atlantic wolffish and Atlantic cod suggest the presence of at least one refugium in the Northeast Atlantic. Diversity is consistently higher in the European populations, and in Atlantic cod the Baltic and Barents Sea samples were found along long branches and in basal clades. Atlantic cod likely persisted on both sides of the Atlantic. Regions of the North Atlantic that have been identified as potentially suitable cod habitat at the LGM by palaeoecological models include the Scotian Shelf, Flemish Cap, and the coasts of Greenland, Iceland, northwestern Europe, and southeastern Europe (Bigg *et al.* 2008). Atlantic and spotted wolffish may have survived in multiple glacial refugia within a single area, likely the Northeast Atlantic, whereas northern wolffish may have persisted on either side of the Atlantic. The precise locations of the refugia are

uncertain, although additional samples from the North, Baltic, and Barents Seas may help to clarify this.

5.2 Population Genetic Structure

Atlantic cod, which separated from their closest sister species (*G. chalcogrammus*) ca. 3.8 mya (Coulson *et al.* 2006), show a relatively deep phylogeographic pattern: eight major clades were found across the majority of populations, including a basal clade H identified here for the first time; there was isolation of the Arctic population; and allele frequency differences were evident between Northwest and Northeast Atlantic populations. There was no evidence to support the current NAFO stock designations. Although the two most abundant and recently-derived clades in the North Atlantic (F and I) have not been found in the Baltic or Barents Seas, diversity was higher in the Northeast Atlantic, as well as in northern Labrador (LBH) due to the presence of clade H. The landlocked Arctic population at Lake Qasigialiminiq on Baffin Island shows a pattern not previously seen in any marine fish population. It comprises three independent clades, two of which are subclades unique to the lake and represented by multiple individuals with identical or near-identical genome sequences. The third clade is represented by a single individual, most closely related to another geographically-distant individual. This supports a hypothesis of multiple invasions followed by a long period of isolation at a small effective population size (N_E).

Sympatric wolffish species (*Anarhichas* spp.), which separated only 1 - 2 mya (Johnstone *et al.* 2007; McCusker & Bentzen 2010b), show correspondingly little within-

species genetic diversity. The diversity present is geographically unstructured, congruent with the lack of population structure seen in many pelagic marine fish species (Ward *et al.* 1994). Sister-species Atlantic and spotted wolffish share biogeographic histories and have similar ecological traits, whereas the northern wolffish are more distinct. The patterns seen in the mitochondrial genomes reflect this relationship – Atlantic and spotted wolffish both comprise three relatively closely-related haplogroups, including multiple individuals from different locations with identical genome sequences in Atlantic wolffish, whereas northern wolffish comprise two much more divergent haplogroups with no shared individual sequences.

In Atlantic cod the basal clades (H, B, D, K, G, and E) were highly polymorphic, with as many as 64 pairwise differences among individuals within a clade, and as many as 89 between the two most distantly related fish (LBH08 and GIL01). These clades had long branches and many individuals that were not closely related. In the two most recent clades, F and I, there were star phylogenies radiating from a basal haplotype. In Atlantic wolffish, a star phylogeny was seen based on clade C, with two smaller divergent groups connected to this central starburst. The greatest within-clade difference was 30 variable sites (in group A), and the largest overall pairwise difference was 38 changes.

The differences are likely due to the difference in the relative age of the species' lineages. Atlantic cod separated from its closest sister species Alaska pollock ca. 3.8 mya (Coulson *et al.* 2006), whereas Atlantic and spotted wolffish separated only ca. 1 - 2 mya (Johnstone *et al.* 2007; McCusker & Bentzen 2010b), which allows the former more time for mutations to accumulate, and/or they did not undergo as severe a bottleneck during the last glaciation and thus retained some of the variation.

5.3 Mitogenomics

Complete mitochondrial genomes provide much greater resolution than is possible in single-locus studies. Atlantic cod showed unique DNA sequences in the marine populations, but had only seven haplotypes among 18 samples in the landlocked fjord population on Baffin Island. Atlantic wolffish comprised 74 DNA sequences in 84 sampled fish, while spotted and northern wolffish each had a unique sequence (n = 17 and n = 19, respectively). This was considerably more variation than seen in single-locus studies, where 39 variable sites, resulting in four high-frequency haplotypes, were found in a 250 bp fragment of cytochrome *b* across 1,278 individuals (Árnason 2004), or in the multi-locus study in Atlantic, spotted, and northern wolffish that found 31, 18, and 11 variable sites in an 1,830 bp concatenated fragment of the control region and ND1 (McCusker & Bentzen 2010b). This increase in variability provides greater detail into the evolutionary history of the species, and how they survived and recolonised following the most recent glaciations.

Mitochondrial and nuclear genetic markers may sometimes show contrasting population genetic patterns. It must always be remembered that these markers are maternal and biparental, respectively. Microsatellite studies have shown evidence of distinctive local populations on scales where mtDNA studies have found panmixia, while mtDNA have shown pan-Atlantic patterns undetected by microsatellites. This may be due to contemporary isolation of the populations or natural selection. Microsatellites are hypervariable markers, mutating by means of replication slippage, and can therefore show recent differences more quickly than more slowly evolving markers (Jarne & Lagoda

1996; Schlötterer 2000). Microsatellites may also hitchhike on nearby loci under selection (Nielsen *et al.* 2006). Early studies of haemoglobin variation in Atlantic cod showed differences among regions, which led to suggestions that stocks were isolated (Sick 1965b, a; Sick 1965c). However, this has since been shown to be due to temperature differences among the regions tested (Mork & Sundnes 1985). Differences found with vertebral averages (Pepin & Carr 1993) and the pantophysin I locus (Pogson & Mesa 2004) have also proven to be temperature-related, and more recent SNP studies have identified a number of markers that show differences along a temperature gradient (Bradbury *et al.* 2010). That microsatellite loci do not consistently support differences, depending on which loci are used, and that not all loci were responsible for the structure seen, indicates that the segregation of North and South within Northern Cod may be a result of a selection gradient for temperature, salinity, and/or depth, rather than true population segregation (Nielsen *et al.* 2006). This is supported by recent genome scan surveys that have identified temperature gradients along latitudes in both the Northeast and Northwest Atlantic (Bradbury *et al.* 2010; Árnason & Halldórsdóttir 2015).

Comparison of mtDNA genomics and microsatellite data in wolffish are more closely in agreement, with three distinct groups found and higher diversity in the Northeast Atlantic. The Atlantic wolffish microsatellite data suggest isolation-by-distance, differentiation between Atlantic Canada, the northern European populations (North Sea, Barents Sea, and Spitsbergen), and the Rockall Bank population (off the west coast of the United Kingdom; McCusker & Bentzen 2010a). RK was isolated from all other populations, likely due to the physical separation of this population from the coast of Ireland by a deepwater channel. Two loci showed significant linkage disequilibrium,

and when either was removed, the separation of Atlantic Canada, Europe, and Rockall Bank became less pronounced, suggesting that the observed differences may be a combination of contemporary isolation and a linkage artefact. In spotted wolffish, only the inshore West Greenland population showed significant population differences (with Atlantic Canada and Iceland-B), and in northern wolffish the Barents Sea population was distinctive, patterns seen as well with the mitochondrial DNA (McCusker & Bentzen 2011).

5.4 Future Directions

The present study includes extensive sampling of most of the commercial range of Atlantic cod in the Northwest Atlantic. Outstanding questions are the number and locations of putative glacial refugia, on either side of the Atlantic Ocean. Additional samples from the extreme south would allow testing for the presence of a southern refugium. The Northeast Atlantic and adjacent waters are under-sampled: fish from the mid-Atlantic (Iceland and Greenland), would show if a secondary contact zone exists, and from more easterly populations (e.g., the North Sea, Faroe Islands, and White Sea) would show if there were one or more eastern refugia and their possible locations. Samples from the three landlocked arctic lake populations would determine if lineage sampling has been the same in each lake, and whether the three lakes have evolved differently over the past 6,000 - 8,000 years.

Increased sample sizes of Atlantic wolffish in both North American and European waters, as well as additional sampling locations in the Northeast Atlantic, will help to

clarify where the refugia may have been, and where the colonisation likely originated. Larger sample sizes and more sampling locations are required in both spotted and northern wolffish to define more precisely their population genetic structure. A comprehensive multispecies approach, comparing the three wolffish species across similar locations with similar numbers, would allow a more thorough comparison as to whether the Atlantic and spotted wolffish persisted in the same areas and recolonised in a similar manner, and whether the northern wolffish did survive on both sides of the Atlantic Ocean. The incorporation of nuclear markers – whether microsatellites, SNPs, or nuclear coding genes – may help to increase resolution and to discriminate between contemporary and historical patterns. A large number of markers should be tested on the same samples analysed with mitogenomics, allowing any suspect of selection to be excluded. This would also allow a direct comparison of nuclear and mitochondrial markers if the same samples and locations are used.

An interesting question in Atlantic cod is whether the population declines of the mid-20th century affected the genetic diversity of the populations, and by how much. Severe population declines often lead to bottlenecks and loss of variability, and can lead to a decreased ability to deal with adverse conditions (Frankham 1995). A temporal comparison between pre-1960s samples (from museums, where available), 1990s samples (our data), and contemporary samples may allow a view of what changes have occurred, and whether there is evidence of loss of diversity.

5.5 General Conclusions

The mitogenomic analysis of At-Risk marine fish species has revealed previously unknown and contrasting patterns of diversity in wolffish and Atlantic cod. Cod and wolffish both show high levels of polymorphism in the mitochondrial genome, with low levels of heterogeneity among populations. Atlantic cod showed much older lineages, with trans-Atlantic differences, suggesting refugia on both sides of the Atlantic, and isolation of a landlocked Arctic population. Atlantic and spotted wolffish have a very different pattern of variation, indicating isolation in multiple glacial refugia – likely in Europe – followed by mixing and recolonisation of the Northwest Atlantic. Northern wolffish showed a similar pattern, although with much greater divergence between two groups, suggesting refugia on both sides of the Atlantic. Both Atlantic cod and Atlantic wolffish show shared genome sequences, but with very different origins. Those in cod are found only in the isolated Arctic fjord population, likely a result of very recent founder effect, while those in wolffish are found across multiple populations, likely the result of a relatively older bottleneck that has since dispersed. The different patterns are consistent with the diverse traits seen in these two taxa: Atlantic cod are highly vagile, suggesting greater gene flow among neighbouring populations, and are an older species. This is seen in the lack of genetic structure in the Northwest Atlantic, and the long branches and high diversity in Europe. Wolffish are more sedentary and constitute a younger species group. This has allowed greater genetic structure to develop through the Pleistocene, despite lower levels of polymorphism and a starburst pattern of variation.

References

Árnason E (2004) Mitochondrial cytochrome b DNA variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. *Genetics* **166**, 1871-1885.

Árnason E, Halldórsdóttir K (2015) Nucleotide variation and balancing selection at the *Ckma* gene in Atlantic cod: analysis with multiple merger coalescent models. *PeerJ* **3**, e786.

Bigg GR, Cunningham CW, Ottersen G, Pogson GH, Wadley MR, Williamson P (2008) Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. *Proceedings of the Royal Society B: Biological Sciences* **275**, 163-173.

Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Snelgrove PVR, Morris CJ, Gregory RS, Hardie DC, Hutchings JA, Ruzzante DE, Taggart CT, Bentzen P (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences* **277**, 3725-3734.

Bremer JRA, Mejuto J, Greig TW, Ely B (1996) Global population structure of the swordfish (*Xiphias gladius* L.) as revealed by analysis of the mitochondrial DNA control region. *Journal of Experimental Marine Biology and Ecology* **197**, 295-310.

Côté CL, Gagnaire P-A, Bourret V, Verreault G, Castonguay M, Bernatchez L (2013) Population genetics of the American eel (*Anguilla rostrata*): $F_{ST} = 0$ and North Atlantic Oscillation effects on demographic fluctuations of a panmictic species. *Molecular Ecology* **22**, 1763-1776.

Coulson MW, Marshall HD, Pepin P, Carr SM (2006) Mitochondrial genomics of gadine fishes: implications for taxonomy and biogeographic origins from whole-genome data sets. *Genome* **49**, 1115-1130.

Dodson J, Tremblay S, Colombani F, Carscadden J, LeComte F (2007) Trans-Arctic dispersals and the evolution of a circumpolar marine fish species complex, the capelin (*Mallotus villosus*). *Molecular Ecology* **16**, 5030-5043.

Frankham R (1995) Conservation genetics. *Annual Review of Genetics* **29**, 305-327.

Gómez A, Lunt D (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: *Phylogeography in southern European refugia: evolutionary perspectives on the origins and conservation of European biodiversity* (eds. Weiss S, Ferrand N). Kluwer Academic Publishers, Dordrecht, The Netherlands.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* **11**, 424-429.

Johnstone KA, Marshall HD, Carr SM (2007) Biodiversity genomics for species at risk: patterns of DNA sequence variation within and among complete mitochondrial genomes of three species of wolffish (*Anarhichas* spp.). *Canadian Journal of Zoology* **85**, 151-158.

McCusker MR, Bentzen P (2010a) Historical influences dominate the population genetic structure of a sedentary marine fish, Atlantic wolffish (*Anarhichas lupus*), across the North Atlantic Ocean. *Molecular Ecology* **19**, 4228-4241.

McCusker MR, Bentzen P (2010b) Phylogeography of 3 North Atlantic wolffish species (*Anarhichas* spp.) with phylogenetic relationships within the family Anarhichadidae. *Journal of Heredity* **101**, 591-601.

McCusker MR, Bentzen P (2011) Limited population structure in northern and spotted wolffishes (*Anarhichas denticulatus* and *A. minor*) despite low apparent dispersal potential. *Marine Biology* **158**, 1869-1878.

McPherson A, O'Reilly P, Taggart C (2004) Genetic differentiation, temporal stability, and the absence of isolation by distance among Atlantic herring populations. *Transactions of the American Fisheries Society* **133**, 434-446.

Mork J, Sundnes G (1985) Haemoglobin polymorphism in Atlantic cod (*Gadus morhua*): allele frequency variation between yearclasses in a Norwegian fjord stock. *Helgolander Meeresuntersuchungen* **39**, 55-62.

Nielsen E, Hansen M, Meldrup D (2006) Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua* L.): implications for inferring population structure in non-model organisms. *Molecular Ecology* **15**, 3219 - 3229.

Pepin P, Carr SM (1993) Morphological, meristic, and genetic analysis of stock structure in juvenile Atlantic cod (*Gadus morhua*) from the Newfoundland shelf. *Canadian Journal of Fisheries and Aquatic Sciences* **50**, 1924-1933.

Pogson G, Mesa K (2004) Positive Darwinian selection at the pantophysin (*Pan* I) locus in marine gadoid fishes. *Molecular Biology and Evolution* **21**, 65-75.

Schlötterer C (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365-371.

Shafer ABA, Cullingham CI, Côté SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology* **19**, 4589-4621.

Sick K (1965a) Haemoglobin polymorphism of cod in the Baltic and the Danish Belt Sea. *Hereditas* **54**, 19-48.

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

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

Vis M, Carr S, Bowering W, Davidson W (1997) Greenland halibut (*Reinhardtius hippoglossoides*) in the North Atlantic are genetically homogeneous. *Canadian Journal of Fisheries and Aquatic Sciences* **54**, 1813-1821.

Ward R, Woodwark M, Skibinski D (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology* **44**, 213-232.

Appendix 1. Sample ID, haplogroup, NAFO division, GenBank ID, sequencing method used, and location information for Atlantic wolffish samples. Each sample was given a project ID with location and number. The 'haplogroup' is the group identified by BAPS v6, as shown in Table 2.3 and Figure 2.1. The GenBank accession numbers are given; for the two-fragment study each sample has two GenBank IDs - the first for the COX1 fragment and the second for the ND5 fragment. The sequencing methods are multispecies microarray (CHIP) or standard dideoxy (DD). Where unavailable, GPS coordinates were approximated from location averages and are shown in grey.

Project ID	Haplogroup	NAFO division	Tag number	GenBank ID	Method	Latitude (°N)	Longitude (°W)
Atlantic wolffish - mitogenomes							
LABC001	A	2H	L7176	KX117921	DD	55.987	59.518
LABS001	B	2J	10457	KX117922	CHIP	53.043	52.430
LABS002	C	2J	10466	KX117923	CHIP	54.407	53.588
LABS003	C	2J	10471	KX117924	CHIP	53.400	53.270
LABS004	C	2J	318	KX117925	CHIP	54.402	53.870
LABS005	B	2J	3809	KX117926	DD	54.402	53.870
LABS006	C	2J	464	KX117927	CHIP	54.402	53.870
LABS007	C	2J	482	KX117928	CHIP	54.402	53.870
LABS008	C	2J	E1384	KX117929	CHIP	52.987	52.505
LABS009	B	2J	Q8565	KX117930	CHIP	52.987	52.505
NENL001	C	3K	4695	KX117931	CHIP	51.542	51.233

NENL002	C	3K	819	KX117932	CHIP	51.320	50.455
NENL003	C	3K	E7117	KX117933	CHIP	49.612	50.880
NENL004	C	3K	E7676	KX117934	CHIP	50.267	51.043
NENL005	C	3K	F00962	KX117935	CHIP	51.468	51.152
NENL006	C	3K	F32021	KX117936	CHIP	50.073	51.355
NENL007	B	3K	F35896	KX117937	CHIP	49.583	51.782
NENL008	A	3K	F45870	KX117938	CHIP	51.810	50.825
NENL009	C	3K	Q1419	KX117939	CHIP	50.010	50.652
NGB001	B	3L	135	KX117940	CHIP	48.485	49.712
NGB002	A	3L	138	KX117941	DD	48.517	49.838
NGB004	C	3L	A26705	KX117943	CHIP	49.028	51.813
NGB005	C	3L	C1143	KX117944	CHIP	48.667	49.830
NGB006	A	3L	C3752	KX117945	CHIP	48.325	49.293
NGB007	A	3L	H2229	KX117946	CHIP	48.667	49.830
NGB008	A	3L	J4141	KX117947	CHIP	46.410	47.330
NGB009	B	3L	L7641	KX117948	CHIP	48.667	49.830
NGB010	A	3L	Q1420	KX117949	CHIP	47.613	47.378
NGB011	A	3L	DAL-NL61	KX117950	DD	47.936	49.057
NGB012	A	3L	DAL-NL76	KX117951	DD	47.936	49.057
NGB013	A	3L	DAL-NL81	KX117952	DD	47.936	49.057
SEGB001	C	3N	2588	KX117953	CHIP	42.880	49.775

SEGB002	A	3N	2600	KX117954	CHIP	42.880	49.775
SEGB003	A	3N	2961	KX117955	CHIP	42.885	49.945
SEGB004	C	3N	488	KX117956	CHIP	43.682	49.147
SEGB005	C	3N	C12115	KX117957	CHIP	44.267	50.985
SEGB006	C	3N	F32798	KX117958	CHIP	42.837	50.397
SEGB007	A	3N	F35198	KX117959	CHIP	44.648	49.852
SEGB008	B	3N	J5577	KX117960	CHIP	42.995	49.718
SEGB009	C	3N	L6569	KX117961	CHIP	43.568	49.862
SEGB010	B	3N	L6589	KX117962	CHIP	43.350	49.582
SEGB011	C	3N	M8043	KX117963	CHIP	43.808	49.072
SEGB012	A	3N	M9824	KX117964	CHIP	44.758	50.083
SEGB013	C	3N	R6995	KX117965	CHIP	42.885	49.945
SWGB001	B	3O	11756	KX117966	CHIP	44.822	52.865
SWGB002	C	3O	4408	KX117967	CHIP	45.017	52.405
SWGB003	A	3O	4413	KX117968	CHIP	44.810	53.018
SWGB004	C	3O	7508	KX117969	CHIP	44.075	51.187
SWGB005	A	3O	8695	KX117970	CHIP	44.535	52.860
SWGB006	A	3O	92672	KX117971	CHIP	44.822	52.462
SWGB007	B	3O	F03116	KX117972	CHIP	44.822	52.865
SWGB008	C	3O	M3060	KX117973	CHIP	43.357	51.643
SWGB009	A	3O	Q1937	KX117974	CHIP	45.182	52.143

SWGB010	C	3O	Q1946	KX117975	CHIP	45.217	54.342
SNL001	A	3P	1128	KX117976	DD	47.335	56.847
SNL002	A	3P	3878	KX117977	CHIP	47.170	57.342
SNL003	A	3P	4446	KX117978	DD	47.133	57.833
SNL004	A	3P	4476	KX117979	DD	45.167	54.535
SNL005	A	3P	4479	KX117980	DD	45.398	55.345
SNL006	C	3P	E06912	KX117981	DD	46.903	57.853
SNL007	A	3P	Q8302	KX117982	DD	47.117	57.585
SNL008	C	3P	R7430	KX117983	DD	45.560	55.477
SNL009	A	3P	T9750	KX117984	DD	47.335	56.847
SNL010	B	3P	F32448	KX117985	CHIP	47.297	58.332
BS001	C	Norway	DAL-BS40	KX117986	DD	N/A	N/A
BS002	A	Norway	DAL-BS42	KX117987	DD	N/A	N/A
EG001	C	E. Greenland	DAL-EG20	KX117988	DD	N/A	N/A
EG002	C	E. Greenland	DAL-EG21	KX117989	DD	N/A	N/A
EG003	B	E. Greenland	DAL-EG22	KX117990	DD	N/A	N/A
IC001	C	Iceland	DAL-IC02	KX117991	DD	N/A	N/A
IC002	C	Iceland	DAL-IC04	KX117992	DD	N/A	N/A
IC003	C	Iceland	DAL-IC05	KX117993	DD	N/A	N/A
NS001	A	North Sea	DAL-NS01	KX117994	DD	N/A	N/A
NS002	B	North Sea	DAL-NS02	KX117995	DD	N/A	N/A

NS003	C	North Sea	DAL-NS03	KX117996	DD	N/A	N/A
RK001	C	Rockall Bank	DAL-RK01	KX117997	DD	N/A	N/A
RK002	C	Rockall Bank	DAL-RK02	KX117998	DD	N/A	N/A
RK003	C	Rockall Bank	DAL-RK04	KX117999	DD	N/A	N/A
SS001	A	4W	DAL-SS01	KX118000	DD	N/A	N/A
SS002	C	4W	DAL-SS04	KX118001	DD	N/A	N/A
SS003	A	4W	DAL-SS05	KX118002	DD	N/A	N/A
WG001	B	W. Greenland	DAL-WG01	KX118003	DD	N/A	N/A
WG002	B	W. Greenland	DAL-WG07	KX118004	DD	N/A	N/A
WG003	C	W. Greenland	DAL-WG08	KX118005	DD	N/A	N/A
Atlantic wolffish - two-fragment study							
LABC002	A	2H	10107	KX118112 KX118192	DD	55.987	59.518
LABS010	C	2J	1183	KX118113 KX118193	DD	54.870	54.452
LABS011	C	2J	1445	KX118114 KX118194	DD	53.920	52.963
LABS012	A	2J	1638	KX118115 KX118195	DD	54.870	54.452
LABS013	A	2J	1690	KX118116	DD	54.870	54.452

				KX118196			
LABS014	C	2J	1924	KX118117 KX118197	DD	52.622	54.988
LABS015	B	2J	2006	KX118118 KX118198	DD	54.062	53.367
LABS016	C	2J	2659	KX118119 KX118199	DD	53.952	53.462
LABS017	B	2J	2730	KX118120 KX118200	DD	54.870	54.452
LABS018	A	2J	341	KX118121 KX118201	DD	54.039	53.618
LABS019	C	2J	3745	KX118122 KX118202	DD	53.013	52.178
LABS020	C	2J	410	KX118123 KX118203	DD	54.039	53.618
LABS021	B	2J	4224	KX118124 KX118204	DD	53.952	53.462
LABS022	C	2J	442	KX118125 KX118205	DD	54.870	54.452
LABS023	C	2J	5462	KX118126 KX118206	DD	53.920	52.963
LABS024	A	2J	6589	KX118127 KX118207	DD	52.987	52.505

LABS025	B	2J	726	KX118128 KX118208	DD	54.870	54.452
LABS026	B	2J	774	KX118129 KX118209	DD	54.870	54.452
NENL010	A	3K	1098	KX118130 KX118210	DD	51.810	50.825
NENL011	A	3K	1248	KX118131 KX118211	DD	51.688	50.868
NENL012	A	3K	1516	KX118132 KX118212	DD	51.320	50.455
NENL013	C	3K	1529	KX118133 KX118213	DD	50.422	51.180
NENL014	C	3K	1748	KX118134 KX118214	DD	51.320	50.455
NENL015	B	3K	1857	KX118135 KX118215	DD	50.073	51.355
NENL016	A	3K	2693	KX118136 KX118216	DD	51.540	51.327
NENL017	B	3K	3030	KX118137 KX118217	DD	51.320	50.455
NENL018	A	3K	4474	KX118138 KX118218	DD	51.485	51.027

NENL019	C	3K	848	KX118139 KX118219	DD	51.320	50.455
NENL020	C	3K	898	KX118140 KX118220	DD	51.975	51.032
NENL021	C	3K	A28141	KX118141 KX118221	DD	51.542	51.233
NENL022	C	3K	A29838	KX118142 KX118222	DD	51.810	50.825
NENL023	C	3K	A29897	KX118143 KX118223	DD	51.320	50.455
NENL024	C	3K	A31109	KX118144 KX118224	DD	51.810	50.825
NENL025	A	3K	A39849	KX118145 KX118225	DD	51.102	50.923
NGB014	A	3L	122	KX118146 KX118226	DD	48.113	48.925
NGB015	B	3L	139	KX118147 KX118227	DD	48.517	49.838
NGB016	C	3L	140	KX118148 KX118228	DD	48.517	49.838
NGB017	A	3L	141	KX118149 KX118229	DD	48.517	49.838

NGB018	B	3L	145	KX118150 KX118230	DD	48.475	49.712
NGB019	C	3L	150	KX118151 KX118231	DD	48.475	49.712
NGB020	C	3L	2856	KX118152 KX118232	DD	46.483	47.305
NGB021	A	3L	2861	KX118153 KX118233	DD	46.483	47.305
NGB022	B	3L	2870	KX118154 KX118234	DD	46.483	47.305
NGB023	A	3L	B03261	KX118155 KX118235	DD	47.936	49.057
NGB024	A	3L	B03265	KX118156 KX118236	DD	47.936	49.057
SEGB014	C	3N	1514	KX118157 KX118237	DD	43.682	49.147
SEGB015	C	3N	1531	KX118158 KX118238	DD	42.832	50.085
SEGB016	C	3N	1551	KX118159 KX118239	DD	43.783	49.040
SEGB017	C	3N	1611	KX118160 KX118240	DD	43.682	49.147

SEGB018	A	3N	2073	KX118161 KX118241	DD	42.995	49.718
SEGB019	C	3N	2222	KX118162 KX118242	DD	44.377	50.407
SEGB020	A	3N	2283	KX118163 KX118243	DD	43.682	49.147
SEGB021	A	3N	2502	KX118164 KX118244	DD	44.147	50.562
SEGB022	A	3N	2505	KX118165 KX118245	DD	43.740	50.410
SEGB023	A	3N	2506	KX118166 KX118246	DD	43.740	50.410
SEGB024	A	3N	2521	KX118167 KX118247	DD	44.193	50.577
SEGB025	B	3N	2542	KX118168 KX118248	DD	42.880	49.775
SWGB011	A	3O	11762	KX118169 KX118249	DD	44.540	52.370
SWGB012	A	3O	11765	KX118170 KX118250	DD	44.905	52.695
SWGB013	C	3O	11767	KX118171 KX118251	DD	43.152	51.212

SWGB014	A	3O	11772	KX118172 KX118252	DD	44.595	52.663
SWGB015	A	3O	11778	KX118173 KX118253	DD	43.858	52.527
SWGB016	C	3O	2381	KX118174 KX118254	DD	44.647	51.587
SWGB017	C	3O	3723	KX118175 KX118255	DD	45.877	53.133
SWGB018	C	3O	4407	KX118176 KX118256	DD	44.810	53.018
SWGB019	C	3O	4409	KX118177 KX118257	DD	45.017	52.405
SWGB020	A	3O	4410	KX118178 KX118258	DD	45.017	52.405
SWGB021	A	3O	4412	KX118179 KX118259	DD	44.810	53.018
SWGB022	A	3O	4415	KX118180 KX118260	DD	44.810	53.018
SWGB023	A	3O	4493	KX118181 KX118261	DD	44.440	53.365
SWGB024	A	3O	4498	KX118182	DD	43.363	51.658

				KX118262			
SWGB025	C	3O	5546	KX118183 KX118263	DD	44.075	51.187
SWGB026	A	3O	6335	KX118184 KX118264	DD	44.075	51.187
SWGB027	C	3O	9560	KX118185 KX118265	DD	43.980	51.285
SNL011	A	3P	4486	KX118187 KX118267	DD	45.470	55.200
SNL012	C	3P	80995	KX118188 KX118268	DD	46.903	57.853
SNL013	B	3P	E04684	KX118189 KX118269	DD	47.538	56.123
SNL014	C	3P	E1440	KX118190 KX118270	DD	47.133	57.833
SNL015	A	3P	M0918	KX118191 KX118271	DD	47.335	56.847
SNL016	A	3P	12260	KX118186 KX118266	DD	47.195	56.235

Appendix 2. Sample ID, haplogroup, NAFO division, GenBank ID, sequencing method used, and location information for spotted and northern wolffish samples. Each sample was given a project ID with location and number. The 'haplogroup' is the group identified by BAPS v6, as shown in Figure 3.1. Where unavailable, GPS coordinates were approximated from location averages and are shown in grey.

Project ID	Haplogroup	NAFO division	Tag number	GenBank ID	Method	Latitude (°N)	Longitude (°W)
Spotted wolffish - mitogenomes							
LAB001	D	2J	10470	KX118006	DD	53.043	52.430
NENL001	E	3K	1532	KX118007	DD	51.688	50.868
NENL002	E	3K	5413	KX118008	DD	50.270	53.698
NENL003	F	3K	806	KX118009	DD	51.468	51.152
NENL004	E	3K	874	KX118010	DD	51.485	51.027
NENL005	E	3K	4361	KX118011	DD	50.472	52.755
NGB001	E	3L	10811	KX118012	DD	47.207	47.512
NGB002	D	3L	14622	KX118013	DD	48.872	49.948
NGB003	F	3L	1533	KX118014	DD	49.040	50.815
NGB004	D	3L	4579	KX118015	DD	48.897	49.972
SEGB001	E	3N	10038	KX118016	DD	42.873	50.397
SEGB002	D	3N	10306	KX118017	DD	43.037	49.652
SEGB003	F	3N	F32761	KX118018	DD	43.648	49.047

SEGB004	D	3N	F33212	KX118019	DD	44.912	49.042
SEGB005	D	3N	L7903	KX118020	DD	44.912	49.042
SEGB006	E	3N	M3921	KX118021	DD	44.912	49.042
SWGB001	D	3O	36579	KX118022	DD	43.225	51.308
Northern wolffish - mitogenomes							
BS002	G	Norway	DAL-BS02	KX118023	DD	N/A	N/A
FC001	H	3M	7777	KX118024	DD	48.508	45.510
FC002	H	3M	866	KX118025	DD	46.332	46.615
LAB001	H	2J	5878	KX118026	DD	52.285	52.763
LAB002	G	2J	6706	KX118027	DD	52.860	51.867
LAB003	H	2J	F16889	KX118028	DD	52.778	51.503
LAB004	G	2J	J5081	KX118029	DD	52.278	50.942
LAB005	H	2J	J5060	KX118030	DD	53.973	52.925
NENL001	G	3K	10522	KX118031	DD	51.153	50.300
SEGB001	H	3N	1186	KX118032	DD	42.885	49.945
SEGB002	H	3N	A35322	KX118033	DD	44.373	49.018
SEGB003	H	3N	E1325	KX118034	DD	42.995	49.718
SEGB004	H	3N	L6544	KX118035	DD	42.998	49.983
SEGB005	H	3N	L6566	KX118036	DD	43.138	49.610
SEGB006	H	3N	L7917	KX118037	DD	45.738	47.950

SWGB001	H	3O	4300	KX118038	DD	44.773	54.440
SWGB002	H	3O	93180	KX118039	DD	43.453	51.972
SWGB003	H	3O	L6679	KX118040	DD	43.453	51.972
SWGB004	G	3O	Q1992	KX118041	DD	45.222	54.127

Appendix 3. Sample ID, haplogroup, NAFO division, GenBank ID, sequencing method used, and location information, listed by population. The 'clade' for Atlantic cod is the major clade identified by Bayesian analysis, as shown in Figure 4.5, 'n/a' signifies that the sample did not fall into one of the eight major clades, with the nearest clade in parentheses. The sequencing methods are multispecies microarray (CHIP), standard dideoxy (DD), or standard dideoxy in the Carr & Marshall (2008) paper with added control region (CM). Where unavailable, GPS coordinates were approximated from location description.

Project ID	Clade	NAFO division	Tag Number	GenBank ID	Method	Latitude (°N)	Longitude (°E / °W)
Norwegian coast, Tromsö, Norway							
NC01	B	Norway	BS4	EU877736 KX432219	CM	70.126	17.805 E
NC02	G	Norway	BS5	EU877737 KX432220	CM	70.126	17.805 E
NC03	G	Norway	BS7	EU877738 KX432221	CM/DD	70.126	17.805 E
NC04	n/a (E)	Norway	BS9	EU877739 KX432222	CM/DD	70.126	17.805 E
NC05	n/a (E)	Norway	BS10	EU877740 KX432223	CM/DD	70.126	17.805 E
NC06	E	Norway	BS11	EU877741 KX432224	CM/DD	70.126	17.805 E
NC07	n/a (E)	Norway	BS12	KX266969	DD	70.126	17.805 E

Baltic Sea, Sopot, Poland							
BAL01	B	Baltic Sea	GD07	KX266970	CHIP	54.569	19.011 E
BAL02	E	Baltic Sea	GD08	KX266971	CHIP	54.569	19.011 E
BAL03	n/a (FG)	Baltic Sea	GD09	KX266972	CHIP	54.569	19.011 E
BAL04	E	Baltic Sea	GD10	KX266973	CHIP	54.569	19.011 E
BAL05	E	Baltic Sea	GD15	KX266974	CHIP	54.569	19.011 E
BAL06	n/a (I)	Baltic Sea	GD16	KX266975	CHIP	54.569	19.011 E
BAL07	E	Baltic Sea	GD18	KX266976	CHIP	54.569	19.011 E
BAL08	E	Baltic Sea	GD19	KX266977	CHIP	54.569	19.011 E
BAL09	D	Baltic Sea	GD21	KX266978	CHIP	54.569	19.011 E
BAL10	B	Baltic Sea	GD22	KX266979	CHIP	54.569	19.011 E
Lake Qasigialiminiq, Baffin Island, Nunavut							
QAS01	E	0B	Pang01	KX266980	DD	65.803	68.172 W
QAS02	B	0B	Pang02	KX266981	DD	65.803	68.172 W
QAS03	B	0B	Pang03	KX266982	DD	65.803	68.172 W
QAS04	B	0B	Pang04	KX266983	DD	65.803	68.172 W
QAS05	E	0B	Pang05	KX266984	DD	65.803	68.172 W
QAS06	E	0B	Pang06	KX266985	DD	65.803	68.172 W
QAS07	B	0B	Pang07	KX266986	DD	65.803	68.172 W
QAS08	E	0B	Pang08	KX266987	DD	65.803	68.172 W
QAS09	E	0B	Pang09	KX266988	DD	65.803	68.172 W

QAS10	E	0B	Pang10	KX266989	DD	65.803	68.172 W
QAS11	E	0B	Pang11	KX266990	DD	65.803	68.172 W
QAS12	E	0B	Pang12	KX266991	DD	65.803	68.172 W
QAS13	E	0B	Pang13	KX266992	DD	65.803	68.172 W
QAS14	E	0B	Pang14	KX266993	DD	65.803	68.172 W
QAS15	F	0B	Pang15	KX266994	DD	65.803	68.172 W
QAS16	E	0B	Pang16	KX266995	DD	65.803	68.172 W
QAS17	B	0B	Pang17	KX266996	DD	65.803	68.172 W
QAS18	B	0B	Pang18	KX266997	DD	65.803	68.172 W
Hopedale, Labrador							
LBH01	I	2H	Hope01	KX266998	DD	55.483	59.979 W
LBH02	K	2H	Hope02	KX266999	DD	55.483	59.979 W
LBH03	n/a (E)	2H	Hope03	KX267000	DD	55.483	59.979 W
LBH05	I	2H	Hope05	KX267001	DD	55.483	59.979 W
LBH06	F	2H	Hope06	KX267002	DD	55.483	59.979 W
LBH07	D	2H	Hope07	KX267003	DD	55.483	59.979 W
LBH08	H	2H	Hope08	KX267004	DD	55.483	59.979 W
LBH09	I	2H	Hope09	KX267005	DD	55.483	59.979 W
LBH10	F	2H	Hope10	KX267006	DD	55.483	59.979 W
LBH11	F	2H	Hope11	KX267007	DD	55.483	59.979 W
LBH12	F	2H	Hope12	KX267008	DD	55.483	59.979 W

LBH13	H	2H	Hope13	KX267009	DD	55.483	59.979 W
Postville, Labrador							
LBP01	I	2J	Post01	KX267010	DD	55.043	59.486 W
LBP02	I	2J	Post02	KX267011	DD	55.043	59.486 W
LBP03	E	2J	Post03	KX267012	DD	55.043	59.486 W
LBP04	F	2J	Post04	KX267013	DD	55.043	59.486 W
LBP05	n/a (EFGI)	2J	Post05	KX267014	DD	55.043	59.486 W
LBP06	F	2J	Post06	KX267015	DD	55.043	59.486 W
LBP07	G	2J	Post07	KX267016	DD	55.043	59.486 W
LBP08	I	2J	Post08	KX267017	DD	55.043	59.486 W
LBP09	F	2J	Post09	KX267018	DD	55.043	59.486 W
LBP10	K	2J	Post10	KX267019	DD	55.043	59.486 W
LBP11	F	2J	Post11	KX267020	DD	55.043	59.486 W
LBP12	I	2J	Post12	KX267021	DD	55.043	59.486 W
LBP13	F	2J	Post13	KX267022	DD	55.043	59.486 W
LBP14	I	2J	Post14	KX267023	DD	55.043	59.486 W
Makkovik, Labrador							
LBM01	I	2J	Makk01	KX267024	DD	55.159	58.783 W
LBM02	I	2J	Makk02	KX267025	DD	55.159	58.783 W
LBM03	F	2J	Makk03	KX267026	DD	55.159	58.783 W

LBM04	I	2J	Makk04	KX267027	DD	55.159	58.783 W
LBM05	E	2J	Makk05	KX267028	DD	55.159	58.783 W
LBM06	F	2J	Makk06	KX267029	DD	55.159	58.783 W
LBM07	I	2J	Makk07	KX267030	DD	55.159	58.783 W
LBM08	I	2J	Makk08	KX267031	DD	55.159	58.783 W
LBM09	G	2J	Makk09	KX267032	DD	55.159	58.783 W
LBM10	F	2J	Makk10	KX267033	DD	55.159	58.783 W
LBM11	E	2J	Makk30	KX267034	DD	55.159	58.783 W
LBM12	E	2J	Makk31	KX267035	DD	55.159	58.783 W
LBM13	I	2J	Makk32	KX267036	DD	55.159	58.783 W
LBM14	I	2J	Makk33	KX267037	DD	55.159	58.783 W
Gilbert Bay, Labrador							
GIL001	F	2J	GIL017	KX267042	DD	52.556	55.934 W
GIL002	F	2J	GIL018	KX267043	DD	52.556	55.934 W
Hawkes Channel, Labrador - NE Newfoundland Shelf							
HC01	I	2J	GA243-30-14	KX267040	DD	53.310	52.380 W
HC02	D	2J	GA243-30-16	KX267041	DD	53.310	52.380 W
HC03	F	2J	GA243-29-01	EU877719 KX432225	CM	53.310	52.380 W

HC04	I	2J	GA243-29-03	EU877720 KX432226	CM	53.310	52.380 W
HC05	K	2J	GA243-29-04	EU877721 KX432227	CM	53.310	52.380 W
HC06	G	2J	GA243-29-05	EU877722 KX432228	CM	53.310	52.380 W
HC07	I	2J	GA243-29-06	EU877723 KX432229	CM	53.310	52.380 W
HC08	F	2J	GA243-29-07	EU877724 KX432230	CM	53.310	52.380 W
HC09	G	2J	GA243-29-09	EU877725 KX432231	CM	53.310	52.380 W
HC10	D	2J	GA243-29-10	EU877726 KX432232	CM	53.310	52.380 W
Random Island, Trinity Bay, Newfoundland - inshore Newfoundland							
RAN01	F	3L	SH226-2-2	KX267048	DD	48.200	53.600 W
RAN02	K	3L	SH226-2-5	KX267049	DD	48.200	53.600 W
RAN03	F	3L	SH226-4-23	KX267050	DD	48.200	53.600 W
RAN04	I	3L	SH226-4-35	KX267051	DD	48.200	53.600 W
RAN05	F	3L	SH226-4-6	KX267052	DD	48.200	53.600 W
RAN06	I	3L	SH226-4-7	KX267053	DD	48.200	53.600 W
RAN07	n/a (I)	3L	SH226-4-9	KX267054	DD	48.200	53.600 W

RAN08	I	3L	SH226-4-10	KX267055	DD	48.200	53.600 W
RAN09	n/a (EFGIK)	3L	SH226-4-11	KX267056	DD	48.200	53.600 W
RAN10	B	3L	SH226-4-12	KX267057	DD	48.200	53.600 W
RAN11	F	3L	SH226-4-13	KX267058	DD	48.200	53.600 W
North Cape, Grand Banks, Newfoundland - NE Newfoundland Shelf							
NGB01	F	3L	GA243-15-11	KX267044	CHIP	48.490	49.460 W
NGB02	I	3L	GA243-15-12	KX267045	CHIP	48.490	49.460 W
NGB03	F	3L	GA243-15-13	KX267046	CHIP	48.490	49.460 W
NGB04	I	3L	GA243-15-14	KX267047	CHIP	48.490	49.460 W
NGB05	F	3L	GA243-14-02	EU877710 KX432233	CM	48.490	49.460 W
NGB06	F	3L	GA243-14-03	EU877711 KX432234	CM	48.490	49.460 W
NGB07	I	3L	GA243-14-04	EU877712 KX432235	CM	48.490	49.460 W
NGB08	F	3L	GA243-14-05	EU877713 KX432236	CM	48.490	49.460 W
NGB09	F	3L	GA243-14-06	EU877714 KX432237	CM	48.490	49.460 W
NGB10	F	3L	GA243-14-07	EU877715 KX432238	CM	48.490	49.460 W

NGB11	I	3L	GA243-14-08	EU877716 KX432239	CM	48.490	49.460 W
NGB12	F	3L	GA243-14-09	EU877717 KX432240	CM	48.490	49.460 W
NGB13	D	3L	GA243-14-10	EU877718 KX432241	CM	48.490	49.460 W
Flemish Cap - offshore seamount							
FC01	I	3M	FC2	EU877727 KX432242	CM	47.250	45.750 W
FC02	F	3M	FC3	EU877728 KX432243	CM	47.250	45.750 W
FC03	F	3M	FC6	EU877729 KX432244	CM	47.250	45.750 W
FC04	F	3M	FC7	EU877730 KX432245	CM	47.250	45.750 W
FC05	B	3M	FC9	EU877731 KX432246	CM	47.250	45.750 W
FC06	E	3M	FC12	EU877732 KX432247	CHIP/CM	47.250	45.750 W
FC07	E	3M	FC14	EU877733 KX432248	CHIP/CM	47.250	45.750 W

FC08	F	3M	FC15	EU877734 KX432249	CHIP/CM	47.250	45.750 W
FC09	I	3M	FC21	KX267039	CHIP	47.250	45.750 W
FC10	F	3M	FC16	EU877735 KX432250	CHIP/CM	47.250	45.750 W
FC11	F	3M	FC17	KX267039	CHIP	47.250	45.750 W
Eastern Scotian Shelf - Scotian Shelf							
ESS01	K	4W	N222-62-222-347687	KX267059	CHIP	45.344	59.957 W
ESS02	F	4W	N222-62-231-347695	KX267060	CHIP	45.344	59.957 W
ESS03	G	4W	N222-62-234-347030	KX267061	CHIP	45.344	59.957 W
ESS04	F	4W	N222-62-242-347034	KX267062	CHIP	45.344	59.957 W
ESS05	n/a (EFGIK)	4W	N222-62-243-347035	KX267063	CHIP	45.344	59.957 W
ESS06	I	4W	N222-62-244-347036	KX267064	CHIP	45.344	59.957 W
ESS07	D	4W	N222-62-246-347038	KX267065	CHIP	45.344	59.957 W
ESS08	K	4W	N222-62-248-347040	KX267066	CHIP	45.344	59.957 W

ESS09	I	4W	N222-62-249-347041	KX267067	CHIP	45.344	59.957 W
ESS10	F	4W	N222-62-250-347042	KX267068	CHIP	45.344	59.957 W
ESS11	F	4W	N222-65-267-347057	KX267069	DD	45.344	59.957 W
Fundy-Brown's-Emerald Bank - Scotian Shelf							
FSS01	F	4X	N221-38-119-348324	KX267070	DD	43.736	63.198 W
FSS02	F	4X	N221-38-124-348314	KX267071	DD	43.736	63.198 W
FSS03	E	4X	N221-38-126-348311	KX267072	DD	43.736	63.198 W
FSS04	I	4X	N221-39-163-348439	KX267073	DD	43.736	63.198 W
FSS05	K	4X	N221-39-167-348448	KX267074	DD	43.736	63.198 W
FSS06	F	4X	N221-39-176-348451	KX267075	DD	43.736	63.198 W
FSS07	F	4X	N221-39-182-348431	KX267076	DD	43.736	63.198 W
FSS08	F	4X	N221-39-184-348438	KX267077	DD	43.736	63.198 W

FSS09	F	4X	N221-39-192-348021	KX267078	DD	43.736	63.198 W
FSS10	n/a (I)	4X	N221-41-233-348471	KX267079	DD	43.736	63.198 W
George's Bank							
GEO01	B	5Ze	N216-24-159	KX267080	CHIP	41.594	69.307 W
GEO02	I	5Ze	N216-30-275	KX267081	CHIP	41.594	69.307 W
GEO03	G	5Ze	N216-30-276	KX267082	CHIP	41.594	69.307 W
GEO04	I	5Ze	N216-30-277	KX267083	CHIP	41.594	69.307 W
GEO05	F	5Ze	N216-30-280	KX267084	CHIP	41.594	69.307 W
GEO06	F	5Ze	N216-30-281	KX267085	CHIP	41.594	69.307 W
GEO07	F	5Ze	N216-30-282	KX267086	CHIP	41.594	69.307 W
GEO08	B	5Ze	N216-30-283	KX267087	CHIP	41.594	69.307 W
GEO09	F	5Ze	N216-30-286	KX267088	CHIP	41.594	69.307 W
GEO10	F	5Ze	N216-30-289	KX267089	CHIP	41.594	69.307 W