Development and application of molecular tools for the study of farmed and wild Atlantic cod (*Gadus morhua*) stocks

Salla Riikka Vartia, MSc

PhD. Dissertation

Supervisors:

Dr. Richard FitzGerald, Dr. Jens Carlsson and Dr. Luca Mirimin

May, 2016

Carna Research Station, Ryan Institute

National University of Ireland, Galway

Table of Contents

Declaration

I hereby certify that this material, which I submit for assessment on the programme of study leading to the award of Doctor of Philosophy (Ph.D.) is all my own work and I have not obtained a degree in this University, or elsewhere, on the basis of the research presented in this submitted work.

Salla Vartia

Date

List of Tables

List of Figures

Chapter 5

To the memory of Vijay

Let us remember to always be kind, curious and silly.

Acknowledgements

I would like to thank my supervisors Dr. Richard FitzGerald, Dr. Jens Carlsson and Dr. Luca Mirimin for supervision throughout the project. Richard, thank you for taking me on board, for your encouragement, flexibility and guidance over the years. Jens, many thanks are owed for welcoming me into your lab group, for being contagiously passionate about what you do, keeping challenging me, for your encouragement and support. Luca, thank you for guiding me through the practicalities of the first years of the project; for your continuing advice and for always reading my drafts. In addition, I would like to thank Prof. Tom Cross for welcoming me into his lab in UCC, for providing access to samples and for insightful conversations on the project.

I would like to thank the staff and fellow students at Carna Research Station (NUIG), UCC and UCD. At Carna Research Station, special thanks are due to Kieran O'Halloran, Ken Maher and Luca Mirimin for acquiring eggs for the establishment of the farmed cod stock; Paul Casburn, Bob Ludgate, Damien O'Keeffe and Majbritt Bolton-Warberg for upkeep of fish and help in sampling events; Joe Casey for excellent technical support in all aspects. In UCC, the technical assistance from Mary Cross, Nettan Carlsson, Dr. Eileen Dillane and Alicia Antoniacomi was much appreciated. In UCD, I received excellent technical support from Dr. Carlotta Sacchi. I would like to thank all members of the Area52 lab group. Jens and Nettan, your support and friendship have been truly appreciated.

I would like to thank all of my co-authors that contributed to publications from Chapters 3 and 5. I am especially indebted to Jose Villanueva-Canas for writing the python script used in Chapter 5 and John Finarelli for help with statistics in Chapter 5. Marine Institute and John Brittain provided wild cod samples used in Chapters 4 and 5. Patrick Collins and Ed Farrell provided feedback on thesis chapter drafts. This work was supported by a PhD studentship via the EIRCOD project which was funded under the Sea Change Initiative with support from the Marine Institute and the National Development Plan 2007 to 2013.

I would like to thank my parents for their continued support and teaching me that I could do whatever I put my mind into. I would like to thank all my friends for their friendship and support in many forms over the years, especially Brian and Mari. Finally, I would like to thank Mark for his love and support. I could not have done it without you.

Abstract

Atlantic cod (*Gadus morhua*) is a commercially and ecologically important fish inhabiting the shelf areas of the North Atlantic. Several genetic resources were developed for the study of cod in this thesis. Fifty-five new microsatellite loci in six multiplex panels were developed for characterising the wild Celtic Sea stock. The existing methodology for validation of novel microsatellite markers was improved by combining three-primer-PCR with multiplex reactions. In addition, the potential of microsatellite genotyping-by-sequencing (GBS) was investigated using a combinatorial barcoding method suitable for implementation in large-scale population genetics studies. The results confirmed a high correspondence between GBS and capillary/gel electrophoresis genotypes, but also revealed cases of homoplasy. Genetic structuring of local wild cod in waters surrounding Ireland was investigated and a baseline of genetic variation was recorded for wild cod from Celtic Sea, Irish Sea and West of Ireland using 53 microsatellite markers. The results showed that cod from the above areas likely form a genetically homogenous population. Genetic markers were employed to monitor the composition of an early phase breeding programme. Parentage analysis revealed large differences in survival between families during rearing for all three sampled year classes (2008, 2009 and 2010). These differences in survival resulted in large variation in family size which was a major contributing factor to a very low effective population size in the farmed stock for the sampled year classes. In addition, F1 broodstock fish contributing to mass spawning events were shown to have unequal reproductive contributions to the F2 cohort. These findings highlight the importance of genetic monitoring in early-phase breeding programs, shed light on the population structure of cod in Irish waters and provide new tools for both gadoid research and other research questions that can be addressed by means of microsatellite marker variation.

Chapter 1. General introduction

1.1. The Atlantic cod

The Atlantic cod (*Gadus morhua,* L.) is an important commercial fisheries species and it has been advanced as a viable aquaculture species, in several countries, over the past two decades. It belongs to the Gadidae family, which also includes species such as Alaska Pollack (*Theragra chalcogramma*), blue whiting (*Micromesistius poutassou*), haddock (*Melanogrammus aeglefinus*), whiting (*Merlangius merlangus*), hake (*Merluccius merluccius*) and other commercial fisheries species (Cohen et al. 1990). Atlantic cod (cod hereafter) is characterized by a stout body, three dorsal fins, two anal fins and a chin barbel (Wheeler 1978) (**Figure 1**). Its body colour is variable from olive green to brown, mottled on back and sides, with a white belly and light lateral line (Wheeler 1978). Cod can reach 95 kg, but an average weight of 4.5 kg is more common (Robins et al. 1986).

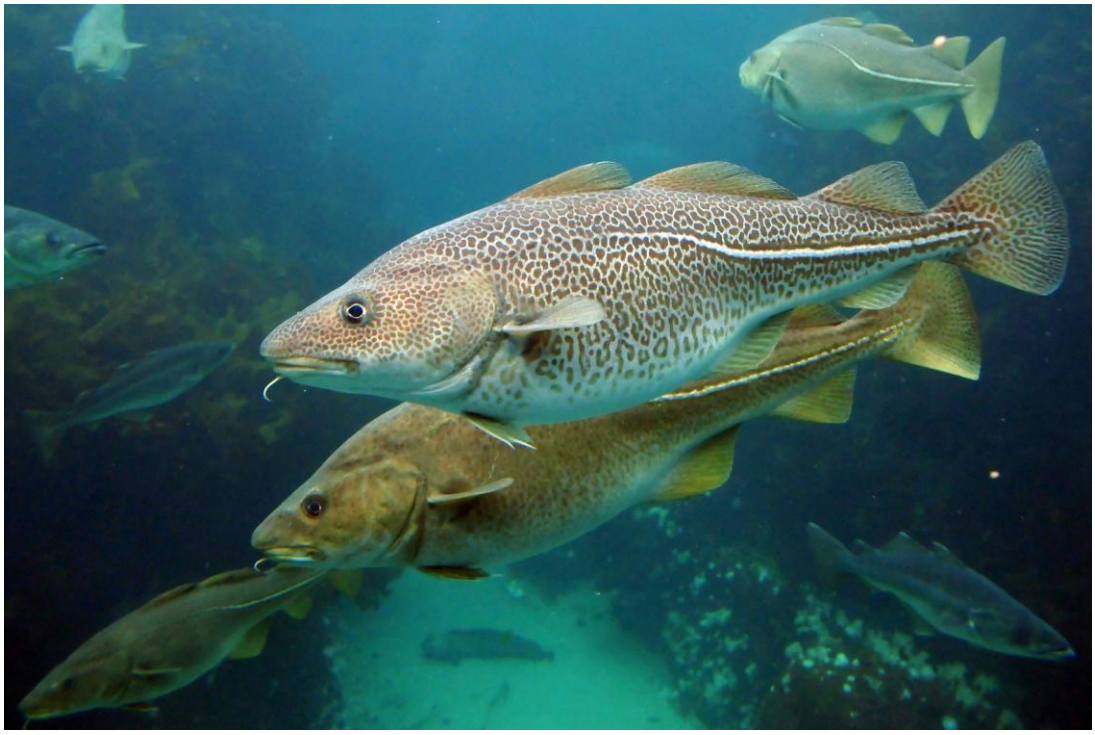

Figure 1 Atlantic cod, *Gadus morhua,* in Atlanterhavsparken, Ålesund, Norway. (Picture by Joachim S. Müller; Available online at https://www.flickr.com/photos/joachim_s_mueller/4500454139).

Figure 2 Atlantic cod caught on the Northeast Atlantic (Picture by the Northeast Fisheries Science Center/National Oceanic and Atmospheric Administration (NEFSC/NOAA); Available online at http://nefsc.noaa.gov/rcb/photogallery/groundfish.html).

The geographical range of Atlantic cod encompasses the continental shelf and coastal waters of the northern North Atlantic Ocean (**Figure 3**). Along the North American coast, cod can be found from Cape Hatteras (North Carolina) to Ungava Bay (northern Quebec). On the coast of Europe, cod's range reaches from the Bay of Biscay to the Barents Sea, including areas around Iceland, Bear Island, Svalbard and Novaya Zemlya (Cohen et al. 1990). Natural climate trends cause fluctuations in the distribution of this species throughout its range, with the most significant distributional changes having been recorded on the coasts of Greenland in recent decades (Drinkwater 2005).

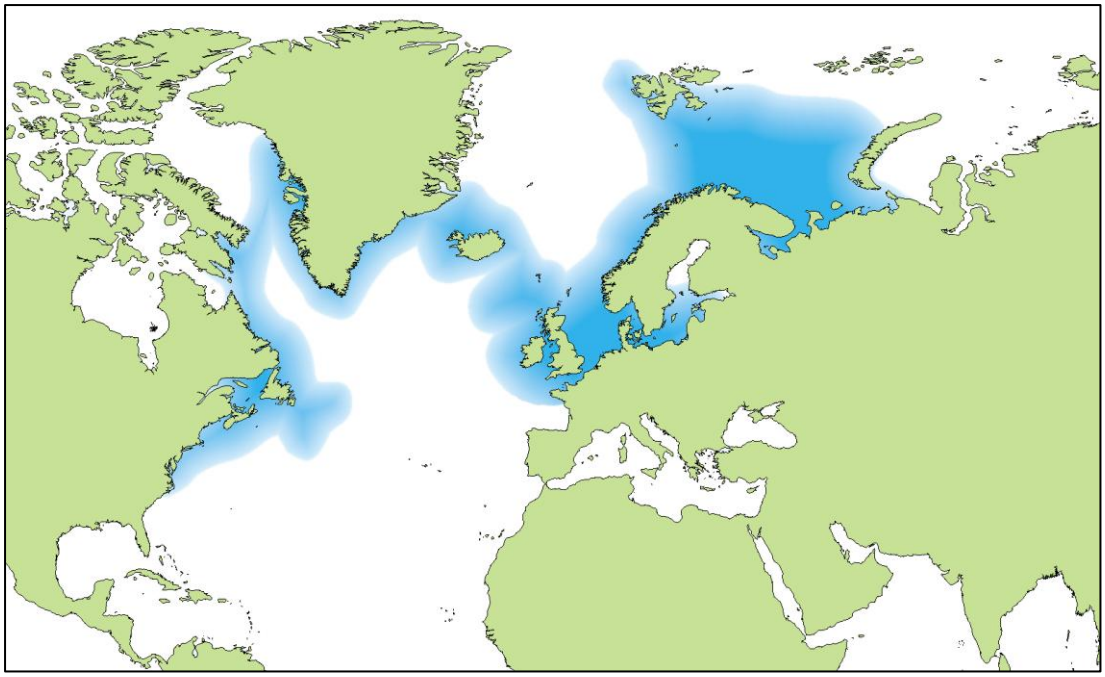

Figure 3 Distribution of Atlantic cod in the Northern Hemisphere.
(adjusted from https://en.wikipedia.org/wiki/Atlantic_cod, Aotearoa 2007).

Cod are a bentho-pelagic species that live in the water column just above the sea floor, although when actively feeding and spawning they become more pelagic in habit (Neat et al. 2006). They are found in varying habitats, from estuaries to the near inshore to the continental shelf and down to depths of 600m (Wheeler 1978). Across such a broad range of habitats, it is evident that cod tolerate salinities ranging from near zero to fully oceanic salinity (Cohen et al. 1990). Likewise, their thermal niche across such an extensive distributional range extends from –1.5 to 19°C; however, a narrower range (1 to 8°C) has been reported during the spawning season (Righton et al. 2010). Moreover, there is evidence that specific, more localized, populations have preferred temperature optima for growth and feeding (Dutil and Brander 2003; Salvanes et al. 2004; Righton et al. 2010).

Cod spawn at different times of the year depending on the geographic location, but most stocks spawn in the period between December and June (Brander 1993). Uniquely, cod on the coast of Maine in the US have two spawning components: winter-spawners are found in October-December and spring-spawners in April-July (Zemeckis et al. 2014). However, most stocks have only a single, distinctive, discrete

spawning period. The Icelandic cod stocks spawn from mid-March to May (Marteinsdóttir and Björnsson 1999) and the Norwegian coastal cod spawn in December-March (Skjæraasen et al. 2011). Both, the Irish Sea and Celtic Sea stocks, spawn from January to April (Brander 1993; Neat et al. 2014) with peak spawning from mid-February to April and February to March, respectively (Marine Institute 2013). Prior to spawning, cod perform migrations of variable distances to reach their dedicated spawning grounds (Jørgensen et al. 2008). The spawning grounds remain spatially stable with fish returning year after year, although the area usage within a spawning ground can vary depending on prevailing ambient conditions, most particularly temperature (Höffle et al. 2014). Once on the spawning grounds, females and males aggregate according to their sex during the day (Dean et al. 2014). Spawning takes place at night when males establish territories and females approach a male of their choice once ready to mate (Dean et al. 2014). There is a courtship behavior between a pair which culminates in the simultaneous release of eggs and milt from a ventral mount (Brawn 1961; Robichaud & Rose 2003). Cod females are capable of producing 500 000 eggs per kilogram of their weight (Cohen et al. 1990); females release their eggs in batches, with two to six days between batches, over an extended spawning period of several weeks (Kjesbu 1989).

1.2. Cod fisheries and stocks

1.2.1. Historic exploitation of cod fisheries

The Vikings fished cod commercially in the North seas as early as the 10th century (Kurlansky 1997). During the 16th century, Spanish, Portuguese, French, English and even Irish fishing fleets started fishing cod annually on the continental shelf off Newfoundland, arriving in the spring, fishing throughout the summer, and returning back home in the autumn with a store of dried or salted fish (Kurlansky 1997). Over the centuries, fishing gear and methods have developed and, in the late 19th century, the first steam trawlers joined the European and local fishing fleets on the continental shelf off the Canadian coast (Kurlansky 1997).

All cod stocks have declined from their maximum biomass since the start of large-scale commercial fisheries (Rose 2004; Brander 2007). The intensification of fishing efforts and continuous development of fishing gear made it possible to fish cod throughout its entire range and over the winter through to the spawning season (Hilborn et al. 2003). In the case of Newfoundland cod stock, fishing had little effect on the stock until the Little Ice Age, a period of cooling water temperatures from the mid- to late 19th century, which depressed the productivity of the stock (Rose 2004). Strong productivity from the early 1900s until about 1960 was able to restore the stock (Rose 2004), and cod landings reached their peak in the 1960s and 1970s, with 1968 being the record year with a 4 000 000 tonne global catch (FAO 2015). By the mid-1960s however, fishing removals rose to >30% which greatly exceeded productivity and lead to the collapse of the stock (Rose 2004). Since then the landings have decreased substantially.

Until the late 1970s, cod fisheries were largely unregulated (Kurlansky 1997), but in 1977, Canada extended its exclusive economic zone (EEZ) to 200 nautical miles from its coastline (Hilborn et al. 2003) and began managing the stock based on scientific information (Hutchings and Rangeley 2011). However, this scientific advice was based on overly-optimistic stock assessments and the foreign fishing effort was replaced by an increased Canadian effort (Hilborn et al. 2003). Eventually, this continuing overexploitation has led to the virtual commercial extinction of Canadian cod stocks (Hutchings and Myers 1994) which until now have still not recovered.

1.2.2. Cod stocks

The International Council for the Exploration of the Sea (ICES), controlling the Eastern Atlantic, and Northwest Atlantic Fisheries Organization (NAFO), regulating the Western Atlantic, currently recognize 25 distinct and individually managed stocks (**Table 1**, **Figure 4**) (Fishbase.org 2016). Historically, a *stock* described a group of fish available for exploitation in a given area (Milton and Shaklee, 1987; Smith et al. 1990). As fisheries science has developed, the stock concept has also evolved in

order to better reflect the actual biological sub-structuring within an exploited species. Thus, a *stock* now refers to a group of fish sustained by recruitment within that group, rather than relying on external sources, and fish within the given stock share similar life history characteristics (Hilborn and Walters 1992). Over the years, considerable effort has been directed at resolving the biological sub-structuring of some commercially important fish species, especially cod, though this information is still lacking for many others. Even if evidence of sub-structuring is available, it is still challenging to change existing management divisions due to political and socio-economic reasons. Hence, the current use of the term *stock* remains a compromise such that stocks are essentially management units that aspire to reflect biological populations, but often fail to do so (Begg and Waldman 1999; Koljonen 2001; Reiss et al. 2009).

Table 1 Cod stocks, the area they occupy, maximum total biomass during historic record, the area of shelf between 0 and 300 m depth and the mean annual shelf bottom temperature.

Stock	Area	Max. biomass (1000 t) (year)[1]	Area (km^2)[2]	Temp. (°C)[2]
Celtic Sea	ICES division VIIe-k	34 (1977)	155 358*	11*
Central and Eastern Baltic	ICES division IIId, subdivisions 25-32	1057 (1982)	216 833**	5**
Faroe Plateau	ICES Subdivision Vb1	155 (1976)	36 236	7.4
Faroe Bank	ICES Division Vb2			
Iceland	ICES division Va	2282 (1955)		
Irish Sea	ICES division VIIa	30 (1973)	48 263	10
North East Arctic - Barents Sea and Norwegian Sea	ICES subareas I and II	4169 (1946)	592 610	4
North Sea, Eastern Channel, Skagerrak	ICES divisions IV, VIId, IIIa West	1146 (1980)	539 823	8.6
West of Scotland	ICES division VIa	54 (1980)	105 666	10
Western Baltic	ICES division IIId, subdivisions 22-24	116 (1972)	43 913	7
Eastern Scotian Shelf	NAFO Div. 4VsW	320 (1985)	102 428	3.75
Flemish Cap	NAFO Div. 3M	113 (1976)		
Georges Bank	NAFO Div. 5Z-6	141 (1981)	102 596	8
Gulf of Maine	NAFO Div. 5Y	41 (1982)	95 582	8
Northern Gulf of St. Lawrence	NAFO Div. 3Pn4RS	603 (1983)	89 041	1
Northern Labrador	NAFO Div. 2GH	-		
Southern Grand Bank	NAFO Div. 3NO	395 (1967)	121 246	1.75
Southern Gulf of St. Lawrence	NAFO Div. 4TVn	474 (1986)	118 343	1.75
Southern Labrador-Eastern Newfoundland	NAFO Div. 2J3KL	2982 (1962)	253 492	0
Southern Newfoundland	NAFO Div. 3Ps	200 (1985)	70 510	2.5
Western Scotian Shelf	NAFO Div. 4X	77 (1983)	77 634	6.75
Norwegian coastal cod	Norwegian coastal waters in ICES subareas I and II			
Rockall	ICES division VIb			
Greenland	ICES Subarea XIV and NAFO Subarea 1	4129 (1949)***		
Kattegat	ICES division Division IIIa East	45 (1973)	17 362	6.5

[1]From Lilly et al. 2008, references within

[2]From Myers et al. 2001

* covers only areas VIIe-f

** covers only the central Baltic, ICES division IIId, subdivisions 25-29

*** covers only West Greenland, NAFO Subarea 1

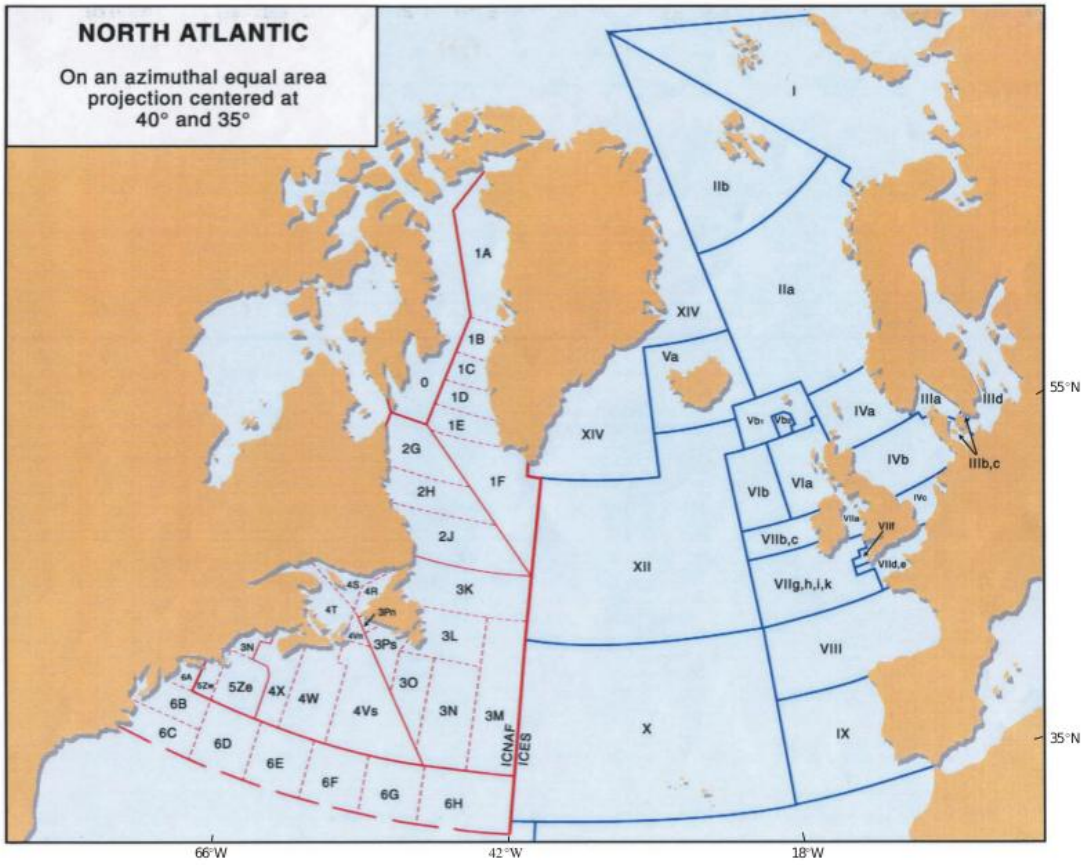

Figure 4 Map of North Atlantic fishery statistical areas. Red squares represent the areas managed by NAFO (formerly the International Commission for the Northwest Atlantic Fisheries (ICNAF)) and blue squares the areas managed by ICES (Horsted 2000).

As noted above, cod stocks are subjected to a wide range of ambient environmental conditions and, consequently, their life histories can vary greatly within the extensive species distribution range. In addition to differences in spawning times as highlighted earlier, the length of their spawning migrations varies considerably (Cohen et al. 1990; Thorsen et al. 2010). For example, the Barents Sea cod typically migrate thousands of kilometres on their spawning migration along the Norwegian cost to the coasts of Finnmark, Lofoten, or further south (Jørgensen et al. 2008). In contrast, several resident populations of cod have been identified on the coastal areas of, for instance, Canada, Iceland and Norway (Ruzzante et al. 2000; Neuenfeldt et al. 2013). As a result of this variation in the length of the pre-spawning migration (from 0km to 1500km) there are differences in reproductively-

associated energy expenditure between the various cod stocks (Jørgensen et al. 2008; Thorsen et al. 2010).

In the same vein, it is also apparent that there is considerable variation in the structuring and composition of ecosystems occupied by these different cod stocks including differences in both the abundance and diversity of prey species and other predatory, piscivorous, finfish species (Lilly et al. 2008). Collectively, these features, in turn, affect the probability of offspring survival, as well as predation pressure encountered at different life stages (Thorsen et al. 2010).

The stocks also differ in size and extent of area occupied (**Table 1**); for example, the maximum biomass of both the West Greenland and Arcto-Norwegian stock were estimated at c. 4 000 000 tonnes each after the Second World War, whereas none of the west of Scotland, Irish Sea, or Celtic Sea stocks have ever exceeded 60 000 tonnes in historic records (Lilly et al. 2008). For some stocks, accurate records begin from 1946, but for west of Scotland, Irish Sea, and Celtic Sea data is only available from 1978, 1968 and 1971, respectively (Lilly et al. 2008). Most stocks have been subjected to high fishing pressure; however, fishing mortality experienced by the stocks does vary due to separate management of stocks (Thorsen et al. 2010).

Collectively, this wide range of environmental conditions and diverse life history features result in significant differences in both productivity and growth rates between the different cod stocks (Dutil and Brander 2003). As a general trend observed in stocks on both sides of the Atlantic, the southerly stocks, including Celtic Sea, show greater growth and higher productivity compared to the northerly stocks (Dutil and Brander 2003). The most productive stocks, Celtic Sea, Irish Sea and West Scotland, are associated with a habitat with higher bottom salinity and temperatures (Dutil and Brander 2003). Temperature has been identified as a major determinant of growth, fecundity, and age and size at maturity (Dutil and Brander 2003; Thorsen et al. 2010; Righton et al. 2010). Cod in warmer waters, such as Celtic Sea stock, have higher growth rates and fecundity but mature at an earlier age and at a smaller size than cod in cold waters (Thorsen et al. 2010).

1.3. Fisheries management issues

1.3.1. Factors effecting recovery after declines in cod stocks

The declines of many cod stocks since the 1970s have been closely examined and are well documented in an extensive literature (e.g. Brander 2007; Rose 2004). It was noted earlier that in the Northwest Atlantic, stocks increased in the 1980s, following the declaration of a 200-nautical-mile fisheries zone in 1977 (Lilly et al. 2008). The exclusion of foreign vessels and the introduction of catch quotas reduced fishing mortality temporarily (Lilly et al. 2008). However, with the development of Canadian fishing industry and overly-optimistic fisheries management, the fishing mortality increased leading to subsequent more severe declines of the stocks in the late 1980s and early 1990s, eventually leading to fishing moratoria on most north-western stocks (Lilly et al. 2008). However, there is increasing evidence that this cessation of directed fishing in the Northwest Atlantic has been insufficient to promote a recovery of stocks (Lilly et al. 2008). In contrast, the declines in the Northeast Atlantic were smaller (in relative terms) and of shorter duration than in Canadian stocks and they did not follow any specific pattern (Brander 2007). Of the Northeast Atlantic stocks, for example, the Northeast Arctic stock and the Icelandic stock have showed strong recovery after a reduction in fishing pressure in recent decades (Hutchings and Kuparinen 2014).

The lack of recovery of the Northwest Atlantic has variously been attributed or linked to, *inter alia*, continued fishing mortality in bycatch; targeted fisheries of some stock components; recreational fisheries; and illegal fishing activity (Hutchings and Rangeley 2011). In addition, intensive fishing has led to reductions in growth rate, age and size at maturity, as well as truncations in age and size structure of stocks (Lilly et al. 2008; Hutchings and Rangeley 2011). Fishing pressure during spawning season can also disrupt the breeding behaviour, delay mating and change operational sex ratios (Rowe and Hutchings 2003). To varying degrees, all these factors have contributed to the reduction in population growth rate concomitant with an overall reduction in population size (total numbers in a given

area). This phenomenon is also called *depensatory population growth* or an *Allee effect* (Courchamp et al. 2008). An Allee effect is essentially a decline in individual fitness at low population size or reduced stock densities (Courchamp et al. 2008). Mechanisms that can cause Allee effects include (altered) mating systems, (modified) predation, and (disrupted) social interactions, such as, the younger fish being unable to learn the migration routes to spawning grounds from the older fish (Courchamp et al. 2008).

The key impacts of these Allee effects give rise to critical population thresholds below which populations are in severe risk of crashing to extinction (Courchamp et al. 2008). Several studies have suggested an Allee effect threshold of 0.10 SSB_{max}, which means that the current spawning biomass (SSB) of a given stock is below 10% of the maximum SSB observed (reviewed in Hutchings 2014). However, it seems that more critical than going below the Allee effect threshold, is the length of time the stock stays under this critical threshold level (Hutchings and Kuparinen 2014). For example, the (now recovered) Northeast Arctic cod stock has fallen under this threshold twice (in 1965 and 1980), but on both occasions it stayed under the threshold for only one year (Hutchings and Kuparinen 2014). In comparison, the (still unrecovered) Northwestern cod stocks continued to be overexploited after collapsing which kept them below the threshold for an extended period and may have resulted in this ongoing lack of recovery (Hutchings and Rangeley 2011).

A further critical factor that must be borne in mind in these considerations is the overall structuring of the fish component within these ecosystems. Prior to their decline in the 1970s, Northwest Atlantic cod stocks were at high densities (estimated at c. 8.8 tonnes per km^2) (Strub and Pauly 2011) and cod was the top predator in the ecosystem (Frank et al. 2005). The fisheries-induced removal of this top predator appears to have caused permanent changes in the ecosystem (Bundy 2005; Brander 2007). The collapse of cod and other large-bodied benthic predator numbers (e.g. haddock) has released their prey, small pelagic fishes and benthic macroinvertebrates, from predator control, leading to drastic increase in the latters' abundance (Frank et al. 2005). These pelagic forage fish often compete with, and

prey on, the young life stages of their larger benthic predators. Increased abundance of pelagic forage fish can therefore suppress the recruitment of cod and other large benthic fish through predation on their eggs and larvae (Fauchald et al. 2010). Thus, there has been a change from a demersal-feeder-dominated system to a pelagic-feeder-dominated system which is contributing to keeping the Northwest cod population at a persistently low level (Bundy 2005).

Finally, in addition to consequent impacts of fishing pressures, the decline of Northwest Atlantic stocks in the 1980s coincided with a cooling of water temperature, known to influence the growth and reproduction of cod (Drinkwater 1996). In summary, then, it is likely that a combination of several factors have and are affecting the declines and recoveries of cod populations. Indeed, a surplus production biomass model that took into account fishery influences, climate influences and the depensatory population growth rate (Allee effect) was able to mimic the documented history of Newfoundland cod (Rose 2004). This model suggested that the collapses in the 1960s were caused by overfishing, while the collapses in the late 1980s were caused by overfishing combined with climate influences, and that recovery in the 2000s was hindered by the depensatory effects of low numbers (Rose 2004).

1.3.2. Structure, status and management of Irish cod stocks

The cod stocks around Ireland are divided into three distinct management entities: Irish Sea (VIIa), Celtic Sea (VIIe-k & VIIbc) and West of Scotland (VIa) (**Figure 5**). In addition, West of Ireland (areas VIIb,c) stock component is recognized, but managed as part of the Celtic Sea stock. However, these management distinctions have yet to be complemented by, and reconciled with, targeted population genetic data. In many cases, the boundaries of biological populations differ from the boundaries of stock management areas (Reiss et al. 2009) and this may well be the case with local cod stocks in Irish waters.

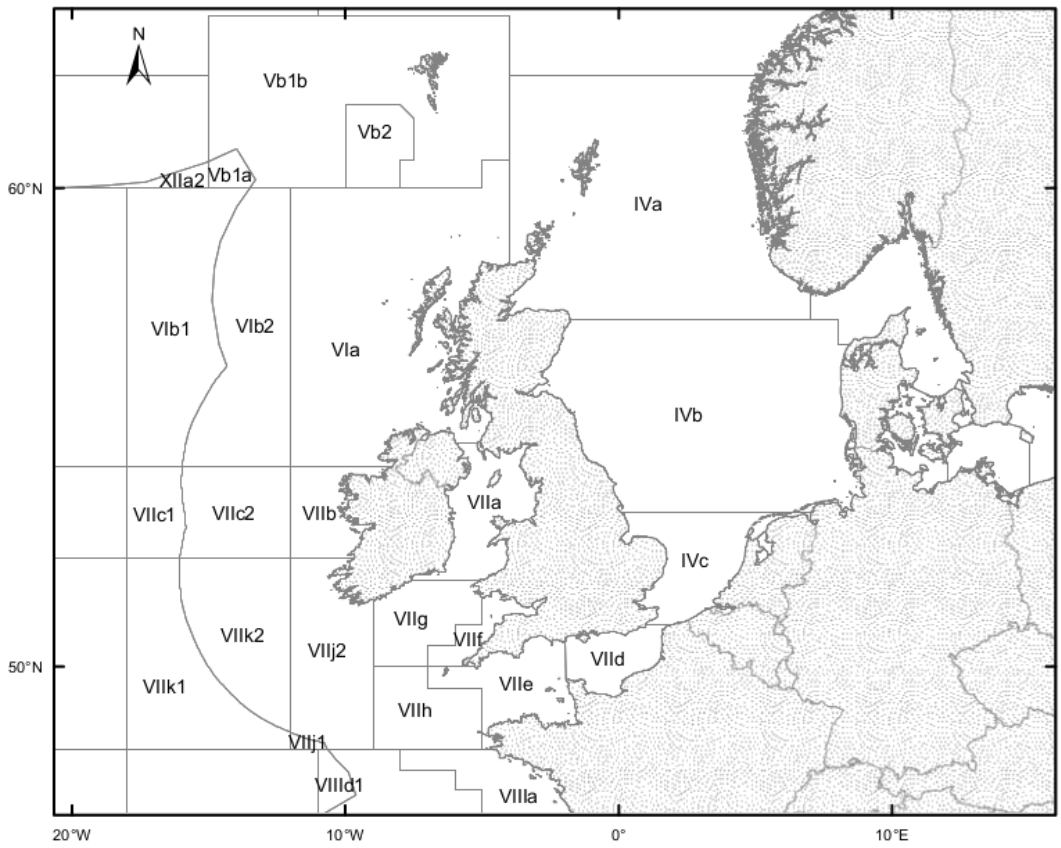

Figure 5 Map of ICES areas in the waters surrounding Ireland. Cod management units are Irish Sea (VIIa), Celtic Sea (VIIe-k & VIIbc) and West of Scotland (VIa).

As noted, ICES conducts ongoing stock assessments of the Northeast Atlantic stocks while NAFO controls the management of the Northwest Atlantic cod stocks (Reiss et al. 2009). Stock management relies on stock assessments that are based on fishery-dependent data, for instance, commercial statistics of catches, catch per effort and fishing effort, and/or fishery-independent data, such as surveys conducted on board research vessels (Cadima 2003). These combined data are used to estimate biological reference points, such as *maximum sustainable yield* (Cadima 2003). *Maximum sustainable yield* is a theoretical concept of the largest (fisheries) yield that can be maintained indefinitely while also maintaining the stock size at the same level (Gulland and Boerema 1973). After stock assessment, the managing authority creates a management plan, for example, in Ireland, the Marine Institute's Fisheries Ecosystems Advisory Services (FEAS) Division formulate the management advice based on their assessments and ICES management advice on the Northeast cod stocks (Marine Institute 2013).

The emergent management plan may include several fisheries management tools. One important measure will be an annual total allowable catch (TAC) (Marine Institute 2013). Other management measures may include bycatch limitations, gear limitations, spatial and temporal closures of specific areas such as spawning areas or nursery grounds (Dankel et al. 2008; Marine Institute 2013).

Of the three cod stocks in the waters surrounding Ireland, both the Irish Sea cod stock (ICES area VIIa, **Figure 5**) and West of Scotland stock (VIa) are currently considered collapsed (Marine Institute 2013). A recovery plan for the Irish Sea cod was put in place in 2000 after the stock reached a record low in 1999 after 20 years of continuous decline (Kelly et al. 2006). The plan, established by the European Commission, included reduction of quotas, technical gear regulations, and a limited closure of spawning grounds during the spawning season (Kelly et al. 2006). Unfortunately, the recovery plan has been unsuccessful in reducing mortality and reversing the decline of the stock (Kelly et al. 2006). The Celtic Sea stock (VIIe-k) continues to be fished at maximum sustainable yield with a TAC of 6,848 tonnes assigned in 2014 (Marine Institute 2013). Since 2005, a temporary area closure in February - March has been in place for most vessels and gear in known spawning grounds within areas VIIg and VIIf (Marine Institute 2013). The stock assessment for the Celtic Sea stock is based only on areas VIIe-k, and does not cover the West of Ireland stock component (areas VIIb,c). However, this area is still included in the Celtic Sea TAC.

Even with a management plan in place, overfishing can occur if management goals are not precautionary. For instance, if mixed stocks are managed as one stock, the population size is overestimated, or TAC is set higher than recommended by scientific advice, management can fail (Lilly et al. 2008; Reiss et al. 2009). In addition, most stock assessments do not take into account natural variation in productivity, which can lead to setting the TAC too high (Morgan et al. 2014). Landings exceeding TAC, high-grading and non-reported catches can also result in overfishing (Lilly et al. 2008). However, a common practice of 15% limit on annual

TAC changes has prevented reactive management needed for stock recovery in the case of the Irish Sea stock (Kelly et al. 2006.)

1.3.3. Summary

All cod stocks have been affected by overfishing in the past decades, but the Northeast Atlantic stocks have shown good recovery following reduction in fishing pressure, whereas Northwest Atlantic stocks have not responded to reduction in harvesting but remain at very low levels. The Irish cod stocks too have been depleted as a result of overfishing, and only the Celtic Sea has a viable TAC and therefore appears to be accessible as a broodstock source for Irish cod aquaculture. Another advantage of choosing a local broodstock, besides accessibility, is that it will be adapted to the local environmental conditions and therefore have an innate performance advantage (Bolton-Warberg et al. 2013). It is important to resolve if fine-scale population structure exists in cod in the Irish waters, both for broodstock acquisition for aquaculture, and for aiming to match stock management with biological populations. If cod aquaculture becomes widespread on the Irish coasts, such knowledge will help also in minimising possible negative effects of farmed-wild interactions.

1.4. Cod aquaculture

1.4.1. Early initiatives in cod rearing

Historically, the first experiments in growing cod were conducted in the 1880s in both Norway and the US (Earll 1880; Moksness et al. 2004) and these included hatching and releasing cod larvae primarily for stock enhancement purposes. In more recent times, the first proper aquaculture endeavors took place nearly a century later. By the 1980s, cod catches had significantly reduced on both sides of the Atlantic (FAO 2015) and these low catches and consequent high market prices for cod sparked interest in cod aquaculture (Brown et al. 2003). The Norwegians were the pioneers in modern cod aquaculture in 1980s and had some initial success

(Brown et al. 2003). However, the early 1990s saw increased wild catches and cod market prices fell, and these efforts in cod aquaculture were abandoned (Kjesbu et al. 2006) as production prices far exceeded the 'spot' price for wild catches. However, within a decade in the late 1990s and early 2000s, interest was once more renewed, again due primarily to decreasing catches and rising prices for cod (Rise et al. 2009). Thus, it is readily apparent that falling wild supplies and rising market prices were (quite logically) the primary drivers forcing interest in cod aquaculture over the past almost four decades (FitzGerald et al. 2010). It is not surprising then that the most recent cycle of cod aquaculture efforts have again wound down due to an increased supply of wild fish largely, on this occasion, from Icelandic and Russian waters and a depressed market price (FitzGerald et al. 2010) (**Figure 6**).

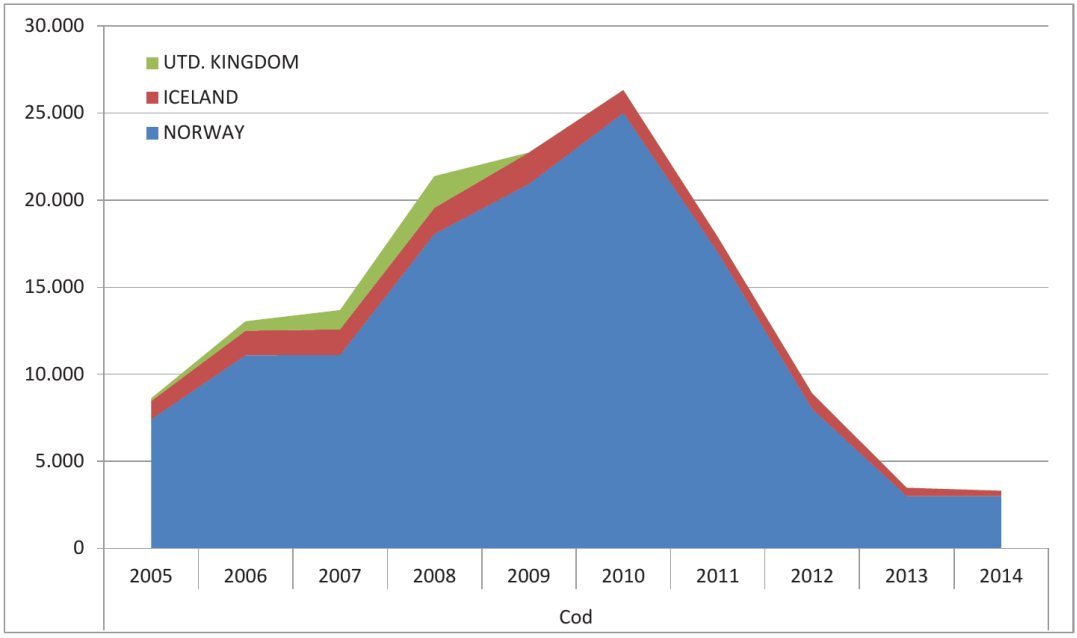

Figure 6 Cod aquaculture production in Europe from 2005 to 2014 (tons) (FEAP 2016).

1.4.2. Recent technical advances in the farming of cod

In the late 1990s, intensive hatchery protocols had been developed for a range of marine species including, turbot (*Scophthalmus maximus*), halibut (*Hippoglossus*

hippoglossus), sea bass (*Dicentrarchus labrax*) and sea bream (*Sparus aurata*) (e.g. Moretti et al. 1999; Imsland et al. 2001; Jackson et al. 2003). Due to the similarity in development of larval stages, these methods, most notably the 'Green Water-Rotifer-Artemia' format, were readily adopted and adapted for cod larval rearing (Shields 2001). Over the past decade, the more specific larval and juvenile rearing conditions required by cod have been further refined and improved by determining the optimal rearing temperatures at different developmental stages (Otterlei et al. 1999; Bolton-Warberg et al. 2015) and enhancing the larval feeding regime by enrichment of live feeds (Shields 2001) and modifying the feeding strategy (Baskerville-Bridges and Kling 2000). Natural or cultivated marine zooplankton, most particularly copepods, have been identified as an alternative diet source to routinely used enriched rotifers, that can have positive long-term effects on juvenile growth (Koedijk et al. 2010; Karlsen et al. 2015).

One of the more interesting facets of this latest round of aquaculture effort has been the availability of growth data, for the first time, on the performance of farmed cod stocks in the marine environment for a variety of different stocks and disparate environments across its natural range, from Ireland in the southern end of the range to northern Norway and Newfoundland. Growth of cod through the full production cycle has been investigated through modeling and empirical (farmed) data (Björnsson and Steinarsson 2002; Björnsson et al. 2007; Bolton-Warberg & FitzGerald 2012). Based on modeling, Irish conditions are favorable for juvenile cod growth and particularly good for younger cod, with a harvest size of 2.5 kg being achievable in 38 months (Bolton-Warberg & FitzGerald 2012). Similar growth rates have indeed been recorded for Celtic Sea and Scottish cod in sea cages in Irish conditions (Bolton-Warberg & FitzGerald 2012).

1.4.3. Breeding concepts and strategies in aquaculture

Globally, apart from some salmonid species, a large part of aquaculture production relies completely on sourcing gametes from wild fish and several finfish aquaculture species are periodically 'refreshed' with wild spawners (Gjedrem 2000). This is in

stark contrast to the situation in agriculture and this practice clearly complicates the concept and long-term course of domestication. Domestication is a process by which the farmed fish become adapted to the captive environment by genetic changes occurring over generations (Price 1984). Domestication leads to, for example, better food conversion rates and tameness in the farmed fish compared to the wild fish. Tameness facilitates animal handling and improves animal welfare (Price 1999). However, using wild spawners will prevent the genetic changes taking place during domestication. This presents problems in terms of animal welfare and suboptimal food conversion rates in the farmed stocks (Gjedrem 2000). For effective domestication, and for improving production traits by selection, the whole life cycle of the species must be completed in captivity. This requires complete control over the species' entire reproductive cycle.

In some cases, the choice of aquaculture base population(s) has been based on the population's performance in their natural habitat (Gjedrem 2010); however, measurements made on wild fish, in their natural habitat, are affected by several factors, and may be a poor predictor of strain performance for economically-important traits in a commercial environment (Holtsmark et al. 2008). Hence, it is important to test the performance of a stock under farming conditions (Bolton-Warberg & FitzGerald 2012).

Ideally, if improvement of certain production traits is desired in an aquaculture population, a breeding programme must be put in place. Breeding programmes define a breeding goal that generally may include some or all of the following characters: enhanced growth rate, increased disease resistance, desired flesh quality and delayed age at sexual maturation (Gjedrem 2000, Kolstad et al. 2006). Selective breeding is used to improve the traits defined in the breeding goal. It is axiomatic but selection can only be carried out on traits that are heritable, that is, they have a genetic component that is passed on from parent to offspring (Frankham et al. 2010). Heritability is defined as the proportion of genetic variation of the total phenotypic variation or, in the narrow sense, as the proportion of additive genetic variation of the total phenotypic variation, which can be estimated

by, for example, comparing the trait in parents with that of their offspring (Frankham et al. 2010).

Genetic variation and its magnitude in these economic traits needs to be investigated in order to choose the optimal selection strategy and predict achievable gains (Gjedrem 2000). For example, if the genetic component of variation in a trait of interest is very small, effort may be better placed in optimizing the environmental conditions of rearing instead. It is also important to resolve any phenotypic and genetic correlations between the traits in question, as selection of one trait may inadvertently affect another (Gjedrem 2000).

In past, aquaculture has relied on traditional quantitative animal breeding models where individual, family or mass selection protocols are utilized over successive generations to select for improved performance in selected traits, such as growth rate among others. More recently, the emergence and development of molecular tools is facilitating a 'genomic' approach to breeding. In the genomic approach, specific traits, especially quantitative trait loci (QTL), can be associated with particular markers, which are more easily assayed for than the trait itself. This process of indirectly selecting for a marker associated with the desired trait is called marker-assisted-selection (MAS). As MAS is an indirect selection method, traits can be selected for even before they are expressed. This is especially useful for traits such as disease resistance. In this way, genomic approach offers the potential to enhance and accelerate the process of selective breeding in aquaculture

A significant consideration in all animal breeding strategies is to maintain a working balance between the conservation of diversity and the avoidance of inbreeding, and improving the farmed stock through selection. Theoretically, inbreeding will accumulate in any 'closed' (and selected) population with the passage of time and with successive generations (Sonesson et al. 2005). Even if the population has large effective population size, the average relatedness between individuals increases every generation (Sonesson et al. 2005). After a certain number of generations, it is not possible to find unrelated individuals to be mated and inbreeding starts to

accumulate (Sonesson et al. 2005). The length of time it takes for inbreeding to start accumulating and the rate of accumulation depend on the effective size of the population in question. The high reproductive capacity of many aquaculture species enables the production of large numbers of offspring from a small number of broodstock. In turn, this can lead to very few families dominating the genetic contributions to the next generation, which will lead to rapid accumulation of inbreeding which can lead to inbreeding depression in important traits, loss of genetic variation available for future selection and genetic erosion (Skaarud et al. 2011). The solution is to retain a large effective population size in the broodstock to ensure that much of the inherent adaptive potential available in the wild populations is captured and retained. In effect, a practical balance must be found between the conservation of both innate diversity and improvement in selected traits.

1.4.4. Examples of cod breeding programmes

In the most recent phase of interest in the 2000s, several countries established aquaculture initiatives for cod. These include Norway, Iceland, Faroe Islands, UK, Canada, USA and, most latterly, Ireland (Björnsson and Steinarsson 2002; Brown et al. 2003; Guðfinnsson 2008). In addition, breeding programmes, at various scales, have been established in most of these jurisdictions and the Norwegian, Icelandic and Canadian programmes have already initiated selective breeding practices (Jørstad et al. 2007; Rise et al. 2009).

The Canadian “Atlantic cod genomics and broodstock development project” (CPG) was an 18.1 million dollar project in 2006-2009 for developing broodstock and genomics tools (Rise et al. 2009). The CPG established family-based selective breeding programmes in three locations (New Brunswick, Newfoundland and Labrador) using local stocks (Rise et al. 2009). Two selective breeding programmes have been initiated in Norway. One of the breeding programmes is funded by the Norwegian Ministry of Fisheries and Coastal affairs and the other one by a commercial company (MarineBreed) (Jørstad et al. 2007). The Norwegian

programmes used both the Northeast Arctic cod and Coastal cod populations to establish their base broodstocks (Jørstad et al. 2007). Iceland's breeding programme managed by a commercial company, Icecod, was funded by both Governmental Agencies and private investment companies (Jørstad et al. 2007).

In addition to full-cycle fish farming, the Icelandic have experimented with ranching of juvenile cod (Björnsson 2011). In this form of ranching, wild juveniles are conditioned to a regular feeding regime at feeding stations (Björnsson 2011). The fish grow faster with exogenous feeding and are caught with little effort and there is no bycatch (Björnsson 2011). Many of the obstacles to ranching are political, legal and social issues, such as the determining the ownership of the fish, conflicts in area usage, as well as the possibility of fish herds leaving the ranching area before capture (Björnsson 2011).

In Canada, 20-30 companies were active in cod farming in the 2000s (Pinfold 2010). In 2007, 523 farms produced 11 087 tonnes of cod in Norway and 13 farms produced 2000 tonnes in Iceland (Paisley et al. 2010). The peak of global cod aquaculture production was 22 729 tonnes in 2009. The production has since dropped to just 4252 tonnes produced in 2013 (FAO 2015), and in 2010, the Canadian cod aquaculture was deemed to be unsustainable in its current state (Pinfold 2010).

The failure of the current cod aquaculture phase has been largely due to market forces. The majority of investments in cod farming, in Norway and other countries, were funded by private investors and therefore the global economic downturn resulted in lack of capital for investment (FitzGerald et al. 2010). This, combined with the remaining biological and technical challenges such as early maturation and disease outbreaks which have not been solved as quickly as initially expected, has essentially halted cod aquaculture efforts in the countries involved (FitzGerald et al. 2010). As a result, all of the national cod breeding programs are now finished, in abeyance or operating at a very low, core broodstock maintenance, level.

1.5. Molecular genetic tools

There are many questions in aquaculture and fisheries studies that cannot be answered by direct observation. For example, solving parentage in a mass spawning event or assigning fish from fishermen's catches to the stock of origin, can be challenging, if not nearly impossible tasks, if one is relying on direct observations alone. Some of these questions may possibly be better answered with the use of genetic tools.

Mutations happen in the genomes of all organisms as a result of normal cellular operations (Liu & Cordes 2004). If these mutations occur in the germline, they can be passed from parent to offspring. As a result of mutation, selection, gene flow and genetic drift, genetic variation will be distributed differently between individuals, species, or higher order taxonomic groups (Liu & Cordes 2004). Genetic markers are locations or structures in the DNA where such genetic variation can be identified.

Popular genetic markers include allozymes, mitochondrial DNA loci, restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellites, single nucleotide polymorphism (SNP), and expressed sequence tag (EST) markers (Liu & Cordes 2004). These markers have different properties regarding their presence in DNA or as proteins, location in the genome (nuclear chromosomes or mitochondria), and in the coding region *vs*. non-coding region. Genotyping assays of the different marker types also differ in their requirements and resolving power (dominant vs. co-dominant), and level of automation. The most important genetic markers utilised today are microsatellites and SNPs (Chistiakov et al. 2006; Seeb et al. 2011).

1.5.1. DNA sequencing

Most genetic markers require prior knowledge of the actual DNA sequence (e.g. microsatellites, SNPs), while others do not (e.g. RAPD, AFLP) (Liu & Cordes 2004). In

order to know the underlying DNA sequence, sequencing techniques need to be applied. DNA sequencing is the procedure of discovering the order of nucleotides within a DNA molecule. The enzymatic di-deoxy chain-termination method, Sanger sequencing, was the principal method of sequencing from the 1980s until the late 2000s (Sanger et al. 1977). Sanger sequencing produces fragments with up to 1100 nt read length with an error rate of <0.01% - 1% (Keith et al. 1993; Noguchi et al. 2006; GATC_Biotech 2015).

New sequencing technologies began developing rapidly in the 2000s. Next generation sequencing (NGS), or massively parallel sequencing, enables the simultaneous sequencing of millions of DNA fragments. There are several commercial platforms that utilize different techniques to produce sequences in such a fashion. For example, Roche/454, Illumina and Ion Torrent are popular NGS platforms. Roche/454 employs pyrosequencing, which relies on bioluminescence (Margulies et al. 2005; 454 Life Sciences, Branford, CT). In contrast, Illumina uses a system of four reversible terminators (A, C, G and T) labeled with different removable fluorophores (Bentley et al. 2008; Illumina, Inc., San Diego, CA). Finally, Ion Torrent determines the sequence by detecting the release of hydrogen ions taking place during nucleotide incorporation by quantitating changes in pH with a coupled silicon detector (Rothberg et al. 2011). A common feature among these platforms is that they include a clonal amplification step in order to magnify the signal resulting from a single sequence (Mardis 2013). The amplification step is then followed by sequencing during synthesis (Mardis 2013). The resulting sequencing data is characterised by a large volume of sequence reads, shorter read lengths compared to traditional Sanger sequencing and higher error rates (Gilles et al. 2011).

Next generation sequencing has made more genetic resources available for researchers. The first implementations of next generation sequencing included sequencing random sequence fragments from DNA extracts without a prior choice of sequenced regions (shotgun sequencing) (Binladen et al. 2007). It was initially used for sequencing relatively small genomes (Margulies et al. 2005; Hofreuter et al.

2006). Since then, the use of NGS technology has vastly expanded and is now commonly applied to, for example, marker discovery, gene expression profiling, whole genome sequencing, target sequencing, chromatin immunoprecipitation sequencing and small RNA sequencing (Wang et al. 2009; Lee et al. 2013). In the same vein, using NGS technology to directly sequence molecular markers offers new possibilities for marker genotyping.

1.5.2. Microsatellites

Microsatellites are 1-6 bp tandem repeats found in the nuclear genomes of most organisms (Selkoe and Toonen 2006). Individual genotypes will vary by having different numbers of repeats in these specific repeat regions. Microsatellites have high mutation rates and are fairly evenly distributed in genomes of nearly all organisms (Ashley and Dow 2013). In eukaryote genomes, they are estimated to occur every few thousand base pairs (Tautz 1989; Brooker et al. 1994). Microsatellites are co-dominant and follow Mendelian inheritance, which makes them useful tools in molecular biology. The most common method of genotyping microsatellites is to amplify the loci in PCR reaction followed by sizing of fragments via automatic capillary electrophoresis system (Guichoux et al. 2011), although other methods employing, for example, mass spectrometry have been developed (Butler 2005; Pitterl et al. 2008). Microsatellites have been used extensively for various applications, such as studies of population structure, parentage analysis, genetic mapping, evolutionary processes and forensics (Bruford and Wayne 1993; Brockmann et al. 1994; Knapik et al. 1998; Goldstein et al. 1999; Primmer et al. 2000). Even though microsatellites are a widely used marker, aspects warranting further research remain, such as their role in maintaining chromosome structure in telomeric and centromeric regions, regulating recombination, and replication as well as regulating gene expression (Chistiakov et al. 2006; Gymrek et al. 2015).

1.5.3. Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are single base pair changes in the DNA caused by point mutations. SNPs can theoretically have four alleles according to the four base pairs in the DNA code; however, they are considered bi-allelic in practical terms due to the low likelihood of more than one point mutation happening in the exact same location. SNPs are more prevalent and widely distributed in the genomes than any other marker, occurring every 200-500 base pairs (Brumfield et al. 2003). Because the difference in the DNA sequence is very small in the case of SNPs, it is challenging to genotype them. Several techniques have been developed for genotyping SNPs at scales of different magnitude; for example, single SNP assays using quantitative PCR, Taqman assays of hundreds of loci and SNP chips with thousands of loci (Chowdhury et al. 2007; LaFramboise 2009; Garvin et al. 2010; Ha et al. 2014). Despite these technical difficulties, SNPs are becoming popular because of good genome coverage, theoretical ease of genotyping and handling of bi-allelic data. SNPs have been used in studies of, for example, population structure, linkage mapping, marker-assisted-selection (Kumar et al. 2012), population assignment, and local adaption (Seeb et al. 2011)

1.6. Molecular genetic tools in fisheries management and aquaculture

1.6.1. Molecular genetic tools in fisheries management

Since the first emergence of genetic tools, they have been used in stock discrimination studies in numerous situations and these methods have been successively and widely used to study the population structure of cod stocks being a major international commercial catch species. Numerous studies have demonstrated population differentiation between the Eastern and Western Atlantic cod using various genetic markers i.e., allozymes (Mork et al. 1985), minisatellites (Galvin et al. 1995), nuclear RFLPs (Pogson et al. 1995; Pogson et al. 2001), microsatellites (Bentzen et al. 1996; Hutchinson et al. 2001; O'Leary et al. 2007; Pampoulie et al. 2008), mitochondrial DNA (Carr and Marshall 2008), single

nucleotide polymorphisms, (O'Leary et al. 2006; Nielsen et al. 2009a; Bradbury et al. 2010). Moreover, several studies have found population differences on smaller geographical scales (Hutchinson 2003; Knutsen et al. 2003; Nielsen et al. 2003) demonstrating the practical utility of these techniques. In addition, microsatellites have also been used to investigate family relationships and effective population size in wild cod stocks (Herbinger et al. 1997). While advances in SNP discovery and genotyping have made it possible to genotype large numbers of SNPs to perform genome scans to detect directional selection and adaptation in selected cod stocks (Moen et al. 2008; Bradbury et al. 2010).

Cod stocks are known to migrate from their breeding grounds to feeding grounds where they mix with other stocks (Jónsdóttir et al. 2007). These aggregations of mixed stocks pose challenges in fisheries management, in very practical terms, but genetic stock identification and mixed stock analysis with several markers have shown promising results (Nielsen et al. 2007; Wennevik et al. 2008; Bradbury et al. 2011). A European project, CodTrace (2001-2004), was established to create a set of tools for assigning cod from fishermen's catches to the stock of origin: among the plethora of techniques assessed, the genetic markers chosen for this were notably microsatellites and SNPs (Higgins et al. 2010).

Different marker types are useful for addressing different questions. Species identification is an important issue in forensic science, taxonomy and ecology (Teletchea 2009). DNA sequencing of mitochondrial and nuclear loci has been employed for studying phylogeny of codfishes (Owens 2015). The barcode of life, cytochrome c oxidase subunit 1 mitochondrial region (COI) has been used for species identification in tracing the origin of fish products (Ward et al. 2005; Miller and Mariani 2010). Microsatellites and SNPs, due to their small size in base pairs, have successfully been used to amplify ancient DNA, which can be highly degraded. Use of ancient DNA samples, for example from historical otolith samples, has made it possible to study demographic changes, such as bottlenecks, that took place in the past (Hutchinson et al. 2003).

1.6.2. Molecular genetic tools in aquaculture

Molecular genetic tools have a very significant role and broad application within key supporting actions for the development of aquaculture. For example, hatchery management of aquaculture species relies on pedigree information (Gjedrem and Baranski 2009) enabling performance of offspring to be traced back to specific parents. In many aquaculture species, tagging of large numbers of young fish with physical tags is both technically challenging and expensive. However, molecular parentage analyses have made reconstructing pedigrees more feasible (Liu and Cordes 2004). Likewise, parentage analysis is particularly important for monitoring inbreeding in aquaculture programs (Waples & Waples 2011).

In molecular parentage analysis, progeny and the potential parents are genotyped for multiple loci. These multilocus genotype data are then used to assign progeny to parents, either by excluding other potential parents, given the rules of Mendelian inheritance or through a probabilistic framework, where the likelihood of each parent or parent pair having produced the offspring in question is determined and and compared, assigning the offspring to the parent/parent pair that achieved the highest likelihood score (Waples and Waples 2011). In cod aquaculture, microsatellites have been successfully used for parentage analysis in communally-reared fish and to determine individual contributions in mass spawning events (Wesmajervi et al. 2006; Herlin et al. 2007).

Pedigree information is also required for estimating genetic parameters for economically-important traits in selective breeding programs (Gjedrem 2010). Typically, the phenotypic variation in production traits, like growth rate, disease resistance or fillet quality, is not controlled by any one gene, but rather by a combination of several genes in conjunction with environmental conditions. *Quantitative genetics* investigates the amount of heritable genetic variation in these quantitative traits, the genetic correlations between the traits and the effects of environmental factors (Falconer and Mackay 1996).

Once the genetic parameters of the traits of interest are known, a selection programme can be designed to improve these traits. In practical commercial terms, the primary trait of interest for aquaculture species is often the growth rate (Gjedrem 2010). Growth rate is a laborious trait to measure directly and, as body weight is significantly correlated with growth rate, body weight is usually used as an indirect (proxy) measure instead (Gjedrem 2010). For example, the heritability of body weight has been estimated to be relatively high (between 0.29 and 0.52) in the current Norwegian and Icelandic cod breeding programmes (Gjerde et al. 2004; Kristjansson and Arnason 2014).

The genomic regions that contain quantitative traits, such as growth rate, are called quantitative trait loci (QTL) (Falconer and Mackay 1996). If a QTL is placed on a genetic or physical map of the organism's genome and a closely linked genetic marker can be identified, this genetic marker can be employed to indirectly select for the desired trait (Lande & Thompson 1990). This type of marker-assisted selection (MAS) has been explored, for example, in the recent Canadian CPG programme (Rise et al. 2009) and can be very useful if the trait of interest is expressed late in an organism's life, or, like disease resistance, is only expressed when challenged with a pathogen (Sonesson 2007).

The first step in developing a working MAS protocol is to create a linkage map or a physical map of the organism's genome. Large numbers of EST-derived microsatellites have been developed for this purpose in cod and two linkage maps have been created for cod (Higgins et al. 2009; Moen et al. 2009; Hubert et al. 2010). The next step is to place the QTLs on the map. Several methods have been developed for this, some specifically to be used with mass spawning species, such as cod (Massault et al. 2008; Massault et al. 2010). An alternative approach, a direct search of SNP markers in the genes of interest has identified 30 SNPs in 18 candidate genes for growth and reproduction in cod that have potential to be used in MAS as well (Hemmer-Hansen et al. 2011).

1.6.3. Future challenges and genomics

Among the many remaining challenges in cod aquaculture are disease management, early maturation, ecological sustainability of farming, and escapes. The use of genomic approaches it is hoped will provide information that can be used to address some of these problems (Goetz et al. 2006).

In this context, the high density of fish and sub-optimal rearing conditions in aquaculture, especially with new aquaculture candidate species, make the spread of pathogens easier than in natural conditions (Bergh et al. 2001). The largest losses due to pathogens usually occur at larval and juvenile stages (Samuelsen et al. 2006). For example, common vibrio species *Vibrio logei*, *V. anguillarum* and *V. splendidus* have been identified as pathogens of cod larvae (Reid et al. 2009) and a bacterial disease, francisellosis (*Francisella noatunensis* subsp. *noatunensis*) has been detected in wild-caught Celtic Sea Atlantic cod juveniles reared in captivity (Ruane et al. 2013). However, whole genome sequencing of the cod genome has revealed a unique immune system structure (Star et al. 2011), lacking MHCII, which may give valuable insights into diverse means of disease control.

Another common problem in cod aquaculture, as with many cultured species, is early maturation, when fish regularly reach maturity at a much earlier age and smaller size than in the wild (Karlsen et al. 2006). In practical terms, as these fish mature and prepare for spawning, their energy resources are directed into producing gametes instead of growth. At the onset of maturation, the fish experience on average a 30% reduction in body weight associated with a reduction in condition and increased mortalities during the spawning season (Kjesbu et al. 2006). Thus, early maturation increases the time needed to reach the harvest size with significant wastage of resources and increased production costs (Kjesbu et al. 2006). One practical approach used to delay maturation is photoperiod manipulation, for example, such that maturation is deferred to a later time (Trippel et al. 2008). In addition, triploid fish have been created and studied as their gonadal development is suppressed, but none of these methods are in routine use for cod

farming (Trippel et al. 2008). Heritability for early maturation, measured as proportion of maturity during the first year, has been estimated at 0.17 in the Icelandic cod breeding programme. In addition, a negative correlation between size and age at first spawning has been identified and this means that selection for larger size will simultaneously select for fish that mature earlier, complicating efforts to delay maturation (Kristjansson and Arnason 2014).

Traditionally, feed formulations of carnivorous marine finfish species are based on fishmeal and fish oil. To ensure sustainable development of aquaculture, there needs to be a shift towards more terrestrial plant products, plant meals and vegetable oils in the feeds (Naylor et al. 2009; Hardy 2010). Promisingly, no major metabolic changes were found in intestinal tissue of juvenile cod in a gene expression study when fish oil in the feed was replaced in different proportions with camelina oil (Morais et al. 2012). Moreover, the environmental impacts of farming cod have to be studied and mitigated (Kjesbu et al. 2006).

Another serious challenge is the environmental issue relating to potential escapes. As the genetic make-up, as well as the phenotype of domesticated fish, can be quite different from their wild counterparts, escapes of domesticated individuals can theoretically threaten not only the integrity, but also the survival of wild populations by changing the genetic make-up of the locally-adapted wild population (Utter 1998; Bekkevold et al. 2006). Cod have a tendency to bite holes in nets and appear, in general, to be more willing to search for escape routes than for example Atlantic salmon (*Salmo salar* L.) (Moe et al. 2007). Experiments with genetic marking have confirmed the escapes and the long-term survival of both adult individuals and eggs during spawning season from sea cages (Jørstad et al. 2008; Jørstad et al. 2013). As cod farming is still a recent phenomenon, the farmed and wild cod are not expected yet to be genetically very different (Jørstad et al. 2013). However, it must be borne in mind that in Atlantic salmon introgression with farmed escapes over past decades has significantly altered the genetic integrity of native populations, indicating loss of local adaptation (Bourret et al. 2011).

Molecular tools can provide a major support for cod aquaculture and be a significant enabling technology in solving many of the problems mentioned above. However, given the recovery of the cod stocks in the Northeast Atlantic and the resulting depression of the 'spot' price of cod, there is currently little interest in investing in cod aquaculture, as seen in the low production numbers (**Figure 6**). The developed resources will however be useful in domestication of other related species, and can be readily utilized in the future if the status of wild stocks declines again.

1.7. EIRCOD

This PhD project is an integral part of the many diverse actions of The National Broodstock and Breeding Programme for Atlantic cod in Ireland (EIRCOD). EIRCOD was established in 2008 with the objective of supporting and facilitating cod aquaculture in Ireland. The project built upon earlier initiatives and the existing facilities, most importantly a marine finfish hatchery, at Carna Research Station of National University of Ireland, Galway, in the west of Ireland. Previously, cod from various backgrounds, often sourced outside of Ireland, were reared in the hatchery and had been on-grown at sea on a pilot commercial farm, Trosc Teo, in a nearby bay since 2004.

From the start of the National Broodstock and Breeding Programme in 2008, the base broodstock population has been derived from wild cod collected in the Celtic Sea. Celtic Sea stock has been found to have the highest levels of production of all cod stocks (Dutil and Brander 2003). As part of EIRCOD, the growth of the Celtic Sea stock has been benchmarked and compared with other stocks in sea cages in Irish conditions (Bolton-Warberg and FitzGerald 2012). Likewise, the optimization of larval and juvenile rearing has been extensively studied in the course of the EIRCOD programme (Bolton-Warberg 2012; Bolton-Warberg et al. 2015).

Many breeding programmes rear families separately until tagging (Gjerde et al. 2004; Hubert et al. 2010). If each family is reared in a separate tank, comparing

performance between families can be confounded due to environmental differences between the rearing units, reflecting 'tank effects' (Speare et al. 1995). This limitation can be overcome if families are mixed and reared communally, and parentage is determined later or progressively through the life stages using genetic markers. In the EIRCOD work programme, a non-lethal, least-invasive, and reliable sampling procedure for post-larval fish was developed by Mirimin et al. (2011). This quick and inexpensive method was adopted for routine monitoring of captive stocks through the critical early life stages.

A comparative desk study on harmful effects of farmed escapes summarised the lessons learned from escape studies on salmon and extrapolated the results to cod farm escapes (Cross et al. 2009). It was highlighted that knowing more about the wild cod population structure is vital for evaluating the magnitude of possible harmful interactions between farmed and wild cod (Cross et al. 2009).

1.8. Aims and objectives of this PhD thesis

This PhD project supports the EIRCOD programme by completing the key genetic aspects of its work programme, specifically the characterisation of wild stocks around Ireland, and monitoring of genetic variability of the farmed stocks throughout their life cycle from egg to harvest, or broodstock. In addition, this thesis develops novel genetic tools and techniques for microsatellite genotyping that are applicaple to a wide range of research questions employing microsatellite markers.

1.8.1. Genetic characterization of aquaculture broodstocks

With few exceptions, aquaculture broodstocks across cultured species are not well characterized genetically (Liu & Cordes 2004). As highlighted earlier, it is recommended that the aquaculture base population should have a broad genetic variation to avoid inbreeding and allow for long-term response to selection (Gjedrem 2005). Therefore, it is important to know how diverse the genetic

backgrounds of aquaculture broodstocks are and how much they differ from their wild counterparts (Liu & Cordes 2004). In addition, genetic variation in the base population and the broodstock must be carefully monitored and maintained in order to minimize inbreeding and maximize variability to retain adaptive potential. One of the aims of this thesis is to increase the knowledge of genetic composition of both the farmed stocks and their wild counterparts.

1.8.2. Developing novel markers and genotyping tools

Many microsatellite and SNP markers currently available for cod originate from EST sequences and are potentially under selection (i.e., Stenvik et al. 2006; Weiss et al. 2007; Westgaard et al. 2007; Higgins et al. 2009; Moen et al. 2008; Hemmer-Hansen et al. 2011). Most population genetics analyses, such as inferences of connectivity, migration rates, are based on the assumption that the loci employed are selectively neutral (Selkoe and Toonen 2006). Therefore, selected markers can bias the estimates (Nielsen et al. 2009b). In addition, it is not possible to distinguish between historical isolation and divergent selection if the markers used to determine genetic structure are subject to divergent selection (Reiss et al. 2009; Nordeide et al. 2011). Many previous studies of population genetics on cod have relied on a small panel of microsatellite markers (i.e. Ruzzante et al. 1998; Hutchinson et al. 2001). With such a limited set of markers, the risk of bias in the parameter estimates is great (Selkoe and Toonen 2006; Nielsen et al. 2009b); using more markers will lead to more accurate parameter estimates, enabling better fisheries management (Selkoe and Toonen 2006).

Therefore, development of more neutral markers for cod is necessary. In order to achieve this, as well as, enhance the genetic resources available for the Celtic Sea cod specifically, next-generation sequencing technology (454) was combined with reduced representation library construction to sequence c. 2% of the cod genome (Carlsson et al. 2013). Over 25 000 potential single-nucleotide polymorphisms (SNP) and >6000 putative microsatellite loci were identified in the resulting sequences (Carlsson et al. 2013). This thesis aimed to develop a large panel of neutral genetic

microsatellite markers from this existing resource and use the panel to investigate the population structuring of the local cod stocks.

1.8.3. More efficient genotyping tools

The aquaculture setting poses specific challenges for using genetic tools as the number of fish to be genotyped can be large and repeated monitoring of the same cohort may be necessary (see Chapter 2). In addition, large panels of markers may be required in order to conclusively resolve the parentage in mass spawning events (Herlin et al. 2007). Hence, medium-to-high-throughput genotyping methods are needed for adequate monitoring of genetic relationships and the level of genetic variability in an aquaculture breeding programme. For genetic markers to become more widely used in aquaculture, fast and economical methods must be developed and are required to employ them.

1.8.4. Objectives and hypotheses

In summary, the objectives of this PhD thesis were:

1. To monitor genetic variability of the farmed stocks in an early-phase breeding programme. This objective is divided into three sub-tasks:

> A) To assess the genetic variability of the wild broodstocks used in the breeding programme.
>
> *Null hypothesis: The wild broodstocks are genetically variable with no major differences between years.*
>
> B) To detect changes in genetic composition of three year classes, from fertilisation up to 3 years of age.
>
> *Null hypothesis: There are no changes in the genetic composition from fertilisation to the time point of sampling.*

C) To resolve parental contributions in a mass spawning event.

Null hypothesis: All parents contribute equally in the production of offspring.

2. To validate new microsatellite markers derived from the Celtic Sea cod stock.

3. To genetically characterise wild cod stocks in the waters surrounding Ireland using the newly-validated markers.

Null hypothesis: There is no genetic structuring in cod in the waters surrounding Ireland.

4. To develop a novel method of genotyping microsatellite markers through the application of next generation sequencing for more efficient application of the markers.

1.9. References

Aotearoa. 2007. "*Gadus morhua*-Atlantic cod" by I, Aotearoa. Licensed under CC BY-SA 3.0 via Commons. https://commons.wikimedia.org/wiki/File:Gadus_morhua-Atlantic_cod.png#/media/File:Gadus_morhua-Atlantic_cod.png.

Ashley, M. V., and Dow, B. D. 2013. The use of microsatellite analysis in population biology: Background, methods and potential applications. Pages 185–201 *in* B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle, editors. Molecular Ecology and Evolution: Approaches and Applications. 2nd edition. Birkhäuser.

Baskerville-Bridges, B., and Kling, L. J. 2000. Early weaning of Atlantic cod (*Gadus morhua*) larvae onto a microparticulate diet. Aquaculture 189:109–117.

Begg, G., and Waldman, J. 1999. An holistic approach to fish stock identification. Fisheries research 43:35–44.

Bekkevold, D., Hansen, M. M., and Nielsen, E. E. 2006. Genetic impact of gadoid culture on wild fish populations: Predictions, lessons from salmonids, and possibilities for minimizing adverse effects. ICES Journal of Marine Science 63:198–208.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, R. K., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M. J., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V, Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M. D., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara E Catenazzi, M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Furey, W. S., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., Hoschler, K., Hurwitz, S., Ivanov, D. V, Johnson, M. Q., James, T., Huw Jones, T. A., Kang, G.-D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ling Ng, B., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Pike, A. C., Chris Pinkard, D., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Chiva Rodriguez, A.,

Roe, P. M., Rogers, J., Rogert Bacigalupo, M. C., Romanov, N., Romieu, A., Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Ernest Sohna Sohna, J., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worsley, G. J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J. C., Hurles, M. E., McCooke, N. J., West, J. S., Oaks, F. L., Lundberg, P. L., Klenerman, D., Durbin, R., and Smith, A. J. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–9.

Bentzen, P., Taggart, C. T., Ruzzante, D. E., and 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.

Bergh, Ø., Nilsen, F., and Samuelsen, O. B. 2001. Diseases, prophylaxis and treatment of the Atlantic halibut *Hippoglossus hippoglossus*: a review. Diseases of Aquatic Organisms 48:57–74.

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., and Willerslev, E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS ONE 2:1–9.

Björnsson, B., and Steinarsson, A. 2002. The food-unlimited growth rate of Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences 59:494–502.

Björnsson, B., Steinarsson, A., and Árnason, T. 2007. Growth model for Atlantic cod (*Gadus morhua*): Effects of temperature and body weight on growth rate. Aquaculture 271:216–226.

Bolton-Warberg, M. 2012. The development of farmed cod in Ireland. PhD thesis. National University of Ireland Galway.

Bolton-Warberg, M., and Fitzgerald, R. D. 2012. Benchmarking growth of farmed Atlantic cod, *Gadus morhua*: A case study in Ireland. Aquaculture Research 43:670–678.

Bolton-Warberg, M., O'Keeffe, D., and FitzGerald, R. D. 2015. Exploring the temperature optima and growth rates of Atlantic cod at the south-easterly limit of its range. Aquaculture Research 46:698–706.

Bourret, V., O'Reilly, P. T., Carr, J. W., Berg, P. R., and Bernatchez, L. 2011. Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. Heredity 106:500–510.

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D. C., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., and 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 of London B: Biological Sciences 277:3725–3734.

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D. C., Borza, T., and Bentzen, P. 2011. Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*. Molecular Ecology Resources 11 Suppl 1:218–25.

Brander, K. 1993. Comparison of spawning characteristics of cod (*Gadus morhua*) stocks in the North Atlantic. NAFO Science Council Studies 18:13–20.

Brander, K. 2007. The role of growth changes in the decline and recovery of North Atlantic cod stocks since 1970. ICES Journal of Marine Science 64:211–217.

Brawn, V. M. 1961. Reproductive behaviour of the cod (*Gadus callarias* L.). Behaviour 18:177–198.

Brockmann, H. J., Colson, T., and Potts, W. 1994. Sperm competition in horseshoe crabs (*Limulus polyphemus*). Behavioral Ecology and Sociobiology 35:153–160.

Brooker, A., and Cook, D. 1994. Organization of microsatellites differs between mammals and cold-water teleost fishes. Canadian Journal of Fisheries and Aquatic Sciences 51:1959–1966.

Brown, J. A., Minkoff, G., and Puvanendran, V. 2003. Larviculture of Atlantic cod (*Gadus morhua*): progress, protocols and problems. Aquaculture 227:357–372.

Bruford, M. W., and Wayne, R. K. 1993. Microsatellites and their application to population genetic studies. Current Opinion in Genetics & Development 3:939–943.

Brumfield, R. T., Beerli, P., Nickerson, D. A., and Edwards, S. V. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18:249–256.

Bundy, A. 2005. Structure and functioning of the eastern Scotian Shelf ecosystem before and after the collapse of groundfish stocks in the early 1990s. Canadian Journal of Fisheries and Aquatic Sciences 62:1453–1473.

Butler, J. M. 2005. High-throughput genotyping of short tandem repeat DNA markers with time-of-flight mass spectrometry. Encyclopedia of Mass Spectrometry 3:1–56.

Cadima, E. L. 2003. Fish stock assessment manual. FAO Fisheries Technical Paper:161.

Carlsson, J., Gauthier, D. T., Carlsson, J. E. L., Coughlan, J. P., Dillane, E., Fitzgerald, R. D., Keating, U., Mcginnity, P., Mirimin, L., and Cross, T. F. 2013. Rapid, economical single-nucleotide polymorphism and microsatellite discovery based on de novo assembly of a reduced representation genome in a non-model organism: A case study of Atlantic cod *Gadus morhu*a. Journal of Fish Biology 82:944–958.

Carr, S. M., and Marshall, H. D. 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 180:381–9.

Chistiakov, D. A., Hellemans, B., and Volckaert, F. A. M. 2006. Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture 255:1–29.

Chowdhury, J., Kagiala, G. V, Pushpakom, S., Lauzon, J., Makin, A., Atrazhev, A., Stickel, A., Newman, W. G., Backhouse, C. J., and Pilarski, L. M. 2007. Microfluidic platform for single nucleotide polymorphism genotyping of the thiopurine S-methyltransferase gene to evaluate risk for adverse drug events. The Journal of Molecular Diagnostics 9:521–529.

Cohen, D. M., Inada, T., Iwamoto, T., and Scialabba, N. 1990. FAO species catalogue. Vol.10. Gadiform fishes of the world (Order Gadiformes). An annotated and illustrated catalogue of cods, hakes, grenadiers and other gadiform fishes known to date. FAO, Rome.

Courchamp, F., Berec, L., and Gascoigne, J. 2008. Allee effects in ecology and conservation. Oxford University Press.

Cross, T. F., McGinnity, P., Carlsson, J., Coughlan, J., Dillane, E., DeEyto, E., Ferguson, A., FitzGerald, R., and Prodöhl, P. 2009. Possible genetic interactions between reared strains and wild populations of cod (*Gadus morhua*): lessons from salmon (*Salmo salar*) common garden field experiments. ICES CM 2009/Q:10.

Dankel, D. J., Skagen, D. W., and Ulltang, Ø. 2008. Fisheries management in practice: review of 13 commercially important fish stocks. Reviews in Fish Biology and Fisheries 18:201–233.

Dean, M. J., Hoffman, W. S., Zemeckis, D. R., and Armstrong, M. P. 2014. Fine-scale diel and gender-based patterns in behaviour of Atlantic cod (*Gadus morhua*) on a spawning ground in the Western Gulf of Maine. ICES Journal of Marine Science 71:1474–1489.

Drinkwater, K. F. 1996. Atmospheric and oceanic variability in the northwest Atlantic during the 1980s and early 1990s. Journal of Northwest Atlantic Fishery Science 18:77–97.

Drinkwater, K. F. 2005. The response of Atlantic cod (*Gadus morhua*) to future climate change. ICES Journal of Marine Science 62:1327–1337.

Dutil, J.-D., and Brander, K. 2003. Comparing productivity of North Atlantic cod (*Gadus morhua*) stocks and limits to growth production. Fisheries Oceanography 12:502–512.

Earll, R. E. 1880. A report on the history and present condition of the shore cod-fisheries of Cape Ann, Mass., together with notes on the natural history and artificial propagation of the species. Report Commissioner of Fish and Fisheries:685–740.

Falconer, D. S., and Mackay, T. F. C. 1996. Introduction to quantitative genetics. Longman, Harlow.

FAO. 2015. Fishery Statistical Collections, Global Production [online] FAO - Fisheries and Aquaculture Information and Statistics Service. Available at: http://www.fao.org/fishery/statistics/global-production/en. [Accessed 24 Aug. 2015]

FEAP. 2016. Production data - FEAP. [online] Feap.info. Available at: http://www.feap.info/Default.asp?SHORTCUT=582 [Accessed 19 Jan. 2016].

Fishbase.org. 2016. Stock information available for Gadus morhua [online] Fishbase.org. Available at: http://www.fishbase.org/summary/stocks/StocksDefinition.php?ID=69&GenusName=Gadus&SpeciesName=morhua&StockCode=79 [Accessed 19 Jan. 2016].

FitzGerald, R. D., Bass, N., Casburn, P., Bolton-Warberg, M., Roche, L., and Watson, L. 2010. An economic assessment of cod farming in Ireland. Report.

Frank, K. T., Petrie, B., Choi, J. S., and Leggett, W. C. 2005. Trophic cascades in a formerly cod-dominated ecosystem. Science 308:1621–1623.

Galvin, P., McGregor, D., and Cross, T. 1995. A single locus PCR amplified minisatellite region as a hypervariable genetic marker in gadoid species. Aquaculture 137:31–40.

Garvin, M. R., Saitoh, K., and Gharrett, A. J. 2010. Application of single nucleotide polymorphisms to non-model species: A technical review. Molecular Ecology Resources 10:915–934.

GATC_Biotech. 2015. [online] GATC Biotech. Available at: https://www.gatc-biotech.com/fileadmin/Kundendaten/Upload_8.0/Dokumente/Lightrun.pdf. [Accessed 24 Aug. 2015]

Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T., and Martin, J.-F. 2011. Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC genomics 12:245.

Gjedrem, T. 2000. Genetic improvement of cold-water fish species. Aquaculture Research 31:25–33.

Gjedrem, T. 2010. The first family-based breeding program in aquaculture. Reviews in Aquaculture 2:2–15.

Gjedrem, T., and Baranski, M. 2009. Selective breeding in aquaculture: an introduction. Springer, New York.

Gjerde, B., Terjesen, B., Barr, Y., Lein, I., and Thorland, I. 2004. Genetic variation for juvenile growth and survival in Atlantic cod (*Gadus morhua*). Aquaculture 236:167–177.

Goetz, F. W., McCauley, L., Goetz, G. W., and Norberg, B. 2006. Using global genome approaches to address problems in cod mariculture. ICES Journal of Marine Science 63:393–399.

Goldstein, D. B., Roemer, G. W., Smith, D. A., Reich, D. E., Bergman, A., and Wayne, R. K. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics 151:797–801.

Guðfinnsson, E. K. 2008. Cod farming in the nordic countries 2008. Conference report.

Gymrek, M., Willems, T., Zeng, H., Markus, B., Daly, M. J., Price, A. L., Pritchard, J., and Erlich, Y. 2015. Abundant contribution of short tandem repeats to gene expression variation in humans. bioRxiv doi: http://dx.doi.org/10.1101/017459. [Early view]

Ha, N.-T., Freytag, S., and Bickeboeller, H. 2014. Coverage and efficiency in current SNP chips. European Journal of Human Genetics 22:1124–1130.

Hardy, R. W. 2010. Utilization of plant proteins in fish diets: Effects of global demand and supplies of fishmeal. Aquaculture Research 41:770–776.

Hemmer-Hansen, J., Nielsen, E. E. G., Meldrup, D., and Mittelholzer, C. 2011. Identification of single nucleotide polymorphisms in candidate genes for growth and reproduction in a nonmodel organism; the Atlantic cod, *Gadus morhua*. Molecular Ecology Resources 11:71–80.

Herbinger, C. M., Doyle, R. W., Taggart, C. T., Lochmann, S. E., Brooker, A. L., Wright, J. M., and Cook, D. 1997. Family relationships and effective population size in a natural cohort of Atlantic cod (*Gadus morhua*) larvae. Canadian Journal of Fisheries and Aquatic Sciences 54:11–18.

Herlin, M., Taggart, J. B., McAndrew, B. J., and Penman, D. J. 2007. Parentage allocation in a complex situation: A large commercial Atlantic cod (*Gadus morhua*) mass spawning tank. Aquaculture 272:S195–S203

Higgins, B., Hubert, S., Simpson, G., Stone, C., and Bowman, S. 2009. Characterization of 155 EST-derived microsatellites from Atlantic cod (*Gadus morhua*) and validation for linkage mapping. Molecular Ecology Resources 9:733–737.

Higgins, R. M., Danilowicz, B. S., Balbuena, J. A., Daníelsdóttir, A. K., Geffen, A. J., Meijer, W. G., Modin, J., Montero, F. E., Pampoulie, C., Perdiguero-Alonso, D., Schreiber, A., Stefánsson, M. Ö., and Wilson, B. 2010. Multi-disciplinary fingerprints reveal the harvest location of cod *Gadus morhua* in the northeast Atlantic. Marine Ecology Progress Series 404:197–206.

Hilborn, R., Branch, T. A., Ernst, B., Magnusson, A., Minte-vera, C. V, Scheuerell, M. D., and Valero, J. L. 2003. State of the world's fisheries. Annual Review of Environment and Resources 28:359–399.

Hilborn, R., and Walters, C. J. 1992. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. First edition. Springer US.

Höffle, H., Solemdal, P., Korsbrekke, K., Johannessen, M., Bakkeplass, K., and Kjesbu, O. S. 2014. Variability of northeast Arctic cod (*Gadus morhua*) distribution on the main spawning grounds in relation to biophysical factors. ICES Journal of Marine Science 71:1317–1331.

Hofreuter, D., Tsai, J., Watson, R. O., Novik, V., Altman, B., Benitez, M., Clark, C., Perbost, C., Jarvie, T., Du, L., and Galán, J. E. 2006. Unique features of a highly pathogenic *Campylobacter jejuni* strain. Infection and Immunity 74:4694–4707.

Holtsmark, M., Klemetsdal, G., Sonesson, A. K., and Woolliams, J. A. 2008. Establishing a base population for a breeding program in aquaculture, from multiple subpopulations, differentiated by genetic drift: I. Effects of the number of subpopulations, heritability and mating strategies using optimum contribution selection. Aquaculture 274:232–240.

Horsted, S. A. 2000. A review of the cod fisheries at Greenland, 1910–1995. Journal of Northwest Atlantic Fishery Science 28:1–112.

Hubert, S., Higgins, B., Borza, T., and Bowman, S. 2010. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). BMC genomics 11:191.

Hutchings, J. A. 2014. Renaissance of a caveat : Allee effects in marine fish. ICES Journal of Marine Science 71:2152–2157.

Hutchings, J. A., and Kuparinen, A. 2014. Ghosts of fisheries-induced depletions: do they haunt us still? ICES Journal of Marine Science 71:1467–1473.

Hutchings, J. A., and Myers, R. A. 1994. What can be learned from the collapse of a renewable resource? Atlantic cod, *Gadus morhua*, of Newfoundland and Labrador. Canadian Journal of Fisheries and Aquatic Sciences 51:2126–2146.

Hutchings, J. A., and Rangeley, R. W. 2011. Correlates of recovery for Canadian Atlantic cod (*Gadus morhua*). Canadian Journal of Zoology 89:386–400.

Hutchinson, W., Carvalho, G., and 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.

Hutchinson, W. F., van Oosterhout, C., Rogers, S. I., and Carvalho, G. R. 2003. Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). Proceedings of the Royal Society of London B: Biological Sciences 270:2125–32.

ICES. 2005. ICES Areas. [online] ICES Available at http://ices.dk/marine-data/maps/Pages/default.aspx. [Accessed 26 Jan. 2012]

Imsland, A. K., Foss, A., Gunnarsson, S., Berntssen, M. H. G., Fitzgerald, R., Bonga, S. W., Nævdal, G., and Stefansson, S. O. 2001. The interaction of temperature and salinity on growth and food conversion in juvenile turbot (*Scophthalmus maximus*). Aquaculture 198:353–367.

Jackson, T. R., Martin-Robichaud, D. J., and Reith, M. E. 2003. Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. Aquaculture 220:245–259.

Jónsdóttir, I. G., Marteinsdottir, G., and Campana, S. E. 2007. Contribution of different spawning components to the mixed stock fishery for cod in Icelandic waters. ICES Journal of Marine Science 64:1749–1759.

Jørgensen, C., Dunlop, E. S., Opdal, A. F., and Fiksen, O. 2008. The evolution of spawning migrations: state dependence and fishing-induced changes. Ecology 89:3436–3448.

Jørstad, K. E., Fjalestad, K. T., Ágústsson, T., and Marteinsdottir, G. 2007. Genetic effects of domestication, culture and breeding of fish and shellfish, and their impacts on wild populations. Atlantic cod – *Gadus morhua*. (T. Svåsand, D. Crosetti, E. García-Vázquez, and E. Verspoor, Eds.) Genetic impact of aquaculture activities on native populations. Genimpact final scientific report (EU contract n. RICA-CT-2005-022802).

Jørstad, K. E., van der Meeren, T., Paulsen, O. I., Thomsen, T., Thorsen, A., and Svåsand, T. 2008. "Escapes" of eggs from farmed cod spawning in net pens: recruitment to wild stocks. Reviews in Fisheries Science 16:285–295.

Jørstad, K. E., Otterå, H., Van Der Meeren, T., Dahle, G., Paulsen, O. I., Bakke, G., and Svåsand, T. 2013. Genetic marking of farmed Atlantic cod (*Gadus morhua* l.) and detection of escapes from a commercial cod farm. ICES Journal of Marine Science 71:574–584.

Karlsen, Norberg, B., Kjesbu, O. S., and Taranger, G. L. 2006. Effects of photoperiod and exercise on growth, liver size, and age at puberty in farmed Atlantic cod (*Gadus morhua* L.). ICES Journal of Marine Science 63:355–364.

Karlsen, Ø., van der Meeren, T., Rønnestad, I., Mangor-Jensen, A., Galloway, T. F., Kjørsvik, E., and Hamre, K. 2015. Copepods enhance nutritional status, growth and development in Atlantic cod (*Gadus morhua* L.) larvae - can we identify the underlying factors? PeerJ 3:e902.

Keith, C. S., Hoang, D. O., Barrett, B. M., Feigelman, B., Nelson, M. C., Thai, H., and Baysdorfer, C. 1993. Partial sequence analysis of 130 randomly selected maize cDNA clones. Plant physiology 101:329–332.

Kelly, C. J., Codling, E. A., and Rogan, E. 2006. The Irish Sea cod recovery plan: some lessons learned. ICES Journal of Marine Science 63:600–610.

Kjesbu, O. S. 1989. The spawning activity of cod, *Gadus morhu*a L. Journal of Fish Biology 34:195–206.

Kjesbu, O. S., Taranger, G. L., and Trippel, E. A. 2006. Gadoid mariculture: Development and future challenges. ICES Journal of Marine Science 63:187–191.

Knapik, E. W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M. C., and Jacob, H. J. 1998. A microsatellite genetic linkage map for zebrafish (*Danio rerio*). Nature Genetics 18:338–343.

Knutsen, H., Jorde, P. E., André, C., and Stenseth, N. C. 2003. Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. Molecular Ecology 12:385–94.

Koedijk, R. M., Folkvord, A., Foss, A., Pittman, K., Stefansson, S. O., Handeland, S., and Imsland, A. K. 2010. The influence of first-feeding diet on the Atlantic cod *Gadus*

morhua phenotype: Survival, development and long-term consequences for growth. Journal of Fish Biology 77:1–19.

Koljonen, M. L. 2001. Conservation goals and fisheries management units for Atlantic salmon in the Baltic Sea area. Journal of Fish Biology 59:269–288.

Kolstad, K., Thorland, I., Refstie, T., and Gjerde, B. 2006. Body weight, sexual maturity, and spinal deformity in strains and families of Atlantic cod (*Gadus morhu*a) at two years of age at different locations along the Norwegian coast. ICES Journal of Marine Science 63:246–252.

Kristjánsson, T., and Arnason, T. 2014. Strong phenotypic and genetic correlation between size and first maturity in Atlantic cod *Gadus morhua* L. reared in commercial conditions. Aquaculture Research 46:2185–219

Kumar, S., Banks, T. W., and Cloutier, S. 2012. SNP discovery through next-generation sequencing and its applications. International Journal of Plant Genomics 2012. Article ID 831460, 15 pages

Kurlansky, M. 1997. Cod: a biography of the fish that changed the world. Walker Publishing Company, Inc.

LaFramboise, T. 2009. Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. Nucleic Acids Research 37:4181–4193.

Lande, R., and Thompson, R. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756.

Lee, C., Chiu, Y., Wang, L., Kuo, Y., Chuang, E. Y., Lai, L.-C., and Tsai, M. 2013. Common applications of next-generation sequencing technologies in genomic research. Translational Cancer Research 2:33–45.

Lilly, G. R., Wieland, K., Rothschild, B. J., Sundby, S., Drinkwater, K. F., Brander, K. M., Ottersen, G., Carscadden, J. E., Stenson, G. B., Chouinard, G. A., Swain, D. A., Daan, N., Enberg, K., Hammill, M. O., Rosing-Asvid, A., Svedäng, H., and Vázquez, A. 2008. Decline and recovery of Atlantic cod (*Gadus morhua*) stocks throughout the North Atlantic. Resiliency of gadid stocks to fishing and climate change, Alaska Sea Grant College Program 39:39–66.

Liu, Z. J., and Cordes, J. F. 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238:1–37.

Mardis, E. R. 2013. Next-generation sequencing platforms. Annual Review of Analytical Chemistry 6:287–303.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V, Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K.

E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., and Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–80.

Marine Institute. 2013. Annual 2013 review of fish stocks with management advice for 2014. Marine Institute Ireland

Marteinsdottir, G., and Björnsson, H. 1999. Time and duration of spawning of cod in Icelandic waters. ICES CM 1999/Y:34 Cod and Haddock recruitment processes - integrating stock and environmental effects.

Massault, C., Bovenhuis, H., Haley, C., and de Koning, D. J. 2008. QTL mapping designs for aquaculture. Aquaculture 285:23–29.

Massault, C., Hellemans, B., Louro, B., Batargias, C., Van Houdt, J. K. J., Canario, A., Volckaert, F. A. M., Bovenhuis, H., Haley, C., and De Koning, D. J. 2010. QTL for body weight, morphometric traits and stress response in European sea bass *Dicentrarchus labrax*. Animal Genetics 41:337–345.

Miller, D. D., and Mariani, S. 2010. Smoke, mirrors, and mislabeled cod: poor transparency in the European seafood industry. Frontiers in Ecology and the Environment 8:517–521.

Milton, D. A., and Shaklee, J. B. 1987. Biochemical genetics and population structure of blue grenadier, *Macruronus novaezelandiae* (Hector) (Pisces : Merluccidae), from Australian waters. Marine and Freshwater Research 38:727–742.

Mirimin, L., O'Keeffe, D., Ruggiero, A., Bolton-Warberg, M., Vartia, S., and Fitzgerald, R. 2011. A quick, least-invasive, inexpensive and reliable method for sampling *Gadus morhua* postlarvae for genetic analysis. Journal of Fish Biology 79:801–805.

Moe, H., Dempster, T., Sunde, L. M., Winther, U., and Fredheim, A. 2007. Technological solutions and operational measures to prevent escapes of Atlantic cod (*Gadus morhua*) from sea cages. Aquaculture Research 38:91–99.

Moen, T., Delghandi, M., Wesmajervi, M. S., Westgaard, J.-I., and Fjalestad, K. T. 2009. A SNP/microsatellite genetic linkage map of the Atlantic cod (*Gadus morhua*). Animal Genetics 40:993–6.

Moen, T., Hayes, B., Nilsen, F., Delghandi, M., Fjalestad, K. T., Fevolden, S.-E., Berg, P. R., and Lien, S. 2008. Identification and characterisation of novel SNP markers in Atlantic cod: evidence for directional selection. BMC Genetics 9:18.

Moksness, E., Kjørsvik, E., and Olsen, Y. 2004. Introduction. Pages 1–7 *in* E. Moksness, E. Kjørsvik, and Y. Olsen, editors. Culture of cold-water marine fish. Blackwell Publishing Ltd, Bath.

Morais, S., Edvardsen, R. B., Tocher, D. R., and Bell, J. G. 2012. Transcriptomic analyses of intestinal gene expression of juvenile Atlantic cod (*Gadus morhua*) fed diets with Camelina oil as replacement for fish oil. Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology 161:283–293.

Moretti, A., Fernandez-Criado, M. P., Cittolin, G., and Guidastri R. 1999. Manual on hatchery production of seabass and gilthead seabream, Volume 1. Food and Agriculture Organization of the United Nations.

Morgan, M. J., Shelton, P. A., and Rideout, R. M. 2014. An evaluation of fishing mortality reference points under varying levels of population productivity in three Atlantic cod (*Gadus morhua*) stocks. ICES Journal of Marine Science 71:1407–1416.

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

Myers, R. A., MacKenzie, B. R., Bowen, K. G., and Barrowman, N. J. 2001. What is the carrying capacity for fish in the ocean? A meta-analysis of population dynamics of North Atlantic cod. Canadian Journal of Fisheries and Aquatic Sciences 58:1464–1476.

Naylor, R. L., Hardy, R. W., Bureau, D. P., Chiu, A., Elliott, M., Farrell, A. P., Forster, I., Gatlin, D. M., Goldburg, R. J., Hua, K., and Nichols, P. D. 2009. Feeding aquaculture in an era of finite resources. Proceedings of the National Academy of Sciences of the United States of America 106:15103–15110.

Neat, F. C., Bendall, V., Berx, B., Wright, P. J., Ó Cuaig, M., Townhill, B., Schön, P.-J., Lee, J., and Righton, D. 2014. Movement of Atlantic cod around the British Isles: implications for finer scale stock management. Journal of Applied Ecology 51:1564–1574.

Neuenfeldt, S., Righton, D., Neat, F., Wright, P. J., Svedäng, H., Michalsen, K., Subbey, S., Steingrund, P., Thorsteinsson, V., Pampoulie, C., Andersen, K. H., Pedersen, M. W., and Metcalfe, J. 2013. Analysing migrations of Atlantic cod *Gadus morhu*a in the north-east Atlantic Ocean: then, now and the future. Journal of Fish Biology 82:741–63.

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

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

Nielsen, E. E., MacKenzie, B. R., Magnussen, E., and Meldrup, D. 2007. Historical analysis of Pan I in Atlantic cod (*Gadus morhua*): temporal stability of allele frequencies in the southeastern part of the species distribution. Canadian Journal of Fisheries and Aquatic Sciences 64:1448–1455.

Nielsen, E., Wright, P., Hemmer-Hansen, J., Poulsen, N., Gibb, I. M., and Meldrup, D. 2009b. Microgeographical population structure of cod *Gadus morhua* in the North Sea and west of Scotland: the role of sampling loci and individuals. Marine Ecology Progress Series 376:213–225.

Noguchi, H., Park, J., and Takagi, T. 2006. MetaGene: Prokaryotic gene finding from environmental genome shotgun sequences. Nucleic Acids Research 34:5623–5630.

Nordeide, J. T., Johansen, S. D., Jørgensen, T. E., Karlsen, B. O., and Moum, T. 2011. Population connectivity among migratory and stationary cod *Gadus morhua* in the Northeast Atlantic - A review of 80 years of study. Marine Ecology Progress Series 435:269–283.

O'Leary, D. B., Coughlan, J., Dillane, E., McCarthy, T. V., and Cross, T. F. 2007. Microsatellite variation in cod *Gadus morhua* throughout its geographic range. Journal of Fish Biology 70:310–335.

O'Leary, D. B., Coughlan, J., McCarthy, T. V., and Cross, T. F. 2006. Application of a rapid method of SNP analysis (glycosylase mediated polymorphism detection) to mtDNA and nuclear DNA of cod *Gadus morhua*. Journal of Fish Biology 69:145–153.

Otterlei, E., Nyhammer, G., Folkvord, A., and Stefansson, S. O. 1999. Temperature-and size-dependent growth of larval and early juvenile Atlantic cod (*Gadus morhua*): a comparative study of Norwegian coastal cod and northeast Arctic. Canadian Journal of Fisheries and Aquatic Sciences 56:2099–2111.

Owens, H. L. 2015. Evolution of codfishes (Teleostei: Gadinae) in geographical and ecological space: evidence that physiological limits drove diversification of subarctic fishes. Journal of Biogeography 42:1091–1102

Paisley, L. G., Ariel, E., Lyngstad, T., Jónsson, G., Vennerström, P., Hellström, A., and Østergaard, P. 2010. An overview of aquaculture in the nordic countries. Journal of the World Aquaculture Society 41:1–17.

Pampoulie, C., Stefánsson, M. Ö., Jörundsdóttir, T. D., Danilowicz, B. S., and Daníelsdóttir, A. K. 2008. Recolonization history and large-scale dispersal in the open sea: the case study of the North Atlantic cod, *Gadus morhua* L. Biological Journal of the Linnean Society 94:315–329.

Pinfold, G. 2010. Factors affecting viability of cod aquaculture in Canada. Genome Atlantic. Report.

Pitterl, F., Niederstätter, H., Huber, G., Zimmermann, B., Oberacher, H., and Parson, W. 2008. The next generation of DNA profiling - STR typing by multiplexed PCR - ion-pair RP LC-ESI time-of-flight MS. Electrophoresis 29:4739–4750.

Pogson, G. H., Mesa, K. A., and Boutilier, R. G. 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.

Pogson, G. H., Taggart, C. T., Mesa, K. A., and Boutilier, R. G. 2001. Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. Evolution 55:131–146.

Primmer, C. R., Koskinen, M. T., and Piironen, J. 2000. The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. Proceedings of the Royal Society B: Biological Sciences 267:1699–1704.

Reid, H. I., Treasurer, J. W., Adam, B., and Birkbeck, T. H. 2009. Analysis of bacterial populations in the gut of developing cod larvae and identification of *Vibrio logei*, *Vibrio anguillarum* and *Vibrio splendidus* as pathogens of cod larvae. Aquaculture 288:36–43.

Reiss, H., Hoarau, G., Dickey-Collas, M., and Wolff, W. J. 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish and Fisheries 10:361–395.

Righton, D. A., Andersen, K. H., Neat, F., Thorsteinsson, V., Steingrund, P., Svedäng, H., Michalsen, K., Hinrichsen, H. H., Bendall, V., Neuenfeldt, S., Wright, P., Jonsson, P., Huse, G., Van Der Kooij, J., Mosegaard, H., Hüssy, K., and Metcalfe, J. 2010. Thermal niche of Atlantic cod *Gadus morhua*: Limits, tolerance and optima. Marine Ecology Progress Series 420:1–13.

Rise, M. L., Gamperl, A. K., Johnson, S. C., Robinson, A., Culver, K., Booman, M., Borza, T., Garber, A., Hubert, S., Tosh, J., Alfonso, L. O. B., Boyce, D., Hamoutene, D., Lush, L., Neill, S., Symonds, J., Butts, I. A. E., Feng, C. Y., Hori, T. S., Perez-Casanova, J., Shine, E., Hall, J. R., Higgins, B., Kimball, J., Simpson, G., Stone, C., Tarrant-Bussey, J., Elliot, J., Moir, J., Nardi, G., Powell, F., Walsh, A., Rise, M., Trippel, E., and Bowman, S. 2009. The Atlantic Cod Genomics and Broodstock Development Project. Bulletin of the Aquaculture Association of Canada 107:21–30.

Robichaud, D., and Rose, G. A. 2003. Sex differences in cod residency on a spawning ground. Fisheries Research 60:33–43.

Robins, C. R., Ray, G. C., and Douglas, J. 1986. A field guide to Atlantic coast fishes. The Peterson field guide series. Houghton Mifflin Co., Boston.

Rose, G. A. 2004. Reconciling overfishing and climate change with stock dynamics of Atlantic cod (*Gadus morhua*) over 500 years. Canadian Journal of Fisheries and Aquatic Sciences 61:1553–1557.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J.,

Williams, A., Roth, G. T., and Bustillo, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348–352.

Rowe, S., and Hutchings, J. a. 2003. Mating systems and the conservation of commercially exploited marine fish. Trends in Ecology and Evolution 18:567–572.

Ruane, N. M., Bolton-Warberg, M., Rodger, H. D., Colquhoun, D. J., Geary, M., Mccleary, S. J., O'Halloran, K., Maher, K., O'Keeffe, D., Mirimin, L., Henshilwood, K., Geoghegan, F., and Fitzgerald, R. D. 2013. An outbreak of francisellosis in wild-caught Celtic Sea Atlantic cod, *Gadus morhua* L., juveniles reared in captivity. Journal of Fish Diseases 38:97–102

Ruzzante, D. E., Taggart, C. T., and 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.

Ruzzante, D. E., Wroblewski, J. S., Taggart, C. T., Smedbol, R. K., Cook, D., and Goddard, S. V. 2000. Bay-scale population structure in coastal Atlantic cod in Labrador and Newfoundland, Canada. Journal of Fish Biology 56:431–447.

Salvanes, A. G. V., Skjæraasen, J. E., and Nilsen, T. 2004. Sub-populations of coastal cod with different behaviour and life-history strategies. Marine Ecology Progress Series 267:241–251.

Samuelsen, O. B., Nerland, A. H., Jørgensen, T., Schrøder, M. B., Svåsand, T., and Bergh, Ø. 2006. Viral and bacterial diseases of Atlantic cod *Gadus morhua*, their prophylaxis and treatment: A review. Diseases of Aquatic Organisms 71:239–254.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74:5463–5467.

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L. W. 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Molecular Ecology Resources 11:1–8.

Selkoe, K. A., and Toonen, R. J. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters 9:615–629.

Shields, R. J. 2001. Larviculture of marine finfish in Europe. Aquaculture 200:55–88.

Skaarud, A., Woolliams, J. A., and Gjøen, H. M. 2011. Strategies for controlling inbreeding in fish breeding programs; an applied approach using optimum contribution (OC) procedures. Aquaculture 311:110–114.

Skjæraasen, J. E., Meager, J. J., Karlsen, Ø., Hutchings, J. A., and Fernö, A. 2011. Extreme spawning-site fidelity in Atlantic cod. ICES Journal of Marine Science 68:1472–1477.

Smith, P. J., Jamieson, A., and Birley, A. J. 1990. Electrophoretic studies and the stock concept in marine teleosts. Journal du Conseil: ICES Journal of Marine Science 47 :231–245.

Sonesson, A. K., Woolliams, J. A., and Meuwissen, T. H. E. 2005. Kinship, relationship and inbreeding. Pages 73–87 *in* T. Gjedrem, editor. Selection and Breeding Programs in Aquaculture. Springer, New York.

Sonesson, A. K. 2007. Within-family marker-assisted selection for aquaculture species. Genetics Selection Evolution 39:301–317.

Speare, D., MacNair, N., and Hammell, K. 1995. Demonstration of tank effect on growth indices of juvenile rainbow trout (*Oncorhynchus mykiss*) during an ad libitum feeding trial. American Journal of Veterinary Research 56:1372–9.

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., Sharma, A., Wetten, O. F., Lanzén, A., Winer, R., Knight, J., Vogel, J.-H., Aken, B., Andersen, O., Lagesen, K., Tooming-Klunderud, A., Edvardsen, R. B., Tina, K. G., Espelund, M., Nepal, C., Previti, C., Karlsen, B. O., Moum, T., Skage, M., Berg, P. R., Gjøen, T., Kuhl, H., Thorsen, J., Malde, K., Reinhardt, R., Du, L., Johansen, S. D., Searle, S., Lien, S., Nilsen, F., Jonassen, I., Omholt, S. W., Stenseth, N. C., and Jakobsen, K. S. 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210.

Stenvik, J., Wesmajervi, M. S., Fjalestad, K. T., Damsgård, B., and Delghandi, M. 2006. Development of 25 gene-associated microsatellite markers of Atlantic cod (*Gadus morhua* L.). Molecular Ecology Notes 6:1105–1107.

Strub, A. M., and Pauly, D. 2011. Atlantic cod: Past and present. The Sea Around Us Project Newsletter March/April:2–5.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17:6463–6471.

Teletchea, F. 2009. Molecular identification methods of fish species: Reassessment and possible applications. Reviews in Fish Biology and Fisheries 19:265–293.

Thorsen, A., Witthames, P. R., Marteinsdóttir, G., Nash, R. D. M., and Kjesbu, O. S. 2010. Fecundity and growth of Atlantic cod (*Gadus morhua* L.) along a latitudinal gradient. Fisheries Research 104:45–55.

Trippel, E. A., Benfey, T. J., Steven, E., and Blanchard, M. J. 2008. Effects of continuous light and triploidy on growth and sexual maturation in Atlantic cod, *Gadus morhua*. Cybium: International Journal of Ichthyology 32:136–138.

Utter, F. 1998. Genetic problems of hatchery-reared progeny released into the wild, and how to deal with them. Bulletin of Marine Science 62:623–640.

Wang, J. 2009. A new method for estimating effective population sizes from a single sample of multilocus genotypes. Molecular Ecology 18:2148–2164.

Waples, R. S., and Waples, R. K. 2011. Inbreeding effective population size and parentage analysis without parents. Molecular Ecology Resources 11:162–171.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. N. 2005. DNA barcoding Australia's fish species. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 360:1847–1857.

Weiss, E., Bennie, M., Hodgins-Davis, A., Roberts, S., and Gerlach, G. 2007. Characterization of new SSR-EST markers in cod, *Gadus morhua*. Molecular Ecology Notes 7:866–867.

Wennevik, V., Jørstad, K. E., Dahle, G., and Fevolden, S. E. 2008. Mixed stock analysis and the power of different classes of molecular markers in discriminating coastal and oceanic Atlantic cod (*Gadus morhua* L.) on the Lofoten spawning grounds, Northern Norway. Hydrobiologia 606:7–25.

Wesmajervi, M. S., Westgaard, J. I., and Delghandi, M. 2006. Evaluation of a novel pentaplex microsatellite marker system for paternity studies in Atlantic cod (*Gadus morhua* L.). Aquaculture Research 37:1195–1201.

Westgaard, J.-I., Tafese, T., Wesmajervi, M. S., Nilsen, F., Fjalestad, K. T., Damsgård, B., and Delghandi, M. 2007. Development of ten new EST-derived microsatellites in Atlantic cod (*Gadus morhua* L.). Conservation Genetics 8:1503–1506.

Wheeler, A. 1978. Key to fishes of northern Europe. Frederick Warne Publishers Ltd, London.

Zemeckis, D. R., Hoffman, W. S., Dean, M. J., Armstrong, M. P., and Cadrin, S. X. 2014. Spawning site fidelity by Atlantic cod (*Gadus morhua*) in the Gulf of Maine : implications for population structure and rebuilding. ICES Journal of Marine Science 71:1356–1365.

Chapter 2. Monitoring genetic variation in the EIRCOD breeding programme

2.1. Introduction

In the early 2000s, following a broad trend across Northern hemisphere countries, Ireland commenced aquaculture trials with cod, *Gadus morhua*. In 2004, the Ryan Institute (RI, National University of Ireland Galway (NUIG)) established a marine finfish hatchery at its Carna Research Station and began a development programme to rear cod from imported fertilized eggs for an emerging Irish cod aquaculture industry. Once suitable size was reached, these juvenile fish were subsequently transferred to sea cages for on-growing until harvest size at a pilot commercial farm, Trosc Teo. These early initiatives in cod farming were supported by a series of strategic grants provided by the key State Agencies, including the Marine Institute, Bord Iascaigh Mhara (BIM) and Údarás na Gaeltachta, as part of wider aquaculture diversification programmes with industry and the Third Level.

In 2007/8, a consortium led by NUIG developed a proposal for a National Broodstock and Breeding Programme for Atlantic cod in Ireland (EIRCOD) and this was funded under the ***Sea Change*** initiative with the support of the Marine Institute and the Marine Research Sub-programme of the National Development Plan 2007-2013 co-funded by the European Regional Development Fund, with the objective of facilitating cod aquaculture in Ireland. Before EIRCOD (2004 to 2007), Ryan Institute sourced ova solely from foreign producers. In contrast, from 2008 onwards, EIRCOD began deliberately acquiring native Irish cod ova through egg collection cruises in the Celtic Sea during the spawning season to establish the base population of the breeding programme.

Of the regional cod stocks in Irish waters, the Celtic Sea stock was chosen as the 'founder' or source population due to its productivity, sustainability and relative ease of access to eggs. This reflects the trend, as implemented in other countries developing cod aquaculture, that local stocks should be employed as the primary source population for a local industry. Celtic Sea stock, along with the other southern, Irish Sea and West of Scotland stocks in Irish coastal waters, have also

been identified as having the highest levels of production in comparison to 15 other stocks across the species' distribution range (Dutil and Brander 2003). The state of the Celtic Sea stock is sustainable according to ICES management recommendations and fishing mortality is at maximum sustainable yield (Marine Institute 2013). In contrast, the Irish Sea and West of Scotland stocks are in a collapsed state with zero TAC assigned (Marine Institute 2013).

2.1.1. Broodstock acquisition and management

In selecting an indigenous native stock, it is held, at its simplest, that such local populations should be better adapted to prevailing, ambient, environmental conditions and various biotic influences and, thereby, make them an optimal choice for establishing a bespoke farmed base broodstock (Salvanes et al. 2004; Bekkevold et al. 2006; Nielsen et al. 2009). Likewise, choosing a local population as the base population for aquaculture may reduce future negative effects on the wild population in case escapes occur from the sea cages (Bekkevold et al. 2006). Conversely, if fish from geographically distant populations breed with fish from the local population, it is believed that this can result in introgression and can lead to outbreeding depression (Bekkevold et al. 2006). Outbreeding depression happens when offspring from a cross between individuals from different populations have lower fitness than offspring from a cross between individuals belonging to the same population (Frankham et al. 2010). Two consequences of outbreeding can cause outbreeding depression. Firstly, offspring can display a phenotype intermediate to that of the parents, the intermediate phenotype being suboptimal in either population's environment (Gilk et al. 2004). Secondly, introgression can disrupt distinct co-adapted gene complexes related to diverse fitness traits such as homing ability in salmon (Gilk et al. 2004). In addition to the risk of escapes, translocations of organisms from geographically distant areas risk introducing pathogens if strict quarantine procedures are not followed.

In theoretical terms, it is recommended that an aquaculture base broodstock population should have a broad genetic variation to avoid inbreeding and allow for

long-term response to selection (Gjedrem 2005). Caution should be used when choosing the source population for establishing the base stock. It has been shown, for example, that genetic variation of collapsed populations may be reduced. If founders are sourced from a collapsed population with reduced genetic variability, the amount of genetic variability can be lower than expected even if a large number of founders are acquired and used. For instance, with cod, a reduction in genetic variation has been observed in the North Sea stock, but no such reduction was detected in the Gulf of St. Lawrence stock (Hutchinson et al. 2003; Therkildsen et al. 2010). The amount of genetic variation can be increased if base populations are founded from several sources, however if source populations are very differentiated, this can cause outbreeding depression in the offspring (Gilk et al. 2004) as highlighted above. Differences in life-histories or genetic variation in neutral molecular markers between populations can give some indication if outbreeding depression is a likely possibility, however, creating crossings between distinct populations and testing their performance is the only reliable way to ascertain the suitability of crosses.

Norway is one the pioneers of modern cod aquaculture and provides some well-described and salutary examples. A minimum of two distinct cod stocks are recognised in Norwegian waters: Norwegian coastal cod (NC) and northeast Arctic cod (NA). The NC and NA strains have different life-histories and genetic differences have been observed between the stocks, therefore, these stocks have generally been managed as two separate (breeding) strains in aquaculture (Gjerde et al. 2004). However, variation or differences in neutral genetic markers are not indicators of performance under farming conditions, and performance of stocks can only be assessed through evaluation in the farm environment (Kolstad et al. 2006). Significantly, differences in larval growth between the NC and NA stocks have been recorded under common rearing conditions (Otterlei et al. 1999). However, a study with these same stocks (NC and NA), but different original, parental, broodstock, detected no significant differences in either body weight or proportion of spinal deformities at harvest between the two stocks reared at different locations off the coast of Norway; this may indicate that the two populations could be merged

(Kolstad et al. 2006). However, maintaining separate breeding programmes for the two strains could be beneficial for a plethora of other reasons, even consumer acceptance (Kolstad et al. 2006).

At the start of a breeding programme, repeated collection of material (eggs, juveniles and adults) from the wild may be needed to create a broad genetic profile in the base population. However, continued reliance on such wild materials for the ongoing propagation of the farmed stock will limit progress in terms of maintaining unwanted traits in the aquaculture line and contributing to market unpredictability as the gamete supply depends on success of, for example, egg collection cruises (Garber et al. 2010). Replacing wild-caught broodstock with domesticated broodstock provides several advantages as selection can be applied to improve specific desired traits, e.g. growth rate and feed conversion efficiencies (Gjedrem 2000). In addition, gamete supply will be more reliable as the wild collections are dependent on wild stock events and annual variations in spawning, recruitment and survival and the success of collection cruises can be highly unpredictable due to episodic weather conditions. Moreover, domestication often increases tameness. As a result, the domesticated fish experience less handling stress allowing for better animal welfare (Garber et al. 2010; Gjedrem 2000).

2.1.2. Fertilization and production methods

While synchronised spawning is readily realisable with, for example, many salmonid species, it is not possible with most marine finfish species which tend to be group-synchronous with only a minor proportion of breeders being active simultaneously and producing discrete egg batches (Murua & Saborida-Rey 2003). As a consequence, most commercial cod hatcheries rely on mass spawning to produce offspring (Herlin et al. 2008). Mass spawning is the simultaneous spawning of a group of males and females in a common environment such as the same tank or cage pen (Gjedrem and Baranski 2009). In the wild, cod naturally form spawning aggregations near the sea floor with males establishing territories and females approaching a male of their choice once ready to spawn (Robichaud & Rose 2003;

Dean et al. 2014). The pair engages in courtship which culminates in a ventral mount and, while the pair swims, they simultaneously release the eggs and the milt (Brawn 1961; Grabowski et al. 2012). Less dominant males may also release their sperm once the couple is spawning, in an attempt to contribute to the fertilisation of the female's eggs (Grabowski et al. 2012). Similar behaviour has been recorded in enclosed spawning tanks/cages (Hutchings et al. 1999; Herlin et al. 2008). The reproductive success of individuals, in communal mass spawning events of cod, can be influenced by size of the individual, aggression of males and female mate-choice (Hutchings et al. 1999; Bekkevold et al 2002). This can lead to disproportionate contribution of eggs and sperm by a few individuals. Such uneven contributions will limit the genetic composition and variability in the offspring. Hence, molecular monitoring of genetic variation is recommended to avoid drastic loss of genetic variability in such mass spawning events (Armitage et al. 2007).

Despite the intricate breeding behaviour of cod, there has been success with natural paired matings, where a male and a female are placed in a tank and allowed to spawn naturally (Gjerde et al. 2004, Lush et al. 2011). A downside to using natural pair matings is that it may take longer time to produce a sufficient number of offspring (Lush et al. 2011). This can lead to large age differences in offspring which may expose the younger (and smaller) offspring to increased cannibalism if mixed with other batches (Garber et al. 2010). In addition, practical technical, operational or economic factors, such as available separate tank units for setting up several paired matings can reduce the number of breeders when using this method.

The third option for acquiring fertilised eggs from the broodstock is (manual) strip spawning. In strip spawning, gametes are manually removed from the fish by applying firm pressure to the ventral surface of the animal. The resulting eggs and milt are combined in a sterile container and fertilization occurs *in vitro* (Hamoutene et al. 2009). This allows for complete control over the family crosses, but involves a risk of causing damage to the broodstock fish. For example, haddock (*Melanogrammus aeglefinus*) easily loses scales and suffers damage to skin and fins

as a result of handling (Thorsen et al. 2003). Cod is slightly more resistant to handling stress, however, enlarged ovary and thin ovarian membrane makes the females more susceptible to internal damage during spawning season (Thorsen et al. 2003). The use of strip spawning can be further complicated by sensitivity in timing of stripping (Kjesbu et al. 2006). Wrong timing (late) can lead to over-ripening of eggs which reduces egg quality necessitating monitoring of individual ovulatory rhythms to achieve the optimal timing for stripping (Kjesbu et al. 2006; Lush et al. 2011).

2.1.3. Monitoring of family survival

Many selective breeding programmes use the paired mating strategy described above and rear the resulting families separately until the fish are large enough to be marked with a physical tag, such as passive integrated transponder (PIT) tags (Gjerde et al. 2004; Hubert et al. 2010). PIT tags are small electronic microchips that can be injected subcutaneously or into the body cavity of the fish allowing for individual identification (Gibbons and Andrews 2004). Separate rearing of families requires one tank per family, hence, the number of tanks can be a practical limitation for number of families within a year-class. In addition, if each family is reared in a separate tank, it is problematic to compare performance between families as family effects are confounded by different environmental effects (i.e. tank effect) (Speare et al. 1995). However, these limitations can be overcome with an alternative approach of rearing families communally, and determining in due course the parentage using genetic markers.

Parentage analysis

Molecular parentage analysis can be divided into two main categories: exclusion methods and categorical allocation methods (Herlin et al. 2007) each of which comes with specific advantages and caveats. Exclusion methods assign progeny to parents by excluding other potential parents based on incompatible genotypes

given the rules of Mendelian inheritance. Exclusion methods are very sensitive to genotyping errors and these errors can cause incompatible genotypes that lead to the exclusion of the true parents. Some parentage analysis software programmes using the exclusion method take this into account by having the option of allowing for a certain number of mismatches between offspring and parent (Jones et al. 2010). Exclusion is considered the gold standard of parentage studies. However, it may be difficult to achieve complete exclusion if the number of offspring and the putative parent group is large.

Categorical allocation methods use likelihood methods to assign parentage to the most likely parent pair. Different parental genotypes can have different probabilities of having produced the genotype of the offspring based on Mendelian transition probabilities (Meagher & Thompson 1986; Marshall et al. 1998). Categorical allocation methods provide a way to choose the single most likely parent from a group of putative parents. If genotyping errors, or mutations, are present in the data set, this can be accommodated by modifying the transition probabilities accordingly (Marshall et al. 1998). Even though categorical allocation methods are popular in parentage analysis, they should be used with caution because it is difficult to predict how the relationships between the mathematical models employed, and the error level set by the user, affect the resultant sensitivity and accuracy of the assignments (Herlin et al. 2007).

Parental reconstruction

Parental reconstruction is a simple way to deduce parental genotypes from offspring data without actual access to parental fish. In parental reconstruction, the genotypes of offspring from full sib or half sib families are used to reconstruct the genotypes of the parents (Jones et al. 2010). The approach gives most accurate results if some of the parental genotypes are known (Jones 2001). However, it can also be performed on a group of offspring without any knowledge of parents (Smith et al. 2001). The offspring are first allocated into sibling groups based on their

genotypes and then the genotypes of the potential parents are reconstructed based on the offspring genotypes (Jones 2001). Especially, if all parental genotypes are unknown, the parental reconstruction relies on equal representation of families in the offspring pool (Smith et al. 2001).

Molecular parentage assignment, and parental reconstruction, require genetic markers with high levels of genetic variation at the individual level. Microsatellites display extremely high levels of genetic variation among individuals compared to other markers, making them especially suitable for parentage analysis (Liu and Cordes 2004). Microsatellites are routinely used for parentage analysis in many fish species both in natural populations and in captive systems, such as aquaculture stocks (Liu and Cordes 2004). In cod aquaculture, parentage analysis has been applied primarily in communally reared fish and in mass spawning events using this marker type (Wesmajervi et al. 2006; Armitage et al. 2007; Herlin et al. 2007; Herlin et al. 2008; Garber et al. 2010).

2.1.4. Detrimental or complicating factors

Most aquaculture species, cod in particular, have a very high fecundity (Gjedrem 2005). As a result, a small number of broodstock can be sufficient to breed enough offspring for the next generation. Due to pragmatic considerations, such as available space, it can seem practical to establish a breeding programme with a small number of founders, and keep a limited number of breeders to produce the following generations. However, a small number of breeders can lead to rapid accumulation of inbreeding leading, in turn, to inbreeding depression (Gjedrem 2005; Dupont-Nivet and Vandeputte 2011). In selective breeding, the breeders to produce the next generation are selected based on performance on certain heritable traits, instead of being chosen randomly. As the trait has a heritable component, it is likely that the chosen individuals will be more related to each other. If high intensity selective breeding is combined with low number of breeders, the risk of fast accumulation of inbreeding is very high (Gjedrem 2005; Dupont-Nivet and Vandeputte 2011). Loss of genetic variation can also reduce future gains

that could be made with selective breeding. For example, the failure in selection experiments for growth rate in carp and tilapia have been attributed to reduced genetic variation in the base population (Moav and Wohlfarth 1973; 1976; Huang and Liao 1990). Indeed, Moav and Wohlfarth (1976) observed several malformations in their experimental line of carp indicating high levels of inbreeding due to reduced genetic variation.

Inbreeding

Inbreeding occurs when related parents interbreed to produce offspring (Gjedrem 2005). Inbreeding can be measured by the inbreeding coefficient, F, which is the probability that two alleles at any locus in an individual are identical by descent (Gjedrem 2005). When a (broodstock) population has a finite size, inbreeding is inevitable over time from a theoretical standpoint (Sonesson et al. 2005). The wild-caught broodstock are presumed to be unrelated, therefore the F of a base population at base generation is considered 0. However, inbreeding causes average genetic relatedness between individuals to increase every generation after the founding generation. Consequently, after a number of generations, it will become impossible to find unrelated individuals for mating, and as a result the accumulation of inbreeding cannot be avoided (Sonesson et al. 2005).

Inbreeding depression can be defined as the reduced fitness of offspring of related individuals compared to the fitness of offspring of unrelated individuals (Leberg and Firmin 2008). It is caused by increased homozygosity which allows deleterious recessive alleles to be expressed in inbred individuals (Leberg and Firmin 2008). Most significantly, inbreeding depression typically affects traits with dominant inheritance such as fitness traits connected with reproductive capacity or physiological efficiency (Gjedrem 2005). For example, reductions in progeny survival, growth, food conversion efficiency and increased frequency of deformities have been reported in fishes (Trippel et al. 2009). It has even been hypothesised that the comparatively late domestication of aquaculture species is due to accumulation of

inbreeding depression in early aquaculture attempts which forced breeders to continuously refresh their stock from the wild which, in turn, prevented improvement of stocks through selection (Gjedrem 2005). For example, in early studies, a reduction of 3 to 6% in body weight per 10% increase in inbreeding has been measured in rainbow trout (*Oncorhynchus mykiss*) (Kincaid 1976; Gjerde et al. 1983). More recently, in cod, eggs of full sibling parents had an 8.5% lower hatch rate than eggs from unrelated parents in the F1 generation fish (Puckrin et al. 2013).

Even though the accumulation of inbreeding cannot be avoided, it is possible to control the rate at which inbreeding accumulates. Different mating strategies such as minimum co-ancestry, compensatory matings, circular matings, avoiding full sib matings and others have been developed and are routinely used (Nomura 1999; Dupont-Nivet and Vandeputte 2011). The choice of selection method and selection intensity also affects the rate of inbreeding. Hence, methods that combine selection choice and mating strategies have been developed (Sonesson and Meuwissen 2000). For example, the Optimum Contribution selection method maximizes the genetic level of the next generation, in relation to the desired traits, while restraining the inbreeding to a pre-set level (Sonesson and Meuwissen 2000; Skaarud et al. 2011). These methods can be efficient, but require specific knowledge and computational skills to implement, and, therefore, may not be accessible to all fish breeders or small breeding programmes. As the rate of inbreeding is closely related to the effective number of breeders, the simplest way to mitigate and reduce the rate of inbreeding is to maximise the effective population size (N_e) of the farmed population. Avoiding inbreeding is very important as selection without monitoring the levels of inbreeding will only lead to short term gains and leave the base population genetically impoverished.

Effective population size

The effective population size (N_e) is the size of that idealised population (conforming to Hardy-Weinberg expectations of a constant population size, equal

sex ratio and no immigration, emigration, mutation or selection) that experiences the same effects of drift or inbreeding as the population under study (Harmon and Braude 2010). In contrast, the census size (N) of a population includes all of the individuals of the population. Effective population size (N_e) is usually considerably smaller than the adult census size (N). In practical applications, such as fish breeding, N_e is sometimes simplified to be the number of breeding individuals in a population. However, this can be dangerously misleading as, for example, unequal sex ratio among parents or variation in family size can make the genetic contributions between parents unequal. As a result, the N_e can be significantly lower than the (total) number of breeding individuals. N_e/N ratios in wild populations are quite low with average of 0.10–0.11 and median of 0.14 being reported from two review studies (Frankham 1995; Palstra and Ruzzante 2008).

In conservation biology, it has been suggested that an N_e of 50 or larger is needed to prevent inbreeding depression and a detectable decrease in viability of wild populations in the short term (Franklin 1980). An N_e of 500 or more is needed to maintain sufficient evolutionary potential in a population, as loss of variance due to genetic drift is balanced by the increase in variance due to new mutations at this size (Franklin 1980). The “50/500 rule” developed by Franklin (1980) has gained considerable popularity as a simple practicable guideline, but it has since been challenged as being far too optimistic. In contrast, Lande (1995) has suggested that N_e of over 5000 is the limit for retaining sufficient genetic variation for a population to be able to respond to natural selection in order to adapt to changing environmental conditions. More recently, it has been suggested that “50/500 rule” by Franklin (1980) should be updated to 100/1000 (Frankham et al. 2014). These generalised guidelines for wild populations have been selectively adopted by fish farmers and hatchery managers. In aquaculture, the number of breeders is regularly limited to 100 individuals or 50 pairs, as a rule of thumb, without establishing *a priori* the relatedness among breeders (Bentsen and Olesen 2002; Gjedrem 2005; Appleyard and Ward 2006), exposing the farmed stock to rapid accumulation of inbreeding and loss of genetic diversity.

The choices made when establishing an aquaculture base population have important consequences for the N_e of the farmed stock. The higher the number of wild-caught broodstock the higher the N_e of the base population will be. However, if the initially sampled wild broodstock are related to each other, the N_e will be lower than expected based on numbers of founding individuals. In the case of sampling wild cod from the Celtic Sea, the assumption that caught individuals will be unrelated seems objectively valid. However, this may not be true for other species, for example, salmonids caught at the river where they spawn may be more likely to be related due to their strong homing instincts (i.e. Carlsson et al. 1999).

Once an N_e of desired size is established by securing sufficient number of founders in the base population, the N_e has to be preserved through management. N_e can be maintained by the simple rule of retaining high number of parents to produce each generation (Bentsen and Olesen 2002). Other feasible, practical, management strategies for sustaining high N_e levels are, for example, balancing the number of sires and dams as well as the number of offspring from each family (Sonesson et al. 2005).

Even if large number of breeders is chosen and female-male ratios are kept even, monitoring the levels of genetic variation is necessary for the success of the programme, because several factors can quickly reduce genetic variation in the first generations of domestication. Factors leading to loss of genetic variation include low and variable survival rates in eggs and larvae (Kjørsvik et al. 2004), and lower number of effective breeders than expected in mass spawning systems (Brown et al. 2005; Fessehaye et al. 2006; Chatziplis et al. 2007; Blonk et al. 2009). Mating can also be controlled by paired mating or strip spawning, however this does not automatically guarantee equal reproductive success between parents due to differences in individual reproductive capacities resulting from age, size, condition and behaviour (for examples see, Kjesbu et al. 1996; Rakitin et al. 2001; Skjæraasen and Hutchings 2010).

2.1.5. Aims and objectives of the chapter

EIRCOD has been building its base population, every year from 2008 to 2013, by catching wild broodstock fish from a defined regional stock and at a known spawning ground in the Celtic Sea. These broodstock fish have been stripped at sea to create full and half sib families which were then returned to the hatchery and laboratories at Carna Research Station, NUIG, for incubation and onrearing.

This chapter aims to investigate the temporal stability of the genetic composition of the spawning aggregations of the Celtic Sea source population and assess the genetic variability of the population by estimating its effective population size (N_e) using microsatellite markers.

Differences in survival between the resultant families during rearing from years 2008, 2009 and 2010 are investigated by means of parentage analysis. The effective population size (N_e) of the wild broodstock is calculated for years 2008, 2009 and 2010 (three cohorts). The resulting values are used to estimate the N_e of the current farmed stock (five cohorts 2008-2012). This allows an assessment of factors affecting the N_e in an aquaculture population. In addition, husbandry practises affecting the N_e can be identified and modified to insure largest retention of genetic variation.

In addition, during 2011, the first F2 offspring were produced by a communal mass spawning event. The ability of the current set of microsatellite loci to resolve parentage is assessed in the mass spawning event that included closely related individuals (full sibs). The reproductive contributions of parents are investigated to assess variability in reproductive success.

2.2. Material and methods

2.2.1. Sample acquisition

Part A: Wild broodstock & Strip spawnings

From 2008 to 2012 cod eggs were sourced from wild fish caught in the Celtic Sea during the spawning season. Wild cod were caught by targeted fishing in a known breeding ground off the south coast of Ireland in March of each year. Several hundred fish taken by purse seine were visually and manually screened by NUIG personnel over several days during each annual collection cruise. A selection of viable, or potential, broodstock were identified and manually stripped and the ova were fertilized at sea to create full and half sib families. The weight and length of each individual used for breeding was recorded and a fin clip or a piece of gill was collected for genotyping.

The numbers of females and males of wild broodstock (P0) used in each of five years, 2008-2012, can be seen in **Figure 1**. In total, the wild broodstock consisted of 130 fish (2008, n=27; 2009, n=23; 2010, n=21; 2011, n=42 and 2012, n=17). In both 2008 and 2010, one of the initially sampled males did not contribute to the subsequent family crosses (2008, n=26 and 2010, n=20). These males were included in the analysis aiming to characterise the source population of the wild broodstock in order to maximise the sample size, but were not included in the parentage and F1 survival analysis.

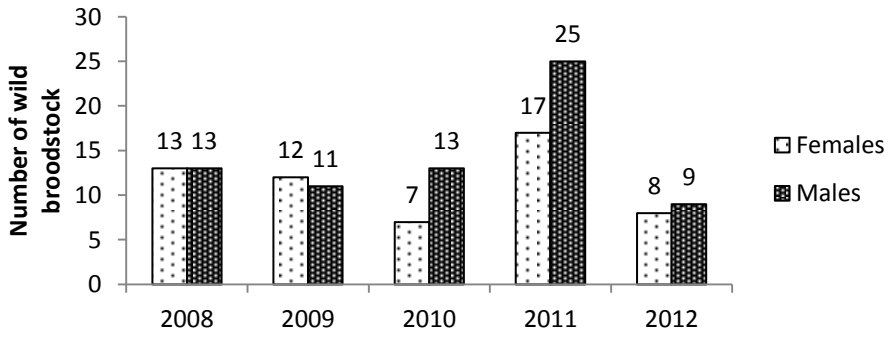

Figure 1 Numbers of females and males of wild cod broodstock for each year from 2008 to 2012.

Part B: Production and on-growing of F1 fish

All the year-classes (F1) resulting from the crosses of wild caught broodstock were reared from egg to juvenile at NUIG's Ryan Institute marine research facilities in Carna, Ireland (53° 19'N, 9° 49'W). During incubation, the eggs from the broodstock crosses were housed in standard 70 l cylindro-conical upwelling vessels with an average water temperature of 8°C and 24 h light. The incubation cones housed between 1-3 families each (~250-300g of eggs per cone). The eggs were transferred to 2000 l circular tanks in the hatchery one day prior to hatch (pre-hatch), except for 2008 year-class which was transferred once 80% of eggs had hatched.

All yearly F1 cohorts (2008, 2009 and 2010) of larvae were fed on a mix of water and 'greenwater' (microalgae; brown algae *Isocrysis galbana* and green algae *Nannochlorpsos oculata*) (from 0 days post hatch (dph)) and rotifers (*Branchionus pilatus*) (from 1-2 dph) followed by weaning onto *Artemia spp.* commencing at 28 dph (2008 F1) or 25 dph (2009 and 2010 F1) (**Figure 2**). Weaning onto dry feed was initiated at 55 dph (2008 F1) or 46 dph (2009 and 2010 F1). Overlap of feeds was allowed for each of the weaning periods. The larvae were fed to excess 3-5 times a day. Once the fish reached an average size of 1 g they were graded by size (small, medium or large) and moved into 2000 l circular tanks in the nursery. Thereafter, the juvenile fish were size graded at regular intervals to reduce stress and cannibalism while kept in the nursery.

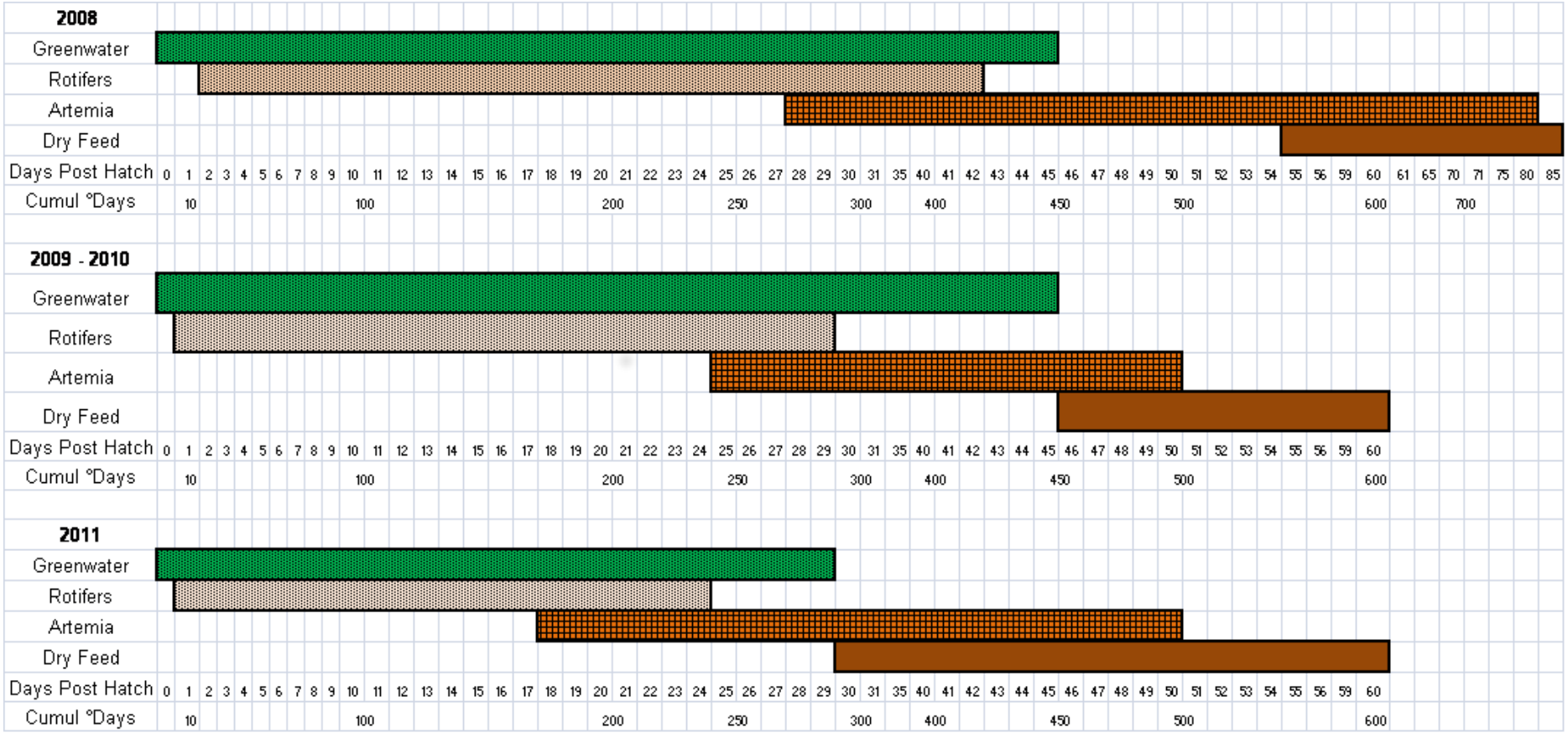

Figure 2 Feeding schedules for larval cod reared in Ireland from 2008 to 2011.

In time, these juvenile fish were transferred to sea cages located in nearby Beirteraghbui Bay (53° 23'N, 9° 49'W) for on-growing. The 2008 cohort was moved to sea in two batches in November 2008 and December 2008 at an average weight of 35 g and 76 g, respectively. The 2009 cohort was moved between June-July 2010 at an average weight of 75 g. For more detailed descriptions of larval, juvenile and adult on-rearing protocols and performance see Bolton-Warberg & FitzGerald (2012).

The survival of full and half sib families, the F1 2008, 2009 and 2010 cohorts, was monitored and sampled at a series of time points in succeeding years.

The resulting fertilisation matrices for the three cohorts are presented in **Tables 1**, **2** and **3**. The majority of crosses were single male-female pairings, except in 2008 when eggs and milt from a group of eight individuals were mixed (four females and four males) in addition to the nine full sib families created that year (**Table 1**). In 2009, 17 pairings were made to create 17 full sib families, 14 of these were nested within half sib families (**Table 2**). In 2010, 13 crosses were made to create full sib families, 9 of these nested in 3 half sib families (**Table 3**). For cohorts 2009 and 2010 the amount of fertilised eggs per family was measured in grams; these data were incomplete for the year 2008.

Table 1 Fertilisation matrix showing the families in the 2008 F1 cohort. Nine crosses were made to create full sib families. In addition, eggs and milt from eight individuals (four females and four males) were mixed to create the 10th egg batch (the lower right corner of the fertilisation matrix). Female parents are in the first row and male parents are in the first column. The light grey boxes represent the crosses between a female and a male; the number inside the square is the amount of fertilised eggs in grams; ? denotes missing data.

2008	A	B	C	D	E	F	G	H	I	J	K	L	M
1	310												
2		300											
3			185										
4				?									
5					1160								
6						?							
7							410						
8								?					
9									?				
10										?	?	?	?
11										?	?	?	?
12										?	?	?	?
13										?	?	?	?

Table 2 Fertilisation matrix showing the families in the 2009 F1 cohort. Female parents are in the first row and male parents are in the first column. The light grey boxes represent the crosses between a female and a male; the number inside the square is the amount of fertilised eggs in grams; * denotes the two males whose genotypes are reconstructed from offspring genotypes.

2009	A	B	C	E	F	I	J	M	N	O	P	Q
1	200											
2		30	230									
4				50								
5					100							
7 *						250						
8						200	50					
9						275	50					
10 *						200						
11								200	125	50	175	
14												250
15												250

Table 3 Fertilisation matrix showing the families in the 2010 F1 cohort. Female parents are in the first row and male parents are in the first column. The light grey boxes represent the crosses between a female and a male; the number inside the square is the amount of fertilised eggs in grams.

2010	A	B	C	E	I	M	N
1	200						
2	200						
3	200						
4	180						
5				100			
7					140		
8					140		
9		50					
10						30	
11							25
12			200				
13			200				
14			140				

Parental genotype reconstruction for the 2009 wild broodstock

The parents of the 2009 cohort included two individual males (number 7 and 10) whose tissue samples were not returned to the laboratory at the conclusion of the collection cruise. The incubation cones with offspring from these males were Cone 1 housing: I7 family (Female I × Male 7) and O11 family (Female O × Male 11); and Cone 2 housing: I10 family (Female I × Male 10) and M11 family (Female M × Male 11). In order to resolve the genotypes of the two males, a random sample of 48 eggs from each of the two incubation cones (96 in total) was used for genotyping.

Part B Sampling for family survival

The 2008 F1 cohort was sampled at harvest at 3 years of age between January and March 2011. In total, 1415 fish from the 2008 cohort were harvested and concurrently sampled for weight and length, while a fin tissue sample was also collected from each fish. A random sample of 96 out of the 1415 fish was drawn for genotyping. In addition, 109 fish from the cohort were selected as broodstock to be kept in a net pen inside the main cage with the intention of producing F2 offspring. The fish were selected based on larger size (indicative of faster growth) and overall body shape and appearance (no deformities or abnormalities evident). The weight and length of each individual was recorded and a fin clip was taken for genotyping. The 109 broodstock fish and 96 sub-sample of the harvest fish constituted the survival sample for the 2008 cohort.

The 2009 F1 cohort was sampled for survival at two time points: at pre-hatch stage and at 2 years old. During incubation, the families were housed in 8 incubation cones containing 1-3 families each. When the eggs were at pre-hatch stage (20th of March 2009) a batch sample was taken from each of the incubation cones and stored in ethanol. To assess the survival of offspring at the pre-hatch stage, 48 eggs were sampled from each of the 6 remaining cones for genotyping. The 48 eggs from the two cones used for the reconstruction of parental genotypes were also included in the survival sample (total eggs genotyped n=384).

Later, having been transferred to the sea site, the 2009 cohort (approximately 8000-9000 fish) was kept in a circular net cage (70 m circumference, 10-14 deep). Due to lack of space a few hundred fish from the same year class, but of Norwegian origin, were housed in the same cage. This combined fish group were sampled on two occasions: 77 fish on 14th April 2011 and 174 fish on 31st August 2011 when the fish were about two years old (approx. 24 and 28 months post-hatch, respectively). To determine survival at 2 years, the 251 fish were fin clipped and subsequently genotyped.

The 2010 F1 cohort was sampled at the age of one year while still at the Carna site. The fish were housed in a series of different tanks based on their size grading: large, medium or small. All the fish were tagged with PIT-tags, measured for length and weight and sampled for tissue for genotyping. A total of 94 fish per size grade (n=282) were sampled for genotyping.

Part C: Mass spawning events

The 2007 and 2008 cohorts were held at sea in a circular net cage (70m circumference, 10-14 m deep). The 2008 cohort consisted of F1 fish from known crosses of wild Celtic Sea broodstock produced via the EIRCOD project. The 2007 cohort consisted of fish from four different backgrounds representing earlier smaller research projects. These included two batches of eggs from Norway, a batch of eggs from F1 fish of broodstock from Isle of Man and batch of eggs from unknown crosses of wild Celtic Sea broodstock (Bolton-Warberg 2012). The fish resulting from the four batches of eggs were housed communally at sea.

In order to complement the 109 fish from the 2008 F1 cohort selected as broodstock, an additional 63 fish from the 2007 F1 cohort were selected as broodstock based on size and lack of deformities. The broodstock fish were held in a net pen inside a bigger polar circle type cage. All the 172 broodstock fish were measured for weight and length, PIT-tagged and sampled for tissue. The fish were also sexed based on visual inspection of the urogenital opening and the body shape,

for example, enlarged bellies during spawning season were considered indicative of ovulating females (**Figure 3**).

Figure 3 Top picture: spawning female cod; Bottom picture: spawning male cod. (Picture from https://maricult.com/en/photo-gallery/aquaculture/atlantic-cod-gadus-morhua/.)

In 2011, the broodstock cage was monitored for presence of spawning activity during the spawning season by visual inspection. Batches of fertilised eggs were collected from the upper sea water column in 9 separate days. A portion of the eggs from each batch was stored in ethanol. The remaining eggs were transferred to the hatchery and reared in the same way as the F1 cohorts. Once hatched, the larvae were fed on a mix of water and 'greenwater' (from 0 dph) and rotifers (from 1-2 dph) followed by weaning onto *Artemia spp.* from 18 dph. Weaning onto dry feed was initiated at 30 dph. Once the fish reached an average size of 1 g they were moved into 2000 l circular tank in the nursery.

Two of the egg collection dates were arbitrarily selected for genetic analysis of parental contributions: 29th March 2011 (Sample 1) and 7th April 2011 (Sample 2). A total of 94 eggs per sampling date (total n=188) were randomly chosen for genotyping.

A summary of all samples subjected to genetic analysis is presented in **Table 4**.

Table 4 A summary of all samples subjected to genetic analysis in Chapter 2.

Dataset	No. of individuals genotyped
Wild broodstock 2008-2012 P0	
2008 **P0**	27
2009 **P0**	23
2010 **P0**	21
2011 **P0**	42
2012 **P0**	17
total	**130**
2009 P0 parental reconstruction from eggs	
2009 **F1** Eggs, Cone 1	48
2009 **F1** Eggs, Cone 2	48
total	**96**
2008 F1 Survival sample	
2008 **F1** Harvest random sample	96
2008 **F1** Broodstock fish	109
total	**205**
2009 F1 Survival sample	
2009 **F1** Eggs, Cone 1	48
2009 **F1** Eggs, Cone 2	48
2009 **F1** Eggs, Cone 3	48
2009 **F1** Eggs, Cone 4	48
2009 **F1** Eggs, Cone 5	48
2009 **F1** Eggs, Cone 6	48
2009 **F1** Eggs, Cone 7	48
2009 **F1** Eggs, Cone 8	48
2009 **F1** 2-year-olds	251
total	**635**
2010 F1 Survival sample	
2010 **F1** Size grade small	94
2010 **F1** Size grade medium	94
2010 **F1** Size grade large	94
total	**282**
Farmed broodstock	
2007 **F1** Broodstock fish	63
2008 **F1** Broodstock fish	109
total	**172**
Mass spawning offspring 2011 F2	
2011 **F2** Day 1 Mass spawning	94
2011 **F2** Day 2 Mass spawning	94
total	**188**

2.2.2. DNA extraction, PCR and gel electrophoresis

DNA extraction

DNA was extracted from eggs, fin and gill tissue samples using proteinase K digestion and a modified Chelex® 100 (Sigma Aldrich) method (Mirimin et al. 2011). Approximately 5-10 mg of tissue or a single egg was used for the DNA extraction. The tissue pieces were immersed in dH_20 and dried over clean white paper tissue three times to remove ethanol. To remove ethanol from the eggs, they were transferred into a beaker of dH_20 and occasionally mixed gently until they sank to the bottom. The tissue samples were cut into smaller pieces; the eggs were burst and incubated in 200 µl of Chelex® (10%) with 0.1 mg of proteinase K for two hours at 56°C shaking occasionally. The samples were then incubated at 99°C for 10 minutes shaking at five minutes. The sample tubes were centrifuged at 8000 g for two minutes to separate the aqueous supernatant containing the DNA from the Chelex® beads and proteins. The supernatant was transferred to a new plate.

PCR amplification

A large range of microsatellite loci are recorded for Atlantic cod and a suite of 13 loci were selected and available for use in this study. The full 13 microsatellite set included loci: *GADM*3 (Hutchinson et al. 2000), *Gmo*2, *Gmo*132 (Brooker et al. 1994), *Gmo*8, *Gmo*19, *Gmo*34, *Gmo*35, *Gmo*37 (Miller et al. 2000), *Gmo*-C18, *Gmo*-C52 (Stenvik et al. 2006), *PGmo*76, *PGmo*94 (Skirnisdottir et al. 2008) and *Mae*211 (Lage & Kornfield 1999). The wild broodstock 2008-2012 were genotyped with all 13 loci. The other data sets were genotyped with varying sub-sets of the 13-loci set (**Table 5**).

Table 5 Microsatellite loci used and the number of individuals genotyped in the different data sets. The 109 F1 broodstock fish from 2008 cohort are displayed twice, in the 2008 F1 survival sample and the farmed broodstock sample that contributed to the F2 generation. The 96 eggs used in the 2009 P0 parental reconstruction are also included in the 384 eggs sampled at pre-hatch

Dataset	No. of loci	No. of individuals genotyped	Names of loci
Wild broodstock 2008-2012 **P0**	13	130	*GADM*3, *Gmo*2, *Gmo*8, *Gmo*19, *Gmo*34, *Gmo*35, *Gmo*37, *Gmo*132, *Gmo*-C18, *Gmo*-C52, *PGmo*76, *PGmo*94, *Mae*211
2009 **P0** parental reconstruction from eggs	13	96	*GADM*3, *Gmo*2, *Gmo*8, *Gmo*19, *Gmo*34, *Gmo*35, *Gmo*37, *Gmo*132, *Gmo*-C18, *Gmo*-C52, *PGmo*76, *PGmo*94, *Mae*211
2008 **F1**	4	96 + 109	*Gmo*19, *Gmo*35, *Gmo*37, *Gmo*132
2009 **F1**	3	384 + 251	*Gmo*8, *Gmo*19, *Gmo*-C52
2010 **F1**	3	282	*Gmo*8, *Gmo*19, *Gmo*132
Farmed broodstock; 2007 **F1** & 2008 **F1**	9	63 + 109	*Gmo*2, *Gmo*19, *Gmo*34, *Gmo*35, *Gmo*37, *Gmo*132, *Gmo*-C18, *Gmo*-C52, *PGmo*94
Mass spawning offspring 2011 **F2**	6	188	*Gmo*2, *Gmo*19, *Gmo*34, *Gmo*-C18, *Gmo*-C52, *PGmo*94

The forward primers were labelled with 5' IRDye 700 or 5' IRDye 800 (Integrated DNA Technologies). For loci *Gmo*2, *Gmo*132, *Gmo*8, *Gmo*19, *Gmo*35, *Gmo*37 and *PGmo*94 the amplifications were performed in 10 µl reactions with 2 µl of DNA, 0.5 × Green GoTag® Flexi buffer (Promega), 0.5 × ColorlessGoTag® Flexi buffer (Promega), 1.5 mM $MgCl_2$ (Promega), 250 µM of dNTP mix (Promega), 10 µM of unlabelled reverse primer, 10 µM of labelled forward primer and 0.5 u of GoTag® Flexi DNA Polymerase (Promega). The PCR was performed with either TC-5000 or TC-512 Thermal Cycler (Techne) and the amplification conditions included an initial denaturation step of 3 min at 95°C, followed by 35 cycles of 40 s at 95°C, 40 s in annealing temperature of 52°C, 40 s at 72°C, followed by a final extension step of 10 min in 72°C. For marker *PGmo*76 the PCR reaction was as above except for 2.5 µM of unlabelled forward primer with T3-tail and 10 µM of universal T3 primer (Jena Bioscience) labelled with IRDye 700. For markers *GADM*3 and *Mae*211, the DNA was diluted 1:5 with distilled water before use and a two-step amplification protocol was used in order to reduce stuttering. The two-step amplification protocol included an initial denaturation step of 3 min at 95°C, followed by 5 cycles of 40 s at 95°C, 40 s at 52°C, 40 s at 72°C and 30 cycles of 40 s at 95°C, 40 s at 56°C,

40 s at 72°C followed by a final extension step of 10 min in 72°C. Markers *Gmo*34, *Gmo*-C18, *Gmo*-C52 were amplified in a 25 µl multiplex reaction with 2 µl of DNA, 0.75 × Green GoTag® Flexi buffer (Promega), 0.75 × Colorless GoTag® Flexi buffer (Promega), 3 mM $MgCl_2$ (Promega), 200 µM of dNTP mix (Promega), 2 µM of each unlabelled reverse primer, 2 µM of each labelled forward primer and 1 u of GoTag® Flexi DNA Polymerase (Promega). The amplification conditions were: an initial denaturation step of 3 min at 95°C, followed by 30 cycles of 15 s at 95°C, 30 s at 60°C, 20 s at 72°C, followed by a final extension step of 10 min in 72°C. The same conditions were used for the multiplex of *Gmo*8, *Gmo*19 and *Gmo*-C52 used to genotype the eggs and 2-year-old fish of the 2009 cohort.

Gel electrophoresis

The PCR products were first diluted 1:50 with distilled water. One µl of this dilution was mixed with 2µl of formamide loading dye (98% formamide, 1 mg/ml bromophenol blue, 1 mg/ml xylene cyanol, 0.1 mM EDTA pH 8) prior to further processing. The PCR products mixed with loading dye were denatured at 99°C for 1 min and, subsequently, kept on ice. The sizes of the PCR products were determined by electrophoresis on 6.5% acrylamide gel (KB Plus 6.5% Gel Matrix (LI-COR), 0.75 mg/ml APS (Sigma Aldrich), 74.25 mg/ml TEMED (Sigma)) on a LI-COR 4300 DNA analyser. When possible the products were pooled with consideration to expected fragment size and dye/channel used for detection. The fragment lengths were determined by manually comparing them to reference ladders. The reference ladders were created by pooling PCR products from individuals possessing different alleles spread evenly across the size range of the locus. The allele sizes in base pairs were estimated by running the reference ladder next to a size standard when first creating the reference ladder.

All genotype data sets were inspected with MICRO-CHECKER 2.2.3 for genotyping errors and presence of null-alleles (van Oosterhout et al. 2004) using default settings.

2.2.3. Data analysis

Data analysis – Part A: Wild broodstock

Number of alleles, allelic richness, allele size ranges, F_{ST} estimates and expected and observed heterozygosity were assessed using MSANALYSER 4.05 (Dieringer and Schlötterer 2003) with standard default settings. Data were analysed for possible departure from Hardy-Weinberg equilibrium, excess and deficiency of homozygotes, and linkage disequilibrium using GENEPOP 4.2, again with standard default settings (Raymond and Rousset 1995; Rousset 2008). Sequential Bonferroni correction was used to correct for multiple comparisons with initial $\alpha = 0.05$ (Rice 1989). Effective population size (N_e) was estimated using the linkage disequilibrium method of Waples and Do (2008), as implemented in NeESTIMATOR V2 (Do et al. 2014). Very rare alleles can bias the estimates of effective population size, however, they can be screened for by using critical values (Waples and Do 2010). Critical value (PCrit in NeESTIMATOR) is an allele frequency threshold, below which alleles are excluded from the analysis. The PCrit values were set separately for each sample in order to set the PCrit high enough to exclude alleles that occur in only a single copy (in one heterozygote) in the sample. This was accomplished by setting that PCrit > 1/(2S), where S is the number of individuals with data at both pairs of loci. For each sample, the S was chosen according to the locus with most missing data. The 95% confidence intervals for N_e estimates were calculated using the parametric method as implemented in NeESTIMATOR V2 (Do et al. 2014).

Data analysis – Part B: Production and on-growing of F1 fish

Genotype reconstruction

PEDIGREE 2.2 was used to recreate the missing parental genotypes (Smith et al. 2001; Herbinger 2005). This approach divides individuals into groups of full or half siblings based on genotypic data, and finds the combination of groups (partition) that has the maximal score obtained with a Markov Chain Monte Carlo (MCMC)

process. It then recreates all the possible combinations of parental genotypes that could have produced the offspring genotypes seen in each of the groups. Defining the best partition to create a plausible pedigree requires adjusting parameter values in an iterative way, as follows. The first six PEDIGREE runs were performed with default settings: 50 000 iterations, 1.0 weight, 10.0 temperature, random 'seeds' and full sib constraint on. Both sampled cones originally contained eggs from two full sib families. As two big groups of full sibling families were expected in both of the datasets, the optimisation process was started with the weight parameter. Runs with different parameters were tested in blocks of six. Values of 5, 10, 15, 20 and 25 were tested for the weight. Higher temperatures from 10 to 50 were tested as well and longer iterations of 10 000 000 were tried for some of the runs. When the number of groups was no longer reduced by adjusting the parameters, a method of nesting the best full sibling families within the best half sibling group (kin group) was employed to detect genotyping errors. When the smallest number of groups was found, the parental genotypes reconstructed by the programme were compared to the genotypes of the real parents. The identification of I7 and I10 families allowed the parental genotypes to be resolved. When the programme gave several options for the parental genotype combinations, the one that was most likely to produce the observed ratios of offspring genotypes was chosen by first by identifying the female genotype and then by comparing the chi^2 values calculated by the programme.

Parentage analysis in the survival data set

Computer software P-LOCI 1.0 (Matson et al. 2008) was used to detect the most informative set of loci for resolving parentage for the cohorts in the survival study (cohorts 2008, 2009 and 2010). This approach simulates offspring genotypes based on parental genotypes and a mating matrix and then ranks the marker loci sets by their discriminatory power. The programme was run with default values and the simulation was done with 500 offspring from every cross in the mating matrix. Out of the sets of markers suggested by P-LOCI, the sets of markers were chosen with additional requirements of ease of scoring and markers' size ranges not overlapping

when run on gel. Parentage analysis was carried out by means of the exclusion method using VITASSIGN 8.2 computer software (Vandeputte et al. 2006).

N_e estimates in the survival data set

N_e was manually calculated for the 2008, 2009 and 2010 cohorts using three different measures as in Naish et al. (2013):

1) The effective size N_e resulting from deviations in the sex ratio (N_{e_sex}) was calculated using equation 4.7a in Hedrick (2011),

where $N_{e\_sex} = (4N_fN_m)/(N_f+N_m)$,

and where N_f and N_m are the number of females and males, respectively.

2) The N_e resulting from variance in reproductive success (N_{e_var}) was calculated using equation 4.9a in Hedrick (2011),

where $N_{e\_var} = (Nk-1)/(k-1+(\sigma^2_k/k))$,

and where k is the average number of progeny and σ^2_k is the variance in the number of progeny.

3) The N_e that accounts for both variance in reproductive success and deviations in sex ratio (N_{e_both}) was calculated using equation 4.11a-c in Hedrick (2011),

where $N_{e\_both} = (4N_{e\_varf}N_{e\_varm})/(N_{e\_varf}+N_{e\_varm})$,

where $N_{e\_varf} = (N_fk_f-1)/(k_f-1+(\sigma^2_{kf}/k_f))$ and

where $N_{e\_varm} = (N_mk_m-1)/(k_m-1+(\sigma^2_{km}/k_m))$,

and where k_m and k_f are the average number of progeny per sex, respectively,

and σ^2_{kf} and σ^2_{km} are the variance in the number of progeny per sex, respectively.

For comparative purpose, N_e was also estimated using the sibship assignment method (Wang 2009) as implemented in COLONY (Jones and Wang 2010; Wang 2012); as well as the linkage disequilibrium method of Waples and Do (2008), as implemented in NeESTIMATOR V2 (Do et al. 2014). In COLONY, the inbreeding model was used in order to estimate an inbreeding coefficient for the samples. Full likelihood method, medium likelihood precision, medium run length of the simulated annealing algorithm and weak sibship prior settings were used. Both the allelic dropout rate and the rate of genotyping errors were set to 0.0005 for all loci. As the exclusion method implemented in VITASSIGN is a more reliable method in parentage assignment than the maximum likelihood method employed in COLONY, the parentage assignments from VITASSIGN were used for the N_e and inbreeding coefficient estimates, instead of using the parentage assigned by COLONY itself. In NeESTIMATOR V2 analysis, the PCrit values were set separately for each sample in order to set the PCrit high enough to exclude alleles that occur in only a single copy (in one heterozygote) in the sample. PCrit was set > 1/(2S) for each sample, where S is the number of individuals with data at both pairs of loci. The Pcrit values used were 0.003 for 2008 cohort and 0.002 for the other cohorts. The 95% confidence intervals for N_e estimates were calculated using the parametric method as implemented in NeESTIMATOR V2 (Do et al. 2014).

Data analysis – Part C: Parentage analysis in mass spawning events

Parentage analysis was conducted with the exclusion method using VITASSIGN and with the maximum likelihood method implemented in COLONY. The parentage analysis in COLONY was performed using the following conditions; parents' sex known, parents' sex unknown and without any parental information. N_e was estimated using the sibship assignment method and the heterozygosity excess method as implemented in COLONY. The inbreeding model was used in order to estimate an inbreeding coefficient. Full likelihood method, medium likelihood

precision, medium run length and weak sibship prior settings were used. Both the allelic dropout rate and the rate of genotyping errors were set to 0.0005 for all loci. The probability for a parent to be included in the group of parents was set to 0.95.

F_{IS} estimates

F_{IS} estimates per population were calculated using GENETIX 4.05 with 10 bootstrap replicates for the wild parents (P0) 2008-2012, three F1 cohorts (2008, 2009 and 2010) and the F2 cohort (2011) (Belkhir et al. 1996-2004).

2.3. Results

2.3.1. Wild broodstock fish

The number of alleles, observed size range and the range of frequencies for the 13 loci in the different parental groups, for 2008, 2009, 2010, 2011 and 2012, are presented in **Table 6**. Due to PCR amplification failure or stuttering, some missing genotypes were observed (18 out of 1690, or 1.07%). MICRO-CHECKER analysis indicated no evidence of scoring error due to stuttering or large allele dropout. An excess of homozygotes was observed in locus *Gmo*8 in the 2010 sample, which could be indicative of null alleles.

Table 6 Summary statistics for 13 microsatellite loci in wild broodstock from years 2008-2012. n, number of individuals; a, number of alleles; R_S, allelic richness per locus and sample; as, allele size range in base pairs; H_E, expected heterozygosity; H_O, observed heterozygosity; HW, probability values of concordance with Hardy-Weinberg expectations. None of the probability estimates were significant after sequential Bonferroni correction for multiple tests (α = 0.05).

Year		*GADM*3	*Gmo*2	*Gmo*8	*Gmo*19	*Gmo*34	*Gmo*35	*Gmo*37	*Gmo*132	*Gmo-C*18	*Gmo-C*52	*Mae*211	*PGmo*76T3	*PGmo*94
2008	n	27	27	27	27	27	27	27	27	27	27	26	25	27
	a	26	11	19	19	8	8	9	19	11	6	24	18	13
	R_S	19.92	9.84	15.11	15.76	7.04	7.24	8.59	14.04	10.19	5.69	19.04	15.09	10.32
	as	107-205	107-143	126-310	122-202	88-116	122-143	236-284	114-170	146-179	283-298	124-192	172-222	270-418
	H_E	0.956	0.877	0.902	0.932	0.660	0.832	0.859	0.927	0.906	0.686	0.962	0.885	0.760
	H_O	0.963	0.852	0.815	0.926	0.630	0.815	0.852	0.963	0.852	0.556	0.885	0.880	0.741
	HW	0.537	0.069	0.210	0.491	0.249	0.791	0.837	0.940	0.334	0.198	0.042	0.816	0.748
2009	n	21	23	23	23	23	23	23	23	23	23	22	23	23
	a	23	9	19	16	7	7	11	17	11	8	19	17	16
	R_S	20.17	8.85	16.56	14.60	6.49	6.66	9.55	13.08	9.63	7.14	16.67	14.89	13.48
	as	121-197	107-125	126-334	138-206	88-116	125-143	136-308	118-170	143-182	283-304	130-192	172-244	268-438
	H_E	0.957	0.866	0.937	0.928	0.632	0.790	0.811	0.906	0.864	0.731	0.947	0.920	0.800
	H_O	1.000	0.826	0.957	0.913	0.522	0.739	0.739	0.913	0.870	0.609	0.955	0.870	0.696
	HW	0.464	0.321	0.990	0.800	0.048	0.445	0.559	0.412	0.734	0.291	0.708	0.024	0.142

Table 6 *Continued.*

Year		*GADM*3	*Gmo*2	*Gmo*8	*Gmo*19	*Gmo*34	*Gmo*35	*Gmo*37	*Gmo*132	*Gmo-C*18	*Gmo-C*52	*Mae*211	*PGmo*76T3	*PGmo*94
2010	n	21	21	21	21	21	21	20	21	21	21	21	20	21
	a	22	8	19	15	7	5	11	13	11	5	18	18	14
	R_S	19.62	7.99	16.77	13.87	6.28	4.81	10.15	11.69	10.39	4.62	15.91	16.31	12.35
	as	125-191	107-121	126-334	142-202	88-116	128-140	228-284	124-164	143-176	283-295	148-188	172-230	268-418
	H_E	0.963	0.873	0.929	0.900	0.719	0.757	0.855	0.923	0.892	0.581	0.945	0.919	0.796
	H_O	0.952	0.714	0.762	0.952	0.619	0.857	0.700	1.000	0.905	0.714	1.000	0.950	0.667
	HW	0.328	0.110	0.007	0.368	0.181	0.442	0.136	0.150	0.621	0.665	0.690	0.964	0.519
2011	n	42	41	41	41	42	42	42	42	42	42	41	42	42
	a	28	10	26	18	7	8	11	20	10	7	24	22	18
	R_S	19.36	8.64	17.34	13.70	5.84	7.12	9.16	12.80	8.91	5.39	17.97	15.32	11.68
	as	117-203	107-141	114-310	126-206	88-116	125-152	236-288	114-166	143-170	280-298	144-194	172-240	266-386
	H_E	0.959	0.827	0.932	0.911	0.682	0.829	0.845	0.905	0.886	0.701	0.960	0.913	0.775
	H_O	1.000	0.756	0.878	0.878	0.619	0.762	0.857	0.929	1.000	0.810	0.976	0.905	0.762
	HW	0.843	0.034	0.063	0.432	0.133	0.192	0.774	0.838	0.690	0.241	0.205	0.065	0.764

Table 6 *Continued.*

Year		*GADM*3	*Gmo*2	*Gmo*8	*Gmo*19	*Gmo*34	*Gmo*35	*Gmo*37	*Gmo*132	*Gmo-C*18	*Gmo-C*52	*Mae*211	*PGmo*76T3	*PGmo*94
2012	n	17	17	17	17	16	17	16	14	17	17	16	17	17
	a	20	10	14	14	6	7	9	13	8	4	18	17	14
	R_S	20.00	10.00	14.00	14.00	6.00	7.00	9.00	13.00	8.00	4.00	18.00	17.00	14.00
	as	125-219	107-143	126-314	122-214	88-120	125-143	228-284	124-162	146-170	283-295	140-196	166-268	270-400
	H_E	0.961	0.809	0.906	0.925	0.661	0.815	0.883	0.926	0.873	0.636	0.962	0.945	0.861
	H_O	1.000	0.765	0.824	0.882	0.563	0.765	1.000	0.929	1.000	0.824	0.938	0.941	0.706
	HW	1.000	0.707	0.050	0.501	0.011	0.654	0.994	0.272	0.807	0.240	0.269	0.837	0.094

F_{ST} *estimates*

Pairwise multi-locus F_{ST} estimates between the broodstock fish from each year (2008-2012) using 13 loci varied from -0.002 to 0.004 and none of them were significant (**Table 7**). Global per locus F_{ST} estimates varied from -0.0063 (*PGmo*94) to 0.0039 (*Gmo*8) with non-significant P-values. The global F_{ST} over all loci was 0.0005 with a non-significant P-value.

Table 7 Pairwise multi-locus F_{ST} values between broodstock fish 2008-2012. Pairwise F_{ST} values are below the diagonal and associated P-values are above the diagonal.

	2008	2009	2010	2011	2012
2008		0.682	0.412	0.429	0.132
2009	-0.002		0.444	0.444	0.224
2010	0.001	0.000		0.477	0.382
2011	0.000	0.000	0.000		0.347
2012	0.004	0.002	0.001	0.001	

N_e estimates

The N_e estimates with the linkage disequilibrium method for the parent samples of 2008-2012 and pooled sample are shown in **Table 8**. The data across samples were pooled to obtain a more reliable estimate of N_e.

Table 8 N_e estimates with the linkage disequilibrium method for broodstock fish 2008-2012. The number of females, males and total number of parents is given. S is the number of individuals with data at both pairs of loci. Pcrit value indicates a threshold for rare allele frequency. 95% confidence interval is given for the N_e estimate.

	Females	Males	total	S	Pcrit value	N_e	95% CI Lower	95% CI Upper
2008	13	14	27	25	0.021	Infinite	232	Infinite
2009	12	11	23	21	0.024	Infinite	228.1	Infinite
2010	7	14	21	20	0.026	1496.6	111.4	Infinite
2011	17	25	42	41	0.013	Infinite	1383.5	Infinite
2012	8	9	17	14	0.036	1529.4	77.2	Infinite
pooled	57	73	130	121	0.005	Infinite	3924	Infinite

These point estimates of N_e were 'infinite' for several of the broodstock samples. In an effort to achieve an unbiased estimate of the N_e, the sampling error based on sample size is subtracted from the raw linkage disequilibrium parameter estimate. If the actual contribution of the sampling error is larger than the expected error, this can lead to a negative N_e value, which is interpreted as an estimate of infinite N_e (Do et al. 2014). Therefore, the infinite point estimate does not necessarily mean that the N_e is approaching infinity, instead it simply means that any linkage disequilibrium observed can be explained by sampling error. As a result, the 95% confidence interval, especially the lower bound, is often more informative than the point estimate. Based on the lower bound of the N_e estimate of the pooled sample, the effective population size for the 'pooled parent group' of the Celtic Sea population is at least 3924 individuals, with 95% confidence.

2.3.2. Survival data sets

Genotype reconstruction of 2009 wild broodstock

In 2009, it was discovered that tissue sample material was not available to confirm the genotypes of two cod (males 7 & 10) so they had to be reconstructed from available progeny data using PEDIGREE.

The 96 eggs from cones I7, O11 (48 eggs) and I10, M11 (48 eggs) were all successfully extracted and genotyped for 13 markers. The other parents contributing to the eggs in the selected cone (Male 11, Females I, O and M) had no missing genotypes for the 13 loci.

The first runs with PEDIGREE resulted in nine groups being identified in both I7, O11 and I10, M11 datasets. Further optimisation by augmenting the weight, temperature and number of iterations reduced the number of groups to four for the I7, O11 dataset and to five I10, M11 dataset. When these "best" full sibling family groups were nested into half sibling groups two groups were seen for the I7, O11 dataset and three groups for the I10, M11 dataset. This allowed the

inspection of individuals with possible genotyping errors or new mutations. In the combined dataset, six individuals had mismatches in one to three loci. One offspring did not match any of the expected parents, but when all the parents of the year-class were included in the parentage assignment it was identified as belonging to the Q15 family. This individual was removed from the analysis and genotypes of the other offspring with mismatches were modified accordingly to allow for their inclusion in the families they matched in the remaining loci. As a result, the I7 family had seven offspring, O11 family had 41, I10 family had 11 and M11 had 36 offspring.

The reconstruction of missing parental genotypes was successful. There were inconsistencies only in the marker *GADM*3 which is difficult to confidently score because of intense stuttering bands. An example of the reconstruction of genotype for Male 7 is shown in **Table 9**.

Table 9 Reconstruction of genotype of Male 7 using PEDIGREE. All possible parental genotypes that could have produced the observed offspring genotypes are shown for each locus. Genotype of Female I is underlined and genotype of Male 7 is in red.

Locus	Possible parents	chi^2	Offspring	#Expected	#Observed
GADM 3	175_187 x 165_191	3.333	187_191	1.5	2
			175_191	1.5	3
			165_175	1.5	1
			165_187	1.5	0
Gmo 2	107_115 x 109_125	6	115_125	1.5	0
			109_115	1.5	1
			107_109	1.5	4
			107_125	1.5	1
Gmo 8	138_314 x 130_138	3.333	130_314	1.5	0
			138_314	1.5	1
			138_138	1.5	3
			130_138	1.5	2
Gmo-C 18	146_155 x 158_161	6	146_158	1.5	1
			155_158	1.5	1
			146_161	1.5	0
			155_161	1.5	4
Gmo 19	142_146 x 158_194	3.333	146_158	1.5	0
			142_194	1.5	2
			146_194	1.5	1
			142_158	1.5	3
Gmo 34	100_100 x 100_100	0	100_100	6	6
	100_100 x 100_XXX	6	100_XXX	3	0
			100_100	3	6
	100_XXX x 100_XXX	18	100_XXX	3	0
			100_100	1.5	6
			XXX_XXX	1.5	0
Gmo 35	128_131 x 137_140	2	131_140	1.5	1
			128_137	1.5	1
			131_137	1.5	1
			128_140	1.5	3
Gmo 37	260_308 x 260_280	0.667	260_280	1.5	2
			280_308	1.5	1
			260_308	1.5	2
			260_260	1.5	1

Locus	Possible parents	chi^2	Offspring	#Expected	#Observed
Gmo-C 52	283_283 x 289_298	0.667	283_289	3	2
			283_298	3	4
	283_298 x 283_289	7.333	289_298	1.5	0
			283_289	1.5	2
			283_298	1.5	4
			283_283	1.5	0
	283_298 x 289_298	7.333	289_298	1.5	0
			283_289	1.5	2
			298_298	1.5	0
			283_298	1.5	4
	283_289 x 289_298	7.333	289_289	1.5	0
			289_298	1.5	0
			283_289	1.5	2
			283_298	1.5	4
	283_XXX x 289_298	7.333	283_289	1.5	2
			283_298	1.5	4
			289_XXX	1.5	0
			298_XXX	1.5	0
PGmo 76	184_186 x 206_232	3.333	184_232	1.5	2
			186_206	1.5	0
			186_232	1.5	3
			184_206	1.5	1
PGmo 94	274_274 x 276_278	0.667	274_278	3	4
			274_276	3	2
	274_278 x 274_276	7.333	274_278	1.5	4
			276_278	1.5	0
			274_274	1.5	0
			274_276	1.5	2
	274_278 x 276_278	7.333	274_278	1.5	4
			278_278	1.5	0
			276_278	1.5	0
			274_276	1.5	2
	274_276 x 276_278	7.333	274_278	1.5	4
			276_276	1.5	0
			276_278	1.5	0
			274_276	1.5	2
	274_XXX x 276_278	7.333	274_278	1.5	4
			276_XXX	1.5	0
			278_XXX	1.5	0
			274_276	1.5	2

Locus	Possible parents	chi^2	Offspring	#Expected	#Observed
Gmo 132	134_144 x 134_136	3.333	134_134	1.5	0
			134_136	1.5	2
			134_144	1.5	1
			136_144	1.5	3
	134_144 x 136_144	3.333	134_136	1.5	2
			134_144	1.5	1
			144_144	1.5	0
			136_144	1.5	3
	134_136 x 136_144	3.333	134_136	1.5	2
			134_144	1.5	1
			136_136	1.5	0
			136_144	1.5	3
Mae 211	150_192 x 152_158	3.333	150_158	1.5	1
			150_152	1.5	0
			158_192	1.5	2
			152_192	1.5	3

Choice of markers for parental allocation

For the 2008 wild broodstock fish, the discriminatory power of individual markers ranged from 65.4% (highest) for *GADM*3 and 4.3% (lowest) for *Gmo*-C52. For the 2009 wild broodstock, the discriminatory power ranged from 72.2% for *GADM*3 and 2.9% for *Gmo*35. For the 2010 wild broodstock the discriminatory power ranged from 84.6% for *GADM*3 and 5.7% for *Gmo*35. For each year, a selection of 3-marker combinations allowed for 100% assignment success. The final loci combinations for parentage analysis were chosen based on non-overlapping allelic size ranges on electrophoresis and clear amplification patterns of loci, allowing for unambiguous genotyping (see **Table 5** for 2008 F1, 2009 F1 and 2010 F1).

Parentage analysis for 2008 Cohort

The random sample of 96 fish at harvest was successfully genotyped for the four selected loci (*Gmo*19, *Gmo*35, *Gmo*37 and *Gmo*132) with no missing genotypes. There was one missing genotype in the 2008 cohort broodstock fish (F1) (1 out of 436 genotypes, or 0.2%; 4 loci X 109 individuals). The wild parents of 2008 (P0; n =

26) were free of missing genotypes for the selected loci. Both the 96 harvested individuals and 109 broodstock fish had a 100% single assignment with perfect match. The offspring were assigned to only 7 families (**Table 10**) of a possible 25 in the putative fertilisation matrix; 5/9 full sib families and 2/16 half sib families.

Parentage analysis for 2009 Cohort

After the reconstruction of genotypes for males 7 and 10, the 2009 wild broodstock (*n*=21) had three missing genotypes (two in *GADM*3 and one in *Mae*211). In addition to the 96 eggs successfully genotyped for the full suite of 13 markers, the 288 eggs from additional cones were extracted and genotyped for only three markers (*Gmo*8, *Gmo*19 and *Gmo-C*52). Out of the 288, two individual eggs were excluded because of suspected cross-sample contamination. The contamination was evident due to the presence of 3 PCR products in some loci. The observed alleles were compatible with the parental alleles of the wild broodstock fish that had contributed to the eggs in the cone. This indicates that the contamination is likely to be from broken eggs in the cone during the storage of the sample.

Of the 17 original crosses, three families did not survive to the incubation phase, leaving 14 families in the incubation cones. The set of eggs for survival investigation (n=382, 3 loci) had no missing genotypes. All the 251 2-year-old fish were successfully extracted and genotyped for the same three markers (*Gmo*8, *Gmo*19 and *Gmo-C*52).

With the chosen marker set, 98.7% of the 382 eggs could be assigned to parents using the exclusion method in VITASSIGN. The assignment was 100% when one mismatch was allowed. There were five samples with one mismatch: one of them was found to be a typographic mistake. After this mistake was corrected, the assignment rate was 99.0%. One individual had a mismatch in *Gmo*8 locus and three in *Gmo*19. These mismatches resulted most likely from null alleles because the non-matching offspring genotypes were all homozygotes for an allele from one

of their parents. All 14 families were present at egg stage with representation varying between 0.7 and 17.1% (**Table 11** and **Figure 4**).

A total of 94.8% of the 2-year-old fish were assigned while 5.2% (13 fish) were left unassigned. There were non-Celtic Sea fish in the same cage, and these individuals did not fit the parents even when one mismatch was allowed. It was concluded that these 13 fish did not originate from the Celtic Sea families but rather came from the smaller contingent of Norwegian stock that had been added as an operational decision on the farm. Eight families were present at 2 years, with their representation varying from 0.4 to 72.3% (**Table 11** and **Figure 4**). The survival of families at three different time points (fertilisation, pre-hatch and 2 years) is shown in **Figure 4**. The number of families present against the logarithm of time is shown in **Figure 5** which shows a steady decline over time.

Parentage analysis for 2010 Cohort

With the 2010 cohort, sampled at 1-year-old, 100% of fish were assigned without mismatches. Of these, 92.6% were single assignments to a broodstock pair in the original fertilisation matrix. The remaining 7.4% matched a pairing in the fertilisation matrix and another pairing not designated in the matrix. As these offspring had a hit to an existing family, the second pairing was considered improbable and these fish were assigned to the pairing specified in the fertilisation matrix. Ten families out of the initial thirteen were present (**Table 12**). All the families were not present in all the size classes. The survival of families across the size classes at two time points (fertilisation & 1 year) is shown in **Figure 6**.

Table 10 Parentage analysis of 2008 cohort. Number of individuals per family in the 2008 F1 cohort at three years of age; (A) broodstock fish, n = 109, (B) random sample of harvest fish, n = 96, (C) combined data set, n = 205. Female parents are in the first row and male parents are in the first column. The light grey boxes represent all family crosses created by strip spawning; the darker grey boxes with numbers represent families present at the time of sampling and the numbers correspond to individuals sampled per family.

(A)	A	B	C	D	E	F	G	H	I	J	K	L	M
1	4												
2		43											
3			17										
4													
5													
6						20							
7													
8													
9									10				
10											14		
11													
12													
13										1			

(B)	A	B	C	D	E	F	G	H	I	J	K	L	M
1	2												
2		48											
3			6										
4													
5													
6						14							
7													
8													
9									11				
10											13		
11													
12													
13										2			

(C)	A	B	C	D	E	F	G	H	I	J	K	L	M
1	6												
2		91											
3			23										
4													
5													
6						34							
7													
8													
9									21				
10											27		
11													
12													
13										3			

Table 11 Parentage analysis of 2009 cohort. Number of individuals per family in the 2009 F1 cohort; (A) at pre-hatch stage, n = 384, (B) at two years of age, n = 238. Female parents are in the first row and male parents are in the first column. The light grey boxes represent all family crosses created by strip spawning; the darker grey boxes with numbers represent families present at the time of sampling and the numbers correspond to individuals sampled per family. *: broodstock fish with reconstructed genotypes

(A)	A	B	C	E	F	I	J	M	N	O	P	Q
1	47											
2		9	39									
4												
5												
7 *						7						
8						43	4					
9						32	15					
10 *						11						
11								78	20	42		
14												27
15												8

(B)	A	B	C	E	F	I	J	M	N	O	P	Q
1	47											
2			2									
4												
5												
7 *												
8												
9						1	4					
10 *						1						
11								172	3			
14												
15												8

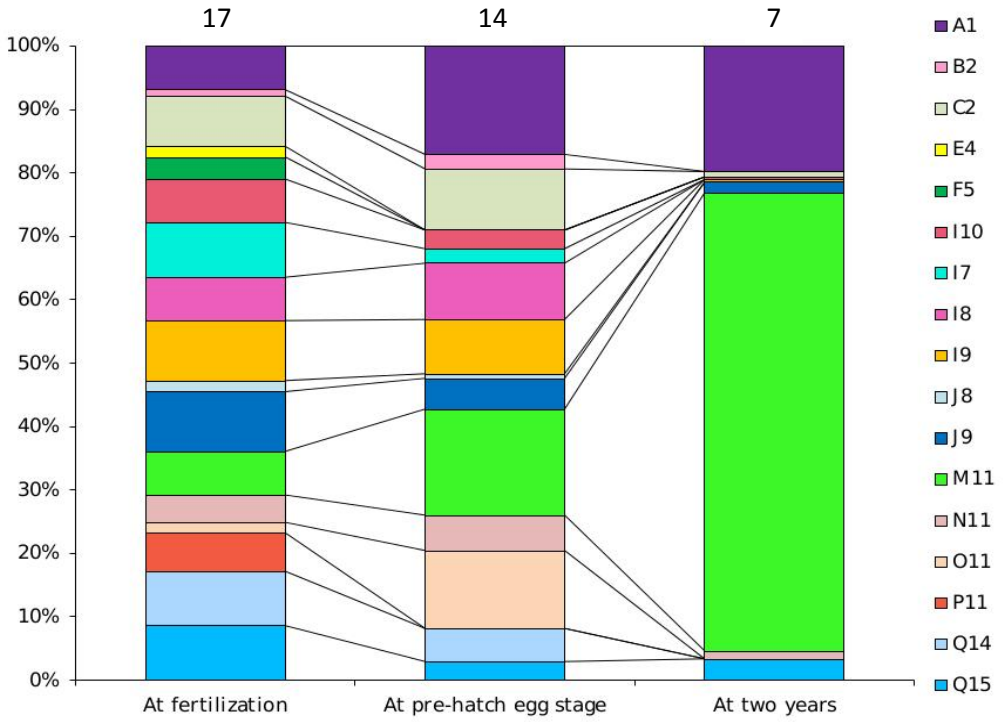

Figure 4 The relative survival of the 2009 cohort F1 families at three different time points. Relative % at fertilization is calculated from weight of fertilized eggs, others are calculated from numbers of sampled individuals. The number of families present is shown on top of each bar.

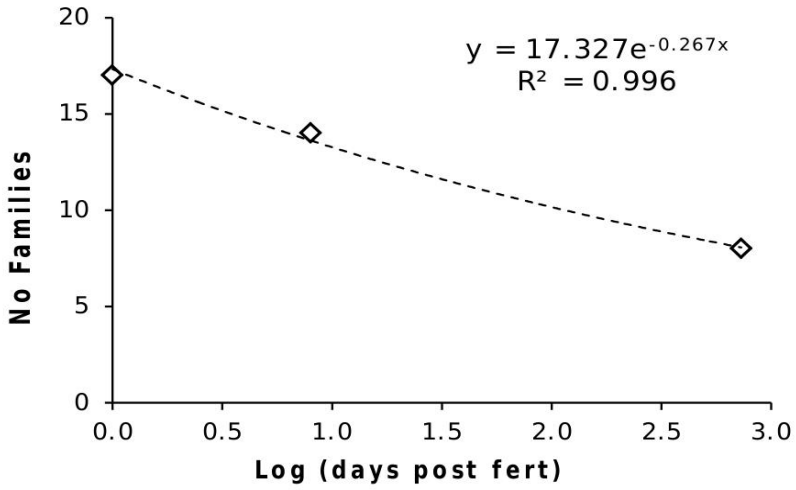

Figure 5 Survival curve for 2009 cohort. Number of families present against the logarithm of time (days post fertilization).

Table 12 Parentage analysis of 2010 cohort. Number of individuals per family in the 2010 F1 cohort at the age of one year; (A) size grade small, n = 94 (B) size grade medium, n = 94, (C) size grade large, n = 94 and (D) all the size grades combined n = 282. Female parents are in the first row and male parents are in the first column. The light grey boxes represent all family crosses created by strip spawning; the darker grey boxes with numbers represent families present at the time of sampling and the numbers correspond to individuals sampled per family.

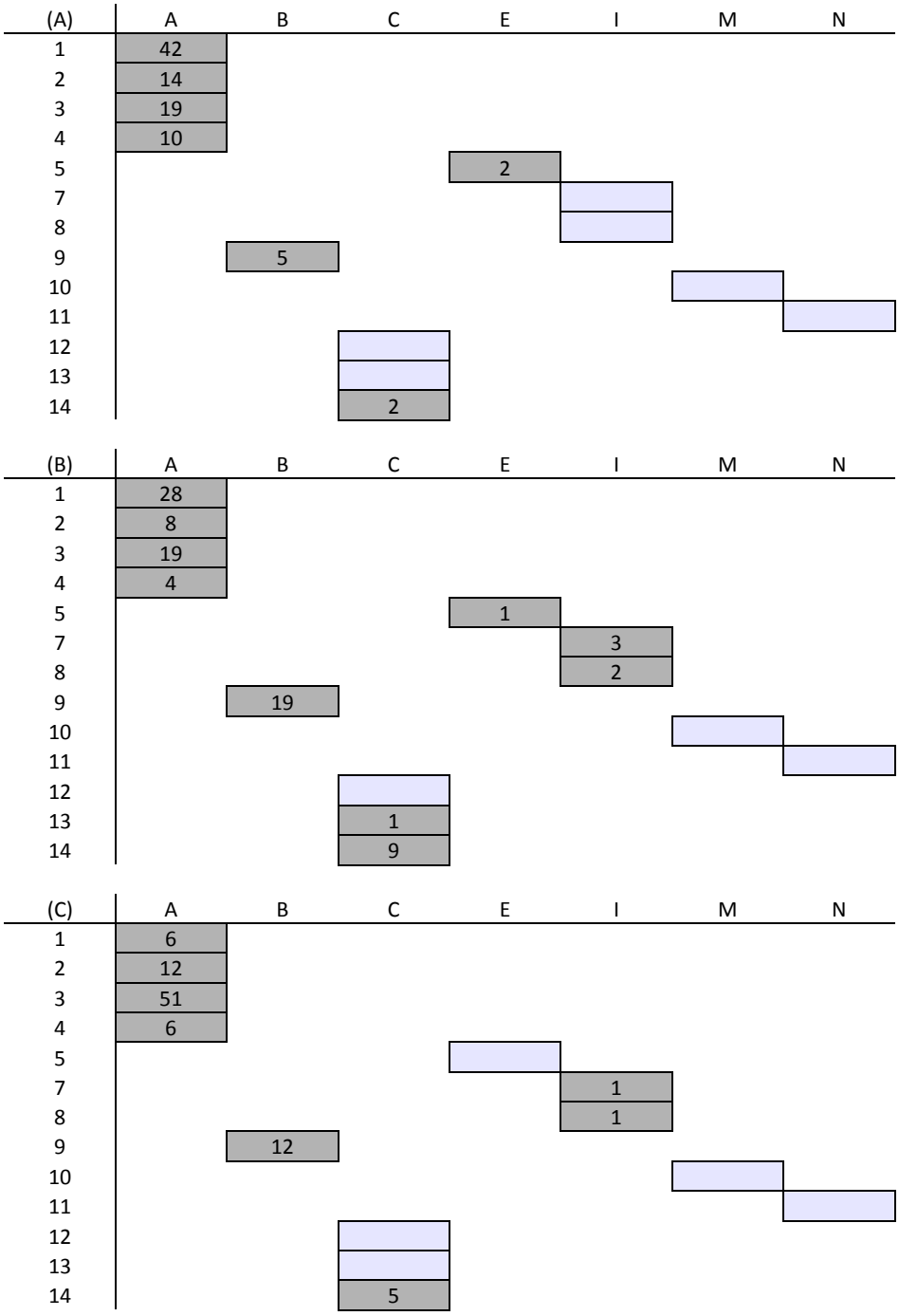

Table 12 *Continued.*

(D)	A	B	C	E	I	M	N
1	76						
2	34						
3	89						
4	20						
5				3			
7					4		
8					3		
9		36					
10							
11							
12							
13			1				
14			16				

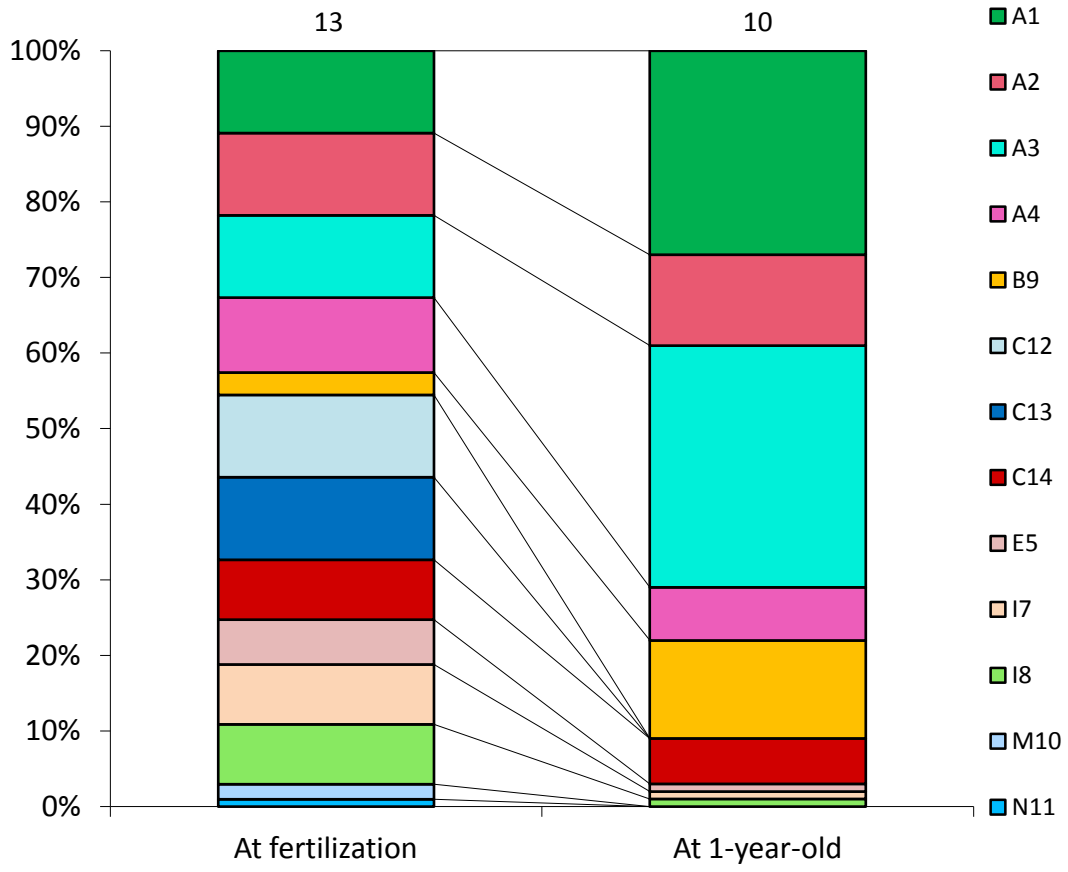

Figure 6 The relative survival of 2010 cohort F1 families at fertilization and at 1-year-old. Relative % at fertilization is calculated from weight of fertilized eggs and at 1-year-old from numbers of sampled individuals. The number of families present is shown on top of each bar.

A summary of number of families present at fertilization and at the time of final sampling is shown in **Table 13** along with the normalized median family size and percentage loss of families during rearing. The average family size and standard deviation of family size for each cohort at the time of final sampling is shown in **Figure 7**.

Table 13 Number of families present at fertilization and at the time of final sampling. Percentage loss of families and normalized median family size are shown. A range of values is given for the 2008 F1 cohort as it consisted of 9 full sib families and a 4 x4 mix of eggs and milt. However, it is not likely that all 16 families from the 4 x 4 cross were successful, so a range of 2 to 16 is given for the mix. Median normalised family size is based on a subsample of 200 offspring.

	2008 F1 cohort	2009 F1 cohort	2010 F1 cohort
Families present at fertilization	11-25	17	13
Families present at sampling	7	8	10
% loss of families	36-72%	53%	23%
Median normalised family size	21	2.5	12.5

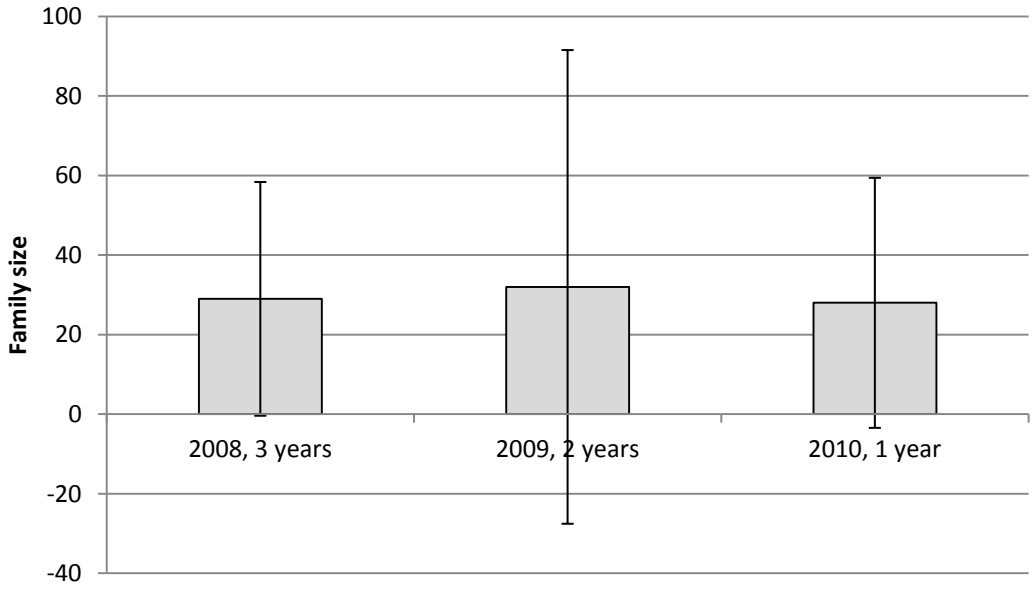

Figure 7 Average family size for each cohort at the time of final sampling. The error bars denote the standard deviation of family size.

N_e estimates from survival data

The three N_e measures calculated for the survival samples 2008, 2009 and 2010 are shown in **Table 14**. The average number of progeny per parent was 34.5 and the variance in the number of progeny per parent was 2525.1, in the combined data set including all of the survival samples (2008, 2009 and 2010). On average, 60% of the parents contributed at moment of final sampling in the F1 cohorts 2008-2010. These values were used in order to estimate N_e measures over the year classes 2008-2012 to assess the N_e of the EIRCOD farmed stock (**Table 15**). In the scenario of 100% wild broodstock contribution, the N_{e_sex} (equal family size) is 124.6. If the number of broodstock contributing to the offspring is reduced to 60%, the N_{e_sex} is lowered to 73.5. When variation in family size is taken into account in the N_e measure, N_{e_both}, the estimates are further reduced to 40.3 and 23.7, for 100% and 60% broodstock contribution, respectively. In the 2009 F1 cohort data set, the number of contributing parents and the effective population size (N_{e_both}) can be calculated at different sampling points. Both the number of contributing parents and the effective population size (N_{e_both}) are reduced over time, from 23 to 13 and 23.0 to 3.1, respectively; but the N_e decreases more severely (**Figure 8)**. When the N_e estimates using the sibship method with both random mating and non-random mating are compared to the N_e measures (N_{e_sex}, N_{e_var} and N_{e_both}) for each F1 cohort (2008-2010), they produce similar values (**Table 16**).

As would be expected using wild broodstock, and despite the significant loss of families, the inbreeding coefficient was 0 for all survival data sets.

Table 14 Three N_e measures calculated for the survival samples. Females and Males indicate the number of breeders, N_f and N_m are the number of females and males that contributed to the families present at the given time point; k is the average number of progeny; σ^2_k is the variance in the number of progeny; k_f, k_m and σ^2_{kf}, σ^2_{km} are the average and variance for progeny per sex; N_{e_sex} is the N_e resulting from deviations in the sex ratio; N_{e_var} is the N_e resulting from variance in reproductive success; N_{e_varf} and N_{e_varm} are N_{e_var} for each sex; N_{e_both} is the N_e accounting for both variance in reproductive success and deviations in sex ratio; N_b is the number of breeders at fertilization; N_{ab} is the actual number of contributing breeders at the time point of sampling; N_{ab}/N_b is the ratio of actual breeders to the initial breeders at the timepoint of sampling.

Year class, Time point	Females	Males	N_f	N_m	k	σ^2_k	k_f	σ^2_{kf}	k_m	σ^2_{km}	**$N_{e\ sex}$**	**$N_{e\ var}$**	$N_{e\ varf}$	$N_{e\ varm}$	**$N_{e\ both}$**	N_b	N_{ab}	N_{ab}/N_b
2008 F1, 3 years	13	13	7	7	29.3	796.5	29.3	862.9	29.3	862.9	14.0	7.4	3.5	3.5	7.1	26	14	0.54
2009 F1, pre-hatch	12	11	9	9	42.4	1138.8	42.4	760.0	42.4	1660.0	18.0	11.2	6.4	4.7	10.9	23	18	0.78
2009 F1, 2 years	12	11	7	6	36.6	3949.1	34.0	3966.3	39.7	4697.5	12.9	3.3	1.6	1.5	3.1	23	13	0.57
2010 F1, 1 year	7	13	5	10	37.7	3228.4	56.6	8405.3	28.2	987.5	13.3	4.6	1.4	4.5	4.2	20	15	0.75
2008-2010 F1 combined	**32**	**37**	**19**	**23**	**34.5**	**2525.1**	**38.2**	**3609.4**	**31.5**	**1731.6**	**41.6**	**13.6**	**5.5**	**8.5**	**13.3**	**69**	**42**	**0.61**

Table 15 Three N_e measures estimated over year classes 2008-2012. The k and σ^2_k calculated based on the combined survival set of 2008-2010 from **Table 14** (k=34.5; σ^2_k=2525.1) are used for k and σ^2_k and k_f, k_m and σ^2_{kf}, σ^2_{km}, respectively. In the upper part, 100% of parents are contributing to the progeny at the moment of sampling and in the lower part 60% of parents are contributing to the progeny at the moment of sampling. The terminology is as in **Table 14**, except $N_{e\_both\_act}$ are the actual calculated values from **Table 16** for comparison

	Year class	Females	Males	N_f	N_m	k	σ^2_k	k_f	σ^2_{kf}	k_m	σ^2_{km}	**N_{e_sex}**	**N_{e_var}**	N_{e_varf}	N_{e_varm}	**N_{e_both}**	**$N_{e\_both\_act}$**
100% of parents contribute	2008	13	13	13	13	34.5	2525.1	34.5	2525.1	34.5	2525.1	26.0	8.4	4.2	4.2	8.4	**7.1**
	2009	12	11	12	11	34.5	2525.1	34.5	2525.1	34.5	2525.1	23.0	7.4	3.9	3.6	7.4	**3.1**
	2010	7	13	7	13	34.5	2525.1	34.5	2525.1	34.5	2525.1	18.2	6.5	2.3	4.2	5.9	**4.2**
	2011	17	25	17	25	34.5	2525.1	34.5	2525.1	34.5	2525.1	40.5	13.6	5.5	8.1	13.1	
	2012	8	9	8	9	34.5	2525.1	34.5	2525.1	34.5	2525.1	16.9	5.5	2.6	2.9	5.5	
											sum	124.6			sum	40.3	

	Year class	Females	Males	N_f	N_m	k	σ^2_k	k_f	σ^2_{kf}	k_m	σ^2_{km}	**N_{e_sex}**	**N_{e_var}**	N_{e_varf}	N_{e_varm}	**N_{e_both}**	**$N_{e\_both\_act}$**
60% of parents contribute	2008	13	13	7.7	7.7	34.5	2525.1	34.5	2525.1	34.5	2525.1	15.3	5.0	2.5	2.5	5.0	**7.1**
	2009	12	11	7.1	6.5	34.5	2525.1	34.5	2525.1	34.5	2525.1	13.5	4.4	2.3	2.1	4.4	**3.1**
	2010	7	13	4.1	7.7	34.5	2525.1	34.5	2525.1	34.5	2525.1	10.7	3.8	1.3	2.5	3.5	**4.2**
	2011	17	25	10.0	14.8	34.5	2525.1	34.5	2525.1	34.5	2525.1	23.9	8.0	3.2	4.8	7.7	
	2012	8	9	4.7	5.3	34.5	2525.1	34.5	2525.1	34.5	2525.1	10.0	3.2	1.5	1.7	3.2	
											sum	73.5			sum	23.7	

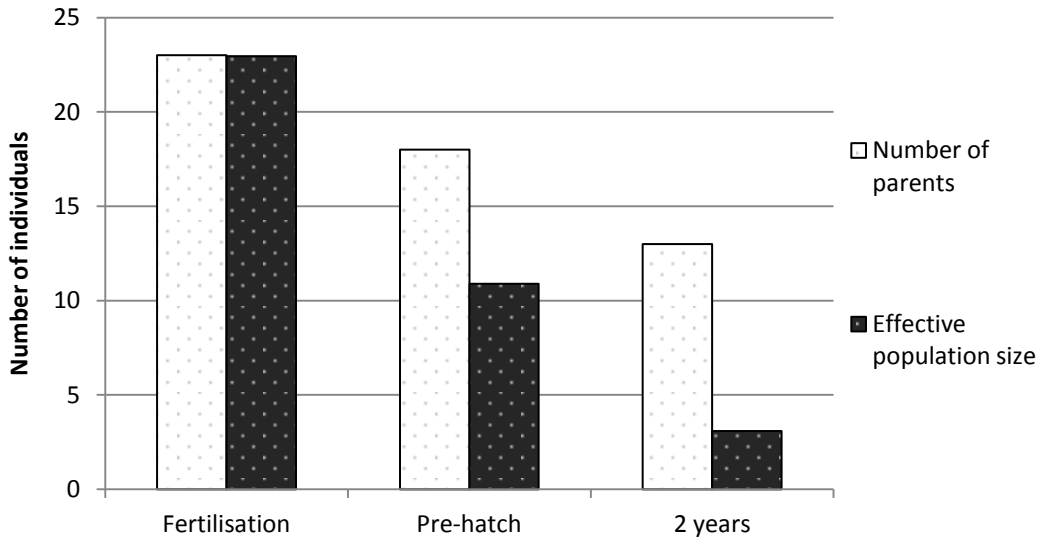

Figure 8 The effective number of parents and the effective population size of the 2009 F1 cohort from fertilization to 2 years of age. The N_e is calculated using the formula 3 (N_{e_both}) taking into account variance in reproductive success and deviations in sex ratio.

Table 16 Three different N_e measures, N_{e_sex}, N_{e_var} and N_{e_both}, and N_e estimates based on the sibship method and the linkage disequilibrium (LD) method. N_{e_sex}, N_{e_var} and N_{e_both} values are from **Table 14**; N_{e_rand} is calculated assuming random mating and $N_{e\_non\text{-}rand}$ is calculated assuming non-random mating; 95% CI signifies the 95% confidence interval.

	Formulas			Sibship method				LD method	
Year class, Time point	N_{e_sex}	N_{e_var}	N_{e_both}	N_{e_rand}	95% CI	$N_{e\_non\text{-}rand}$	95% CI	N_{e_LD}	95% CI
2008 F1, 3 years	14.0	7.4	7.1	8.0	4 - 24	8	4 - 24	5.5	3.8 - 7.3
2009 F1, pre-hatch	18.0	11.2	10.9	11	6 - 26	12	7 - 26	11.5	9.4 - 13.8
2009 F1, 2 years	12.9	3.3	3.1	4	2 - 20	5	2 - 20	8.4	6.6 - 10.3
2010 F1, 1 year	13.3	4.6	4.2	5	2 - 20	11	6 - 28	10.4	8.3 - 12.6

2.3.3. Mass spawning

The 2008 and 2007 broodstock fish (n=172) were successfully genotyped for 9 microsatellite loci (see **Table 5**). There was one missing genotype for *Gmo*34, *Gmo*-C18 and *Gmo*-C52, and eight for *PGmo*94. MICRO-CHECKER analysis indicated an excess of homozygotes at *Gmo*2 locus. A possible cause for this is the presence of large number of related individuals in the 2008 cohort. Indeed, when the number of individuals from the same 2008 family was restricted to a maximum of two, there was no evidence of homozygote excess. Based on sexing by visual inspection there were 69 females, 95 males and 8 fish with indeterminable sex.

The PCR amplification of egg DNA was poor which made genotyping of the eggs challenging. Gel electrophoresis of the genomic DNA revealed that the quantity of DNA was low and when DNA was visible it was fragmented, hence, of poor quality. Similarly, three loci (*Gmo*35, *Gmo*37 and *Gmo*132) did not amplify consistently for the eggs. After excluding any offspring with more than two missing genotypes, the genotyping was successful for 125 eggs (59 for sample 1 and 66 for sample 2) and using six loci. The resulting data set had three missing genotypes for *Gmo*2, thirty-four for *Gmo*19, two for *Gmo*34, seven for *Gmo*-C18, five for *Gmo*-C52 and two for *PGmo*94. MICRO-CHECKER analysis indicated homozygote excess at *Gmo*2 with stuttering or null alleles as possible causes. However, as noted earlier, this could also be caused by the presence of related individuals.

To resolve the parentage of the individuals resulting from the mass spawning based on the available genotype data, a series of reconstruction protocols were attempted using two software packages; VITASSIGN and COLONY. The outputs and findings from these evaluations are presented in the following accounts.

VITASSIGN results

Genotypic exclusion using VITASSIGN allowed the resolution of parental origin of only 12.8% of offspring with a single assignment with perfect match, leaving 50.4%

of offspring unassigned. When one mismatch was allowed, additional 17.6% of offspring obtained a single assignment with one mismatch, and 12.8% were left unassigned. The remaining 56.8% of offspring matched multiple parent pairs with either perfect match (36.8%) or one mismatch (20%). The number of possible parent pairs with perfect match ranged from 1 to 327, with an average of 11.4. The number of possible parent pairs with one mismatch ranged from 1 to 790, with an average of 28.2.

COLONY results with known sex of parents

The first COLONY run was carried out with known sex of parents based on visual sexing of broodstock fish. COLONY divided the offspring into 72 full sib families with the probability of inclusion >80%. 59 of the families had only one member, the remaining 13 ranged from 2 to 20 in family size. Out of these 72 families, 9 had missing paternal parent and 30 had missing maternal parent when the most likely parent pair was examined: two families had both parents missing. One parent with undefined sex was being identified as a father for some individuals and a mother for some individuals. As the visual sexing may not have been very reliable, the parentage analysis using COLONY was repeated with unknown sex of parents and without any parental information.

COLONY results with parents of unknown sex

A total of 73 full sib families were identified. 57 of the families had only one member, the remaining 16 ranged from 2 to 20 in family size. When the most likely parent pair was examined, 43 out of the 73 families had a missing paternal parent, 38 had a missing maternal parent and 28 families had both parents missing.

COLONY results without parental information

When no parental information was prescribed, the programme identified 88 full sib families. 79 of these were single offspring families and 9 had multiple offspring

ranging from 2 to 20 offspring. COLONY identified 31 mothers and 39 fathers. COLONY reconstructed the genotypes of all possible parents and assigned likelihood scores for each genotype. The genotypes with the highest likelihoods were chosen for each locus and this genotype was compared to the actual broodstock genotypes. There were two matches, Male 32 with most offspring (*n*=21) matched a 2007 broodstock fish visually sexed as a male (2011_006_009M) and Female 6 with most offspring (*n*=41) matched a 2007 cohort broodstock fish visually sexed as a male (2011_006_023M). These two fish were the parents of the largest full sib family of 20 individuals. **Figures 9** and **10** show the numbers of offspring per reconstructed parent, for females and males, respectively.

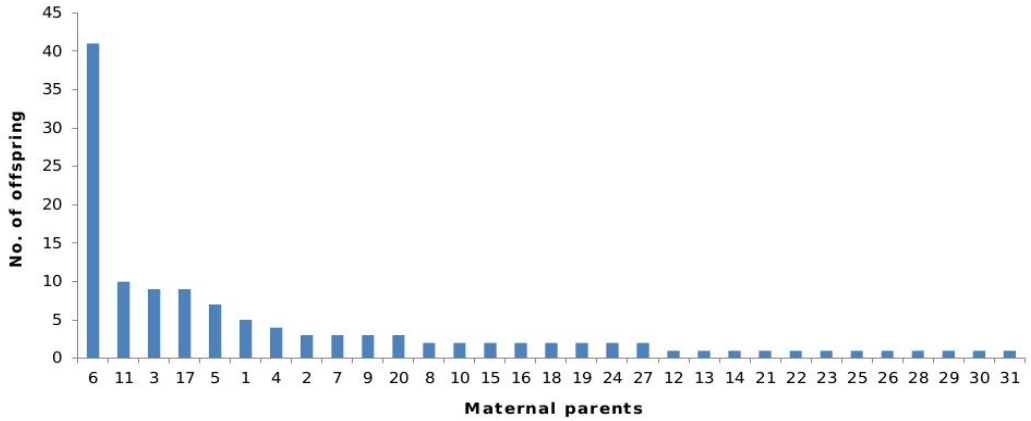

Figure 9 Number of offspring per maternal parent. The maternal parents are reconstructed from the COLONY run with no parental information. The number below the x-axis represents the identity of the reconstructed parent.

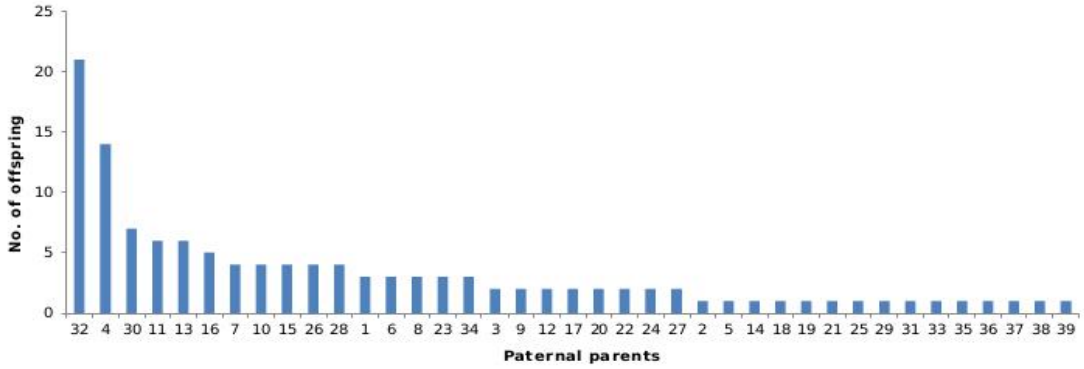

Figure 10 Number of offspring per paternal parent. The paternal parents are reconstructed from the COLONY run with no parental information. The number below the x-axis represents the identity of the reconstructed parent.

Comparing the COLONY results

When families with more than one offspring resulting from the three COLONY approaches were compared, many of the families were identified by more than one approach and sometimes by all three approaches (**Table 17**). The VITASSIGN parentage assignment of the offspring in the families identified by more than one COLONY approach was inspected. For three families, the parent individuals identified by COLONY were also found in the matching parent pairs identified by VITASSIGN (**Table 18**).

Table 17 Comparison of families identified by the different approaches in COLONY. Only families with more than one offspring are shown. The families identified by more than one approach are marked by coloured background. T33 in the offspring name refers to Sample 1 of sampling and T41 to Sample 2.

Sex of parent known		Sex of parent unknown		No parental information	
Family_01	T33egg001 T33egg033 T33egg058 T33egg065 T33egg077	Family_01	T33egg001 T33egg058		
		Family_41	T33egg065 T33egg077		
Family_08	T33egg011 T33egg018 T33egg026 T33egg037 T33egg068 T33egg075	Family_08	T33egg011 T33egg018 T33egg026 T33egg037 T33egg068 T33egg075	Family_27	T33egg037 T33egg068 T33egg075
Family_10	T33egg015 T33egg053	Family_10	T33egg015 T33egg053		
Family_27	T33egg046 T33egg073	Family_27	T33egg046 T33egg073		
Family_38	T33egg063 T33egg069	Family_13	T33egg021 T41egg064	Family_14	T33egg021 T41egg064
Family_42	T33egg070 T41egg058	Family_11	T33egg016 T33egg040	Family_18	T33egg025 T41egg035
Family_03	T33egg004 T33egg049	Family_09	T33egg013 T33egg071		
		Family_49	T41egg001 T41egg075		
Family_48	T41egg002 T41egg004 T41egg008 T41egg011 T41egg018 T41egg022 T41egg032 T41egg047 T41egg052 T41egg055 T41egg063 T41egg070 T41egg073	Family_50	T41egg002 T41egg004 T41egg008 T41egg011 T41egg018 T41egg022 T41egg032 T41egg047 T41egg052 T41egg055 T41egg063 T41egg070 T41egg073	Family_59 Family_62	T41egg002 T41egg004 T41egg018 T41egg022 T41egg047 T41egg052 T41egg055 T41egg063 T41egg070 T41egg073 T41egg008 T41egg011 T41egg032

Table 17 *Continued.*

Sex of parent known		Sex of parent unknown		No parental information	
Family_49	T41egg003	Family_51	T41egg003	Family_60	T41egg003
	T41egg005		T41egg005		T41egg005
	T41egg006		T41egg006		T41egg006
	T41egg009		T41egg009		T41egg009
	T41egg012		T41egg012		T41egg012
	T41egg013		T41egg013		T41egg013
	T41egg017		T41egg017		T41egg017
	T41egg021		T41egg021		T41egg021
	T41egg024		T41egg024		T41egg024
	T41egg028		T41egg028		T41egg028
	T41egg033		T41egg033		T41egg033
	T41egg041		T41egg041		T41egg041
	T41egg043		T41egg043		T41egg043
	T41egg049		T41egg049		T41egg049
	T41egg051		T41egg051		T41egg051
	T41egg059		T41egg059		T41egg059
	T41egg062		T41egg062		T41egg062
	T41egg065		T41egg065		T41egg065
	T41egg066		T41egg066		T41egg066
	T41egg068		T41egg068		T41egg068
Family_50	T41egg007	Family_52	T41egg007	Family_61	T41egg007
	T41egg025		T41egg025		T41egg042
	T41egg042		T41egg042		
	T41egg057				
		Family_55	T41egg020		
			T41egg037		
Family_55	T41egg027				
	T41egg071	Family_56	T41egg023		
			T41egg034		
Family_59	T41egg035	Family_57	T41egg035		
	T41egg061		T41egg061		
Family_60	T41egg036	Family_58	T41egg036	Family_72	T41egg036
	T41egg038		T41egg038		T41egg048
	T41egg048		T41egg048	Family_17	T33egg024
	T41egg072		T41egg072		T41egg072

Table 18 VITASSIGN parentage assignment of the offspring in the families in common with more than one COLONY approach. The colours are the same as in **Table 17**. For the three families the parents identified by COLONY matched parent pairs identified by VITASSIGN. The individual codes of these parents are given in brackets.

Family	No. of offspring/ Day of Sampling	Male parent	Female parent
Family 1	5 Day 1	2007 cohort male (2011_006_058)	2007 cohort female (2011_006_001)
Family 2	6 Day 1	2008 cohort 2008 Family 10	2007 cohort
Family 3	2 Day 1	Unresolved	Unresolved
Family 4	2 Day 1	2007 cohort male (2011_006_051)	2007 cohort female (2011_006_032)
Family 5	13 Day 2	2007 cohort male	2007 cohort female
Family 6	2 Day 1&2	Unresolved	Unresolved
Family 7	20 Day 2	2007 cohort male	2007 cohort female
Family 8	4 Day 2	2007 cohort male	2007 cohort female
Family 9	2 Day 2	2007 cohort male (2011_006_012)	2008 cohort female (2011_011_007)
Family 10	5 Day 2	Unresolved	Unresolved

N_e estimates & Inbreeding

The N_e estimates calculated in COLONY with the sibship method with the three conditions (sex of the parents known, the sex of the parents unknown and with no parental information) are shown in **Figure 11**. The inbreeding coefficient was estimated to be 0.0353 when the parents were included, both sex known and sex unknown. When the putative parents were excluded from the analysis the inbreeding coefficient was 0.0385.

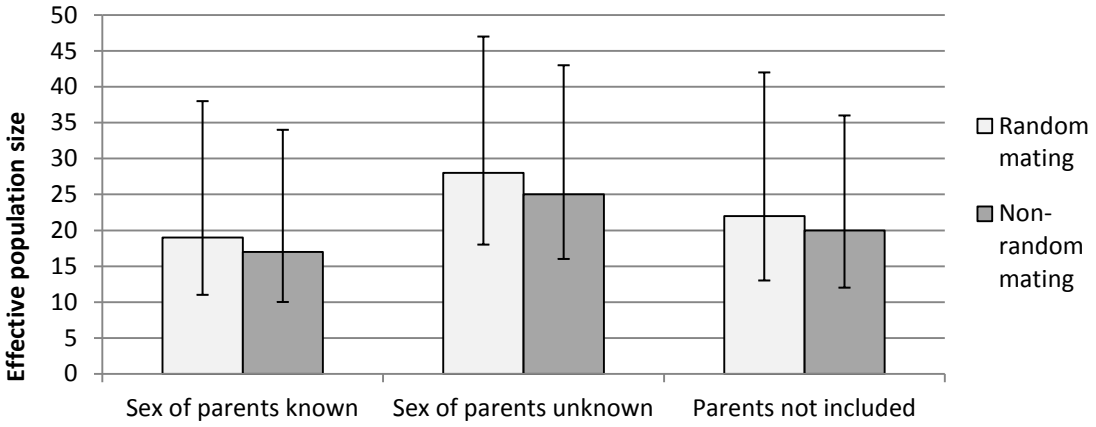

Figure 11 The N_e estimates calculated in COLONY with the sibship method with the sex of the parents known, the sex of the parents unknown and with no parental information. For each set, the N_e was estimated both with random mating assumed and non-random mating assumed. The error bars show the 95% confidence intervals of the estimates.

F_{IS} estimates

The per population F_{IS} estimates calculated using GENETIX 4.05 for the wild parents 2008-2012 (P0), the three F1 cohorts (2008, 2009 and 2010) and the F2 cohort (2011) are presented with 95% confidence intervals in **Table 19**. The values ranged from 0.04440 in 2009 P0 to -0.19765 in 2009 F1 cohort sampled at 2 years. For the positive values, the 95% confidence intervals included 0 indicating that these values were not signifigantly different from 0.

Table 19 Per population F_{IS} estimates for P0 wild parents 2008-2012, F1 cohorts 2008-2010 and F2 cohort 2011 with 95% confidence intervals.

	F_{IS}	95% Confidence interval
2008 P0	0.03790	-0.01359 - 0.05058
2009 P0	0.04440	-0.01195 - 0.05119
2010 P0	0.02413	-0.00818 - 0.02604
2011 P0	-0.00061	-0.03727 - 0.01189
2012 P0	0.00262	-0.08063 - 0.01881
2008 F1	-0.10635	-0.13106 - -0.09643
2009 F1, eggs	-0.07119	-0.10080 - -0.06416
2009 F1, 2-years	-0.19765	-0.21391 - -0.18908
2010 F1	-0.17027	-0.18913 - -0.15453
2011 F2	0.00465	-0.04073 - 0.03779

2.4. Discussion

2.4.1. Wild broodstock

The genotypes of the various wild broodstock parental groups from 2008 to 2012 were effectively characterised using the selected panel of 13 microsatellite markers (**Table 6**). The 13 markers were variable for all years and suitable for parentage analysis. In many cases, as few as three markers provided full assignment of the offspring of these parents. Pairwise F_{ST} values between the different years were all non-significant giving some indication that the sampled source population is temporally stable. These findings concur with Pampoulie et al. (2008) who found that the Celtic Sea spawning aggregations were temporally stable between spring 2002 and spring 2003. Even though the sample sizes were small in the present chapter, the temporal stability opinion is supported by the findings of Chapter 4. N_e estimate for the pooled sample was 'infinite' with a lower bound of the 95% confidence interval at 3924. According to the ICES stock assessment, the spawning stock biomass of the Celtic Sea stock was 17 206 tonne in 2014. If this is divided by the average weight of 3-year-old and 4-year-old Celtic sea cod, 4.3 kg and 7.3 kg respectively (Brander 1995), the census size can be very roughly estimated to be between 2 and 4 M individuals. With this crude estimate, the N_e/N ratio would be at least 0.001-0.002. These estimates are higher than the N_e/N ratio of 4×10^{-5}

estimated for North Sea cod (Hutchinson et al. 2003) and the average N_e/N ratio of $10^{-3.99}$ reported for marine organisms (Hauser and Carvalho 2008).

The large 95% confidence intervals around the N_e point estimates for different sampling years indicate that the method may not be very sensitive. However, the lower end of the confidence interval of nearly 4000 in the pooled sample gives a strong indication that the Celtic Sea population has large enough genetic base to retain its adaptive potential in the face of environmental changes according to the updated 100/1000 rule of Frankham et al. (2014). It is also approaching the 5000 limit suggested by Lande (1995) as the limit for retaining evolutionary potential. As broad genetic base is required for an aquaculture base population, the Celtic Sea stock is a suitable source for acquiring wild broodstock having satisfied these more stringent requirements.

2.4.2. Differences in survival

Drastic differences were observed over time in the survival between different family crosses created by strip spawning from these wild broodstock. The number of families was severely reduced from the time of fertilization to the time of sampling in all three year classes, and variation in family size was considerable.

The large differences in survival of families affected the N_e measures in each year class for which survival was assessed. The N_e estimates based on the sibship method were in concordance with the calculated N_e measures with N_{e_var} having the most similar values. Comparison of the different N_e measures showed that variance of the number of offspring produced by a parent affected the N_e much more than unequal sex ratios in broodstock. Similar observations have been made in an Atlantic halibut (*Hippoglossus hippoglossus*) breeding programme and steelhead trout (*Oncorhynchus mykiss*) hatchery stock (Jackson et al. 2003; Naish et al. 2013).

In order to assess how the high variability in family size and its impact on N_e affect the whole breeding programme, N_e measures were estimated for all five years from 2008 to 2012. The average family size and its variance from the three cohorts

sampled for survival was used in the estimates for all years to assess the accuracy of the estimates. The N_e estimates assuming 60% of parents contributing were closer to the N_e measures based on real data than the estimates assuming 100% contribution. The results have serious consequences for management of breeding programmes. In the case of EIRCOD, if equal parental contributions are presumed and N_e is estimated with N_{e_sex} based on the number and sex of broodstock, N_e of 126 is expected over the five years (2008-2012). However, when the loss of families and unequal family size are taken into account (N_{e_both}, 60% parents contributing), the N_e could be as low as 24.

Unequal survival and loss of specific families could have resulted from variability in the initial quality of the fertilised eggs. The quality of eggs could be affected by maternal effects, genetic effects associated with the particular combination of parents and environmental effects (Garber et al. 2010; Hansen et al. 2014). The use of strip spawning may also have affected the egg quality as stripping eggs at the suboptimal moment can cause immature or over-ripe eggs to be used (Lush et al. 2011). The strip spawning took place on board a ship with little possibility to monitor the optimal moment of stripping for individual fish. As a result, the timing of gamete collection may have been more optimal for some crosses than others.

The amount of sampled time points was limited in the present chapter, but it seems that biggest losses occur during egg and larval stages in early life. The egg and larval stages have been previously identified as bottlenecks in cod aquaculture (Kjesbu et al. 2006). Cod is a species with high mortality rates at egg and larvae stages in its natural environment. As a result, it will be challenging to increase the survival of offspring in these stages.

Pooling offspring for communal rearing at different developmental stages may affect the survival of families, possibly through competition and cannibalism between individuals and families. In the present study, the eggs were mixed during incubation with 1-3 families per incubation cone. After reaching an average size of 1 g, the juveniles were size graded and pooled. Garber et al. (2010) observed lower survival of families and lower numbers of individuals per family in families that

were mixed at the egg stage in comparison with families mixed at larvae and juvenile stages.

Low and variable survival rates among the families from pooling to the time of selection of parents have been noted as a potential issue in communal rearing (Gjerde 2005). Indeed, it has been suggested that in species with high mortality rates at egg and larval stages, full sib groups should be reared separated until the survival has stabilized (Bentsen and Olesen 2002). Communal rearing has been preferred over separate rearing as it requires more tank space and hinders comparisons between families due to confounding environmental effects (tank effects) (Speare et al. 1995). However, the unequal family survival and large variation in family size in the present study make possibilities for comparison between families very limited. In addition, with the serious implications of variable family size on N_e, increasing the survival of families should be the main priority in future EIRCOD cohorts. Hence, separate rearing of each family is recommended and the suitable stage of pooling should be investigated. The possibility of pooling a portion of each family in a communal tank to assess family effects should also be considered, however, due to variable survival this may not lead to generation of a useful data set unless the variation in survival can be reduced.

As biggest losses seem to happen at egg and larval stages it seemed appropriate to presume that survival had stabilised for the 2009 and 2010 cohorts, even though they were sampled at 2 and 1 years of age, respectively. It is of course possible that by 3 years more families will have been lost in these cohorts, especially those that were under represented, either through random chance, or husbandry practises, such as size grading. In the 2010 F1 cohort, the sample from the size grade medium had 2-3 more families present than the other two size grades (small and large). If breeders were to cull smaller size grades, families could be lost as a result. Indeed, Frost et al. (2006) have found that size grading had a significant negative effect on the genetic diversity of commercially reared barramundi (*Lates calcarifer*) juvenile cohorts.

It is evident that having 100 breeders or 50 breeding pairs as base population founders is not an adequate number to protect the farmed population from inbreeding. The results highlight the importance of continued genetic monitoring within a breeding programme. The use of 50 breeding pairs as a rule of thumb is especially harmful in hatcheries operated for restocking natural populations, as their very intent is to release these inbred fish into the wild were they hybridise with the local populations.

Mass spawning

Of the 9-loci marker set chosen for parentage assignment, three loci failed to amplify consistently reducing the final data set to six loci. The 6-loci data set was not sufficient to conclusively assign parentage in the mass spawning event. With the exclusion method of VITASSIGN parentage of only 17.6% of offspring could be unambiguously resolved. Herlin et al. (2007) had 77% single match assignment success in a cod mass spawning event that included 99 potential parents with a 5-loci dataset. The single match assignment success rose to 84% when three additional loci were genotyped resulting in an 8-loci dataset (Herlin et al. 2008). In contrast, when assignment was attempted with only four loci the assignment success was only 70% (Herlin et al. 2007). These studies indicate that the number of loci is an important factor in the achieved assignment rate. The parentage assignment in the current study was further complicated by the relatedness of the parents involved. Seven full sib families were present in the 2008 F1 component of the broodstock and unknown number of families in the 2007 F1 component. Resolving parentage in this kind of situation likely requires more loci than regularly used in parentage analysis.

Large skew in the reproductive success of broodstock was observed. On day 2, one pair produced 30% and another pair produced 20% of the offspring in the sample. On day 1, the skew was smaller with the largest families comprising 10% and 8% of the offspring in the sample. This is similar to the results of Herlin et al. (2008) who found that one parental pair contributed to 25% of the offspring in their sample and nine parents contributed to 90% of offspring in single day of mass spawning. High

variance in family size contributed to low N_e in gilthead seabream (*Sparus aurata*) reproduced by mass spawning (Brown et al. 2005). In a study by Bartley et al. (1992) N_e was estimated at 10 from progeny originating from a mass spawning tank with 25 white seabass (*Atractoscion nobilis*) broodstock using a linkage disequilibrium method.

In general, the 2007 F1 component of the broodstock was more successful in breeding than the 2008 F1 component. This indicates that second-time spawners were more successful at reproducing than first-time spawners perhaps due to their larger size and age. The N_e estimates of the mass spawning data set were likely to reflect the contribution of 2007 F1 fish from mixed background.

F2 production by mass spawning

The viability of the F2 eggs was low and few of them survived to become juveniles (data not shown). The poor quality of DNA from the eggs could be connected to the poor viability. The 2008 F1 component of the broodstock were first-time spawners, which could negatively affect egg quality. In addition, the broodstock was receiving regular pelleted feed and was not on a specialized broodstock diet. Commercial broodstock diet is enriched with vitamins A and E, and contains squid meal, for improved reproductive success (Hamoutene et al. 2013). Especially vitamin E has an important role in the control of reproduction in mammals and fish (Pavlov et al. 2004).

The mass spawning broodstock included several full sib families and mating between siblings would result in inbred offspring. Lower percent hatch rate has been recorded in inbred offspring resulting from F1 full sib matings in cod (Puckrin et al. 2013), meaning this could have been a contributor to the low viability. Similarly, one generation of brother–sister matings had a negative effect on survival of eyed eggs in rainbow trout (*Oncorhynchus mykiss*) (Gjerde et al. 1983). In addition, the frequency of crippled fry, feed conversion efficiency, fry survival and growth rate were all negatively affected in rainbow trout after just one generation of full sib matings (Kincaid 1976). Therefore, it is possible that inbreeding may have

contributed to the low survival of eggs and larvae in our case. The reductions in hatch rate after one generations of full sib mating observed by Puckrin et al. (2013) and Gjerde et al. (1983) were modest, 8.5% and 8.9% respectively, and it is likely that for example the broodstock diet was a major factor in the low survival of eggs and larvae. The lacking broodstock diet and inbreeding may also have caused compounded effects on egg survival.

Broodstock diet has been identified as one of the main factors affecting egg quality in cod (Thorsen et al. 2003; Hamoutene et al. 2013). Commercially available broodstock diets may also need re-evaluation as broodstock fed a diet supplemented with bait fish produced better quality eggs than broodstock fed a pelleted broodstock diet (Hamoutene et al. 2013). Broodstock nutrition and management in order to maintain high gamete quality and fish health is one of the challenges that future cod aquaculture needs to overcome (Kjesbu et al. 2006).

Concluding remarks

Both wild and farmed broodstock have unequal reproductive capacities as was observed in the mass spawning event and survival data of F1 cohorts. Unequal contributions from parents are further accentuated by differences in survival of offspring during the on-rearing. As a result, the N_e of each year class was severely lower than the number of parent fish chosen to breed. It is recommended that EIRCOD continues to source eggs from wild broodstock until a much higher N_e is achieved. Separate rearing of families at least until the early larval life is completed is recommended to ameliorate the highly variable and low survival of families. It is evident that the retention of genetic variation needs to be consistently monitored to prevent bottlenecks and fast accumulation of inbreeding.

2.5. References

Appleyard, S. A., and Ward, R. D. 2006. Genetic diversity and effective population size in mass selection lines of Pacific oyster (*Crassostrea gigas*). Aquaculture 254:148–159.

Armitage, D., Maguire, J., Treasurer, J., and Cross, T. F. 2007. Parental assignment in farmed cod (*Gadus morhua*) progeny in mass spawning using microsatellite DNA loci reveals major disparity in family representation. Abstracts / Aquaculture 272S1:S238–S321.

Bartley, D., Bagley, M., Gall, G., and Bentley, B. 1992. Use of linkage disequilibrium data to estimate effective size of hatchery and natural fish populations. Conservation Biology 6:365–375.

Bekkevold, D., Hansen, M. M., and Loeschcke, V. 2002. Male reproductive competition in spawning aggregations of cod (*Gadus morhua*, L.). Molecular Ecology 11:91–102.

Bekkevold, D., Hansen, M. M., and Nielsen, E. E. 2006. Genetic impact of gadoid culture on wild fish populations: Predictions, lessons from salmonids, and possibilities for minimizing adverse effects. ICES Journal of Marine Science 63:198–208.

Belkhir K., Borsa P., Chikhi L., Raufaste N. and Bonhomme F. 1996-2004 GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Bentsen, H. B., and Olesen, I. 2002. Designing aquaculture mass selection programs to avoid high inbreeding rates. Aquaculture 204:349–359.

Blonk, R. J. W., Komen, J., Kamstra, A., Crooijmans, R. P. M. A., and van Arendonk, J. A. M. 2009. Levels of inbreeding in group mating captive broodstock populations of Common sole, (*Solea solea*), inferred from parental relatedness and contribution. Aquaculture 289:26–31.

Bolton-Warberg, M. 2012. The development of farmed cod in Ireland. PhD Thesis. National University of Ireland Galway.

Bolton-Warberg, M., and FitzGerald, R. D. 2012. Benchmarking growth of farmed Atlantic cod, *Gadus morhua*: a case study in Ireland. Aquaculture Research 43:670–678.

Brawn, V. M. 1961. Reproductive behaviour of the cod (*Gadus callarias* L.). Behaviour 18:177–198.

Brooker, A., and Cook, D. 1994. Organization of microsatellites differs between mammals and cold-water teleost fishes. Canadian Journal of Fisheries and Aquatic Sciences 51:1959–1966.

Brown, R. C., Woolliams, J. A., and McAndrew, B. J. 2005. Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. Aquaculture 247:219–225.

Carlsson, J., Olsen, K. H., Nilsson, J., Overli, O., and Stabell, O. B. 1999. Microsatellites reveal fine-scale genetic structure in stream-living brown trout. Journal of Fish Biology 55:1290–1303.

Chatziplis, D., Batargias, C., Tsigenopoulos, C. S., Magoulas, A., Kollias, S., Kotoulas, G., Volckaert, F. A. M., and Haley, C. S. 2007. Mapping quantitative trait loci in European sea bass (*Dicentrarchus labrax*): The BASSMAP pilot study. Aquaculture 272:S172–S182.

Dean, M. J., Hoffman, W. S., Zemeckis, D. R., and Armstrong, M. P. 2014. Fine-scale diel and gender-based patterns in behaviour of Atlantic cod (*Gadus morhua*) on a spawning ground in the Western Gulf of Maine. ICES Journal of Marine Science 71:1474–1489.

Dieringer, D., and Schlötterer, C. 2003. (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3:167–169.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., and Ovenden, J. R. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. Molecular Ecology Resources 14:209–214.

Dupont-Nivet, M., and Vandeputte, M. 2011. Is avoiding full sib matings really worthwhile in a breeding program? Aquaculture 314:277–281.

Dutil, J.-D., and Brander, K. 2003. Comparing productivity of North Atlantic cod (*Gadus morhua*) stocks and limits to growth production. Fisheries Oceanography 12:502–512.

Fessehaye, Y., El-bialy, Z., Rezk, M. A., Crooijmans, R., Bovenhuis, H., and Komen, H. 2006. Mating systems and male reproductive success in Nile tilapia (*Oreochromis niloticus*) in breeding hapas: A microsatellite analysis. Aquaculture 256:148–158.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: a review. Genetics Research 66:95–107.

Frankham, R., Ballou, J., and Briscoe, D. 2010. Introduction to conservation genetics. 2nd edition. Cambridge University Press.

Frankham, R., Bradshaw, C. J. A., and Brook, B. W. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation 170:56–63.

Franklin, I. R. 1980. Evolutionary change in small populations. Pages 135–150 *in* M. E. Soule and B. A. Wilcox, editors. Conservation Biology: An Evolutionary-Ecological Perspective. Sinauer, Sunderland.

Frost, L. A., Evans, B. S., and Jerry, D. R. 2006. Loss of genetic diversity due to hatchery culture practices in barramundi (*Lates calcarifer*). Aquaculture 261:1056–1064.

Garber, A. F., Tosh, J. J., Fordham, S. E., Hubert, S., Simpson, G., Symonds, J. E., Robinson, J. A. B., Bowman, S., and Trippel, E. A. 2010. Survival and growth traits at harvest of

communally reared families of Atlantic cod (*Gadus morhua*). Aquaculture 307:12–19.

Gibbons, J. W., and Andrews, K. M. 2004. PIT tagging: simple technology at its best. BioScience 54:447.

Gilk, S. E., Wang, I. A., Hoover, C. L., Smoker, W. W., Taylor, S. G., Gray, A. K., and Gharrett, A. J. 2004. Outbreeding depression in hybrids between spatially separated pink salmon, *Oncorhynchus gorbuscha*, populations: Marine survival, homing ability, and variability in family size. Environmental Biology of Fishes 69:287–297.

Gjedrem, T. 2000. Genetic improvement of cold-water fish species. Aquaculture Research 31:25–33.

Gjedrem, T. 2005. Selection and breeding programs in aquaculture. Springer. New York.

Gjedrem, T., and Baranski, M. 2009. Selective breeding in aquaculture: an Introduction. Springer, New York.

Gjerde, B. 2005. Design of breeding programs. Page 378 *in* T. Gjedrem, editor. Selection and Breeding Programs in Aquaculture. Springer, Dordrecht, The Netherlands.

Gjerde, B., Gunnes, K., and Gjedrem, T. 1983. Effect of inbreeding on survival and growth rate. Aquaculture 34:327–332.

Gjerde, B., Terjesen, B. F., Barr, Y., Lein, I., and Thorland, I. 2004. Genetic variation for juvenile growth and survival in Atlantic cod (*Gadus morhua*). Aquaculture 236:167–177.

Grabowski, T. B., Boswell, K. M., McAdam, B. J., Wells, R. J. D., and Marteinsdóttir, G. 2012. Characterization of Atlantic cod spawning habitat and behavior in Icelandic coastal waters. PloS ONE 7:e51321.

Hamoutene, D., Lush, L., Drover, D., and Walsh, A. 2009. Investigation of the temporal effects of spawning season and maternal and paternal differences on egg quality in Atlantic cod *Gadus morhua* L. broodstock. Aquaculture Research 40:1668–1669.

Hansen, Ø. J., Puvanendran, V., and Bangera, R. 2014. Do maternal age and experience contribute to better growth, survival and disease resistance of offspring in Atlantic cod (*Gadus morhua*)? Aquaculture International 23:1157–1164.

Harmon, L. J., and Braude, S. 2010. Chapter 12: Conservation of small populations: effective population sizes, inbreeding, and the 50/500 rule. *in* S. Braude and B. S. Low, editors. An Introduction to Methods and Models in Ecology, Evolution, and Conservation Biology. Princeton University Press.

Hauser, L., and Carvalho, G. R. 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9:333–362.

Hedrick, P. 2011. Genetics of Populations. 4th edition. Jones and Bartlett, Sudbury, Massachusetts.

Herbinger, C. M. 2005. PEDIGREE. Available from http://herbinger.biology.dal.ca:5080/Pedigree/.

Herlin, M., Delghandi, M., Wesmajervi, M., Taggart, J. B., McAndrew, B. J., and Penman, D. J. 2008. Analysis of the parental contribution to a group of fry from a single day of spawning from a commercial Atlantic cod (*Gadus morhua*) breeding tank. Aquaculture 274:218–224.

Herlin, M., Taggart, J. B., McAndrew, B. J., and Penman, D. J. 2007. Parentage allocation in a complex situation : A large commercial Atlantic cod (*Gadus morhua*) mass spawning tank. Aquaculture 272S1:S195–S203.

Huang, C., and Liao, I. 1990. Response to mass selection for growth rate in *Oreochromis niloticus*. Aquaculture 85:199–205.

Hubert, S., Higgins, B., Borza, T., and Bowman, S. 2010. Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). BMC Genomics 11:191.

Hutchings, J. A., Bishop, T. D., and McGregor-Shaw, C. R. 1999. Spawning behaviour of Atlantic cod, *Gadus morhua*: evidence of mate competition and mate choice in a broadcast spawner. Canadian Journal of Fisheries and Aquatic Sciences 56:97–104.

Hutchinson, W. F., van Oosterhout, C., Rogers, S. I., and Carvalho, G. R. 2003. Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). Proceedings of the Royal Society of London B: Biological Sciences 270:2125–32.

Jackson, T. R., Martin-Robichaud, D. J., and Reith, M. E. 2003. Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. Aquaculture 220:245–259.

Jones, A. G. 2001. GERUD1.0: A computer program for the reconstruction of parental genotypes from progeny arrays using multilocus DNA data. Molecular Ecology Notes 1:215–218.

Jones, A. G., Small, C. M., Paczolt, K. A., and Ratterman, N. L. 2010. A practical guide to methods of parentage analysis. Molecular Ecology Resources 10:6–30.

Jones, O. R., and Wang, J. 2010. COLONY: A program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources 10:551–555.

Kincaid, H. L. 1976. Effects of inbreeding on rainbow trout populations. Transactions of the American Fisheries Society 105:273–280.

Kjesbu, O. S., Solemdal, P., Bratland, P., and Fonn, M. 1996. Variation in annual egg production in individual captive Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences 53:610–620.

Kjesbu, O. S., Taranger, G. L., and Trippel, E. A. 2006. Gadoid mariculture: Development and future challenges. ICES Journal of Marine Science 63:187–191.

Kjørsvik, E., Pittman, K., and Pavlov, D. 2004. From fertilisation to the end of metamorphosis—functional development. Page 546 *in* E. Moskness, E. Kjørsvik, and Y. Olsen, editors. Culture of cold-water marine fish. Blackwell Publishing Ltd, Bath.

Kolstad, K., Thorland, I., Refstie, T., and Gjerde, B. 2006. Genetic variation and genotype by location interaction in body weight, spinal deformity and sexual maturity in Atlantic cod (*Gadus morhua*) reared at different locations off Norway. Aquaculture 259:66–73.

Lande, R. 1995. Mutation and conservation. Conservation Biology 9:782–791.

Leberg, P. L., and Firmin, B. D. 2008. Role of inbreeding depression and purging in captive breeding and restoration programmes. Molecular Ecology 17:334–343.

Liu, Z. J., and Cordes, J. F. 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238:1–37.

Lush, L., Hamoutene, D., Drover, D., Walsh, A., and Puvanendran, V. 2011. Gamete collection method and egg quality comparison in Atlantic cod, *Gadus morhua* L. mating pairs and its importance in selective breeding. Aquaculture 315:407–409.

Marine Institute. 2013. The Stock book: Annual 2013 review of fish stocks with management advice for 2014.

Marshall, T. C., Slate, J., Kruuk, L. E. B., and Pemberton, J. M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639–655.

Matson, S. E., Camara, M. D., Eichert, W., and Banks, M. A. 2008. P-Loci: a computer program for choosing the most efficient set of loci for parentage assignment. Molecular Ecology Resources 8:765–8.

Meagher, T. R., and Thompson, E. 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theoretical Population Biology 29:87–106.

Miller, K. M., Le, K. D., and Beacham, T. D. 2000. Development of tri- and tetranucleotide repeat microsatellite loci in Atlantic cod (*Gadus morhua*). Molecular ecology 9:238–9.

Mirimin, L., O'Keeffe, D., Ruggiero, A., Bolton-Warberg, M., Vartia, S., and Fitzgerald, R. 2011. A quick, least-invasive, inexpensive and reliable method for sampling *Gadus morhua* postlarvae for genetic analysis. Journal of Fish Biology 79:801–805.

Moav, R., and Wohlfarth, G. 1973. Carp breeding in Israel. Pages 295–318 Agricultural Genetics—Selected Topics. Wiley New York, NY.

Moav, R., and Wohlfarth, G. 1976. Two-way selection for growth rate in the common carp (*Cyprinus carpio* L.). Genetics 82:83–101.

Murua, H., and Saborido-Rey, F. 2003. Female reproductive strategies of marine fish species of the North Atlantic. Journal of Northwest Atlantic Fishery Science 33:23–31.

Naish, K. A., Seamons, T. R., Dauer, M. B., Hauser, L., and Quinn, T. P. 2013. Relationship between effective population size, inbreeding and adult fitness-related traits in a steelhead (*Oncorhynchus mykiss*) population released in the wild. Molecular Ecology 22:1295–1309.

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

Nomura, T. 1999. A mating system to reduce inbreeding in selection programmes: Theoretical basis and modification of compensatory mating. Journal of Animal Breeding and Genetics 116:351–361.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. 2004. Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535–538.

Otterlei, E., Nyhammer, G., Folkvord, A., and Stefansson, S. O. 1999. Temperature-and size-dependent growth of larval and early juvenile Atlantic cod (*Gadus morhua*): a comparative study of Norwegian coastal cod and northeast Arctic. Canadian Journal of Fisheries and Aquatic Sciences 56:2099–2111.

Palstra, F. P., and Ruzzante, D. E. 2008. Genetic estimates of contemporary effective population size: What can they tell us about the importance of genetic stochasticity for wild population persistence? Molecular Ecology 17:3428–3447.

Pampoulie, C., Stefánsson, M. Ö., Jörundsdóttir, T. D., Danilowicz, B. S., and Daníelsdóttir, A. K. 2008. Recolonization history and large-scale dispersal in the open sea: the case study of the North Atlantic cod, *Gadus morhua* L. Biological Journal of the Linnean Society 94:315–329.

Puckrin, O. A., Purchase, C. F., and Trippel, E. A. 2013. Using purposeful inbreeding to reduce outbreeding depression caused by escaped farmed Atlantic cod. Aquaculture Environment Interactions 4:207–221.

Rakitin, A., Ferguson, M. M., and Trippel, E. A. 2001. Male reproductive success and body size in Atlantic cod *Gadus morhua* L. Marine Biology 138:1077–1085.

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

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223-225.

Robichaud, D., and Rose, G. A. 2003. Sex differences in cod residency on a spawning ground. Fisheries Research 60:33–43.

Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8:103–106.

Salvanes, A. G. V., Skjæraasen, J. E., and Nilsen, T. 2004. Sub-populations of coastal cod with different behaviour and life-history strategies. Marine Ecology Progress Series 267:241–251.

Skaarud, A., Woolliams, J. A., and Gjøen, H. M. 2011. Strategies for controlling inbreeding in fish breeding programs; an applied approach using optimum contribution (OC) procedures. Aquaculture 311:110–114.

Skirnisdottir, S., Pampoulie, C., Hauksdottir, S., Schulte, I., Olafsson, K., Hreggvidsson, G. O., and Hjorleifsdottir, S. 2008. Characterization of 18 new microsatellite loci in Atlantic cod (*Gadus morhua* L.). Molecular Ecology Resources 8:1503–5.

Skjæraasen, J., and Hutchings, J. 2010. Shifting reproductive success in a shoal of Atlantic cod, *Gadus morhua* L. Environmental Biology of Fishes 88:311–318.

Smith, B. R., Herbinger, C. M., and Merry, H. R. 2001. Accurate partition of individuals into full-sib families from genetic data without parental information. Genetics 158:1329–38.

Sonesson, A. K., and Meuwissen, T. H. 2000. Mating schemes for optimum contribution selection with constrained rates of inbreeding. Genetics, Selection, Evolution 32:231–248.

Sonesson, A. K., Woolliams, J. A., and Meuwissen, T. H. E. 2005. Kinship, relationship and inbreeding. Pages 73–87 *in* T. Gjedrem, editor. Selection and Breeding Programs in Aquaculture. Springer, New York.

Speare, D., MacNair, N., and Hammell, K. 1995. Demonstration of tank effect on growth indices of juvenile rainbow trout (*Oncorhynchus mykiss*) during an ad libitum feeding trial. American Journal of Veterinary Research 56:1372–9.

Stenvik, J., Wesmajervi, M. S., Fjalestad, K. T., Damsgård, B., and Delghandi, M. 2006. Development of 25 gene-associated microsatellite markers of Atlantic cod (*Gadus morhua* L.). Molecular Ecology Notes 6:1105–1107.

Therkildsen, N. O., Nielsen, E. E., Swain, D. P., and Pedersen, J. S. 2010. Large effective population size and temporal genetic stability in Atlantic cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. Canadian Journal of Fisheries and Aquatic Sciences 67:1585–1595.

Thorsen, A., Trippel, E. A., and Lambert, Y. 2003. Experimental methods to monitor the production and quality of eggs of captive marine fish. Journal of Northwest Atlantic Fishery Science 33:55–70.

Trippel, E. A., Rideout, R. M., O'Reilly, P. T., Herbinger, C. M., Neil, S. R. E., and Hamilton, L. 2009. Communal spawning leads to high potential for inbreeding in gadoid aquaculture. Aquaculture 296:27–35.

Vandeputte, M., Mauger, S., and Dupont-Nivet, M. 2006. An evaluation of allowing for mismatches as a way to manage genotyping errors in parentage assignment by

exclusion. Molecular Ecology Notes 6:265–267.

Wang, J. 2009. A new method for estimating effective population sizes from a single sample of multilocus genotypes. Molecular Ecology 18:2148–2164.

Wang, J. 2012. Computationally efficient sibship and parentage assignment from multilocus marker data. Genetics 191:183–194.

Waples, R. S., and Do, C. 2008. ldne: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8:753–756.

Waples, R. S., and Do, C. 2010. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evolutionary Applications 3:244–262.

Wesmajervi, M. S., Westgaard, J. I., and Delghandi, M. 2006. Evaluation of a novel pentaplex microsatellite marker system for paternity studies in Atlantic cod (*Gadus morhua* L.). Aquaculture Research 37:1195–1201.

Chapter 3. Multiplexing with three-primer PCR for rapid and economical microsatellite validation[1]

[1] Published as Vartia, S., Collins, P.C., Cross, T.F., FitzGerald, R.D., Gauthier, D.T., McGinnity, P., Mirimin, L. & Carlsson J. (2014). Multiplexing with three-primer PCR for rapid and economical microsatellite validation. Hereditas, 151 (2-3), pp. 43-54

3.1. Introduction

Thirty years after their discovery in the 1980s, microsatellite-based genetic markers are still extensively used in studies of population structure, parentage analysis, genetic mapping, evolutionary processes and forensics (Bruford and Wayne 1993; Brockmann et al. 1994; Knapik et al. 1998; Goldstein et al. 1999; Primmer et al. 2000). These markers have a wide application due to high allelic diversity and co-dominance of alleles (e.g. Chambers and MacAvoy 2000). Many methodologies have been established in order to discover such markers, but it is only following the recent advent of next generation sequencing (NGS) technology that large amounts of markers can be increasingly rapidly and economically developed from non-model organisms (Gardner et al. 2011; Zalapa et al. 2012; Schoebel et al. 2013). NGS approaches allow the fast discovery of large amounts of microsatellite-containing sequences, however mining such data for suitable DNA fragments and validation of candidate markers are still posing challenges prior to the utilisation of fully operating new markers.

The most common approach for *de novo* microsatellite marker development includes creation of repeat-enriched DNA libraries, fragment replication by cloning, and Sanger sequencing of clones containing potential microsatellites (Zane et al. 2002). These processes are laborious and time consuming, and typically have low marker yield, with the percentage of positive clones averaging 2-3% (Ashworth et al. 2004). The final marker yield is even lower with a large portion of markers discarded during the isolation–characterisation process (Squirrell et al. 2003). Alternatively, microsatellite-containing sequences can be mined from existing molecular data such as genomic DNA or Expressed Sequence Tag (EST) sequences (Li et al. 2004). These approaches are limited by the paucity of data on non-model organisms. EST-linked microsatellites can be relatively easy to identify but have a higher probability of being affected by selective processes, and hence may not be suitable for population analyses that assume that loci are selectively neutral (Ellis & Burke 2007). Microsatellite markers developed for one species may also be applied to closely related species (Schlötterer et al. 1991). However, this approach is limited

by varying levels of successful cross-species amplification between species (Moore et al. 1991). Even when cross-species amplification is successful, levels of variability tend to be lower compared with the species for which the markers were developed (Primmer et al. 1996). Because of these limitations, it may be preferable to develop markers *de novo* for a species or population of interest to ensure optimal power of newly discovered markers (Carlsson et al. 2013).

Recently, several approaches have been presented for discovery of microsatellites using next generation sequencing (NGS)-generated data (e.g. Abdelkrim et al. 2009; Allentoft et al. 2009). Large amounts of sequence data can be generated from either genomic DNA or microsatellite enriched libraries (Guichoux et al. 2011; Malausa et al. 2011) and then mined for microsatellite loci. With this approach, thousands of potential loci can be rapidly identified (Guichoux et al. 2011). Large-scale microsatellite identification has several advantages, including the ability to rigorously screen individual loci for presence of optimal primer-binding sites in flanking regions (Guichoux et al. 2011; Zalapa et al. 2012; Fernandez-Silva et al. 2013). Additionally, deployment criteria (e.g. high levels of variability, neutrality and linkage equilibrium) are study-specific and cannot be assessed until the markers have been validated (Selkoe and Toonen 2006). Validation of a large number of markers enables selection of most suitable loci (Selkoe and Toonen 2006). Even with the NGS approach, validation is labour and cost intensive, resulting in a limiting bottleneck in microsatellite marker development (Squirrell et al. 2003; Malausa et al. 2011; Fernandez-Silva et al. 2013).

Typically, methods for validation and genotyping of microsatellites involve capillary gel electrophoresis with fluorescence-based amplicon detection (Edwards et al. 1991; Guichoux et al. 2011), where either the forward or reverse primer of each primer pair combination is fluorescently labelled. A three-primer PCR method can be used to reduce the expense associated with fluorescently labelled primers (*sensu* Schuelke 2000; Diniz et al. 2007; Rubin et al. 2009). In fact, in three-primer PCR, the primer set comprises an unlabelled forward primer with a universal tail attached to its 5’ end, a labelled universal primer matching the tail sequence and an

unlabelled reverse primer (Steffens et al. 1993; Oetting et al. 1995; Neilan et al. 1997; Schuelke 2000). The labelled universal primer can be used in combination with any appropriately tailed forward primer, thereby eliminating the need to synthesize a fluorescently labelled forward primer for every unique locus during the validation phase, in which a large proportion of loci may be excluded because of problems with amplification.

Microsatellite markers can be amplified in single PCR amplification or combined into a multiplex PCR containing multiple markers (Guichoux et al. 2011). Improvement of the traditional multiplex PCR technique (i.e. Missiaggia and Grattapaglia 2006) employed human microsatellite primer sequences as universal tails and combined three universal tails with three dyes in a true multiplex PCR (*sensu* Guichoux et al. 2011). However, despite the obvious cost benefits of the three primer PCR approach in combination with multiplexing, few studies have employed such an approach (Langen et al. 2011; Blacket et al. 2012). This may be attributable to poor amplification or poor quality chromatograms, resulting in difficulty in accurate genotyping of individuals (e.g. Hagell et al. 2013).

This thesis has been undertaken as part of the work carried out under the Irish Cod Broodstock and Breeding Programme (EIRCOD). The EIRCOD programme has established a base population and core broodstock using wild caught cod from the Celtic Sea, one of the most southern cod stocks. There are existing microsatellite markers available for cod (Chapter 2), however, a large proportion of them are EST-linked (i.e. Higgins et al. 2009) and, as such, are not optimal for resolving population structure. This study is, thus, the first large scale validation of markers specifically designed for the Celtic Sea cod after the initial small scale validation described in Carlsson et al. (2013).

The current chapter presents the development and application of a method for rapid validation and genotyping of novel microsatellites in cod using the three primer approach with multiplex PCR. In addition to cod, this fast and economic development and deployment of microsatellite multiplexes from raw NGS data is

applicable to studies on a wide range of non-model organisms and, equally, the facilitation of applied breeding programmes.

Aims:

- Development of a rapid and economical approach for the validation and deployment microsatellite markers
- Validation of novel microsatellite markers for cod from the Celtic Sea

3.2. Materials and methods

3.2.1. Sampling

Tissue samples were obtained from cod specimens collected by trawling in 2009 and 2011 from the Celtic Sea, south of Ireland (*n* = 7 and *n* = 46, respectively) and in 1996 from the Scotian Shelf, off Nova Scotia in eastern Canada (*n* = 46). Previous research has shown that these two populations are genetically differentiated (Hutchinson et al. 2001; O'Leary et al. 2007). Fin clip samples were preserved in 100% ethanol.

3.2.2. DNA Extraction

DNA was extracted from fin clips using a Chelex® protocol as described in Mirimin et al. (2011). DNA from the Scotian Shelf samples was extracted using a standard phenol-chloroform method (see O'Leary et al. 2007). DNA was quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific) and normalised to a concentration of 50 ng/µl.

3.2.3. Microsatellite selection

The sequence data used here were generated in a previous study (Carlsson et al. 2013), in which microsatellite containing sequences were obtained from five of the 2009 Celtic Sea individuals, also used in the present study, using 454

pyrosequencing of a reduced representation library. Carlsson et al. (2013) identified a total of 11 341 microsatellite containing sequences as suitable for primer design using the Primer3 plug-in (Rozen and Skaletsky 2000) for MISA v.1.0 (http://pgrc.ipk-gatersleben.de/misa). Of these, 6424 were estimated to be unique. These microsatellite-containing sequences were used in the present study.

To avoid excessive homoplasy (alleles identical in state but not in descent c.f. Estoup et al. 2002) and to ensure ease of genotyping, complex repeat motifs (i.e. compound and imperfect motifs) were excluded. To ensure sufficient space for primer design, reads that had less than 50bp of sequence before and after the repeat-containing region were removed. In addition, to avoid excessively large allele size ranges, repeat sequences of more than 100bp, and penta- and hexanucleotide repeats were excluded. A subsample of the remaining microsatellite sequences (n = 1309, see Results section) were visually inspected for primer design.

3.2.4. Primer design

Primers were designed using PRIMER3PLUS (Rozen and Skaletsky 2000; Untergasser et al. 2007) with optimal primer length of 20bp and an optimal T_m of 60°C. Two sets of three size classes were used: the first set was separated by 30 bp (100-150, 180-250, 280-450 bp) and the second set separated by 50 bp (100-150, 200-250, 300-450 bp). Equal numbers of markers were designed for each size class. Only primer pairs with a T_m difference of less than 1°C were accepted in order to facilitate PCR multiplexing.

Designed primers were cross-referenced with the original sequence data set to identify primers that annealed to multiple regions (not unique) or originated from redundant sequences (different reads of the same sequence). Redundant sequences not detected in the previous steps (due to sequencing error in the primer regions) were identified by performing a *de novo* assembly with the remaining candidate loci sequences using GENEIOUS 6.1.5 (created by Biomatters; available from http://www.geneious.com), CAP3 plug-in (default settings; i.e. min

overlap length = 40 bp, min overlap identity = 90%; Huang and Madan 1999). If two or more reads assembled together they were considered redundant and only one of them was kept for future analysis.

To minimise the risk that primer sequences were derived from contaminants, expressed cod gene regions or previously published cod microsatellites, microsatellite-containing sequences and primers were subjected to a BLAST search in the GenBank nucleotide database (Altschul et al. 1990). For possible contaminants, a match with ≥95% coverage and 100% identity was considered as a threshold for excluding reads. No threshold was employed for matches on cod sequences. When such matches were encountered, primers were excluded from further analyses. In addition, validated primer sequences were subjected to BLAST searches against the cod genome (Accession number CAEA00000000.1, Star et al. 2011) in the whole-genome shotgun contigs database in GenBank.

Universal Primers

The universal dye-labelled primers used were T3: PET-5' AATTAACCCTCACTAAAGGG 3', M13 Reverse: NED-5' GGATAACAATTTCACACAGG 3' (Diniz et al. 2007), Hill: 6FAM-5' TGACCGGCAGCAAAATTG 3' (Tozaki et al. 2001) and Neomycin rev: VIC-5' AGGTGAGATGACAGGAGATC 3'. Each forward primer had one of the above universal primer sequences added to its 5' end. PIG-tails were added to the 5' end of all the reverse primers. PIG-tailing leads to an addition of a non-templated adenosine nucleotide to the 3' end on nearly 100% of PCR products which reduces stuttering caused by random addition of dATP (Brownstein et al. 1996). The tails were matched with the primers using OLIGOANALYZER 3.1 (Available from http://www.idtdna.com) to ensure the least amount of different secondary structures. Equal numbers of primers were paired with each of the four different universal primers.

3.2.5. Microsatellite validation

Primers were combined into twelve multiplex PCR reactions containing 12 markers each (12-plex) and validated using all seven 2009 Celtic Sea individuals. Loci that amplified successfully and showed polymorphism were combined into further multiplexes. The construction of multiplexes was done by means of successive attempts of adding and removing loci from sets of markers that had amplified together in the initial test panels until at least six loci were successfully combined in a panel. When amplified loci were monomorphic, the procedure was repeated on seven Scotian Shelf individuals to test for variability over a larger geographic scale.

Multiplex PCRs were performed in 5 µL reactions with 50 ng template DNA, 1× Multiplex PCR Master Mix (Qiagen), 0.2 µM of each reverse primer, 0.05 µM of each unlabelled forward primer (modified with the appropriate universal tail) and 0.2 µM of the respective labelled universal primer. Further adjustments made to optimize concentrations of primers in the PCR reactions are given in **Table 1**. PCR thermal cycling conditions were as follows: 1 × 95°C (15 min); 30 × (94°C (30 s), 60°C (90 s), 72°C (60 s)); 8 × (94°C (30 s), 53°C (90 s), 72°C (60 s)); 1 × 60°C (30 min). Controls with no template were included to monitor for potential contamination.

A total of 1 µL of the multiplex PCR product was added to 9 µL of Super-DI™ Formamide (MCLAB) with 0.01 µL of Orange DNA Size Standard (MCLAB) and run on an ABI 3130xl Genetic Analyzer according to manufacturer's recommendations. GENEMARKER v.1.97 (Available from http://www.softgenetics.com) was used for fragment length analysis.

Table 1 Six multiplex panels for amplifying 55 loci. Underlined sequence in the forward primer signifies the universal tail sequence, whereas underlined sequence in the reverse primer signifies the PIG-tail sequence. Adjustment to primer concentration applies for all three primers used to amplify a locus (forward, reverse and universal primer). SRA accession number for the sequences is SRP041380.

Multiplex	Name	Motif	Size range	Adjustment to primer concentration	Forward primer	Reverse primer
1	A08_T3	tetra	163–187		AATTAACCCTCACTAAAGGGATCTCGAGTGGCGCAGTAG	GTTTCTTGCACGCTGACAAGTGAAGAG
	A16_M13	tetra	156–194		GGATAACAATTTCACACAGGCGTTCCCTCAGCTTGTCTCT	GTTTCTTCGTCCAACAGATCTATGCAATC
	A43_T3	tetra	309–369		AATTAACCCTCACTAAAGGGCCACTTTAACCTGCGGTTTC	GTTTCTTGCCTGCTTGTATACGCTGA
	B19_T3	tri	242–320		AATTAACCCTCACTAAAGGGAAAAGGATCTGCTTGCCTCA	GTTTCTTCGTGAGCTCAGTTTTGGCTA
	B38_Neo	tri	388–415		AGGTGAGATGACAGGAGATCGAATTGAGGAGGCATGGGTA	GTTTCTTGGTTAATTCCAGCCGTAGAGG
	C15_Hill	tetra	169–205		TGACCGGCAGCAAAATTGCCTTTCGTTCTCTCCGTCAG	GTTTCTTGAGGATTTGGTGGGATGAT
	C28_Neo	tri	255–282		AGGTGAGATGACAGGAGATCCAGCACAAGTGGTAGGGTCA	GTTTCTTGCGATCAGAAGTTGTGCTT
	C36_Hill	tri	319–352		TGACCGGCAGCAAAATTGGTTGGCTCACACAATCATCG	GTTTCTTATCCTTCAAACAGCCCTCAA
	C40_M13	tri	276–277		GGATAACAATTTCACACAGGGGTCTTTGGGAGGTCTTCCT	GTTTCTTCATCCTGCTTGCGGACTTAT
	D14_Hill	tetra	228–276		TGACCGGCAGCAAAATTGGGCGATAATCTGCCATTTTG	GTTTCTTCGTGACAAGTGTCGATTTGC
	D30_M13	tetra	317–333		GGATAACAATTTCACACAGGTTCACAAACGGGAACTACGA	GTTTCTTGACAAGTCCAAGATGTGTCATCA
2	A18_M13	tetra	203–243		GGATAACAATTTCACACAGGGACTGTCCGTTGAGGGTGTT	GTTTCTTGGGTCGAACTGGTCTGGTTA
	A19_T3	tetra	242–274		AATTAACCCTCACTAAAGGGTCCTGGTTCCAACACATGAC	GTTTCTTCTAGCCAATGGTGCAAGT
	A34_M13	tetra	297–321		GGATAACAATTTCACACAGGTCCTTAACGACAGGCACCTT	GTTTCTTCCTGACTTGTGTCGTTCCAG
	A37_Neo	tetra	265–489		AGGTGAGATGACAGGAGATCCCGGCAGTACAGCTAATGAA	GTTTCTTAAATGCTCAACCCATTGGAC
	B03_T3	di	147–161		AATTAACCCTCACTAAAGGGCGATAATAGCGTTCCCATCC	GTTTCTTGGGGTACCTTGTGACCTGT
	B12_T3	di	346–398	2x	AATTAACCCTCACTAAAGGGGCTTTGGCAACACTGTTTGA	GTTTCTTGTCGAGCAGACCAGAAGACC
	B30_Neo	tri	151–169		AGGTGAGATGACAGGAGATCTTGTACGGACAGGAAGTCCA	GTTTCTTGCGAACAGTGTGTAAATTGAA

Table 1 *Continued.*

Multiplex	Name	Motif	Size range	Adjustment to primer concentration	Forward primer	Reverse primer
2	C01_M13	tetra	115–195		GGATAACAATTTCACACAGGACCAGGAGGTTGGATCAGTG	GTTTCTTCCATTATTCATTCGTCATCCA
	C14_Hill	tetra	146–166		TGACCGGCAGCAAAATTGAGGTTCAGCCAGAAGCTGAT	GTTTCTTCCATTGGTTGTCGGTGATTA
	C20_Neo	tri	227–239		AGGTGAGATGACAGGAGATCCCGCCTATCACCCTAAATCTG	GTTTCTTCGTCTACATGTCGTGGTAGGG
	D12_Hill	tetra	251–291		TGACCGGCAGCAAAATTGAACGGCTCCTCAAGACAAAC	GTTTCTTAGGCATCTGCGTCCATACTC
3	A33_M13	tetra	233–297		GGATAACAATTTCACACAGGAGACACTGAGCTCGACAGCA	GTTTCTTATCAATGATCCCAGGCAAAC
	A39_Hill	tetra	356–416		TGACCGGCAGCAAAATTGCCTGTCCAAATGCACACAAG	GTTTCTTAGTGCTTGGATGGTGTGATG
	B01_Hill	di	129–213		TGACCGGCAGCAAAATTGTAGACTCTGGGGCTGGGTAA	GTTTCTTGCATCCGAGACTCTTGTTCC
	B15_T3	di	170–178		AATTAACCCTCACTAAAGGGCGATGCGATTCTTGGTAAATG	GTTTCTTGGTGTCTCATCCCTCTTCA
	B29_Neo	tri	262–281		AGGTGAGATGACAGGAGATCGGGAAAGAGCCGGAAAAGTA	GTTTCTTGCTAATGTTGGCAGAACCA
	C13_Neo	tri	158–176		AGGTGAGATGACAGGAGATCGGGTGATTGAGGTTGCGATA	GTTTCTTGCTCACACATCCTACGAGCA
	C17_M13	tetra	154–190		GGATAACAATTTCACACAGGCTTCTCGATGGCATGTTTCC	GTTTCTTCCTGCACAATGATCTGCAT
	C22_Hill	tetra	225–273		TGACCGGCAGCAAAATTGGGCTTGTCGTTGGTTCCTT	GTTTCTTGTTGAATGCAACCCCTCAGT
	C30_Neo	tri	337–343		AGGTGAGATGACAGGAGATCAGGTGGTCGCAGTGAAGAAG	GTTTCTTGGGTGAATGCCTCTTAATCG
	C42_M13	tri	345–358		GGATAACAATTTCACACAGGGCTGAGGGGATGCGATAATA	GTTTCTTAGCCAAGGGTGAAGTGTTGT
	D37_T3	tetra	313–333		AATTAACCCTCACTAAAGGGATGTGACACCGAATCACAGC	GTTTCTTACCCGTCCTGTACGTGAACT
4	A11_Hill	tetra	163–175		TGACCGGCAGCAAAATTGCGACAGGGAGGCATAAAGAC	GTTTCTTGTTCACCTCCCTGGCTCTT
	A22_Neo	tetra	214–243		AGGTGAGATGACAGGAGATCGGTGAGGTTCTTGAGGGTCA	GTTTCTTGATTATTTCCCCCTGCTG
	A31_T3	tetra	314–386		AATTAACCCTCACTAAAGGGGGATATGTGGGGATGAGCAC	GTTTCTTATGGGTCCTTCTCCTTTGGT
	B33_M13	tri	120–135	0.5x	GGATAACAATTTCACACAGGCTACAGCAGGGGTTCCTCAG	GTTTCTTGTTTGTTGCTCCGATGGACT
	C08_T3	tetra	145–185		AATTAACCCTCACTAAAGGGCTCGGACCCAGAGATCAAAA	GTTTCTTGCAGCATCTGAACTGAAACG

Table 1 *Continued.*

Multiplex	Name	Motif	Size range	Adjustment to primer concentration	Forward primer	Reverse primer
4	D15_Hill	tetra	244–256		TGACCGGCAGCAAAATTGTGACTCAACGGAGGTACGTG	GTTTCTTCCATCAGGATCAGGACCACT
	D21_Neo	tetra	337–389		AGGTGAGATGACAGGAGATCAACACGCTTGCTGGGACTAC	GTTTCTTCACTGGAGTGTACGGTCTCTGA
	D46_Neo	tetra	139–155		AGGTGAGATGACAGGAGATCCCTCCCTAATACCATGTCACCA	GTTTCTTCGTCTGTTCACGGATGCAC
5	A04_Neo	tetra	143–211		AGGTGAGATGACAGGAGATCACAATCAACCCTCCAACTCG	GTTTCTTCAGGTCCCGAATATCAAGG
	B07_M13	di	163–187		GGATAACAATTTCACACAGGTGGACAATTACATTGAAAATCACAG	GTTTCTTCCTGAACTGCCTGTCAATG
	B28_Neo	tri	226–238		AGGTGAGATGACAGGAGATCCCCACCCCTTAATGTTTCAA	GTTTCTTGGCGTCATTCTCTTTGATG
	D10_Neo	tetra	308–312	2x	AGGTGAGATGACAGGAGATCCGCCAATGCAAATCTCTTTT	GTTTCTTATCTGAGTGCGTCGAGTGC
	D35_M13	tetra	316–360		GGATAACAATTTCACACAGGTCCACACTTGGTCGATGAAA	GTTTCTTGACGAGTGTCAGCAGGTGTG
	D43_T3	tetra	260–292	2x	AATTAACCCTCACTAAAGGGGTGCCGCTCACGCTACTAAT	GTTTCTTCGTGATCGCTCTCGATTC
6	A03_Hill	tetra	158–170		TGACCGGCAGCAAAATTGGAGCGTGTTGAACGACTTGA	GTTTCTTCCTGAGCAGTGGAGTGACAA
	A20_Neo	tetra	164–182		AGGTGAGATGACAGGAGATCCGAGGCTACACAGCCTGTAA	GTTTCTTACTGTGGGCATGTAACAGCA
	A30_T3	tetra	245–269		AATTAACCCTCACTAAAGGGAGTTGAACTGCGGGTTCTGT	GTTTCTTGCACGATGTCACAGCTGATT
	B36_Hill	tri	384–396		TGACCGGCAGCAAAATTGCCCGCCCAGACATAATAAGA	GTTTCTTCAGTCTCAGCCTCCACATCA
	C31_Neo	tri	332–341		AGGTGAGATGACAGGAGATCGCCAAGACAAGCATTTCCAT	GTTTCTTCGAGCCAGCGTTTACTTCTC
	C35_M13	tri	233–348		GGATAACAATTTCACACAGGGGCAATGTCGTACACCTCAA	GTTTCTTCTGAACGGCAACACTTCGTA
	D05_Hill	tetra	230–270		TGACCGGCAGCAAAATTGACTGCCCCTGATAACAATGC	GTTTCTTAGGCATCGACCATTTGTAGC
	D16_T3	tetra	424–436		AATTAACCCTCACTAAAGGGCCAGCAGCTTTCTGGGTAGT	GTTTCGAAGCGTTACTGCAGACAG

3.2.6. Microsatellite genotyping

All 46 Celtic Sea and 46 Scotian Shelf samples were genotyped with multiplex panels. Genotype data were inspected with MICRO-CHECKER 2.2.3 for genotyping errors and presence of null-alleles (van Oosterhout et al. 2004) using default settings. The 99% confidence interval was used when checking for null alleles to avoid false positives resulting from multiple tests. MSANALYSER 4.05 (Dieringer and Schlötterer 2003), using default settings, was used to assess the number of alleles, allelic richness, allele size ranges, F_{ST} estimates and expected and observed heterozygosity. Data were analysed for possible departure from Hardy-Weinberg equilibrium, linkage disequilibrium, and excess and deficit of heterozygotes using GENEPOP 4.2 with default settings (Raymond and Rousset 1995; Rousset 2008). False discovery rate (FDR) was used to correct for multiple comparisons (Benjamini and Yekutieli 2001) with initial $\alpha = 0.05$. LOSITAN (Antao et al. 2008) was used to detect loci that could be under positive or balancing selection (settings ‘‘Neutral’ mean F_{ST}’ and ‘Force mean F_{ST}’ with 10 000 simulations were used under both the infinite allele model (IA) and stepwise mutation model (SMM)).

3.2.7. F_{ST} replicate sampling

The current study purposefully aimed to validate more markers than required for accurate evaluation of population differentiation (i.e. multilocus F_{ST}). To estimate the number of microsatellite loci future studies on cod population structure may require, it was investigated how many markers were needed to accurately estimate multilocus F_{ST}. Data sets were generated by randomly drawing 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 loci from the 55-locus dataset with each condition (number of loci) replicated ten times. Average F_{ST} and 95% confidence interval of the ten replicates were calculated and plotted to visualise the variability of average F_{ST} estimates as a function of numbers of markers.

3.3. Results

3.3.1. Primer design

Of 1309 candidate microsatellite loci, 559 were determined to be suitable for primer design upon visual inspection. A total of 349 primer pairs were rejected based on T_m difference, sequence redundancy or secondary structure with the 5' tail. Another 18 were excluded due to a BLAST match (one match to a cod microsatellite, nine to a cod gene and eight to a possible contaminant). The remaining 192 primer pairs were chosen for validation. Of those, 51 failed to amplify and 45 were excluded due to low scorability. Unambiguous amplification of PCR products in the expected size range was successful in 96 of the 192 markers tested (50%), of which 73 showed polymorphism (38%). The 73 polymorphic loci were used to build multiplex panels. Of these, 13 were not included in the final multiplexes because of incompatible size, associated fluorochrome or failure to amplify with the other markers in a panel. As a result, 60 markers were combined into six multiplex panels ranging between eight and twelve loci per panel. Five markers were not used in the final analysis due to ambiguous genotyping leading to high chance of scoring errors, resulting in a final panel of 55 polymorphic markers combined into six multiplexes (**Table 1**). The results of the BLAST search on the validated primers against the cod genome are presented in Appendix 1.

3.3.2. Application of markers to test populations

The mean allelic richness (R_S) was 7.4 (sd = 4.11) in the Celtic Sea sample and 7.5 (sd = 4.24) in the Scotian Shelf sample. The minimum number of alleles was two for both the Celtic Sea and Scotian Shelf samples; the maximum number of alleles was 21 and 25, respectively (**Table 2**). MICRO-CHECKER analyses indicated no genotyping errors. However, ten loci had a different repeat pattern than the motif originally identified from the raw sequence (**Table 3**). Null alleles were observed in 13 loci in the Celtic Sea

sample and 11 in the Scotian Shelf sample (**Table 4**). Twelve and eleven loci deviated significantly from Hardy-Weinberg equilibrium (after FDR correction) in Celtic Sea and Scotian Shelf samples, respectively. Linkage disequilibrium was observed (after FDR correction) in locus pair A43_T3 x C01_M13 in the Celtic Sea sample; and locus pairs A43_T3 x B19_T3 and C15_Hill x C17_M13 in the Scotian Shelf sample. Lositan identified loci A11_Hill, C40_M13, C42_M13 and D14_Hill as being potentially affected by positive selection, (both under IA and SMM). Only C40_M13 (global F_{ST} = 0.581), C42_M13 (global F_{ST} = 0.301) and D14_Hill (global F_{ST} = 0.246) remained significant after correction for multiple comparisons.

Global multilocus F_{ST} was estimated at 0.067 (P = 0.001). After the three outlier loci identified by Lositan were excluded, the global F_{ST} was estimated at 0.043 (P = 0.001). Global F_{ST} values of individual loci are presented in **Figure 1**.

Table 2 Summary statistics for 55 microsatellite loci in two samples of Atlantic cod. n, number of individuals; a, number of alleles; R_S, allelic richness per locus and sample; *as*, allele size range in base pairs; H_E, expected heterozygosity; H_O, observed heterozygosity; HW, probability values of concordance with Hardy-Weinberg expectations. Values in bold type are significant probability estimates after false discovery rate (FDR) correction for multiple tests (α = 0.05).

	Celtic Sea							Scotian Shelf						
Locus name	n	a	R_S	as	H_E	H_O	HW	n	a	R_S	as	H_E	H_O	HW
A03_Hill	45	4	4.00	158–170	0.687	0.733	0.885	46	4	4.00	158–170	0.586	0.652	0.217
A04_Neo	45	16	16.00	143–203	0.903	**0.711**	**0.004**	46	18	17.91	143–211	0.925	**0.674**	**0.000**
A08_T3	46	6	6.00	163–183	0.698	0.630	0.187	46	7	7.00	163–187	0.569	0.587	0.468
A11_Hill	46	4	3.98	163–175	0.532	0.478	0.819	45	4	4.00	163–175	0.186	0.200	1.000
A16_M13	46	12	12.00	156–194	0.864	**0.652**	**0.000**	46	10	10.00	170–194	0.650	**0.587**	0.045
A18_M13	46	9	9.00	211–243	0.833	0.804	0.308	44	8	8.00	203–239	0.839	0.727	0.084
A19_T3	46	6	5.93	242–274	0.257	**0.152**	**0.010**	45	7	7.00	250–274	0.266	0.267	0.362
A20_Neo	45	2	2.00	178–182	0.022	0.022	–	46	5	4.98	164–182	0.323	0.304	0.344
A22_Neo	46	5	4.98	214–239	0.338	**0.217**	**0.006**	45	5	5.00	214–243	0.428	0.422	0.927
A30_T3	45	7	7.00	245–269	0.784	0.578	0.023	44	7	7.00	245–269	0.840	**0.545**	**0.000**
A31_T3	46	7	6.96	318–374	0.331	0.304	0.056	45	11	11.00	314–386	0.306	0.244	**0.004**
A33_M13	46	12	11.90	233–277	0.729	**0.435**	**0.000**	42	15	15.00	241–297	0.884	**0.643**	**0.001**
A34_M13	46	7	6.96	297–321	0.699	0.674	0.058	44	6	6.00	297–317	0.636	0.682	0.415
A37_Neo	46	21	20.52	265–489	0.844	0.739	**0.004**	44	25	25.00	289–441	0.929	0.886	0.465
A39_Hill	46	12	12.00	356–408	0.739	0.674	0.092	46	12	12.00	364–416	0.685	0.696	0.144
A43_T3	46	15	14.89	309–369	0.754	0.630	0.374	45	9	9.00	331–359	0.754	0.689	0.556
B01_Hill	46	13	13.00	129–181	0.735	0.717	0.391	46	14	14.00	129–213	0.838	0.826	0.463
B03_T3	46	6	5.91	147–161	0.394	0.435	1.000	44	4	4.00	151–161	0.226	0.250	1.000
B07_M13	46	9	9.00	163–181	0.602	**0.435**	0.081	46	9	9.00	165–187	0.488	**0.326**	0.019
B12_T3	45	11	10.96	346–398	0.741	**0.467**	**0.000**	44	10	10.00	346–398	0.653	**0.409**	**0.001**
B15_T3	46	4	4.00	170–178	0.478	0.500	0.562	46	4	4.00	170–178	0.382	0.435	0.839

Table 2 *Continued.*

	Celtic Sea							Scotian Shelf						
Locus name	n	a	R_S	as	H_E	H_O	HW	n	a	R_S	as	H_E	H_O	HW
B19_T3	46	13	12.93	242–290	0.806	0.783	0.677	45	19	19.00	242–320	0.799	0.867	0.206
B28_Neo	46	4	4.00	226–238	0.536	0.391	0.116	46	4	4.00	226–235	0.507	0.565	0.213
B29_Neo	46	9	8.87	262–281	0.738	0.783	0.894	44	5	5.00	265–278	0.509	0.545	0.612
B30_Neo	46	6	6.00	154–169	0.742	0.783	0.250	45	7	7.00	151–169	0.693	**0.622**	0.222
B33_M13	46	5	4.96	120–135	0.354	0.413	0.690	45	6	6.00	120–135	0.463	0.533	0.896
B36_Hill	41	2	2.00	384–393	0.048	**0.000**	0.013	41	3	3.00	390–396	0.291	**0.098**	**0.000**
B38_Neo	46	3	3.00	409–415	0.365	0.370	1.000	46	4	4.00	388–415	0.065	0.065	1.000
C01_M13	46	14	13.65	127–195	0.844	**0.543**	**0.000**	39	8	8.00	115–151	0.833	**0.410**	**0.000**
C08_T3	46	9	8.98	149–185	0.757	0.761	0.765	45	10	10.00	145–181	0.855	0.800	0.041
C13_Neo	46	5	5.00	158–173	0.276	0.239	0.300	46	6	6.00	158–176	0.242	0.261	1.000
C14_Hill	45	6	6.00	146–166	0.751	**0.644**	0.062	43	4	4.00	146–158	0.631	0.535	0.425
C15_Hill	45	18	18.00	171–205	0.903	**0.667**	**0.000**	46	15	14.91	169–187	0.796	0.761	0.178
C17_M13	46	8	8.00	154–190	0.749	**0.609**	0.059	46	7	7.00	154–182	0.788	0.609	0.083
C20_Neo	45	4	3.96	227–236	0.188	**0.089**	**0.001**	43	5	5.00	227–239	0.276	**0.140**	**0.000**
C22_Hill	46	9	9.00	237–273	0.785	0.739	0.624	46	9	9.00	225–273	0.799	0.848	0.522
C28_Neo	46	7	7.00	258–282	0.580	0.543	0.338	46	7	7.00	255–279	0.480	0.500	0.951
C30_Neo	46	2	2.00	337–340	0.505	0.435	0.389	46	3	3.00	337–343	0.473	0.370	0.229
C31_Neo	45	4	4.00	332–341	0.516	0.556	1.000	46	3	3.00	332–338	0.389	0.391	0.657
C35_M13	44	9	9.00	241–348	0.605	0.591	0.972	46	6	6.00	233–281	0.329	0.326	0.832
C36_Hill	46	6	6.00	319–352	0.347	0.326	0.827	46	7	7.00	325–349	0.260	0.283	1.000
C40_M13	46	2	2.00	276–277	0.410	0.435	1.000	46	2	2.00	276–277	0.161	0.174	1.000
C42_M13	45	2	2.00	354–357	0.481	0.511	0.756	46	8	7.96	345–358	0.711	0.674	0.754
D05_Hill	45	8	8.00	230–262	0.478	0.489	0.526	46	10	9.89	230–270	0.659	0.717	0.984
D10_Neo	46	2	2.00	308–312	0.436	0.500	0.498	46	2	2.00	308–312	0.452	0.370	0.318

Table 2 *Continued.*

	Celtic Sea							Scotian Shelf						
Locus name	n	a	R_S	as	H_E	H_O	HW	n	a	R_S	as	H_E	H_O	HW
D10_Neo	46	2	2.00	308–312	0.436	0.500	0.498	46	2	2.00	308–312	0.452	0.370	0.318
D12_Hill	46	7	6.98	259–279	0.711	0.717	0.055	45	8	8.00	251–291	0.369	0.333	0.049
D14_Hill	46	12	12.00	228–276	0.792	0.739	0.374	46	5	5.00	228–244	0.557	0.565	0.272
D15_Hill	46	3	2.98	244–252	0.144	0.152	1.000	45	4	4.00	244–256	0.169	0.178	1.000
D16_T3	44	4	4.00	424–436	0.340	0.341	0.504	45	4	4.00	424–436	0.207	0.222	1.000
D21_Neo	46	4	3.96	337–343	0.125	0.130	1.000	45	5	5.00	341–389	0.297	0.244	0.128
D30_M13	46	4	4.00	321–333	0.521	0.565	0.939	45	5	5.00	317–333	0.602	0.467	0.043
D35_M13	45	12	11.93	316–360	0.847	**0.667**	**0.002**	44	10	10.00	316–360	0.709	**0.455**	**0.000**
D37_T3	46	4	4.00	321–333	0.572	0.522	0.372	46	5	5.00	313–333	**0.567**	0.870	**0.000**
D43_T3	45	8	8.00	264–292	0.831	**0.444**	**0.000**	41	8	8.00	260–288	0.814	**0.341**	**0.000**
D46_Neo	46	4	3.96	143–155	0.164	0.174	1.000	45	4	4.00	139–151	0.406	0.444	0.337
Average across loci		**7.40**	**7.37**		**0.57**	**0.50**			**7.49**	**7.48**		**0.54**	**0.48**	

Table 3 Loci that exhibited a different repeat motif than initially identified from the raw sequence.

Locus	Expected motif	Detected motif
A16_M13	tetra	di
A22_Neo	tetra	di-tetra
A43_T3	tetra	di
B29_Neo	tri	mono
C15_Hill	tetra	mono
C40_M13	tri	mono
C42_M13	tri	mono
D05_Hill	tetra	di
D21_Neo	tetra	di
D35_M13	tetra	di-tetra

Table 4 Loci showing signs of presence of null alleles

Celtic Sea	Scotian Shelf
A04_Neo	A04_Neo
A16_M13	A30_T3
A19_T3	A33_M13
A22_Neo	B07_M13
A30_T3	B12_T3
A33_M13	B36_Hill
B07_M13	C01_M13
B12_T3	C17_M13
C01_M13	C20_Neo
C15_Hill	D35_M13
C20_Neo	D43_T3
D35_M13	
D43_T3	

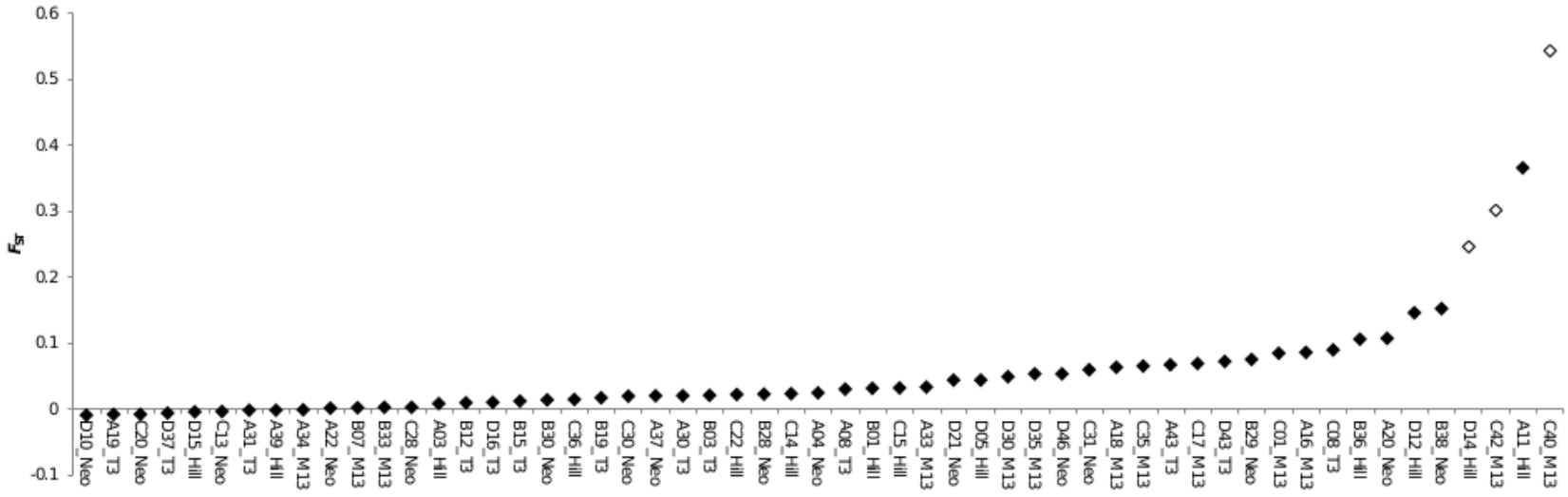

Figure 1 Global F_{ST} –values of individual loci. Unfilled squares signify the loci identified as potentially affected by positive selection.

Replicate sampling of loci to visualise the effect of increasing numbers of loci on F_{ST} estimates and their variances is presented in **Figure 2**. This was done in order to see how many loci were needed to reach F_{ST} point estimates with low variances to accurately describe the level of genetic variability between the Celtic Sea and Scotian Shelf samples. The procedure was performed both with and without loci under potential selection. In both cases, increased number of markers reduced the variation in multilocus F_{ST} estimates.

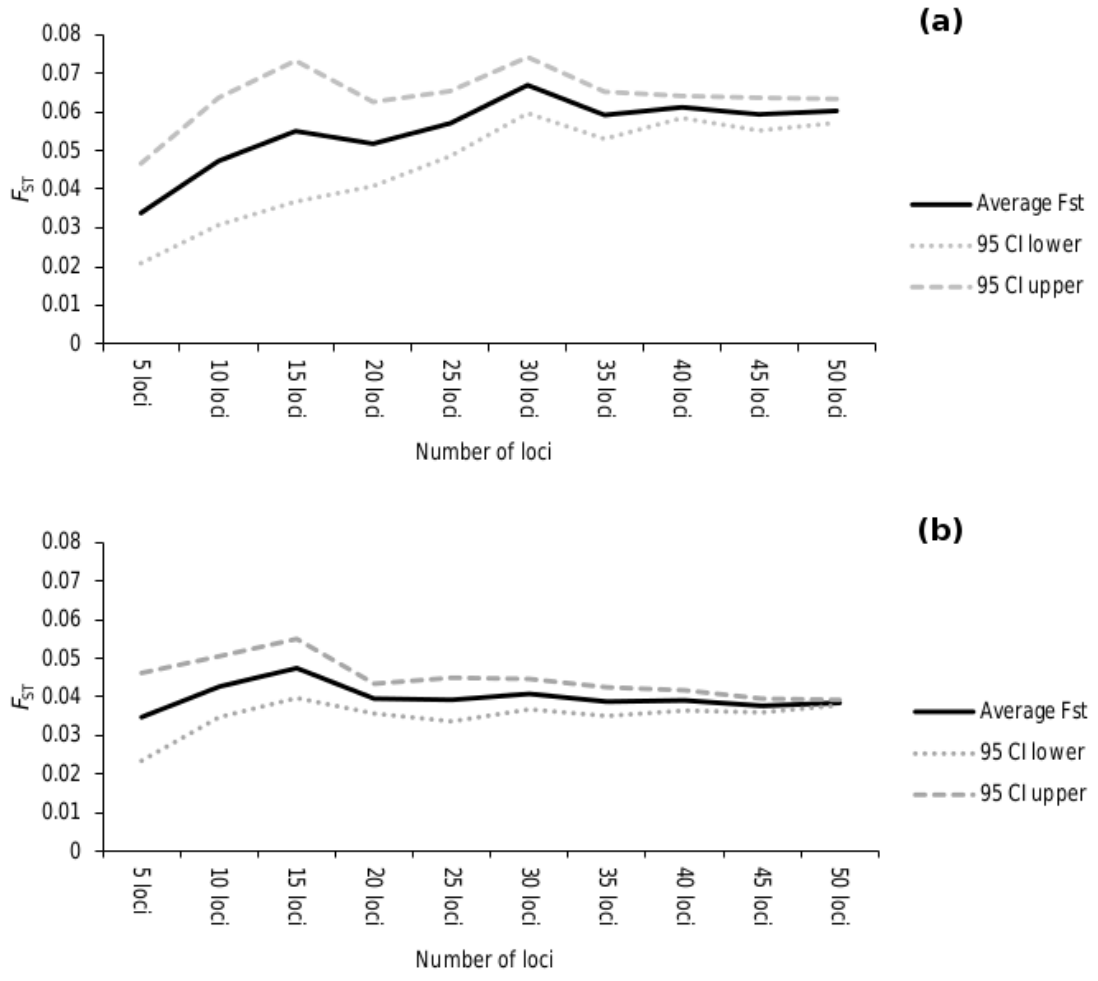

Figure 2 F_{ST} replicate sampling with (a) all 55 loci and (b) with outliers excluded. 95% confidence interval is displayed.

3.4. Discussion

Since the initial reports of three-primer PCR (Steffens et al. 1993; Oetting et al. 1995; Neilan et al. 1997; Schuelke 2000) the approach has gained wide acceptance, particularly for initial validation while using conventional two-primer PCR for genotyping (e.g. Guichoux et al. 2011; Hunter and Hart 2013; Skirnisdottir et al. 2013; Olafsdottir et al. 2013). Similarly, multiplex amplification of microsatellites is now commonly employed. In a few instances, these two techniques have been combined for microsatellite deployment (Missiaggia and Grattapaglia 2006; Langen et al. 2011; Blacket et al. 2012). However, the combined three primer/multiplex PCR approach, as used here for both microsatellite development and deployment, has not to our knowledge been previously reported. The lack of such studies may reflect conservative views on multiplexing and/or the limited availability of suitable universal primers (Guichoux et al. 2011; Blacket et al. 2012). The three-primer/multiplex PCR approach for validation and genotyping has several characteristics that facilitate cost savings (consumables and labour) relative to other approaches. Fluorescently labelled primers are typically an order of magnitude more expensive than unlabelled primers. Therefore, direct modification of locus specific primers substantially increases project costs, especially when markers must be excluded due to low scorability and/or bias (Selkoe and Toonen 2006). The use of fluorescently labelled universal primers avoids these potential complications and further decreases project costs, as a limited number of these primers can be purchased at large synthesis scales. Further multiplexing six to eleven amplicons per ABI capillary lane reduces PCR and genotyping costs as well as labour effort.

Primers were designed to amplify loci in three non-overlapping allele size ranges per dye as reported by Neff et al. (2000), however, larger size separations between ranges were employed. Overlapping size ranges have a disadvantage in that only one marker can be used per dye (see Miller et al. 2013a; Miller et al. 2013b). Because actual allele

sizes were not known in advance of capillary separation, two gap sizes (30 bp and 50 bp) between marker class size ranges were used to minimize overlap chances within a dye set. In practice, only a single overlap between markers was observed in the combined set of gap sizes, while the remaining markers were separated by at least 8 bp. However, microsatellites generated using the 50 bp gap size were more easily combined in multiplex PCR. It is therefore recommended that marker size classes are set apart by at least 50 bp during primer design (e.g. 100-150, 200-250, 300-450 bp) to facilitate combining loci in multiplexes.

Both raw reads and contigs have been used with similar success for microsatellite discovery (see **Table 5**). Contigs can yield more robust primers because increased sequencing depth can be used to detect sequencing errors or genetic variation in the primer binding region (Fernandez-Silva et al. 2013; Zalapa et al. 2012). However, repeat-containing reads may fail to assemble during contig construction, preventing discovery of some valid microsatellite loci (*sensu* Cavagnaro et al. 2010). Also, if the assembler is not able to distinguish the repeat and uses it as the basis for alignment, the unique flanking regions can easily be erroneously collapsed (Treangen and Salzberg 2012). Raw reads were used to maximise microsatellite yield in this study and a 50% amplification success rate was achieved for trialled primers.

Table 5 Comparison of previous studies using either contigs or raw reads in microsatellite discovery via 454 pyrosequencing
The studies employed 454 GS-FLX Titanium chemistry, apart from the publications marked with * which used the 454 GS-FLX chemistry.

Data	Taxon name	Primers screened	Polymorphic	Primer-to-polymorphic marker proportion	Reference		
Contig	*Neophoca cinerea*	28	12	0.43	Ahonen et al. 2013	Average	0.45
	Cyanoramphus malherbi	35	18	0.51	Andrews et al. 2013*	Median	0.43
	Catha edulis	63	27	0.43	Curto et al. 2013*		
	Stylissa carteri	96	12	0.13	Giles et al. 2013		
	Python molurus bivittatus	26	18	0.69	Hunter & Hart 2013		
	Popenaias popeii	28	20	0.71	Inoue et al. 2013		
	Isoodon obesulus	46	9	0.20	Li et al. 2013		
	Antilocapra americana sonoriensis	100	14	0.14	Munguia-Vega et al. 2013		
	Scomber scombrus	80	30	0.38	Olafsdottir et al. 2013		
	multiple species	16–81	8–25	0.15–0.88	Schoebel et al. 2013		
	Unio crassus	77	11	0.14	Sell et al. 2013*		
	Cyclopterus lumpus	48	22	0.46	Skirnisdottir et al. 2013		
	Kunzea pulchella	27	10	0.37	Tapper et al. 2013		
Raw	*Gadus morhua*	15	6	0.40	Carlsson et al. 2013	Average	0.44
	Mulloidichthys flavolineatus	24	23	0.96	Fernandez-Silva et al. 2013	Median	0.38
	Pleuromamma xiphias	15	8	0.53	Fernandez-Silva et al. 2013		
	Brachyptera braueri	30	5	0.17	Geismar & Nowak 2013		
	Euastacus bispinosus	40	15	0.38	Miller et al. 2013a*		
	Neophema chrysogaster	55	14	0.25	Miller et al. 2013b*		
	Prionace glauca	100	12	0.12	Taguchi et al. 2013		
	Silurus asotus	70	47	0.67	Xu et al. 2013		

The design of a multiplex panel usually starts with evaluation of loci in single locus PCR reactions (Neff et al. 2000; Guichoux et al. 2011). For this study, evaluating 192 loci in single PCR reactions would have required 192 additional PCRs on the validation panel of seven individuals and a negative control, and the analyses of the resulting 1536 amplicons via capillary electrophoresis. The elimination of this step reduced primer validation time and lowered consumable and labour costs. The present approach will yield markers for use in multiplex panels. However, it is recognised that some markers that would amplify in single locus PCR may fail in multiplex, therefore potentially lowering the conversion proportion from tested loci to polymorphic loci. Nevertheless, it is contended that the increase in speed outweighs the possible loss of potential markers.

The conversion proportion from tested loci to polymorphic loci in this study was 38% (73/192). This value was similar to the conversion proportion of 40% (6/15) observed in the initial small scale validation by Carlsson et al. (2013) using the same data, and is consistent with recent studies using raw reads from 454 GS-FLX sequencing of genomic DNA (**Table 5**). Considerable variability of primer-to-polymorphic marker proportion has been observed among studies (**Table 5**). A portion of this variation can be attributed to differences in the genome composition of the study organisms (e.g. Schoebel et al. 2013), for example, PCR amplification success is lower in organisms with comparatively large genomes (Garner 2002; Schoebel et al. 2013). This can be due to larger genomes typically harbouring more repetitive elements (Hancock 2002). Microsatellite discovery strategies, such as differences in search parameters and algorithms, or using contigs versus raw sequence reads, can possibly affect the conversion from tested loci to polymorphic loci as well. Variation is also likely caused by different strategies in selection of loci to be validated (see Fernandez-Silva et al. 2013).

3.4.1. Implementation in cod

As a proof of concept, the described approach was applied to samples of cod from the Celtic Sea and the Scotian Shelf. Cod from the Celtic Sea were used for initial microsatellite development (Carlsson et al. 2013) and the Scotian Shelf cod form a genetically distinct population from the Eastern Atlantic (Hutchinson et al. 2001; O'Leary et al. 2007). The present study estimated F_{ST} between Celtic Sea and Scotian Shelf cod at 0.067 when 55 loci were employed. After exclusion of three loci that were potentially under positive selection, F_{ST} was estimated at 0.043. The reduction in F_{ST} is consistent with previous studies that have demonstrated that inclusion of outlier loci that are potentially under selection can markedly affect F_{ST} estimates (Allendorf et al. 2010; Nielsen et al. 2006). The presence and scale of population structure between Celtic Sea and Scotian Shelf cod in the present study concurs with previous studies that examined these populations (Hutchinson et al. 2001; O'Leary et al. 2006, 2007; Pampoulie et al. 2008), and with additional studies that demonstrated population differentiation between the Eastern and Western Atlantic cod (i.e. allozymes, Mork et al. 1985; minisatellite, Galvin et al. 1995; nuclear RFLPs, Pogson et al. 1995; Pogson et al. 2001; microsatellites, Bentzen et al. 1996; Hutchinson et al. 2001; O'Leary et al. 2007; Pampoulie et al. 2008; single nucleotide polymorphisms, O'Leary et al. 2006; Nielsen et al. 2009; Bradbury et al. 2010).

The combined microsatellite validation and genotyping approach presented here was designed to be a fast and cost-effective means for developing and deploying large numbers of microsatellite markers. Using larger numbers of genetic markers confers considerable advantages of increased precision and statistical power when assessing intra- and inter-population genetic parameters such as population structure and gene flow, as well as when inferring demographic parameters, such as effective population size, population expansions and bottlenecks (Nei and Tajima 1981; Ryman et al. 2006). This improved precision allows for more robust and trustworthy management advice based on genetic data. In the present case, the rate of reduction of multilocus F_{ST}

variability decreased after 20 – 30 loci suggesting that this is the point where using more loci only slightly improves the precision of the multilocus F_{ST} estimate. The point of diminishing returns may not be the same for other populations, other geographic scales or other organisms. The advantage of the method presented here is that more loci can be effectively genotyped, ensuring that the point of diminishing returns has in fact been reached and the most precise estimate of population genetic parameter acquired.

The current study combines three-primer PCR with multiplexing to allow for more economical, rapid development and deployment of microsatellite markers discovered from high throughput sequencing data. Fifty-five polymorphic cod microsatellites were combined into six PCR multiplexes, which allowed for determination of F_{ST} between two populations with high precision. This approach is transferable to any species, including those for which extensive sequence resources are not available, and will allow for large and robust population genetic studies while reducing cost and data collection time.

3.5. References

Abdelkrim, J., Robertson, B. C., Stanton, J. A. L., and Gemmell, N. J. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. BioTechniques 46:185–192.

Ahonen, H., Lowther, A. D., Goldsworthy, S. D., Harcourt, R. G., and Stow, A. J. 2013. Characterization of 12 novel microsatellite loci and cross-amplification of four loci in the endangered Australian sea lion (*Neophoca cinerea*). Conservation Genetics Resources 5:283–285.

Allendorf, F. W., Hohenlohe, P. A., and Luikart, G. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 11:697–710.

Allentoft, M. E., Schuster, S. C., Holdaway, R. N., Hale, M. L., McLay, E., Oskam, C., Gilbert, M. T. P., Spencer, P., Willerslev, E., and Bunce, M. 2009. Identification of microsatellites from an extinct moa species using high-throughput (454) sequence data. BioTechniques 46:195–200.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. Journal of Molecular Biology 215:403–10.

Andrews, B. J., Hale, M. L., and Steeves, T. E. 2013. Characterisation of microsatellite loci in the critically endangered orange-fronted kākāriki (*Cyanoramphus malherbi*) isolated using genomic next generation sequencing. Conservation Genetics Resources 5:235–237.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., and Luikart, G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinformatics 9:323.

Ashworth, V. E. T. M., Kobayashi, M. C., De La Cruz, M., and Clegg, M. T. 2004. Microsatellite markers in avocado (*Persea americana* Mill.): Development of dinucleotide and trinucleotide markers. Scientia Horticulturae 101:255–267.

Benjamini, Y., and Yekutieli, D. 2001. The control of the false discovery rate in multiple testing under depencency. The Annals of Statistics 29:1165–1188.

Bentzen, P., Taggart, C. T., Ruzzante, D. E., and 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.

Blacket, M. J., Robin, C., Good, R. T., Lee, S. F., and Miller, A. D. 2012. Universal primers for fluorescent labelling of PCR fragments – an efficient and cost-effective approach to genotyping by fluorescence. Molecular Ecology Resources 12:456–463.

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D. C., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., and 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 277:3725–34.

Brockmann, H. J., Colson, T., and Potts, W. 1994. Sperm competition in horseshoe crabs (*Limulus polyphemus*). Behavioral Ecology and Sociobiology 35:153–160.

Brownstein, M., Carpten, J., and Smith, J. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. Biotechniques 20:1004–1010.

Bruford, M. W., and Wayne, R. K. 1993. Microsatellites and their application to population genetic studies. Current Opinion in Genetics & Development 3:939–943.

Carlsson, J., Gauthier, D. T., Carlsson, J. E. L., Coughlan, J. P., Dillane, E., Fitzgerald, R. D., Keating, U., McGinnity, P., Mirimin, L., and Cross, T. F. 2013. Rapid, economical single-nucleotide polymorphism and microsatellite discovery based on de novo assembly of a reduced representation genome in a non-model organism: A case study of Atlantic cod *Gadus morhua*. Journal of Fish Biology 82:944–958.

Cavagnaro, P. F., Senalik, D. A., Yang, L., Simon, P. W., Harkins, T. T., Kodira, C. D., Huang, S., and Weng, Y. 2010. Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). BMC genomics 11:569.

Chambers, G. K., and MacAvoy, E. S. 2000. Microsatellites: Consensus and controversy. Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology 126:455–476.

Curto, M. A., Tembrock, L. R., Puppo, P., Nogueira, M., Simmons, M. P., and Meimberg, H. 2013. Evaluation of microsatellites of *Catha edulis* (qat; Celastraceae) identified using pyrosequencing. Biochemical Systematics and Ecology 49:1–9.

Dieringer, D., and Schlötterer, C. 2003. (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3:167–169.

Diniz, F. M., Iyengar, A., da Costa Lima, P. S., Maclean, N., and Bentzen, P. 2007. Application of a double-enrichment procedure for microsatellite isolation and the use of tailed primers for high throughput genotyping. Genetics and Molecular Biology 30:380–384.

Edwards, A., Civitello, A., Hammond, H. A., and Caskey, C. T. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. American Journal of Human Genetics 49:746–756.

Ellis, J. R., and Burke, J. M. 2007. EST-SSRs as a resource for population genetic analyses. Heredity 99:125–132.

Estoup, A., Jarne, P., and Cornuet, J. M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology 11:1591–1604.

Fernandez-Silva, I., Whitney, J., Wainwright, B., Andrews, K. R., Ylitalo-Ward, H., Bowen, B. W., Toonen, R. J., Goetze, E., and Karl, S. A. 2013. Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. PLoS ONE 8.

Galvin, P., McGregor, D., and Cross, T. 1995. A single locus PCR amplified minisatellite region as a hypervariable genetic marker in gadoid species. Aquaculture 137:31–40.

Gardner, M. G., Fitch, A. J., Bertozzi, T., and Lowe, A. J. 2011. Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development. Molecular Ecology Resources 11:1093–1101.

Garner, T. W. 2002. Genome size and microsatellites: the effect of nuclear size on amplification potential. Genome 45:212–215.

Geismar, J., and Nowak, C. 2013. Isolation and characterisation of new microsatellite markers for the stonefly *Brachyptera braueri* comparing a traditional approach with high throughput 454 sequencing. Conservation Genetics Resources 5:413–416.

Giles, E. C., Saenz-Agudelo, P., Berumen, M. L., and Ravasi, T. 2013. Novel polymorphic microsatellite markers developed for a common reef sponge, *Stylissa carteri*. Marine Biodiversity 43:237–241.

Goldstein, D. B., Roemer, G. W., Smith, D. A., Reich, D. E., Bergman, A., and Wayne, R. K. 1999. The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics 151:797–801.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., and Petit, R. J. 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources 11:591–611.

Hagell, S., Whipple, A. V., and Chambers, C. L. 2013. Population genetic patterns among social groups of the endangered Central American spider monkey (*Ateles geoffroyi*) in a human-dominated landscape. Ecology and Evolution 3:1388–1399.

Hancock, J. M. 2002. Genome size and the accumulation of simple sequence repeats: Implications of new data from genome sequencing projects. Genetica 115:93–103.

Higgins, B., Hubert, S., Simpson, G., Stone, C., and Bowman, S. 2009. Characterization of 155 EST-derived microsatellites from Atlantic cod (*Gadus morhua*) and validation for linkage mapping. Molecular Ecology Resources 9:733–737.

Huang, X., and Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Research 9:868–877.

Hunter, M. E., and Hart, K. M. 2013. Rapid microsatellite marker development using next generation pyrosequencing to inform invasive Burmese python - *python molurus bivittatus* - management. International Journal of Molecular Sciences 14:4793–4804.

Hutchinson, W., Carvalho, G., and 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, K., Lang, B. K., and Berg, D. J. 2013. Development and characterization of 20 polymorphic microsatellite markers for the Texas hornshell, *Popenaias popeii* (Bivalvia: Unionidae), through next-generation sequencing. Conservation Genetics Resources 5:195–198.

Knapik, E. W., Goodman, A., Ekker, M., Chevrette, M., Delgado, J., Neuhauss, S., Shimoda, N., Driever, W., Fishman, M. C., and Jacob, H. J. 1998. A microsatellite genetic linkage map for zebrafish (*Danio rerio*). Nature Genetics 18:338–343.

Langen, K., Schwarzer, J., Kullmann, H., Bakker, T. C. M., and Thünken, T. 2011. Microsatellite support for active inbreeding in a cichlid fish. PLoS ONE 6.

Li, Y. C., Korol, A. B., Fahima, T., and Nevo, E. 2004. Microsatellites within genes: Structure, function, and evolution. Molecular Biology and Evolution 21:991–1007.

Li, Y., Lancaster, M. L., Cooper, S. J. B., Packer, J. G., and Carthew, S. M. 2013. Characterization of nine microsatellite loci from the endangered southern brown bandicoot (*Isoodon obesulus*) using 454 pyrosequencing. Conservation Genetics Resources 5:105–107.

Malausa, T., Gilles, A., Meglécz, E., Blanquart, H., Duthoy, S., Costedoat, C., Dubut, V., Pech, N., Castagnone-Sereno, P., Délye, C., Feau, N., Frey, P., Gauthier, P., Guillemaud, T., Hazard, L., Le Corre, V., Lung-Escarmant, B., Malé, P. J. G., Ferreira, S., and Martin, J. F. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. Molecular Ecology Resources 11:638–644.

Miller, A. D., Good, R. T., Coleman, R. A., Lancaster, M. L., and Weeks, A. R. 2013a. Microsatellite loci and the complete mitochondrial DNA sequence characterized through next generation sequencing and de novo genome assembly for the critically endangered orange-bellied parrot, *Neophema chrysogaster*. Molecular Biology Reports 40:35–42.

Miller, A. D., Van Rooyen, A., Sweeney, O. F., Whiterod, N. S., and Weeks, A. R. 2013b. The development of 10 novel polymorphic microsatellite markers through next generation sequencing and a preliminary population genetic analysis for the endangered Glenelg spiny crayfish, *Euastacus bispinosus*. Molecular Biology Reports 40:4415–4419.

Mirimin, L., O'Keeffe, D., Ruggiero, A., Bolton-Warberg, M., Vartia, S., and Fitzgerald, R. 2011. A quick, least-invasive, inexpensive and reliable method for sampling *Gadus morhua* postlarvae for genetic analysis. Journal of Fish Biology 79:801–805.

Missiaggia, A., and Grattapaglia, D. 2006. Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers. Genetics and Molecular Research 5:72–8.

Moore, S. S., Sargeant, L. L., King, T. J., Mattick, J. S., Georges, M., and Hetzel, D. J. 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. Genomics 10:654–660.

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

Munguia-Vega, A., Klimova, A., and Culver, M. 2013. New microsatellite loci isolated via next-generation sequencing for two endangered pronghorn from the Sonoran Desert. Conservation Genetics Resources 5:125–127.

Neff, B., Fu, P., and Gross, M. 2000. Microsatellite multiplexing in fish. Transactions of the American Fisheries Society 129:584–593.

Nei, M., and Tajima, F. 1981. Genetic drift and estimation of effective population size. Genetics 98:625–640.

Neilan, B. A., Wilton, A. N., and Jacobs, D. 1997. A universal procedure for primer labelling of amplicons. Nucleic Acids Research 25:2938–9.

Nielsen, E. E., Hansen, M. M., and 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 15:3219–29.

Nielsen, E. E., Hemmer-Hansen, J., Poulsen, N. A., Loeschcke, V., Moen, T., Johansen, T., Mittelholzer, C., Taranger, G.-L., Ogden, R., and Carvalho, G. R. 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'Leary, D. B., Coughlan, J., Dillane, E., McCarthy, T. V., and Cross, T. F. 2007. Microsatellite variation in cod *Gadus morhua* throughout its geographic range. Journal of Fish Biology 70:310–335.

O'Leary, D. B., Coughlan, J., McCarthy, T. V., and Cross, T. F. 2006. Application of a rapid method of SNP analysis (glycosylase mediated polymorphism detection) to mtDNA and nuclear DNA of cod *Gadus morhua*. Journal of Fish Biology 69:145–153.

Oetting, W. S., Lee, H. K., Flanders, D. J., Wiesner, G. L., Sellers, T. A., and King, R. A. 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers.

Olafsdottir, G., Olafsson, K., Skirnisdottir, S., Oskarsson, G. J., Kohlbach, D., Franklinsdottir, H., Kvaavik, C. E. K., Morneau, R., Chevrier, A., Pampoulie, C., Helyar, S., and Danielsdottir, A. K. 2013. Isolation and characterization of thirty microsatellite loci for Atlantic mackerel (*Scomber scombrus* L.). Conservation Genetics Resources 5:491–494.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. 2004. Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535–538.

Pampoulie, C., Stefánsson, M. Ö., Jörundsdóttir, T. D., Danilowicz, B. S., and Daníelsdóttir, A. K. 2008. Recolonization history and large-scale dispersal in the open sea: the case study of the North Atlantic cod, *Gadus morhua* L. Biological Journal of the Linnean Society 94:315–329.

Pogson, G. H., Mesa, K. A., and Boutilier, R. G. 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.

Pogson, G. H., Taggart, C. T., Mesa, K. A., and Boutilier, R. G. 2001. Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. Evolution 55:131–146.

Primmer, C. R., Koskinen, M. T., and Piironen, J. 2000. The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. Proceedings of the Royal Society B: Biological Sciences 267:1699–1704.

Primmer, C. R., Møller, A. P., and Ellegren, H. 1996. A wide-range survey of cross-species microsatellite amplification in birds. Molecular Ecology 5:365–378.

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

Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8:103–6.

Rozen, S., and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology 132:365–386.

Rubin, B. E., Makarewich, C. A., Talaba, A. L., Stenzler, L., Bogdanowicz, S. M., and Lovette, I. J. 2009. Isolation and characterization of microsatellite markers from the acacia-ant *Crematogaster mimosae*. Molecular Ecology Resources 9:1212–4.

Ryman, N., Palm, S., André, C., Carvalho, G. R., Dahlgren, T. G., Jorde, P. E., Laikre, L., Larsson, L. C., Palmé, A., and Ruzzante, D. E. 2006. Power for detecting genetic divergence:

Differences between statistical methods and marker loci. Molecular Ecology 15:2031–2045.

Schlötterer, C., Amos, B., and Tautz, D. 1991. Conservation of polymorphic simple sequence loci in cetacean species. Nature 354:63–65.

Schoebel, C. N., Brodbeck, S., Buehler, D., Cornejo, C., Gajurel, J., Hartikainen, H., Keller, D., Leys, M., Říčanová, Š., Segelbacher, G., Werth, S., and Csencsics, D. 2013. Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. Journal of Evolutionary Biology 26:600–611.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments – A poor man's approach to genotyping for research and high-throughput diagnostics. Nature Biotechnology 18:233–234.

Selkoe, K. A., and Toonen, R. J. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters 9:615–629.

Sell, J., Mioduchowska, M., Kaczmarczyk, A., and Szymańczak, R. 2013. Identification and characterization of the first microsatellite loci for the thick-shelled river mussel *Unio crassus* (Bivalvia: Unionidae). Journal of Experimental Zoology Part A: Ecological Genetics and Physiology 319:113–116.

Skirnisdottir, S., Olafsdottir, G., Olafsson, K., Jendrossek, T., Lloyd, H. A. D., Helyar, S., Pampoulie, C., Danielsdottir, A. K., and Kasper, J. M. 2013. Twenty-two novel microsatellite loci for lumpfish (*Cyclopterus lumpus*). Conservation Genetics Resources 5:177–179.

Squirrell, J., Hollingsworth, P. M., Woodhead, M., Russell, J., Lowe, A. J., Gibby, M., and Powell, W. 2003. How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology 12:1339–1348.

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., Malmstrøm, M., Gregers, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., Sharma, A., Wetten, O. F., Lanzén, A., Winer, R., Knight, J., Vogel, J.-H., Aken, B., Andersen, O., Lagesen, K., Tooming-Klunderud, A., Edvardsen, R. B., Tina, K. G., Espelund, M., Nepal, C., Previti, C., Karlsen, B. O., Moum, T., Skage, M., Berg, P. R., Gjøen, T., Kuhl, H., Thorsen, J., Malde, K., Reinhardt, R., Du, L., Johansen, S. D., Searle, S., Lien, S., Nilsen, F., Jonassen, I., Omholt, S. W., Stenseth, N. C., and Jakobsen, K. S. 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210.

Steffens, D. L., Sutter, S. L., and Roemer, S. C. 1993. An alternate universal forward primer for improved automated DNA sequencing of M13. BioTechniques 15:580–582.

Taguchi, M., Shigenobu, Y., Ohkubo, M., Yanagimoto, T., Sugaya, T., Nakamura, Y., Saitoh, K., and Yokawa, K. 2013. Characterization of 12 polymorphic microsatellite DNA loci in the

blue shark, *Prionace glauca*, isolated by next generation sequencing approach. Conservation Genetics Resources 5:117–119.

Tapper, S. L., Byrne, M., and Gardner, M. G. 2013. Isolation and characterisation of ten microsatellite markers for the tetraploid *Stypandra glauca* R.Br. (Hemerocallidaceae) identified using next generation sequencing. Conservation Genetics Resources 5:529–531.

Tozaki, T., Mashima, S., Hirota, K., Miura, N., Choi-Miura, N. H., and Tomita, M. 2001. Characterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods. DNA Research 8:33–45.

Treangen, T. J., and Salzberg, S. L. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics 13.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research 35:W71–4.

Xu, D., Zhang, Y., and Peng, Z. 2013. High-throughput microsatellite marker development in Amur catfish (*Silurus asotus*) using next-generation sequencing. Conservation Genetics Resources 5:487–490.

Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., and Simon, P. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. American Journal of Botany 99:193–208.

Zane, L., Bargelloni, L., and Patarnello, T. 2002. Strategies for microsatellite isolation: A review. Molecular Ecology 11:1–16.

Chapter 4. Population structure of cod in Irish waters

4.1. Introduction

4.1.1. Evolutionary forces affecting genetic structuring

Populations are defined as primarily self-reproducing units that are sustained by recruitment within the unit, rather than relying on migration and recruitment from outside of the unit for survival (Knutsen et al. 2003). Sometimes populations are separated by physical barriers that prevent migration of individuals between populations; however, barriers may also only reduce migration between populations, and not prevent it entirely. In the marine environment, physical barriers are usually less obvious or less defined and migration is often limited by an organism's dispersal ability (Jones and Wang 2012). A range of biotic and abiotic factors, including oceanic trenches, gyres, temperature, salinity and habitat/nutrient distribution, can form barriers to dispersal at different life-stages of organisms in the marine environment (Bentzen et al. 1996; Hastie et al. 2004; Jørgensen et al. 2005; Banks et al. 2007; Hauser and Carvalho 2008; Selkoe et al. 2010). Populations of variable sizes are distributed within geographical space where they are subjected to evolutionary forces of genetic drift, gene flow, mutation and selection. The interplay of these evolutionary forces causes the development of genetic differentiation between populations.

Genetic drift is change in allele frequencies caused by random sampling of reproducing individuals that pass on their allelic variants, combined with the role of chance in survival of offspring to reproduction. Genetic drift can be the most important evolutionary force, particularly in small populations (Frankham et al. 2010). Genetic drift tends to increase genetic differentiation between populations over time.

Gene flow occurs if migrants reproduce in their recipient population and their alleles become part of the gene pool of the new population. Gene flow will homogenize genetic variation between populations reducing differences caused by

drift (Slatkin 1987). The amount of gene flow depends on physical and natural barriers, such as organism's dispersal ability. Sometimes populations can utilise the same area, and still have minimal gene flow due to reproductive isolation mechanisms. For example, the northeast Arctic cod and Norwegian coastal cod (AC and NC) experience some overlap in spawning areas around the Lofoten Islands, yet there is very limited interbreeding between the two stocks (Wennevik et al. 2008). Extreme spawning-site fidelity has been identified as one of the mechanisms maintaining reproductive isolation in marine fish (Skjæraasen et al. 2011).

Mutation is the only evolutionary force that can create new genetic variation. Generally, most mutations happen randomly on the genome. Like genetic drift, mutation will increase genetic differences between populations as different parts of the genome will mutate in different populations. Compared to drift, however, mutation is much slower in creating genetic differences between populations (Frankham et al. 2010). Mutations are frequently classified as non-synonymous or synonymous depending on whether they change the coded amino acid sequence, or not (Loewe and Hill 2010). Synonymous mutations are often neutral or have such a small effect on the fitness of the organism that they can be considered neutral (Loewe and Hill 2010). Most non-synonymous mutations are detrimental to the organism's fitness, yet a small proportion may be advantageous (Loewe and Hill 2010).

Natural selection happens when individuals that carry certain gene variants have higher reproductive success due to the expression of those genes and, as a result, the beneficial gene variants will become more common in the gene pool of the population (positive selection). Similarly, gene variants that negatively affect their carriers' survival and success in reproducing will become less abundant (negative selection). Sometimes the intermediate phenotype is the most successful while individuals at both extremities have lower fitness (balancing selection). Natural selection is imposed by the environmental conditions experienced by the organism during its life time. Local environmental conditions vary between areas occupied by

populations. These differences in environmental conditions create regional differences in natural selection, which in turn, generate local adaptation, which can be observed as spatial differences in allele frequencies of the selected genes.

4.1.2. Weak genetic structuring in marine environments

Many marine species have considerable dispersal potential as adults can be highly mobile and many species have pelagic egg and larval stages that can be passively carried by currents (Ward et al. 1994). This has led to the view that most marine species have large panmictic populations with no genetic structuring (Hauser and Carvalho 2008). Early genetic studies supported such notion as they failed to show genetic differentiation even across wide geographic scales (Árnason et al. 1992; Árnason and Pálsson 1996). It was also assumed that high migration combined with large movements of individuals across different environments during their lifetime would result in few opportunities for local adaptation (Hauser and Carvalho 2008).

More recently, genetic studies have demonstrated population subdivision in marine fish species on a geographical scale ranging from tens to hundreds of kilometres (Hauser and Carvalho 2008), indicating that previous studies lacked power to detect population structuring. F_{ST} is a commonly used estimator of population differentiation. It indicates the proportion of total genetic variation in a data set explained by differences among populations (Wright 1969). The studies that have detected genetic structuring in marine species have yielded consistently lower F_{ST} values than studies covering comparable distances in freshwater species and terrestrial animals (Ward et al. 1994). As the observed F_{ST} values are low in the marine environment, their biological significance may need to be confirmed by additional measures (Waples 1998). For example, an F_{ST} of less than 0.01 means that less than 1% of total genetic variation in the data is explained by population differentiation. One step in confirming biological significance is establishing the temporal stability of the detected differences by temporal sampling (Hauser and Carvalho 2008). Small F_{ST} estimates can reflect short evolutionary history of many marine populations. For example, the high-latitude habitats in the North Atlantic

have only been re-colonised by major marine vertebrate species in the 10 000–15 000 years since the last glaciation (Pampoulie et al. 2008). This is a short evolutionary time scale to accumulate genetic differences in species with relatively high migration rates (Hauser and Carvalho 2008). The ability to detect differences that have occurred on such short time scale can also be due to low effective population size/census size (N_e/N) ratios in marine populations (Hauser and Carvalho 2008).

Effective population size is the size of an idealised population (following Hardy-Weinberg expectations of a constant population size, equal sex ratio and no immigration, emigration, mutation or selection) that would experience the same effect of drift or inbreeding as the population under study (Harmon and Braude 2010). It is the measure of the population's genetic behaviour compared to that of the ideal population (Frankham et al. 2010). When assessing the viability of populations, N_e is more informative than N, which is the census size including all individuals regardless of their age or reproductive status. It has been suggested that N_e of over 50 is needed to avoid extinction risk due to inbreeding depression and N_e of 500 is needed to maintain evolutionary potential (Franklin 1980). It has since been suggested the numbers of 50/500 need to be updated to 100/1000 (Frankham et al. 2014) or 5000 for retaining adaptive potential (Lande 1995).

N_e/N ratios in the wild can be quite low with average of 0.10–0.11 and median of 0.14 being reported by two review studies of wild populations across diverse species (Frankham 1995; Palstra and Ruzzante 2008). Even lower N_e/N ratios have been reported in marine organisms (Frankham 1995). It is likely that the most important factor reducing N_e in marine species is high variance in reproductive success between individuals (Hauser and Carvalho 2008). Many marine species exhibit a type III survivorship curve where initial fecundity is high, but high mortality occurs during the pelagic larval phase (Hauser and Carvalho 2008). During the pelagic egg and larval phase survival may not be random, but depends on time and place of spawning and hatching. Local environmental conditions can lead to random

'sweepstake recruitment' events (Hedgecock 1994). For example, a shift in time of larval production and food availability can result in large variability in survival between batches of eggs and larvae with whole batches being lost if food is not available (Cushing 1990). In addition, the mating behaviour of some species, such as the large reproductive contribution of few dominant males in cod (Bekkevold et al. 2002), may increase the variance in reproductive success and thus also contribute in reducing the N_e/N ratio.

4.1.3. Population structure in cod

Cod eggs are pelagic and usually hatch one month post fertilization, whereas the larvae may stay in the pelagic stage for up to three months (Cohen et al. 1990). Adult cod can migrate over large distances (Jørgensen et al. 2008), but many coastal populations tend to be more residential (Ruzzante et al. 2000; Neuenfeldt et al. 2013). Significant genetic differentiation has been detected between northwest Atlantic and northeast Atlantic cod using various genetic markers (i.e. allozymes, Mork et al. 1985; minisatellite, Galvin et al. 1995; nuclear RFLPs, Pogson et al. 1995; Pogson et al. 2001; microsatellites, Bentzen et al. 1996; Hutchinson et al. 2001; O'Leary et al. 2007; Pampoulie et al. 2008; single nucleotide polymorphisms, O'Leary et al. 2006; Nielsen et al. 2009; Bradbury et al. 2010). Population structure has also been demonstrated within both the northwest and northeast Atlantic (Ruzzante et al. 1998; Hutchinson et al. 2001; Beacham et al. 2002; O'Leary et al. 2007; Pampoulie et al. 2008b). Patterns of isolation-by-distance between cod populations have been identified (Hutchinson et al. 2001; Beacham et al. 2002; O'Leary et al. 2007), but deep ocean trenches (Bentzen et al. 1996) and oceanic gyres have also been identified as physical barriers that can confine movements of eggs and juveniles (Hauser and Carvalho 2008).

Genetic differentiation has also been detected at smaller geographical scales such as between neighbouring fjords (Knutsen et al. 2003; Nielsen et al. 2003), however these studies have been criticized for having used loci under selection that can bias the estimates of differentiation (Nielsen et al. 2009). Because genetic

differentiation is weak in the marine environment, any detected microscale population structure should ideally be corroborated by ecological data such as results of tagging studies (Nielsen et al. 2009). Indeed, extreme homing behaviour within a spatial scale of less than a kilometre (Skjæraasen et al. 2011) and retention of eggs within spawning grounds have been identified as mechanisms facilitating microscale genetic structuring in marine environments (Knutsen et al. 2007; Ciannelli et al. 2010).

Existence of microscale genetic structuring would mean that local adaptation is possible also in marine species. Macro- and micro-geographic patterns have been detected in pantophysin (Pan I) allele frequencies (Case et al. 2005) with evidence of the locus being under selection (Pogson and Fevolden 2003). In addition, evidence of range-wide adaptive differentiation was provided by identification of 40 gene-associated single nucleotide polymorphisms displaying allele frequencies that showed parallel temperature-associated clines in the eastern and western north Atlantic (Bradbury et al. 2010). Hutchings et al. (2007) studied local adaption in cod by means of a common garden experiment assessing larval growth, survival and phenotypic expression in four populations. They observed divergent phenotypic plasticity in both growth and survival between the four populations that had been undistinguishable in microsatellite analysis. Another common garden experiment revealed a counter gradient variation among coastal cod off the Norwegian coast indicating local adaption and population structure (Salvanes et al. 2004). The northern cod displayed higher growth potential and an increase in condition when compared to the southern cod. The more active feeding behaviour and larger energy allocation to storage tissues of northern cod represents adaptations to high-latitude environment where maximising growth during a short growing season may be a selective advantage (Salvanes et al. 2004).

4.1.4. Farmed-wild interactions

The majority of wild cod stocks are depleted and some are considered collapsed (Marteinsdottir et al. 2005). In 1996, cod was listed on the IUCN Red List of

Threatened Species as “Vulnerable” (Sobel 1996). The state of wild stocks has resulted in an increased interest in cod aquaculture. As a result, several countries throughout the species’ distribution range have initiated cod breeding programmes (Björnsson and Steinarsson 2002; Brown et al. 2003; Paisley et al. 2010; Rise et al. 2009). However, care should be taken because, while the presence of farmed products in the market might alleviate pressure on wild stocks, at the same time unregulated and unsustainable farming can cause unforeseen threats to wild populations.

Escape events occur in nearly all species that are grown in aquaculture (Jensen et al. 2010). In Norway, 3.93 million farmed Atlantic salmon (*Salmo salar*) and 1.05 million farmed cod escaped between 2001 and 2009 (Jensen et al. 2010). Large escape events (>10 000 fish) are usually caused by structural failures of equipment, whereas smaller escape events are often caused by operational, external or biological causes (Jensen et al. 2010). Cod have a higher risk of escaping compared to other species. For example, 1.02% of Norwegian farmed cod escaped in 2001-2009 compared to 0.19% of salmon and 0.40% of trout (Jensen et al. 2010). Due to its biology as a bottom feeder, cod exhibits searching and net biting behaviour and general willingness to escape through holes in the netting, which have been documented both in laboratory experiments and in the field (Moe et al. 2007; Hansen et al. 2008). Unlike the salmonids which mostly spawn in fresh water, farmed cod can spawn in the sea cages. The released fertilised eggs can be carried by currents to suitable nursery grounds and the larvae can get recruited into the natural population. Indeed, field experiments with genetically marked cod have confirmed adult escape events during normal farming activities and the long term survival of both adult individuals and eggs during spawning season from sea cages (Jørstad et al. 2008; Jørstad et al. 2013).

With the growing evidence of local adaptation in marine species, there is a serious concern that fish escaping from large-scale farming operations will interbreed with wild fish leading to introgression. Introgression can lead to outbreeding depression

lowering the fitness of wild populations in terms of survival and recruitment (Bekkevold et al. 2006). Outbreeding depression occurs if introgression leads to offspring displaying sub-optimal phenotypes, or disruption of distinct co-adapted gene complexes (Gilk et al. 2004). In Atlantic salmon, introgression with farmed escapes has significantly altered the genetic integrity of native populations resulting in loss of local adaptation (Bourret et al. 2011). The concerns regarding introgression are two-fold. Firstly, wild populations can introgress with domesticated fish that are maladapted to the wild environment (Araki et al. 2007). Secondly, if the farmed population was founded from a geographically distant population, escapes will lead to mixing of individuals from populations adapted to possibly very different environments (Bekkevold et al. 2006).

As more evidence of local adaptations in cod populations is becoming available (Salvanes et al. 2004; Hutchings et al. 2007), introgression between local cod and escaped farmed cod originating from a geographically distant population may indeed result in outbreeding depression breaking down co-adapted gene complexes. So far, most studies on outbreeding depression have been carried out in salmonids (Gilk et al. 2004; Houde et al. 2011; Côte et al. 2014). In outbred Atlantic salmon, the break-up of co-adapted gene complexes became evident in early developmental stages of second generation hybrid crosses (McGinnity et al. 2003). A study of oxygen stress during embryonic development in wild populations of Atlantic salmon revealed that the effects of outbreeding depression can be stronger under unfavourable environmental conditions (Côte et al. 2014).

McGinnity et al. (2003) also noted that the overall survival of hatchery reared and introgressed fish was lower than that of pure wild fish, however the hatchery fish were larger and performed better than wild fish at specific juvenile life stages. This resulted in hybrids and farmed escapes competitively displacing the wild fish in their juvenile habitat. Competitive displacement of wild fish by escapes can occur at several life stages (Bekkevold et al. 2006). The limiting factors in individual life

stages are still poorly understood and can vary between populations and habitats (Myers 2001).

4.1.5. Cod farming in Ireland

Cod farming in Ireland has a short history. In 2002, cod was named as one of the new candidate species for Irish aquaculture by the Minister for Marine and Natural Resources (NUIG 2002). In 2004, a marine finfish hatchery began rearing cod at the Carna Research Station of National University of Ireland, Galway, in West of Ireland. Farmed stocks of various backgrounds (Scottish, Irish Sea, Norwegian and Celtic Sea) were reared on a pilot commercial farm in a nearby bay. In 2008, the National Broodstock and Breeding Programme for Atlantic cod in Ireland (EIRCOD) was established with cod from the Celtic Sea stock as the source population. In addition to the Celtic Sea strain, a Norwegian reference strain was reared at the facilities for comparison purposes. Over the years, a minimum of 67 704 cod were reared in the sea cages (6542 individuals were transferred from land facilities to sea cages in 2004, 41 907 in 2005, no transfers in 2006, unrecorded number in 2007, 11 374 in 2008 and 7881 in 2009) (Bolton-Warberg 2012). It was envisaged that more cod farms would soon be established in the coastal areas around Ireland (Cawley et al. 2006). However, due to global economic downturn, which began in 2008, and rising wild catches, cod farming has since become financially unviable in the current economic environment in Ireland (FitzGerald et al. 2010). Other cod farming countries were similarly affected, and cod production is currently at a very low level (see Figure 6 of the Introduction).

4.1.6. Evidence for population structuring in the waters around Ireland

One of the objectives of the EIRCOD programme was to genetically characterise the wild cod stocks in the waters surrounding Ireland. The cod stocks around Ireland are divided into three distinct management units by ICES: Irish Sea (VIIa), Celtic Sea (VIIe-k & VIIbc) and West of Scotland (VIa) (**Figure 1**). Within the Celtic Sea stock, VIIbc represents West of Ireland which is part of the total allowable catch (TAC) of

Celtic Sea area, accounting for around 1% of the total TAC, however, it is not part of the stock assessment area (Marine Institute 2013). A study investigating population structure in northwest European waters using SNPs has suggested that the management distinctions are not justified by population genetic data (Heath et al. 2014).

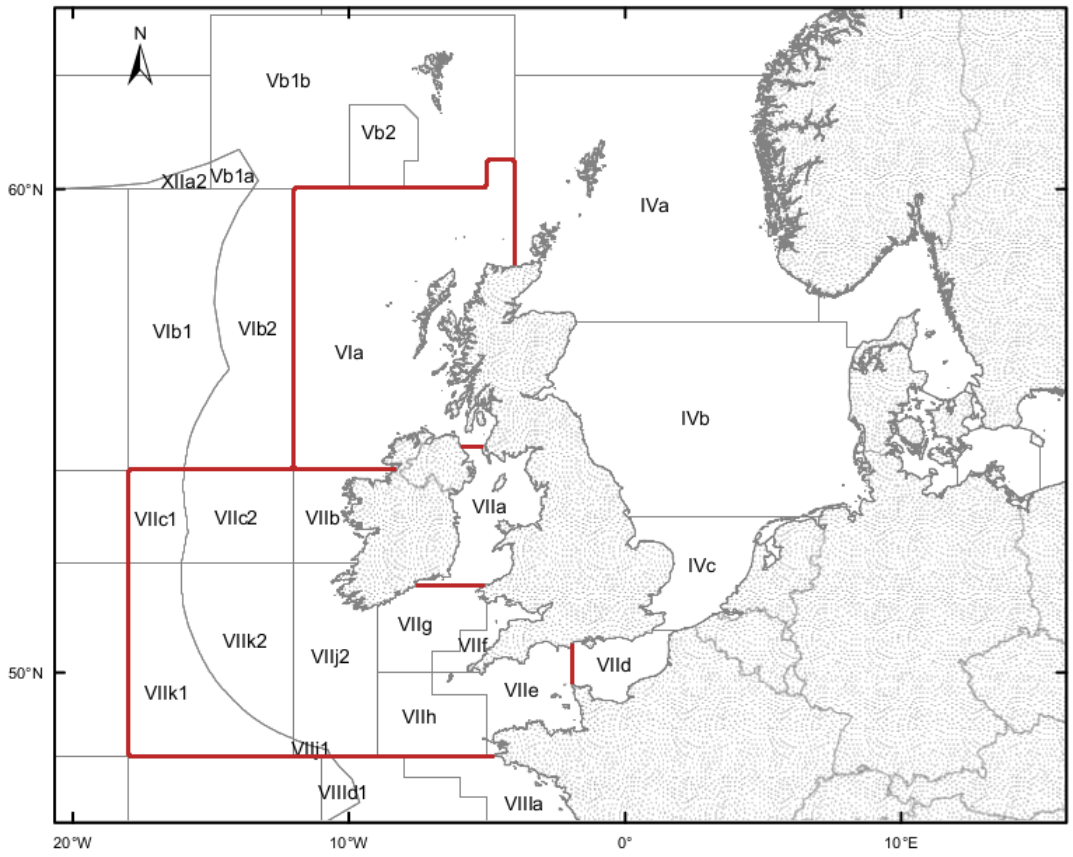

Figure 1 Map of ICES areas in the waters surrounding Ireland. Cod management units are Irish Sea (VIIa), Celtic Sea (VIIe-k & VIIbc) and West of Scotland (VIa).

Previous studies of cod population structure have included samples from either the Irish Sea, Celtic Sea or both (Hutchinson et al. 2001; O'Leary et al. 2006, 2007; Pampoulie et al. 2008). Hutchinson et al. (2001) found no evidence of sub-structuring between the Irish and Celtic Sea with five microsatellite markers. To the contrary, a study using eight microsatellite markers did show a small but significant difference between the Irish and Celtic Sea (Pampoulie et al. 2008). In addition, Pampoulie et al. (2008) found that the Celtic Sea spawning aggregations were temporally stable between spring 2002 and spring 2003 whereas those of the Irish Sea sampled at the same time were not.

Tagging survey targeting spawning fish in Scottish coastal waters showed that >90% of cod tagged in West of Scotland (ICES area VIa) had a home-range of 80 km throughout the year, while cod from the north of Scotland spawning grounds moved around more extensively (Wright et al. 2006). West of Scotland, Irish Sea and Celtic Sea, were included in a tagging study covering nine sea areas in northwest European waters and spanning from 1999 to 2010 (Neat et al. 2014). The lack of tagged individuals in the West of Scotland prevented comparisons of mobility between this and the other areas in this study (Neat et al. 2014). For Celtic Sea the average daily distance travelled (10.1 ± 2.0 km) was slightly higher than for Irish Sea (7.3 ± 1.2 km) (Neat et al. 2014). West of Scotland, Irish Sea and Celtic Sea did not exchange significant numbers of migrants with the neighbouring North Sea populations in the tested time frame (Neat et al. 2014). Most data sets had <1 year between tagging and recapture, which limited conclusions about homing behaviour (Neat et al. 2014).

4.1.7. Aims

A better understanding of population dynamics of cod in the waters surrounding Ireland will be useful for identification of distinct populations and will help in minimising and monitoring possible negative effects of farmed-wild interactions. If cod aquaculture will become more widespread along the Irish coasts such knowledge will be useful for, for example, acquiring fish for broodstock. Cod farming is still a recent phenomenon, for example EIRCOD produced its first F2 generation fish in 2011 and majority of its farmed stock are F1 generation fish. Therefore, the farmed and wild cod are not expected to display big differences in fitness (Jørstad et al. 2013). However, it is very important to record the baseline of genetic variation in the local populations in view of potential establishment of large scale aquaculture, so that the effects of potential escapes can be assessed in the future. Hence, this chapter has three aims: 1) to establish baseline data regarding local genetic variation, 2) assess potential population structure in the waters surrounding Ireland and 3) estimate the effective population sizes of the sampled putative populations.

4.2. Materials and Methods

4.2.1. Sampling

Atlantic cod were collected from five locations at different time points: Celtic Sea was sampled in February of 2011 (*n* = 46) and January of 2012 (*n* = 45), Irish Sea in March of 2011 (*n* = 45) and February-March of 2012 (*n* = 45), West of Ireland in July of 2011 (*n* = 45) and West of Scotland in September-November of 2011 (**Figure 2**). In addition, a sample from the Scotian Shelf, off Nova Scotia in eastern Canada sampled in 1996 (*n* = 46) was included. Fin clip samples were preserved in 100% ethanol.

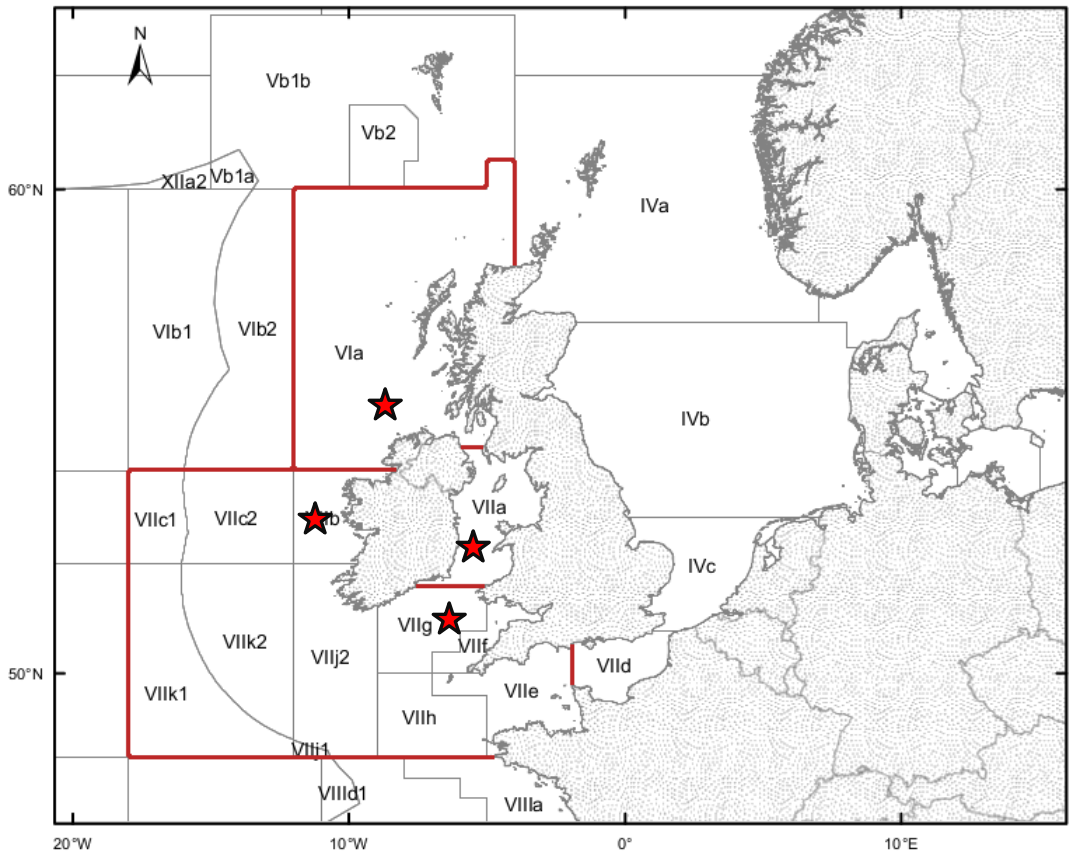

Figure 2 Map of sampling locations in the waters surrounding Ireland (marked with a star).

4.2.2. DNA Extraction

DNA was extracted from fin clips using a Chelex® protocol as described in Mirimin et al. (2011). DNA from the Scotian Shelf samples was extracted using a standard phenol-chloroform method (see O'Leary et al. 2007). DNA was quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific) and normalised to a concentration of 50 ng/µl.

4.2.3. Microsatellite genotyping

The samples were genotyped with 55 microsatellite loci in six multiplex panels developed in Chapter 3 (Table 1 of Chapter 3). Multiplex PCRs were performed as described in Chapter 3. A total of 1 µL of the multiplex PCR product was added to 9 µL of Super-DI™ Formamide (MCLAB) with 0.01 µL of Orange DNA Size Standard (MCLAB) and run on an ABI 3130xl Genetic Analyzer according to manufacturer's recommendations. GeneMarker v.1.97 (Available from http://www.softgenetics.com) was used for fragment length analysis.

When the microsatellite markers had been developed in Chapter 3, allelic ladders were built in GeneMarker v.1.97 for resolving allele size. These allelic ladders typically display a pattern where each allele is separated from the next by one repeat motif. Microvariant alleles are alleles that vary by less than the consensus repeat unit and they are observed between the alleles of the allelic ladder (NCSBI 2007). In non-human population genetic studies it is common practise to bin microvariant alleles to the closest "on-ladder" allele (Ellis et al. 2011). Concurrent analysis of microsatellite sequence data from Chapter 5 revealed that the microvariant alleles were not the result of electrophoresis anomalies, but were true alleles. Therefore, the allelic ladders from Chapter 3 were modified to include microvariant alleles as separate alleles.

4.2.4. Data analysis

Individuals and loci that had more than 20% of genotypes missing were excluded from the data set in order not to bias the analysis. Genotype data were inspected with MICRO-CHECKER 2.2.3 for genotyping errors and presence of null-alleles (95% CI) using default settings (van Oosterhout et al. 2004). MSANALYSER 4.05 (default settings) was used to assess the number of alleles, allele size ranges, F_{ST} estimates and expected and observed heterozygosities (Dieringer and Schlötterer 2003). FSTAT V2.9.3.2 was used to calculate allelic richness (Goudet 1995). Loci were analysed for possible departure from Hardy-Weinberg equilibrium, linkage disequilibrium, and excess and deficit of heterozygotes using GENEPOP 4.2 with default settings (Raymond and Rousset 1995; Rousset 2008). Sequential Bonferroni correction was used to correct for multiple comparisons with initial $\alpha = 0.05/50$ (Rice 1989). LOSITAN was used to detect loci that could be under positive or balancing selection (settings ''Neutral' mean F_{ST}' and 'Force mean F_{ST}' with 50 000 simulations were used under both the infinite allele model and stepwise mutation model) (Antao et al. 2008). Loci identified as being under selection were excluded from further analysis, unless otherwise stated.

Effective population size (N_e) was estimated for (1) all samples and (2) a pooled sample that included all Irish samples, using the linkage disequilibrium method of Waples and Do (2008), as implemented in NeESTIMATOR V2 (Do et al. 2014). Rare alleles can bias the estimates of effective population size with this method, however, they can be screened for by using critical values (Waples and Waples 2011). Critical values are allele frequency thresholds, below which alleles are excluded from the analysis. The critical values were set separately for each sample in order to to exclude alleles that occur in only a single copy (in one heterozygote) in the sample (Waples and Waples 2011). This was accomplished by setting the critical value (PCrit) > 1/(2S), where S is the number of individuals with data at both alleles of loci. For each sample, the S was chosen according to the locus with most missing data. The Pcrit values used were 0.013 for Celtic Sea 2011 and West of Ireland, 0.014 for Celtic Sea 2012 and Scotian Shelf, and 0.022 for the Irish Sea. The

95% confidence intervals of the N_e estimates were calculated using the parametric method as implemented in NeESTIMATOR V2 (Do et al. 2014).

MIGRATE-N 3.6.8 was used to investigate gene flow between the putative populations and as an alternative method of estimating effective population size (Beerli 2009). MIGRATE-N was used to estimate theta, Θ, and M using the Maximum likelihood method (Beerli 1998; Beerli and Felsenstein 1999, 2001). Θ is equal to $4N_e\mu$, where N_e is the long term inbreeding effective population size and μ is the mutation rate and M equals m/μ, where m is the migration rate per generation and μ is the mutation rate. As the mutation rates of the loci were unknown it was not possible to directly estimate N_e, however it was presumed that the mutation rates would be similar across samples and Θ could be used as a relative measure of N_e. The number of effective migrants per generation (N_m) was calculated based on ΘM/4, where Θ is that of the donor population. All analyses were performed under the Brownian motion model. Each run consisted of 10 short chains (sampling 10 000 trees) and four long chains (sampling 1 000 000 trees) with a burn-in period of 10 000 trees. Θ and numbers of migrants were estimated for (1) all locations with the Celtic Sea temporal samples pooled together and for (2) a pooled sample that included all Irish samples *vs*. Scotian Shelf.

STRUCTURE 2.3.4 was used to infer population structure between samples (Pritchard et al. 2000; Falush et al. 2003). The analysis was first carried out without using prior information on sampling location, and then repeated using the information on sampling location as this can help detect weak structure. Admixture ancestry model and correlated allele frequency models were used in both analyses. After carrying out test runs to insure convergence of parameter estimates, the length of burn-in period was set to 50 000 and the number of MCMC reps after burn-in to 500 000. K values from 1 to 5 were tested with five replicates each. Posterior probabilities of K values and delta K values (Evanno et al. 2005) were calculated using STRUCTURE HARVESTER ver. 0.9.94 (Earl and von Holdt 2012).

In the case that one population is quite divergent from the others, removing that population could lead to better inference of the structure between the remaining populations. As the Scotian Shelf sample was known to be differentiated from the Irish samples, the STRUCTURE analysis was repeated with only the Irish samples. In addition, if population structure is very weak, using a large number of loci including many markers with low F_{ST} values could obscure patterns of genetic structure. Choosing a limited set of loci with the most informative markers (high-grading) is commonly practised in population genetic analysis with single nucleotide polymorphism (SNP) markers (Ozerov et al. 2013). High-grading of markers was carried out to further investigate population structure within the Irish samples. The two loci potentially affected by selection were included in the analysis in order to ensure the highest information content. Neutral markers are limited in their capacity to detect differentiation on recent time scales, even when ecological differences exist between populations (Taylor et al. 2000). In contrast, assigning individuals to ecotypes has been successful with selected loci (Russello et al. 2011). The loci with highest detection power were selected based on their locus specific global F_{ST} values. In order to avoid high-grading bias, loci selection was cross-validated by the split sample method (Waples 2010). Each of the Irish samples was divided into two sub-samples with individuals randomly assigned between the sub-samples. Per locus global F_{ST} values were calculated for all loci in both subsamples using MSANALYSER and ranked from highest to lowest. The ten loci with highest global F_{ST} values were compared between the two subsamples. Only the loci that were present in the top ten of both subsamples were selected for the high-grading dataset. This reduced dataset include the selected loci and all the Irish samples. Pairwise F_{ST} values were calculated using MSANALYSER and STRUCTURE analysis was repeated for the reduced dataset with K from 1 to 4.

Discriminant Analysis of Principal Components (DAPC; Jombart et al. 2010) using adegenet 2.0.0 package (Jombart 2008) for the R software (R Development Core Team 2011) was used to identify genetic clusters in a reduced data set containing 26 loci and 27 West of Scotland individuals. The method was applied according to the tutorial of Jombart & Collins (2015). The raw data was transformed using PCA

and all principal components (PCs) were retained. Then k-means, a clustering algorithm which finds a given number (k) of groups maximizing the variation between groups was run with k-values from 1 to 40. The different clustering solutions were compared using Bayesian Information Criterion (BIC). The numbers of groups with the lowest BIC values were further analysed by DAPC.

4.3. Results

All individuals in the Celtic Sea 2011, Celtic Sea 2012, West of Ireland and Scotian Shelf samples had over 80% genotyping success. A total of 35 individuals from the Irish Sea 2011 sample and 10 individuals from the Irish Sea 2012 sample had over 80% of loci successfully genotyped. The Irish Sea samples were pooled together due to low number of individuals in the 2012 sample and as a result temporal stability could not be tested in the Irish Sea. Only one individual in the West of Scotland sample had over 80% genotyping success, hence this sample was left out of the following analysis. As a result, the samples used in further analysis were Celtic Sea 2011 (n = 46), Celtic Sea 2012 (n = 45), Irish Sea 2011/2012 (n = 45), West of Ireland 2011 (n = 45) and Scotian Shelf 1996 (n = 46).

Out of 55 loci in the original multiplex panels, 5 loci that had more than 20% missing genotypes (A33_M13, B12_T3, D10_Neo, D37_T3 and D43_T3) were excluded from further analyses. Loci B12_T3, D10_Neo and D37_T3 displayed large PCR product sizes and weak amplification possibly due to PCR competition which could have led to the large amount of missing data in these loci. A33_M13 was characterized by a complex allelic pattern and D43_T3 displayed weak amplifications, which hampered genotyping. Summary statistics for the 50 loci per each sample can be found in the Appendix 2. The average number of genotyped individuals, number of alleles, allelic richness, expected and observed heterozygosity over all 50 loci are shown in **Table 1** for each sample.

West of Scotland samples

Due to low genotyping success the West of Scotland samples had to be excluded from the main data set. In an effort to salvage some information from these samples, a limited data set was created for use in DAPC analysis. This data set contained a reduced set of 26 loci and included 27 West of Scotland individuals, in addition to the other samples included in the main 50 loci data set.

Table 1 The average values over 50 microsatellite loci for five samples.
n, number of individuals; a, number of alleles; R_S, allelic richness based on minimum sample size of 24 individuals; H_E, expected heterozygosity; H_O, observed heterozygosity.

	n	a	R_S	H_E	H_O
Celtic Sea 2011	45.6	8.2	6.743	0.575	0.499
Celtic Sea 2012	44.3	8.2	6.772	0.559	0.508
Irish Sea 2011/2012	42.2	7.6	6.387	0.541	0.510
West of Ireland 2011	44.8	8.2	6.749	0.572	0.515
Scotian Shelf 1996	45.1	8.3	6.751	0.536	0.479

MICRO-CHECKER detected possible null alleles and stuttering in some of the remaining loci. The loci with null alleles are shown in **Table 2**. LOSITAN identified C40_M13 and D14_Hill as outliers possibly affected by selection under both mutation models, leaving 48 loci in the final data set.

Table 2 Loci showing signs of presence of null alleles. The loci denoted with * also showed signs of stuttering. The locus highlighted in bold showed evidence for null alleles in all samples.

Celtic Sea 2011	Celtic Sea 2012	Irish Sea 2011/2012	West of Ireland 2011	Scotian Shelf 1996
A04_Neo	**A04_Neo**	**A04_Neo**	**A04_Neo**	**A04_Neo**
A08_T3	A16_M13	A16_M13	A16_M13	A18_M13
A16_M13	A37_Neo	A18_M13	A18_M13	A30_T3
A19_T3	A39_Hill	A37_Neo	A22_Neo	A31_T3 *
A22_Neo	B07_M13	A43_T3	A31_T3 *	B07_M13
A31_T3 *	C01_M13	C14_Hill	A37_Neo	B36_Hill
A37_Neo	C15_Hill *	C15_Hill *	A43_T3	C01_M13
A43_T3	C17_M13	C17_M13	B36_Hill	C20_Neo
B07_M13	C20_Neo	D35_M13	C15_Hill	D21_Neo
B28_Neo	D35_M13		C17_M13	D30_M13
B36_Hill			C20_Neo	D35_M13
C01_M13			C35_M13	
C15_Hill *			D35_M13	
C17_M13			D46_Neo *	
C20_Neo				
D12_Hill				
D21_Neo *				
D35_M13				

Population structure

Pairwise F_{ST} values between samples showed that Scotian Shelf was significantly genetically differentiated from all Irish samples (**Table 3**). The F_{ST} estimates for the West of Ireland sample did not differ significantly from either of the Celtic Sea samples, however it was significantly different from the pooled Irish Sea sample. The two Celtic Sea samples from consecutive years showed a small but significant F_{ST} difference between years. The Irish Sea sample was significantly different from the Celtic Sea 2011 sample, but not from the Celtic Sea 2012 sample. The F_{ST} values between the Irish samples were an order of magnitude lower than the values between Scotian Shelf and the Irish samples.

Table 3 Pairwise F_{ST} values between all samples. Pairwise F_{ST} values are below the diagonal and associated P-values are above the diagonal. Values in bold type are significant probability estimates after Bonferroni correction for multiple tests (initial α = 0.05).

	Celtic Sea 2011	Celtic Sea 2012	Irish Sea 2011/2012	West of Ireland 2011	Scotian Shelf 1996
Celtic Sea 2011		**0.000**	**0.000**	0.067	**0.000**
Celtic Sea 2012	0.006		0.008	0.009	**0.000**
Irish Sea 2011/2012	0.006	0.004		**0.000**	**0.000**
West of Ireland 2011	0.002	0.004	0.007		**0.000**
Scotian Shelf 1996	0.048	0.054	0.054	0.051	

The STRUCTURE analysis suggested that K value of two best described the division of genetic variation among samples both with and without prior information on sampling location. The Scotian Shelf sample formed its own cluster, which was clearly separated from the Irish samples (**Figure 3**). No population structure was detected among the Irish samples using this approach.

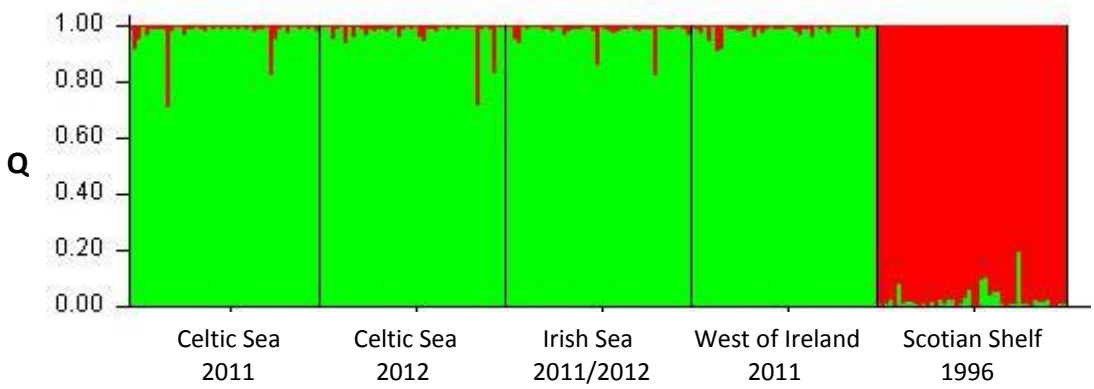

Figure 3 Genetic variation partitioned between samples in STRUCTURE without using sampling location information (optimal partition K=2). Q is the admixture proportions between the two groups for each individual. Each vertical bar represents one individual.

Effective population size

Effective population size estimates calculated with the linkage disequilibrium method (NeESTIMATOR) showed that Celtic Sea 2011, Celtic Sea 2012, West of Ireland and Scotian Shelf had an N_e estimate of 500 or higher, whereas the Irish Sea sample had lower estimate of approximately 270 (**Figure 4**). The N_e estimate of the pooled Irish sample was approximately 8000 (**Figure 5**). In contrast, the Θ values which are relative measures of N_e were lower for all Irish samples as compared to Scotian Shelf (**Figure 6**)**.** The relative N_e estimate of the pooled Irish sample was slightly higher than the estimates of single samples, but still approximately 50% lower than the Θ of Scotian Shelf (**Figure 7**).

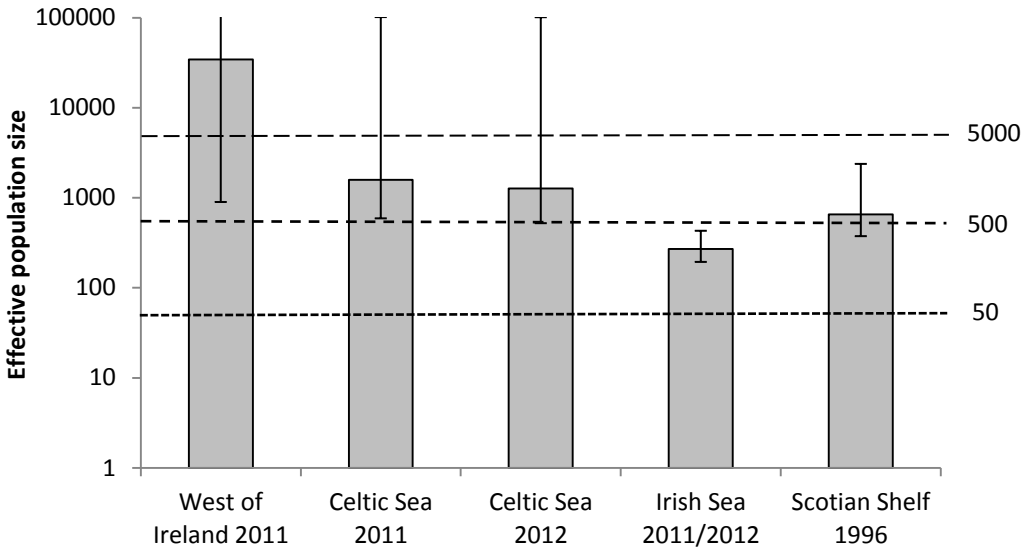

Figure 4 Effective population size estimates for the five samples. N_e was calculated with the linkage disequilibrium model. The error bars represent 95% confidence intervals calculated using the parametric method. The Y-axis is in log scale. The N_e values of 50, 500 and 5000 are marked by the dashed lines to facilitate comparison of estimates to the N_e thresholds of Franklin (1980) and Lande (1995).

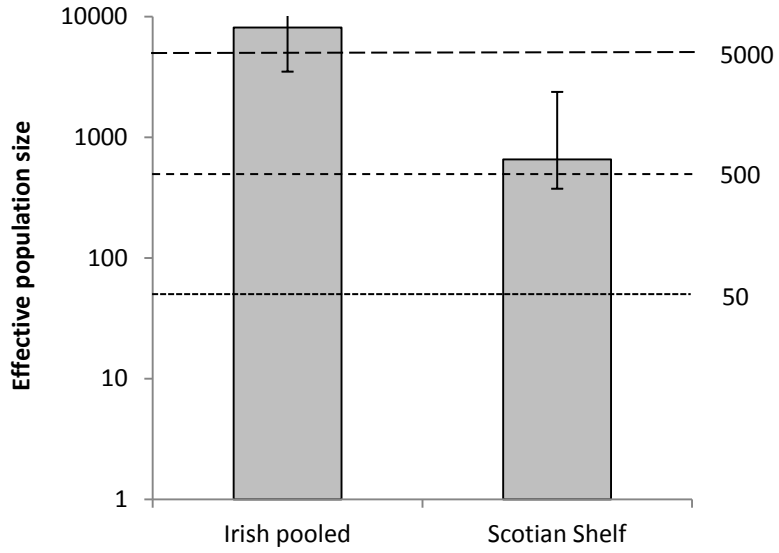

Figure 5 Effective population size estimates for the Irish pooled sample and Scotian Shelf sample. N_e was calculated with the linkage disequilibrium model with lowest allele frequency used 0.050. 95% confidence intervals were calculated using the parametric method. The Y-axis is in log scale. The N_e values of 50, 500 and 5000 are marked by the dashed lines to facilitate comparison of estimates to the N_e thresholds of Franklin (1980) and Lande (1995).

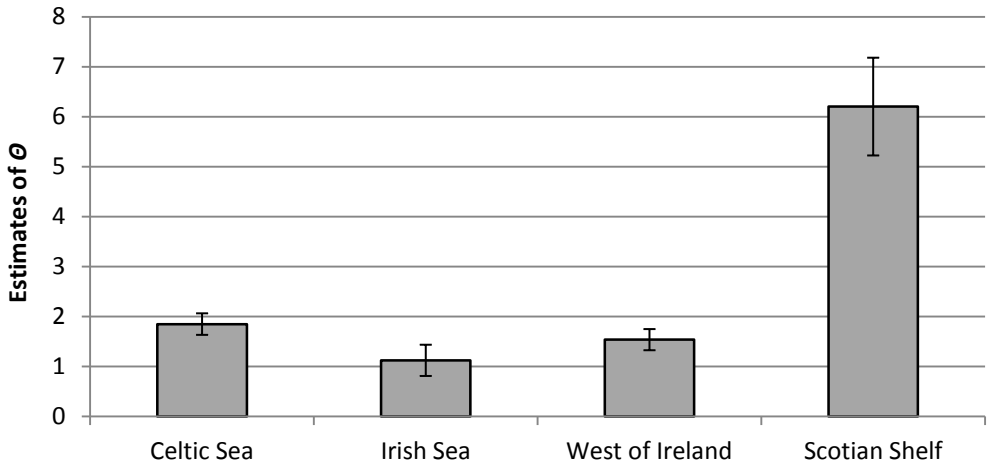

Figure 6 Mutation-scaled effective population sizes (Θ) estimated with MIGRATE-N. If mutation rate (μ) is known, N_e can be calculated based on Θ/4μ. The columns represent an average of five runs and the error bars depict a 95% confidence interval of the mean.

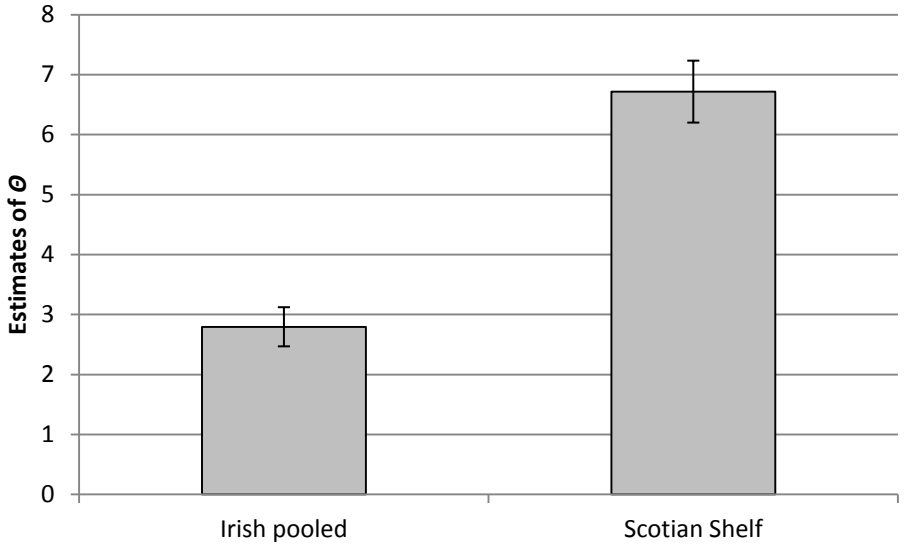

Figure 7 Mutation-scaled effective population sizes (Θ) for pooled Irish sample and Scotian Shelf estimated with MIGRATE-N. If mutation rate (μ) is known, N_e can be calculated based on Θ/4μ. The columns represent an average of five runs and the error bars depict a 95% confidence interval of the mean.

Direction of gene flow

The numbers of effective migrants per generation from Scotian Shelf to all Irish samples were higher (~20-25) than the migrant numbers from these samples to Scotian Shelf (<3) (**Figure 8**). The same pattern was evident with the pooled Irish sample as well (**Figure 9**). Celtic Sea was contributing slightly more migrants to Irish Sea and West of Ireland than it received from them (**Figure 8**)

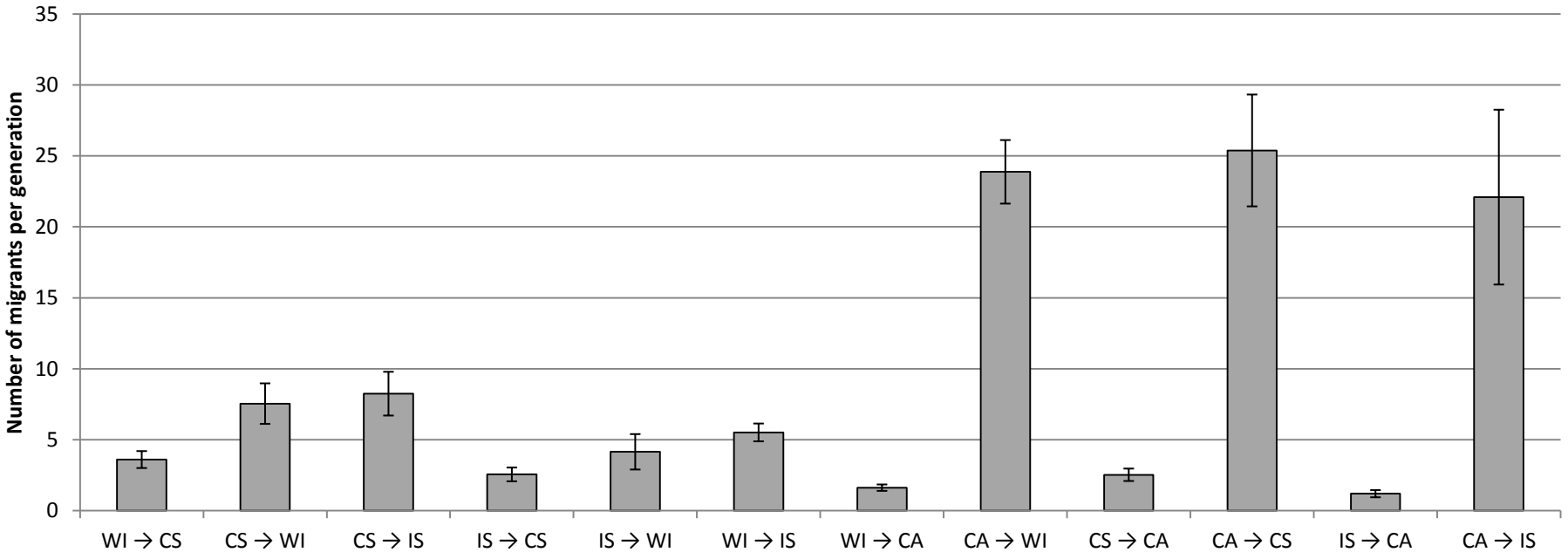

Figure 8 Numbers of migrants between four sampling locations estimated by MIGRATE-N. Celtic Sea 2011 and Celtic Sea 2012 were pooled for this analysis. The number of effective migrants per generation (N_m) was calculated based on ΘM/4, where Θ is that of the donor population. The columns represent an average of five runs and the error bars depict a 95% confidence interval of the mean. IS = Irish Sea, CS = Celtic Sea, WI = West of Ireland and CA = Scotian Shelf

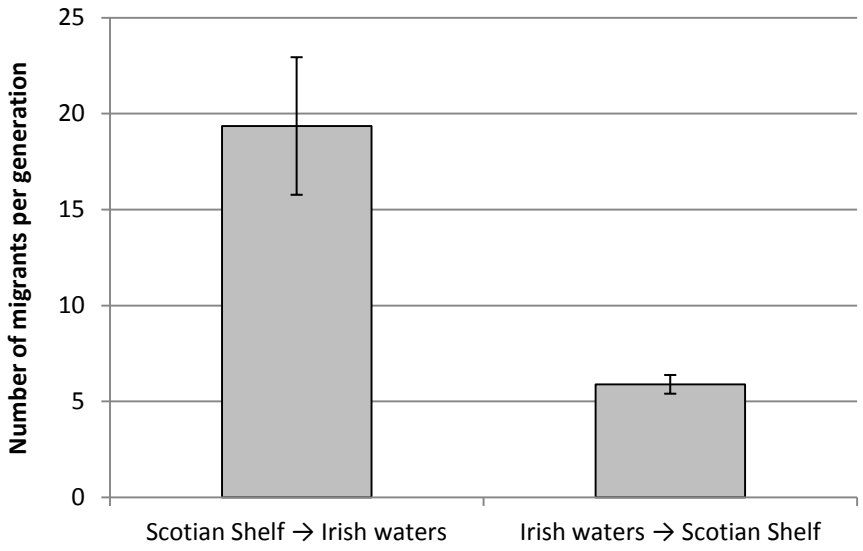

Figure 9 Numbers of migrants per generation between different sampling locations. The number of effective migrants per generation (N_m) was calculated based on ΘM/4, where Θ is that of the donor population. The columns represent an average of five runs and the error bars depict a 95% confidence interval of the mean.

Structure within Irish samples

STRUCTURE analyses of the Irish samples with 48 loci showed no genetic structuring with and without prior information on sampling location. K = 2 was the most likely outcome according to deltaK, however all individuals were a mix of the two putative groups making this result biologically meaningless. Cross validation of loci high-graded based their global F_{ST} values identified 6 loci that appeared in the top ten of both subsets of the data. The six loci were: A22_Neo, A31_T3, B28_Neo, C31_Neo, C40_M13 and D21_Neo. Of these, C40_M13 had been identified as potentially under selection.

Allowing STRUCTURE analysis to use prior information on sampling location can reveal patterns of weak structure that would not otherwise be detected, yet the approach is robust when no population structure exist (Hubisz et al. 2009). Interestingly, when information of sampling location was used in the inference, the reduced dataset revealed a pattern where K=2 was the best outcome and the data were divided into two clusters, one containing 8 individuals in the Celtic Sea 2011

sample and the other containing all the other individuals (**Figure 10**). These 8 individuals could be possible migrants from a divergent population not sampled in this study. Indeed, when these 8 individuals were removed from the original 48 loci dataset including all sampling locations, the Celtic Sea 2011 sample was no longer significantly different from the other Irish samples (**Table 4**). Pairwise F_{ST} values with the reduced 6-loci dataset with putative migrants removed showed that the Irish Sea was significantly different from all of the other samples (**Table 5**).

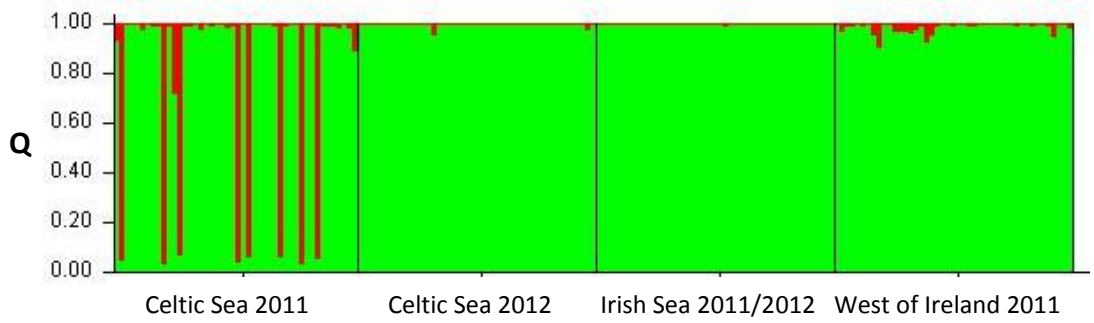

Figure 10 Genetic variation partitioned between the four Irish samples in the reduced 6-loci dataset analysed with STRUCTURE using sampling location information (optimal partition K=2). Q is the admixture proportions between the two groups for each individual. Each vertical bar represents one individual.

Table 4 Pairwise multilocus F_{ST} values between all samples after removing eight putative migrant fish from the Celtic Sea 2011 sample. Pairwise multilocus F_{ST} values are below the diagonal and associated P-values are above the diagonal. Values in bold type are significant probability estimates after Bonferroni correction for multiple tests (initial α = 0.05).

	Celtic Sea 2011	Celtic Sea 2012	Irish Sea 2011/2012	West of Ireland 2011	Scotian Shelf 1996
Celtic Sea 2011		0.018	0.015	0.246	**0.000**
Celtic Sea 2012	0.004		0.006	0.007	**0.000**
Irish Sea 2011/2012	0.004	0.004		**0.000**	**0.000**
West of Ireland 2011	0.001	0.004	0.007		**0.000**
Scotian Shelf 1996	0.048	0.054	0.054	0.051	

Table 5 Pairwise multilocus F_{ST} values between the four Irish samples based on the reduced 6-loci dataset after removing eight putative migrant fish from the Celtic Sea 2011 sample. Pairwise multilocus F_{ST} values are below the diagonal and associated P-values are above the diagonal. Values in bold type are significant probability estimates after Bonferroni correction for multiple tests (initial α = 0.05).

	Celtic Sea 2011	Celtic Sea 2012	Irish Sea 2011/2012	West of Ireland 2011
Celtic Sea 2011		0.212	**0.006**	0.505
Celtic Sea 2012	0.005		**0.000**	0.203
Irish Sea 2011/2012	0.025	0.039		**0.000**
West of Ireland 2011	-0.002	0.004	0.052	

DAPC with West of Scotland samples

The 26 loci included in the reduced data set were: A04_Neo, A08_T3, A11_Hill, A16_M13, A18_M13, A20_Neo, A30_T3, B01_Hill, B07_M13, B15_T3, B19_T3, B33_M13, C01_M13, C08_T3, C13_Neo, C14_Hill, C15_Hill, C17_M13, C20_Neo, C28_Neo, C40_M13, D05_Hill, D14_Hill, D15_Hill, D30_M13 and D46_Neo. Genetic clusters were identified with pre-set number of groups (k) set from 1 to 40. When the BIC of these 1 to 40 groups was compared, the clustering solutions with 2 and 3 groups had the lowest BIC value indicating that they best described the data (**Figure 11**). **Table 6** shows the division of individuals into clusters under both clustering solutions. In the k=2 model, Scotian Shelf forms the majority of group 2, while nearly all the other individuals are assigned to group 1, apart from few stray individuals assigned to group 2. In the k=3 model, Scotian Shelf still forms the group 2, however, the other individuals are now divided between groups 1 and 3. This division looks artificial as individuals from each sampling location seem to be divided into the two groups in more or less equal proportions. It is likely that the division into two genetic clusters better describes the data. These results would suggest that the West of Scotland is part of the same interbreeding unit as Celtic Sea, Irish Sea and West of Ireland.

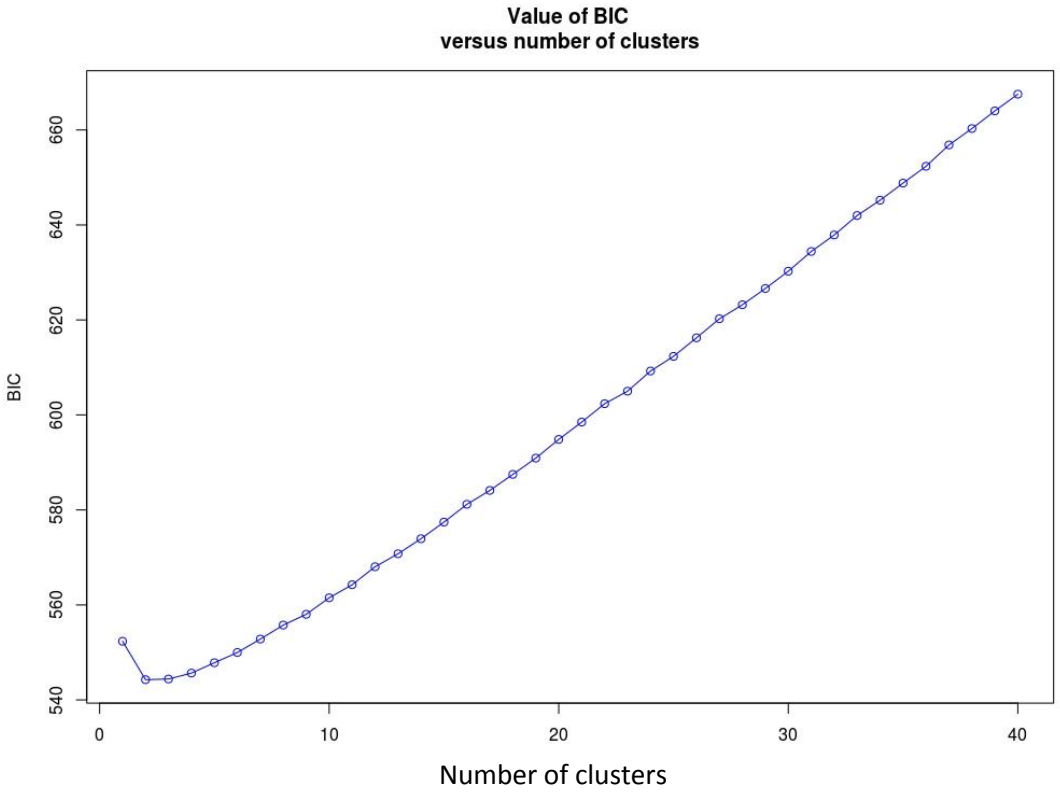

Figure 11 Bayesian Information Criterion against the number of genetic clusters identified in the data.

Table 6 Division of individuals into DAPC identified groups under a) a two group and b) a three group clustering solution.

a) Two groups

	Group 1	Group 2
Celtic Sea 2011	35	3
Celtic Sea 2011, migrants	8	0
Celtic Sea 2012	43	2
Irish Sea 2011/2012	42	3
West of Ireland 2011	43	2
West of Scotland 2011	26	1
Scotian Shelf 1996	0	46

b) Three groups

	Group 1	Group 2	Group 3
Celtic Sea 2011	13	3	22
Celtic Sea 2011, migrants	3	0	5
Celtic Sea 2012	21	2	22
Irish Sea 2011/2012	30	0	15
West of Ireland 2011	22	1	22
West of Scotland 2011	13	1	13
Scotian Shelf 1996	0	46	0

4.4. Discussion

4.4.1. Population structure between the northwest and northeast Atlantic

The set of 48 microsatellite loci detected population structure between the samples from the Irish waters and the Scotian Shelf sample. The pairwise F_{ST} estimate between all the Irish samples and the Scotian Shelf sample was very similar to the F_{ST} estimate between the Celtic Sea sample and the Scotian Shelf sample as analysed in Chapter 3. In chapter 3, the F_{ST} was estimated at 0.067 with all 55 loci and at 0.043 without potentially selected loci (52 loci). The F_{ST} estimates in the present chapter are slightly higher (~0.05-0.06) than the 0.043 from chapter 3. This could be due to the adjustment in the scoring protocol to treat microvariants as separate alleles.

According to the maximum likelihood analysis of migration, the Scotian Shelf is giving more migrants to the Irish samples than it receives. Adult cod exhibit a variety of migratory behaviour from sedentary to homing, and dispersing (Robichaud and Rose 2004). The longest recorded cod migrations span approximately 2000 km (Jørgensen et al. 2008), which is less than the >4000 km needed to cross the Atlantic Ocean. Migration of adult cod across the Atlantic is an improbable explanation for the observed results. It is more likely that the suggested migration events represent a historical pattern of gene flow rather than contemporary migration. Due to recent colonization of contemporary habitats since the end of the last ice age cod may not have had sufficient time to reach migration – drift equilibrium which can distort genetic estimates of connectivity (Nichols & Freeman 2004; Hauser & Carvalho 2008).

Most of cod's current range was covered by ice sheet or was otherwise unsuitable habitat during the last glacial maximum circa 21 000 years ago (Bigg et al. 2008). A paleoclimate model indicated that limited areas of suitable habitat for cod did remain on both sides of the Atlantic during the last glacial maximum. This was

supported by genetic results which indicated the persistence of genealogically continuous populations on both sides of the North Atlantic since before the last ice age (Bigg et al. 2008). This Atlantic basin level differentiation of cod populations has been confirmed by microsatellite analysis by Pampoulie et al. (2008a) who dated the origin of the genetic pattern to the last glacial maximum. Although Pampoulie et al. (2008a) found evidence of isolation by distance indicating contemporary gene flow, this alone was not enough to explain the overall low observed genetic differentiation which was more indicative of postglacial expansion of cod following the retreat of glacial coverage (Pampoulie et al. 2008a). A study using complete mitochondrial genomes identified the most basal (oldest) clade of cod to be from the western Atlantic but their test for directionality of cod movement did not reach statistical signifigance due to low number of samples and limited geographic sampling (Carr & Marshall 2008).

4.4.2. Effective population size

When N_e was estimated with the linkage disequilibrium method, the point estimate was lower for Scotian Shelf (~650) than the pooled Irish sample (~8000). In contrast, the mutation scaled effective population size (Θ) calculated with the maximum likelihood method of MIGRATE-N, estimated the N_e of the Scotian Shelf to be twice as large as that of the pooled Irish sample, or three times as large if the single sample estimates were compared.

The other studies that have attempted to estimate effective population size in cod populations have given variable estimates. Herbinger et al. (1997) studied sibship relationships in a wild larval cohort in the Western Bank of Scotian Shelf to determine the size of a minimum parental population that had produced it. This method established that the lower bound of the effective population size was at least 2800 individuals (Herbinger et al. 1997). Studies on the North Sea, Baltic Sea (Poulsen et al. 2006), southern Gulf of St. Lawrence (Therkildsen et al. 2010) and several locations across the Atlantic (Pampoulie et al. 2008) have given estimates of effective population sizes that ranged in the order of hundreds or thousands,

generally exceeding 500 individuals and displaying wide confidence intervals. These studies employed a range of estimators including the linkage disequilibrium method (Therkildsen et al. 2010), the coalescence-based method (Pampoulie et al. 2008) and the temporal method (Poulsen et al. 2006; Therkildsen et al. 2010).

In contrast, a study of the southern North Sea population using a temporal method gave very low estimates of only 69 individuals (30 –222, 95% CI) between the years 1954 and 1960 and 121 individuals (51– 426, 95% CI) between 1960 and 1970 (Hutchinson et al. 2003). The North Sea study employed only three microsatellite loci and may have included a locus that has been shown to be affected by hitch-hiking selection, which may have biased the estimates (Nielsen et al. 2006).

The study of Pampoulie et al. (2008) estimated the N_e of the Celtic Sea and Irish Sea to be 16 157 and 17 117, respectively, using the coalescence-based method (implemented in MIGRATE-N) assuming a mutation rate (μ) of 10^{-5}. If the same mutation rate was assumed in the current study, and N_e calculated based on $\Theta/4\mu$, the corresponding estimates are higher: Celtic Sea 46 237, Irish Sea 28 094, West of Ireland 38 455 and Scotian Shelf 155 061. However, as different loci were used differences in their mutation rates hamper direct comparison of the values. The coalescence-based N_e estimates are highly reliant on the assumed mutation rate and therefore it is more appropriate to use Θ as a within-study relative measure of effective population size (Carlsson et al. 2004).

The linkage disequilibrium method used in the present chapter is known to be less accurate, the larger the actual effective population size is (Waples and Do 2010). Unfortunately, the other methods of estimating effective population size suffer from the same issue (Therkildsen et al. 2010). N_e estimators rely on detecting drift, therefore they perform poorly for populations with large N_e which experience little drift making the signal-to-noise ratio very small (Nei and Tajima 1981; Waples 1989).

The larger effective population size estimate of the Scotian Shelf sample achieved with the maximum likelihood method in the present study may be a reflection of the historical population size of the Scotian Shelf stock. The spawning stock biomass (SSB) of the eastern Scotian Shelf cod population was at 16 000 t (10 million mature individuals) at the time of sampling in 1996 (DFO 2011). However, in the 1960s the SSB of the population was 110 000t (50 million mature individuals). The SSB of the Celtic Sea stock has been consistently smaller over the years (11 726 t in 2011 and 10 100 t in 1970) (Marine Institute 2013).

4.4.3. Cod in the waters surrounding Ireland

Initial analysis of genetic structuring in the Irish waters detected no structure among the Irish samples based on the clustering method (STRUCTURE approach). In contrast, pairwise F_{ST} estimates showed small but significant structuring between some of the samples, however one of these differences was between the two temporal samples from the Celtic Sea. Both samples were taken from known spawning grounds at spawning time which is often recommended as the optimal time to sample/observe differences due to less mixing of populations (Hutchinson et al. 2001). In addition, the temporal stability of Celtic Sea spawning aggregation in the same area had been previously confirmed by another study (Pampoulie et al. 2008).

Further analysis with the high-graded loci set revealed that 8 migrants from a divergent population in the Celtic Sea 2011 sample were responsible for most of the differences detected by the pairwise F_{ST} values. After removing these individuals from the data set, differences between the Irish samples were reduced and Celtic Sea spawning aggregations were confirmed temporally stable. In the full 48 loci dataset, the only remaining difference was a small but significant pairwise F_{ST} value (multiplocus pairwise F_{ST}: 0.007) between the West of Ireland and Irish Sea. In the reduced 6 loci dataset, the small but significant differences between the Irish Sea and the other samples persisted (pairwise F_{ST} value with Celtic Sea 2011: 0.025; with Celtic Sea 2012: 0.039; and with West of Ireland: 0.052). Apart from the slight

differentiation of the Irish Sea, the results are in concordance with a recent study using SNPs that identified three distinct reproductively isolated population units of cod in the northwest European shelf (Heath et al. 2014). One of these single breeding units was the “Celtic” unit which included the western Channel, Celtic Sea, Irish Sea and Firth of Clyde (located southwest of Scotland) (Heath et al. 2014). The other two units were the “Dogger” unit on Dogger Bank and the southern and western North Sea and the “Viking” unit in the northern North Sea (Heath et al. 2014). A tagging study supported the findings by reporting very limited mixing between these three groups of cod (Neat et al. 2014).

Only locations around the island of Ireland were included in the present study (apart from Scotian Shelf). As a result, it is not possible to determine the origin of the eight migrant fish in the Celtic Sea sample. The genetic differences were subtle enough not to be detected with the full 48 loci dataset. Even though tagging studies have reported very limited mixing between the “Celtic” unit and the other two adjacent groups of cod (Wright et al. 2006; Righton et al. 2010; Neat et al. 2014), cod are capable of large migrations and do not always return to the region where they were tagged (Robichaud and Rose 2004; Neat et al. 2014). In a review of tagging studies, Robichaud and Rose (2004) categorized cod groups as sedentary, accurate homers, inaccurate homers or dispersers. The dispersers were defined as groups that ranged over large areas without recognizable return migrations (Robichaud and Rose 2004). The Celtic Sea 2011 fish had been caught in six hauls over two consecutive days. Migrant individuals were present in four of the hauls, and on both days. The eight fish were likely part of a migrating shoal that was mixing with the resident group at the time of the sampling. Indeed, Pampoulie et al. (2008) speculated that the lack of temporal stability in the Irish Sea spawning aggregation in their study was due to large movements of individuals occurring from year to year resulting from a diversity of migratory patterns of cod in Scottish waters (Wright et al. 2006).

The Celtic Sea, Irish Sea and West of Scotland stocks clustered together in a comparison of productivity levels between 15 cod stocks across the Atlantic showing the highest levels of productivity over all compared stocks (Dutil and Brander 2003). In addition, the three stocks showed similar temporal trends in production experiencing a peak in 1987 and declining since then (Dutil and Brander 2003). High production levels in this stock cluster were associated with higher bottom salinity and temperature (Dutil and Brander 2003). If the three putative stocks share a similar environment, they are likely experiencing similar selection pressures. This supports the argument that they could form a single evolutionary unit.

Heath et al. (2014) had identified west of Scotland (apart from Firth of Clyde) to be part of the "Dogger" unit found on Dogger Bank and the southern and western North Sea. Due to the poor amplification of the west of Scotland samples in the present study, it was not possible to confirm this division. In addition, the poor DNA extraction success prevented the assessment of the temporal stability of the Irish Sea samples.

A common feature of the Irish Sea and West of Scotland samples was a low ethanol to tissue ratio in the sample tubes. A minimum ratio of 2:1 is recommended to ensure preservation of DNA in genetic samples (Stein et al. 2013), however ratios of 0.5:1 and lower were observed in the sample tubes. Low ethanol to tissue ratio can lead to incomplete penetration of the tissue by ethanol. This will leave the interior of the tissue in an enzymatically active state allowing nuclease enzymes to degrade DNA (Spigelman et al. 2001). Upon the receipt of the samples, the ethanol to tissue ratios in the tubes were adjusted to 2:1, but presumably degradation had already taken place.

Using a large panel of loci has more power making it more likely to obtain a significant result even though it may be spurious or biologically meaningless. Waples (1998) has argued that as the genetic signal of population structure is weak

in many marine species, biases and errors in estimating population genetic parameters can have a relatively greater role in these species. He suggested that for example sample size will affect the amount of noise in the F_{ST} parameter (intralocus sampling error) and suggests a bias correction for this error by subtracting the error (approximately 1/(2S), where S is the sample size) from the raw value of the statistic (Waples 1998). If this correction were applied in the present study, the differences between the Irish Sea and the other samples from Irish waters would disappear.

When migration between the Irish locations was investigated, Celtic Sea was contributing ~7-8 migrants per generation to the other Irish sample locations and receiving <2 migrants. These numbers of detected migrants may be enough to prevent population differentiation in neutral loci even if structuring did exist. A simulation study showed that when migration rate (m) increased from negligible to low (from 0.04 to 0.1), population structure was no longer detected by any of the conventional methods including F_{ST} and the assignment method (used in STRUCTURE) (Jones and Wang 2012). The effective population sizes of the sampled locations will determine what kind of a migration rate (m) will the <10 migrants represent. The large confidence intervals of point estimates calculated with the linkage disequilibrium method reduce the accuracy of the estimates. The pooled Irish Sea sample had a narrower confidence interval indicating that its N_e is between ~200-430 individuals. With this kind of N_e, <10 migrants would result in m between 0.02-0.05 which should not hinder the detection of population structuring, even if weak structuring existed (Jones and Wang 2012)

Most population genetics analyses, such as inferences of connectivity, migration rate, F_{ST}, are based on the assumption that the loci employed are selectively neutral (Selkoe and Toonen 2006). Therefore, selected markers or markers experiencing hitch-hiking selection due to close physical proximity to a selected locus cannot be used as they could bias the estimates (Nielsen et al. 2009). Another issue with using selected loci is that strong selection can theoretically cause differences in the

genetic composition of fish from adjacent areas even without reproductive isolation (O'Leary et al. 2007; Nordeide et al. 2011). However, adding selected SNPs to population assignment panels has been shown to yield significantly higher resolution in population assignment over using only neutral loci (Ackerman et al. 2011).

Detecting no differentiation based on neutral genetic markers does not preclude differences in selected markers or genes between populations. In fact, local adaptation would first be seen in selected loci. In common garden experiments, local adaptation has been recorded in traits related to larval growth, larval survival (Hutchings et al. 2007), and growth and condition of 1-year old fish (Salvanes et al. 2004).

4.4.4. Conclusions

The present chapter showed that cod in waters surrounding Ireland are reproductively homogeneous with some indication of weak isolation of the Irish Sea. The differentiation of the Irish Sea was not confirmed by the clustering analysis. The potential structuring was not very strong and its biological significance is yet to be confirmed. In this regard, the ova from the broodstock bred by the EIRCOD programme would be suitable for on-growing on cod farms around the Irish coast. Cod farming in Ireland is a recent phenomenon, hence, the Irish farmed cod and the local wild cod are not expected to be genetically much differentiated yet (Jørstad et al. 2013). However, if cod farming were to become more widely established in Ireland, there is no existing regulation preventing fish farmers from purchasing fertilised eggs from foreign producers which would result in introgression between local and foreign stocks if escapes occurred.

Interest for cod farming has momentarily decreased due to economic downturn, however it is likely that cod aquaculture will become profitable in the future. Therefore, it is very important to have a record of the baseline of local genetic

variation prior to large scale farming taking place. Regardless of the origin of the farmed stock, escape events should be minimised as competitive displacement of wild fish can occur and distress the local population regardless of introgression taking place.

4.5. References

Ackerman, M. W., Habicht, C., and Seeb, L. W. 2011. Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. Transactions of the American Fisheries Society 140:865–881.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., and Luikart, G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323.

Araki, H., Cooper, B., and Blouin, M. S. 2007. Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. Science 318:100–103.

Árnason, E., and 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., and 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.

Beacham, T. 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.

Beerli, P. 1998. Estimation of migration rates and population sizes in geographically structured populations. Advances in Molecular Ecology 306:39–53.

Beerli, P. 2009. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? Pages 42–79 *in* G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi, editors. Population Genetics for Animal Conservation. Cambridge University Press.

Beerli, P., and Felsenstein, J. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152:763–773.

Beerli, P., and Felsenstein, J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences of the United States of America 98:4563–4568.

Bekkevold, D., Hansen, M. M., and Loeschcke, V. 2002. Male reproductive competition in spawning aggregations of cod (*Gadus morhua*, L.). Molecular Ecology 11:91–102.

Bekkevold, D., Hansen, M. M., and Nielsen, E. E. 2006. Genetic impact of gadoid culture on wild fish populations: Predictions, lessons from salmonids, and possibilities for minimizing adverse effects. ICES Journal of Marine Science 63:198–208.

Bentzen, P., Taggart, C. T., Ruzzante, D. E., and 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.

Bigg, G.R, Cunningham, C.W., Ottersen, G., Pogson, G.H., Wadley, M.R., and Williamson, P. 2008. Ice-age survival of Atlantic cod: agreement between palaeoecology models and genetics. Proceedings of the Royal Society of London B: Biological Sciences 275:163–173.

Björnsson, B., and Steinarsson, A. 2002. The food-unlimited growth rate of Atlantic cod (*Gadus morhua*). Canadian Journal of Fisheries and Aquatic Sciences 59:494–502.

Bolton-Warberg, M. 2012. The development of farmed cod in Ireland. PhD Thesis. National University of Ireland Galway.

Bourret, V., O'Reilly, P. T., Carr, J. W., Berg, P. R., and Bernatchez, L. 2011. Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. Heredity 106:500–510.

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D. C., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., and 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 of London B: Biological Sciences 277:3725–3734.

Brown, J. A., Minkoff, G., and Puvanendran, V. 2003. Larviculture of Atlantic cod (*Gadus morhua*): progress, protocols and problems. Aquaculture 227:357–372.

Carlsson, J., McDowell, J. R., Díaz-Jaimes, P., Carlsson, J. E. L., Boles, S. B., Gold, J. R., and Graves, J. E. 2004. Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. Molecular Ecology 13:3345–3356.

Carr, S.M., and Marshall H.D. 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.

Case, R. A. J., Hutchinson, W. F., Hauser, L., Van oosterhout, C., and Carvalho, G. R. 2005. Macro- and micro-geographic variation in pantophysin (Pan I) allele frequencies in NE Atlantic cod *Gadus morhua*. Marine Ecology Progress Series 301:267–278.

Cawley, N., Murrin, J., and O'Bric, R. 2006. Steering a new course – Strategy for a restructured, sustainable and profitable Irish seafood industry 2007-2013, Report of the Seafood Industry Strategy Review Group.

Ciannelli, L., Knutsen, H., Olsen, E. M., Espeland, S. H., Asplin, L., Jelmert, A., Knutsen, J. A., and Stenseth, N. C. 2010. Small-scale genetic structure in a marine population in relation to water circulation and egg characteristics. Ecology 91:2918–2930.

Cohen, D. M., Inada, T., Iwamoto, T., and Scialabba, N. 1990. FAO species catalogue. Vol.10. Gadiform fishes of the world (Order Gadiformes). An annotated and illustrated catalogue of cods, hakes, grenadiers and other gadiform fishes known to date. FAO, Rome.

Cushing, D. 1990. Plankton production and year-class strength in fish populations: an update of the match/mismatch hypothesis. Advances in Marine Biology 26.

DFO. 2013. Stock assessment of northern (2J3KL) cod in 2013. DFO Canadian Science Advisory Secretariat Science Advisory Report 2013/014:1–27.

Dieringer, D., and Schlötterer, C. 2003. (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3:167–169.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., and Ovenden, J. R. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. Molecular Ecology Resources 14:209–214.

Dutil, J.-D., and Brander, K. 2003. Comparing productivity of North Atlantic cod (*Gadus morhua*) stocks and limits to growth production. Fisheries Oceanography 12:502–512.

Earl, D. A., and Von Holdt, B. M. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359–361.

Ellis, J. S., Gilbey, J., Armstrong, A., Balstad, T., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coughlan, J., Cross, T. F., Crozier, W., Dillane, E., Ensing, D., de Leániz, C. G., García-Vázquez, E., Griffiths, A. M., Hindar, K., Hjorleifsdottir, S., Knox, D., Machado-Schiaffino, G., McGinnity, P., Meldrup, D., Nielsen, E. E., Olafsson, K., Primmer, C. R., Prodohl, P., Stradmeyer, L., Vähä, J. P., Verspoor, E., Wennevik, V., and Stevens, J. R. 2011. Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar* L.). Genetica 139:353–367.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molecular Ecology 14:2611–2620.

Falush, D., Stephens, M., and Pritchard, J. K. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.

FitzGerald, R. D., Bass, N., Casburn, P., Bolton-Warberg, M., Roche, L., and Watson, L. 2010. An economic assessment of cod farming in Ireland. Report.

Frankham, R. 1995. Effective population size/adult population size ratios in wildlife: a review. Genetics Research 66:95–107.

Frankham, R., Ballou, J., and Briscoe, D. 2010. Introduction to conservation genetics. 2nd

edition. Cambridge University Press.

Frankham, R., Bradshaw, C. J. A., and Brook, B. W. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation 170:56–63.

Franklin, I. R. 1980. Evolutionary change in small populations. Pages 135–150 *in* M. E. Soule and B. A. Wilcox, editors. Conservation Biology: An Evolutionary-Ecological Perspective. Sinauer, Sunderland.

Galvin, P., McGregor, D., and Cross, T. 1995. A single locus PCR amplified minisatellite region as a hypervariable genetic marker in gadoid species. Aquaculture 137:31–40.

Galway National University of Ireland. 2002. Minister identifies new Fish Species suitable for Commercial Farming while launching new Marine Research Programme. Press release. [online] Nuigalway.ie Available at: http://www.nuigalway.ie/about-us/news-and-events/news-archive/2002/may2002/minister-identifies-new-fish-species-suitable-for-commercial-farming-while-launc-1.html. [Accessed 11 Sep. 2015]

Gilk, S. E., Wang, I. A., Hoover, C. L., Smoker, W. W., Taylor, S. G., Gray, A. K., and Gharrett, A. J. 2004. Outbreeding depression in hybrids between spatially separated pink salmon, *Oncorhynchus gorbuscha*, populations: Marine survival, homing ability, and variability in family size. Environmental Biology of Fishes 69:287–297.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86:485–486.

Hansen, L. A., Dale, T., Damsgård, B., Uglem, I., Aas, K., and Bjørn, P. A. 2008. Escape-related behaviour of Atlantic cod, *Gadus morhua* L., in a simulated farm situation. Aquaculture Research 40:26–34.

Harmon, L. J., and Braude, S. 2010. Chapter 12: Conservation of small populations: effective population sizes, inbreeding, and the 50/500 rule. *in* S. Braude and B. S. Low, editors. An Introduction to Methods and Models in Ecology, Evolution, and Conservation Biology. Princeton University Press.

Hauser, L., and Carvalho, G. R. 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9:333–362.

Heath, M. R., Culling, M. A., Crozier, W. W., Fox, C. J., Gurney, W. S. C., Hutchinson, W. F., Nielsen, E. E., Sullivan, M. O., Preedy, K. F., Righton, A., Speirs, D. C., Taylor, M. I., Wright, P. J., and Carvalho, G. R. 2014. Combination of genetics and spatial modelling highlights the sensitivity of cod (*Gadus morhua*) population diversity in the North Sea to distributions of fishing. ICES Journal of Marine Science 71:794–807.

Hedgecock, D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms? Pages 122–134 *in* A. R. Beaumont, editor. Genetics and Evolution of Aquatic Organisms. Chapman and Hall, New York.

Herbinger, C. M., Doyle, R. W., Taggart, C. T., Lochmann, S. E., Brooker, A. L., Wright, J. M.,

and Cook, D. 1997. Family relationships and effective population size in a natural cohort of Atlantic cod (*Gadus morhua*) larvae. Canadian Journal of Fisheries and Aquatic Sciences 54:11–18.

Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9:1322–1332.

Hutchings, J. A., Swain, D. P., Rowe, S., Eddington, J. D., Puvanendran, V., and Brown, J. A. 2007. Genetic variation in life-history reaction norms in a marine fish. Proceedings of the Royal Society B: Biological Sciences 274:1693–9.

Hutchinson, W., Carvalho, G., and 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.

Hutchinson, W. F., van Oosterhout, C., Rogers, S. I., and Carvalho, G. R. 2003. Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). Proceedings of the Royal Society of London B: Biological Sciences 270:2125–32.

Jensen, Ø., Dempster, T., Thorstad, E., Uglem, I., and Fredheim, a. 2010. Escapes of fishes from Norwegian sea-cage aquaculture: causes, consequences and prevention. Aquaculture Environment Interactions 1:71–83.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jombart T., Devillard S., Balloux F. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics 11:94.

Jombart T. and Collins C. 2015. A tutorial for Discriminant Analysis of Principal Components (DAPC) using *adegenet* 2.0.0. Vignette June 23, 2015. [online] adegenet.r-forge.r-project.org Available at: http://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf [Accessed 14 Oct. 2016]

Jones, O. R., and Wang, J. 2012. A comparison of four methods for detecting weak genetic structure from marker data. Ecology and Evolution 2:1048–1055.

Jørgensen, C., Dunlop, E. S., Opdal, A. F., and Fiksen, O. 2008. The evolution of spawning migrations: state dependence and fishing-induced changes. Ecology 89:3436–3448.

Jørstad, K. E., van der Meeren, T., Paulsen, O. I., Thomsen, T., Thorsen, A., and Svåsand, T. 2008. “Escapes” of eggs from farmed cod spawning in net pens: recruitment to wild stocks. Reviews in Fisheries Science 16:285–295.

Jørstad, K. E., Otterå, H., van der Meeren, T., Dahle, G., Paulsen, O. I., Bakke, G., and Svåsand, T. 2013. Genetic marking of farmed Atlantic cod (*Gadus morhua* l.) and detection of escapes from a commercial cod farm. ICES Journal of Marine Science 71:574–584.

Knutsen, H., Jorde, P. E., André, C., and Stenseth, N. C. 2003. Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. Molecular Ecology 12:385–94.

Knutsen, H., Olsen, E. M., Ciannelli, L., Espeland, S. H., Knutsen, J. A., Simonsen, J. H., Skreslet, S., and Stenseth, N. C. 2007. Egg distribution, bottom topography and small-scale cod population structure in a coastal marine system. Marine Ecology Progress Series 333:249–255.

Lande, R. 1995. Mutation and conservation. Conservation Biology 9:782–791.

Marine Institute. 2013. The Stock book: Annual 2013 review of fish stocks with management advice for 2014. Marine Institute Ireland.

McGinnity, P., Prodöhl, P., Ferguson, A., Hynes, R., Maoiléidigh, N. ó, Baker, N., Cotter, D., O'Hea, B., Cooke, D., Rogan, G., Taggart, J., and Cross, T. 2003. Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. Proceedings of the Royal Society of London B: Biological Sciences 270:2443–2450.

Mirimin, L., O'Keeffe, D., Ruggiero, A., Bolton-Warberg, M., Vartia, S., and FitzGerald, R. 1995. A quick, least-invasive, inexpensive and reliable method for sampling *Gadus morhua* postlarvae for genetic analysis. Journal of Fish Biology 79:801–805.

Moe, H., Dempster, T., Sunde, L. M., Winther, U., and Fredheim, A. 2007. Technological solutions and operational measures to prevent escapes of Atlantic cod (*Gadus morhua*) from sea cages. Aquaculture Research 38:91–99.

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

Myers, R. A., MacKenzie, B. R., Bowen, K. G., and Barrowman, N. J. 2001. What is the carrying capacity for fish in the ocean? A meta-analysis of population dynamics of North Atlantic cod. Canadian Journal of Fisheries and Aquatic Sciences 58:1464–1476.

NCSBI. 2007. Forensic Biology Section Quality Assurance Manual- Revision 08 Appendix F – STR interpretation guidelines. [online] Ncids.com Available at: (http://www.ncids.com/forensic/sbi/Biology/QA/Molecular_Genetics_QA_MANUAL_10.pdf) [Accessed 5 Aug. 2015]

Neat, F. C., Bendall, V., Berx, B., Wright, P. J., Ó Cuaig, M., Townhill, B., Schön, P.-J., Lee, J., and Righton, D. 2014. Movement of Atlantic cod around the British Isles: implications for finer scale stock management. Journal of Applied Ecology 51:1564–1574.

Nei, M., and Tajima, F. 1981. Genetic drift and estimation of effective population size. Genetics 98:625–640.

Neuenfeldt, S., Righton, D., Neat, F., Wright, P. J., Svedäng, H., Michalsen, K., Subbey, S., Steingrund, P., Thorsteinsson, V., Pampoulie, C., Andersen, K. H., Pedersen, M. W., and

Metcalfe, J. 2013. Analysing migrations of Atlantic cod *Gadus morhua* in the north-east Atlantic Ocean: then, now and the future. Journal of Fish Biology 82:741–63.

Nichols, R.A., and Freeman, K.L.M. 2004. Using molecular markers with high mutation rates to obtain estimates of relative population size and to distinguish the effects of gene flow and mutation: a demonstration using data from endemic Mauritian skinks. Molecular Ecology 13:775–787.

Nielsen, E. E., Hansen, M. M., and 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 15:3219–29.

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

Nielsen, E., Wright, P., Hemmer-Hansen, J., Poulsen, N., Gibb, I. M., and Meldrup, D. 2009. Microgeographical population structure of cod *Gadus morhua* in the North Sea and west of Scotland: the role of sampling loci and individuals. Marine Ecology Progress Series 376:213–225.

Nordeide, J. T., Johansen, S. D., Jørgensen, T. E., Karlsen, B. O., and Moum, T. 2011. Population connectivity among migratory and stationary cod *Gadus morhua* in the Northeast Atlantic — A review of 80 years of study. Marine Ecology Progress Series 435:269–283.

O'Leary, D. B., Coughlan, J., Dillane, E., McCarthy, T. V., and Cross, T. F. 2007. Microsatellite variation in cod *Gadus morhua* throughout its geographic range. Journal of Fish Biology 70:310–335.

O'Leary, D. B., Coughlan, J., McCarthy, T. V., and Cross, T. F. 2006. Application of a rapid method of SNP analysis (glycosylase mediated polymorphism detection) to mtDNA and nuclear DNA of cod *Gadus morhua*. Journal of Fish Biology 69:145–153.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. 2004. Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4:535–538.

Ozerov, M., Vasemägi, A., Wennevik, V., Diaz-Fernandez, R., Kent, M., Gilbey, J., Prusov, S., Niemelä, E., and Vähä, J. P. 2013. Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. PLoS ONE 8:e82434.

Paisley, L. G., Ariel, E., Lyngstad, T., Jónsson, G., Vennerström, P., Hellström, A., and Østergaard, P. 2010. An overview of aquaculture in the nordic countries. Journal of the World Aquaculture Society 41:1–17.

Palstra, F. P., and Ruzzante, D. E. 2008. Genetic estimates of contemporary effective population size: What can they tell us about the importance of genetic stochasticity

for wild population persistence? Molecular Ecology 17:3428–3447.

Pampoulie, C., Stefánsson, M. Ö., Jörundsdóttir, T. D., Danilowicz, B. S., and Daníelsdóttir, A. K. 2008a. Recolonization history and large-scale dispersal in the open sea: the case study of the North Atlantic cod, *Gadus morhua* L. Biological Journal of the Linnean Society 94:315–329.

Pampoulie, C., Steingrund, P., Stefánsson, M. Ö., and Daníelsdóttir, A. K. 2008b. Genetic divergence among East Icelandic and Faroese populations of Atlantic cod provides evidence for historical imprints at neutral and non-neutral markers. ICES Journal of Marine Science 65:65–71.

Pogson, G. H., and Fevolden, S.-E. 2003. Natural selection and the genetic differentiation of coastal and Arctic populations of the Atlantic cod in northern Norway: a test involving nucleotide sequence variation at the pantophysin (PanI) locus. Molecular ecology 12:63–74.

Pogson, G. H., Mesa, K. A., and Boutilier, R. G. 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.

Pogson, G. H., Taggart, C. T., Mesa, K. A., and Boutilier, R. G. 2001. Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. Evolution 55:131–146.

Poulsen, N. A. A., Nielsen, E. E., Schierup, M. H., Loeschcke, V., and Grønkjaer, P. 2006. Long-term stability and effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). Molecular Ecology 15:321–31.

Pritchard, J. K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

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

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43:223-225.

Righton, D. A., Andersen, K. H., Neat, F., Thorsteinsson, V., Steingrund, P., Svedäng, H., Michalsen, K., Hinrichsen, H. H., Bendall, V., Neuenfeldt, S., Wright, P., Jonsson, P., Huse, G., Van Der Kooij, J., Mosegaard, H., Hüssy, K., and Metcalfe, J. 2010. Thermal niche of Atlantic cod *Gadus morhua*: Limits, tolerance and optima. Marine Ecology Progress Series 420:1–13.

Rise, M. L., Gamperl, A. K., Johnson, S. C., Robinson, A., Culver, K., Booman, M., Borza, T., Garber, A., Hubert, S., Tosh, J., Alfonso, L. O. B., Boyce, D., Hamoutene, D., Lush, L., Neill, S., Symonds, J., Butts, I. A. E., Feng, C. Y., Hori, T. S., Perez-Casanova, J., Shine, E.,

Hall, J. R., Higgins, B., Kimball, J., Simpson, G., Stone, C., Tarrant-Bussey, J., Elliot, J., Moir, J., Nardi, G., Powell, F., Walsh, A., Rise, M., Trippel, E., and Bowman, S. 2009. The Atlantic Cod Genomics and Broodstock Development Project. Bulletin of the Aquaculture Association of Canada 107:21–30.

Robichaud, D., and Rose, G. A. 2004. Migratory behaviour and range in Atlantic cod: inference from a century of tagging. Fish and Fisheries 5:185–214.

Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources 8:103–106.

Russello, M. A., Kirk, S. L., Frazer, K. K., and Askey, P. J. 2012. Detection of outlier loci and their utility for fisheries management. Evolutionary Applications 5:39–52.

Ruzzante, D. E., Taggart, C. T., and 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.

Ruzzante, D. E., Wroblewski, J. S., Taggart, C. T., Smedbol, R. K., Cook, D., and Goddard, S. V. 2000. Bay-scale population structure in coastal Atlantic cod in Labrador and Newfoundland, Canada. Journal of Fish Biology 56:431–447.

Salvanes, A. G. V., Skjæraasen, J. E., and Nilsen, T. 2004. Sub-populations of coastal cod with different behaviour and life-history strategies. Marine Ecology Progress Series 267:241–251.

Selkoe, K. A., and Toonen, R. J. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecology Letters 9:615–629.

Skjæraasen, J. E., Meager, J. J., Karlsen, Ø., Hutchings, J. A., and Fernö, A. 2011. Extreme spawning-site fidelity in Atlantic cod. ICES Journal of Marine Science 68:1472–1477.

Slatkin, M. 1987. Gene flow and the geographic structure of natural populations. Science 236 :787–792.

Sobel, J. 1996. *Gadus morhua*. [online] The IUCN Red List of Threatened Species 1996: e.T8784A12931575. Available at http://dx.doi.org/10.2305/IUCN.UK.1996.RLTS.T8784A12931575.en. [Accessed 23 Jan. 2016]

Spigelman, M., Barnes, I., Holton, J., Vaira, D., and Thomas, M. G. 2001. Long-term DNA survival in ethanol-preserved archival material. Annals of the Royal College of Surgeons of England 83:283–284.

Stein, E. D., White, B. P., Mazor, R. D., Miller, P. E., and Pilgrim, E. M. 2013. Evaluating ethanol-based sample preservation to facilitate use of DNA barcoding in routine freshwater biomonitoring programs using benthic macroinvertebrates. PLoS ONE 8:1–7.

Taylor, E. B., Kuiper, A., Troffe, P. M., Hoysak, D. J., and Pollard, S. 2000. Variation in

developmental biology and microsatellite DNA in reproductive ecotypes of kokanee, *Oncorhynchus nerka*: Implications for declining populations in a large British Columbia lake. Conservation Genetics 1:231–249.

Therkildsen, N. O., Nielsen, E. E., Swain, D. P., and Pedersen, J. S. 2010. Large effective population size and temporal genetic stability in Atlantic cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. Canadian Journal of Fisheries and Aquatic Sciences 67:1585–1595.

Waples, R. S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics 121:379–391.

Waples, R. S. 2010. High-grading bias: Subtle problems with assessing power of selected subsets of loci for population assignment. Molecular Ecology 19:2599–2601.

Waples, R. S., and Do, C. 2008. LDNE: A program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources 8:753–756.

Waples, R. S., and Do, C. 2010. Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evolutionary Applications 3:244–262.

Waples, R. S., and Waples, R. K. 2011. Inbreeding effective population size and parentage analysis without parents. Molecular Ecology Resources 11:162–171.

Ward, R. D., Woodwark, M., and Skibinski, D. O. 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. Journal of Fish Biology 44:213–232.

Wennevik, V., Jørstad, K. E., Dahle, G., and Fevolden, S. E. 2008. Mixed stock analysis and the power of different classes of molecular markers in discriminating coastal and oceanic Atlantic cod (*Gadus morhua* L.) on the Lofoten spawning grounds, Northern Norway. Hydrobiologia 606:7–25.

Wright, P. J., Galley, E., Gibb, I. M., and Neat, F. C. 2006. Fidelity of adult cod to spawning grounds in Scottish waters. Fisheries Research 77:148–158.

Wright, S. 1969. Evolution and the genetics of populations. Volume 2: The theory of gene frequencies. The University of Chicago Press, Chicago.

Chapter 5. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding[2]

[2] Published as Vartia, S., Villanueva-Cañas, J.L., Finarelli, J., Farrell, E.D., Collins, P.C., Hughes, G.M., Carlsson, J.E.L., Gauthier, D.T., McGinnity, P., Cross, T.F., FitzGerald, R.D., Mirimin, L., Crispie F, Cotter, P.D. & Carlsson J. (2016). A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. Royal Society Open Science, 3: 150565.

5.1. Introduction

This chapter examines the potential of next generation sequencing based '*Genotyping By Sequencing'* (GBS) of microsatellite loci for rapid and cost-effective genotyping in large-scale population genetic studies. The advent of next generation sequencing (NGS) technologies has fundamentally changed how genetic sequence data are generated (Mardis 2013). While NGS was primarily introduced to substantially increase sequence yield for genome projects (Hofreuter et al. 2006; Oh et al. 2006; Wheeler et al. 2008), it has in addition enabled high-throughput genotyping that can be used for genetic studies, including population genetics, by utilising a range of protocols (e.g. RAD: (Baird et al. 2008); ddRAD: (Peterson et al. 2012); 2bRAD: (Wang et al. 2012)). This is collectively known as '*Genotyping By Sequencing'* (GBS) (Davey et al. 2011; Narum et al. 2013).

5.1.1. Genotyping-by-sequencing

The primary advantage of GBS for population genetic studies is the generation of increased quantities of data that allows for improved statistical power and high genome representation (Narum et al. 2013). The concurrent development of single nucleotide polymorphism (SNP) genotyping platforms (e.g. SNP-chip: (LaFramboise 2009; Ha et al. 2014) microfluidic TaqMAN assays (Chowdhury et al. 2007)) and the persistent problems associated with microsatellite genotyping has led to a shift from using microsatellites to SNPs as the preferred marker for genetic studies (Seeb et al. 2011). The main problems with capillary and gel based microsatellite studies include fragment size-homoplasy, poor levels of inter-laboratory calibration, the genotype not including the underlying sequence information and inherently laborious genotyping (Delmotte et al. 2001; Pasqualotto et al. 2007). However, there are also unresolved problems regarding SNPs, such as ascertainment bias, transferability among SNP genotyping platforms and the requirement for high template DNA quality (Kuhner et al. 2000; Nielsen 2000; Helyar et al. 2011). While SNPs do not suffer many of the issues associated with microsatellite genotyping, the major advantage of microsatellites over SNP based approaches for population

analyses is the higher statistical power per locus (Anderson & Garza 2006; Morin et al. 2009; Haasl & Payseur 2011). Additionally, microsatellites are preferred to SNPs in forensic, parentage and kinship studies due to their higher mutation rates and polyallelic nature (Clayton et al. 1998; Gill 2001; Glaubitz et al. 2003; Morin et al. 2004).

Many of the issues associated with microsatellite-based population studies could be mitigated using a GBS approach. The effects of size-homoplasy (Estoup et al. 2002) can be effectively eliminated because the genotype incorporates the underlying sequence information. The difficulties of inter-laboratory calibration would be significantly reduced, as GBS considers actual base-pair lengths of the alleles, and not the estimated allele size sequence length (Pasqualotto et al. 2007). Additionally, the elimination of time consuming capillary or gel electrophoresis runs (Kan et al. 2004) can reduce genotyping time, thereby increasing efficiency. Further, GBS has the potential to utilise established microsatellite multiplex panels, enabling calibration with existing data sets and facilitating inter-laboratory collaboration. However, few GBS studies based on microsatellites have been published and all were in the forensic sciences (Fordyce et al. 2011; Scheible et al. 2011; Van Neste et al. 2012).

5.1.2. Aims

The current chapter aims to assess the potential of microsatellite GBS using cod as the study organism. The primary objective was to examine the potential of microsatellite GBS using *de novo* and existing capillary/gel electrophoresis-based multiplex marker panels. The secondary objective was to develop a rapid and cost effective method for microsatellite GBS, that can take advantage of modern NGS platforms, using combinatorial barcoding for implementation in large-scale population genetics studies.

5.2. Material and Methods

5.2.1. Sample material

A total of 64 cod were collected from the Celtic Sea (n = 32) and the West of Ireland (n = 32) in 2011. Tissue samples were preserved in molecular grade ethanol. DNA was extracted using a Chelex® protocol following Mirimin et al. (2011). DNA was quantified using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and diluted to a concentration of 50 ng/µl. These samples had previously been genotyped in six original PCR multiplexes at 53 microsatellite loci using a conventional ABI capillary-based approach (Vartia et al. 2014, Chapters 3 and 4) thus providing a platform for direct comparison with the current study.

5.2.2. Microsatellite multiplex panels

Six multiplex PCR panels comprising 53 microsatellite loci of Vartia et al. (2014, Chapter 3) were analysed to test the capability of microsatellite GBS on all individuals. Each of the six original multiplex panels included loci in three different size classes (class I: 115-213 bp, class II: 203-320 bp and class III: 265-416 bp, see **Table 1**). To assess the preferential amplification of shorter fragments during PCR and NGS sequencing (Wattier et al. 1998; Van Neste et al. 2012), weighted ratio mixtures of fluorescently labelled amplicons were visualised on an ABI 3130xl Genetic Analyzer (Applied Biosystems; conditions according to Vartia et al. 2014) suggesting an optimal size-class amplicon-ratio of 1(I):1(II):8(III).

Table 1 Multiplex panels used in the three experimental conditions with the size ranges of markers based on Vartia et al. 2014.

Size-class-combination 1		**Size-class-combination 2**		**Size-class-combination 3**	
3 PCR		**6 PCR**		**18 PCR**	
Multiplex 1	**Size-class I**	**Multiplex 1**	**Size-class mixed**	**Multiplex 1a**	**Size-class I**
A08_T3	163–187	A08_T3	163–187	A08_T3	163–187
A16_M13	156–194	A16_M13	156–194	A16_M13	156–194
C15_Hill	169–205	C15_Hill	169–205	C15_Hill	169–205
B03_T3	147–161	B19_T3	242–320		
B30_Neo	151–169	C28_Neo	255–282	**Multiplex 1b**	**Size-class II**
C01_M13	115–195	C40_M13	276–277	B19_T3	242–320
C14_Hill	146–166	D14_Hill	228–276	C28_Neo	255–282
B01_Hill	129–213	A43_T3	309–369	C40_M13	276–277
B15_T3	170–178	B38_Neo	388–415	D14_Hill	228–276
C13_Neo	158–176	C36_Hill	319–352		
C17_M13	154–190	D30_M13	317–333	**Multiplex 1c**	**Size-class III**
A11_Hill	163–175			A43_T3	309–369
B33_M13	120–135	**Multiplex 2**	**Size-class mixed**	B38_Neo	388–415
C08_T3	145–185	B03_T3	147–161	C36_Hill	319–352
D46_Neo	139–155	B30_Neo	151–169	D30_M13	317–333
A04_Neo	143–211	C01_M13	115–195		
B07_M13	163–187	C14_Hill	146–166	**Multiplex 2a**	**Size-class I**
A03_Hill	158–170	A18_M13	203–243	B03_T3	147–161
A20_Neo	164–182	A19_T3	242–274	B30_Neo	151–169
		C20_Neo	227–239	C01_M13	115–195
Multiplex 2	**Size-class II**	D12_Hill	251–291	C14_Hill	146–166
B19_T3	242–320	A34_M13	297–321		
C28_Neo	255–282	A37_Neo	265–489	**Multiplex 2b**	**Size-class II**
C40_M13	276–277	B12_T3	346–398	A18_M13	203–243
D14_Hill	228–276			A19_T3	242–274
A18_M13	203–243	**Multiplex 3**	**Size-class mixed**	C20_Neo	227–239
A19_T3	242–274	B01_Hill	129–213	D12_Hill	251–291
C20_Neo	227–239	B15_T3	170–178		
D12_Hill	251–291	C13_Neo	158–176	**Multiplex 2c**	**Size-class III**
A33_M13	233–297	C17_M13	154–190	A34_M13	297–321
B29_Neo	262–281	A33_M13	233–297	A37_Neo	265–489
C22_Hill	225–273	B29_Neo	262–281	B12_T3	346–398
A22_Neo	214–243	C22_Hill	225–273		
D15_Hill	244–256	A39_Hill	356–416	**Multiplex 3a**	**Size-class I**
D43_T3	260–292	C30_Neo	337–343	B01_Hill	129–213
B28_Neo	226–238	C42_M13	345–358	B15_T3	170–178
A30_T3	245–269	D37_T3	313–333	C13_Neo	158–176
C35_M13	233–348			C17_M13	154–190
D05_Hill	230–270				
				Multiplex 3b	**Size-class II**
				A33_M13	233–297
				B29_Neo	262–281

Table 1 *Continued.*

Multiplex 3	**Size-class III**
A43_T3	309–369
B38_Neo	388–415
C36_Hill	319–352
D30_M13	317–333
A34_M13	297–321
A37_Neo	265–489
B12_T3	346–398
A39_Hill	356–416
C30_Neo	337–343
C42_M13	345–358
D37_T3	313–333
A31_T3	314–374
D21_Neo	337–389
D10_Neo	308–312
D35_M13	316–360
B36_Hill	384–396
C31_Neo	332–341

Multiplex 4	**Size-class mixed**
A11_Hill	163–175
B33_M13	120–135
C08_T3	145–185
D46_Neo	139–155
A22_Neo	214–243
D15_Hill	244–256
A31_T3	314–374
D21_Neo	337–389

Multiplex 5	**Size-class mixed**
A04_Neo	143–211
B07_M13	163–187
D43_T3	260–292
B28_Neo	226–238
D10_Neo	308–312
D35_M13	316–360

Multiplex 6	**Size-class mixed**
A03_Hill	158–170
A20_Neo	164–182
A30_T3	245–269
C35_M13	233–348
D05_Hill	230–270
B36_Hill	384–396
C31_Neo	332–341

C22_Hill	225–273

Multiplex 3c	**Size-class III**
A39_Hill	356–416
C30_Neo	337–343
C42_M13	345–358
D37_T3	313–333

Multiplex 4a	**Size-class I**
A11_Hill	163–175
B33_M13	120–135
C08_T3	145–185
D46_Neo	139–155

Multiplex 4b	**Size-class II**
A22_Neo	214–243
D15_Hill	244–256

Multiplex 4c	**Size-class III**
A31_T3	314–374
D21_Neo	337–389

Multiplex 5a	**Size-class I**
A04_Neo	143–211
B07_M13	163–187

Multiplex 5b	**Size-class II**
D43_T3	260–292
B28_Neo	226–238

Multiplex 5c	**Size-class III**
D10_Neo	308–312
D35_M13	316–360

Multiplex 6a	**Size-class I**
A03_Hill	158–170
A20_Neo	164–182

Multiplex 6b	**Size-class II**
A30_T3	245–269
C35_M13	233–348
D05_Hill	230–270

Multiplex 6c	**Size-class III**
B36_Hill	384–396
C31_Neo	332–341

5.2.3. Experimental conditions

Three PCR size-class-combinations for GBS were evaluated (**Table 1**). The first *'de novo'* multiplex panels were weighted for amplicon size by combining all markers of the same size classes (3-PCR) from the six original multiplexes using a subset of 16 individuals from the Celtic Sea. The second size-class-combination consisted of the existing multiplexes (6-PCR) using the same subset of 16 individuals from the Celtic Sea (*sensu* Vartia et al. 2014). The third PCR size-class-combination comprised the six existing multiplex panels, subdivided into three smaller multiplex panels according to the three size classes using all 64 individuals (18-PCR).

5.2.4. Barcoding of amplicons

The commonly used three-primer PCR approach (Schuelke 2000; Vartia et al. 2014) was modified into a four-primer PCR to produce unique individual combinatorially barcoded amplicons suitable for pooled amplicon NGS in population genetics studies. Each barcode comprised 10 bp of synthesised DNA sequence (MID sequences; Roche 2009). The modified protocol consisted of locus specific primers (forward and reverse) that were adapted to include universal primer sequence (**Figure 1, Table 2**). Two barcoded universal primers were included to incorporate two barcodes into each of the resulting amplicons. In total, twelve forward and eight reverse DNA barcodes allowed for the recovery of 96 unique individual combinatorial barcodes (**Table 3**).

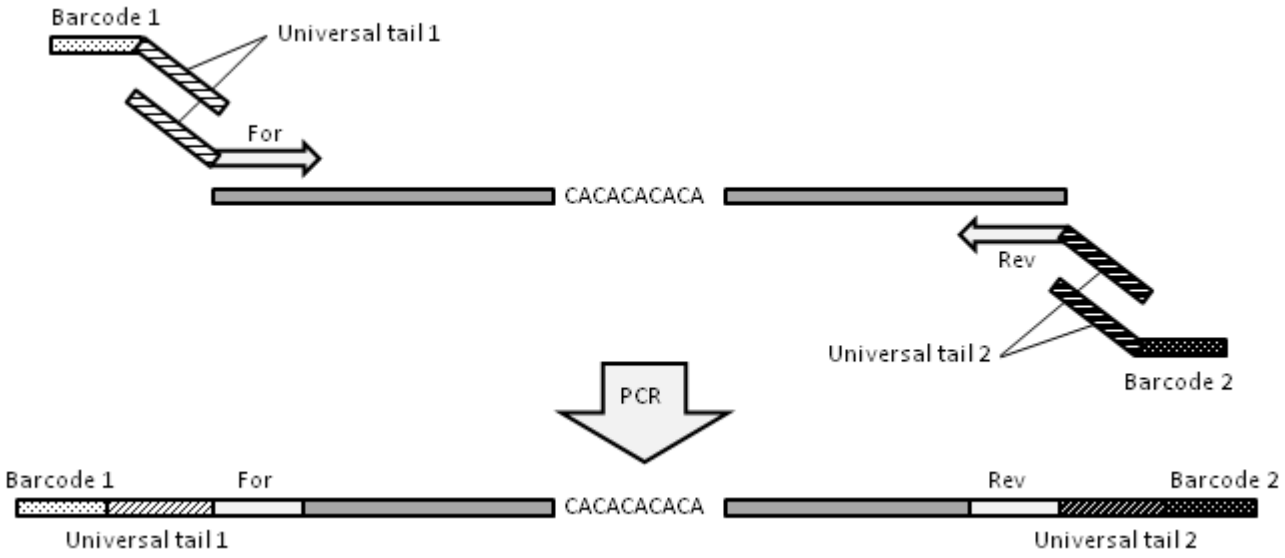

Figure 1 Diagram of the 4-primer PCR and the structure of the resulting amplicon.

Table 2 Barcode and universal primer sequences.

Oligonucleotide type	Name	Sequence 5' → 3'
Forward universal primer	T3	AATTAACCCTCACTAAAGGG
Forward universal primer	M13 Reverse	GGATAACAATTTCACACAGG
Forward universal primer	Hill	TGACCGGCAGCAAAATTG
Forward universal primer	Neomycin rev	AGGTGAGATGACAGGAGATC
Reverse universal primer	CAG	CAGTCGGGCGTCATCA
Barcode_F1	MID-1	ACGAGTGCGT
Barcode_F2	MID-2	ACGCTCGACA
Barcode_F3	MID-3	AGACGCACTC
Barcode_F4	MID-4	AGCACTGTAG
Barcode_F5	MID-5	ATCAGACACG
Barcode_F6	MID-6	ATATCGCGAG
Barcode_F7	MID-7	CGTGTCTCTA
Barcode_F8	MID-10	TCTCTATGCG
Barcode_F9	MID-11	TGATACGTCT
Barcode_F10	MID-13	CATAGTAGTG
Barcode_F11	MID-14	CGAGAGATAC
Barcode_F12	MID-15	ATACGACGTA
Barcode_R1	MID-16	TCACGTACTA
Barcode_R2	MID-17	CGTCTAGTAC
Barcode_R3	MID-19	TGTACTACTC
Barcode_R4	MID-21	CGTAGACTAG
Barcode_R5	MID-22	TACGAGTATG
Barcode_R6	MID-23	TACTCTCGTG
Barcode_R7	MID-24	TAGAGACGAG
Barcode_R8	MID-27	ACGCGAGTAT

Table 3 The barcode combinations used to identify individual amplicons. CS = Celtic Sea; WI = West of Ireland; 3-PCR is size-class combination 1; 6-PCR is size-class combination 2; and 18-PCR is size-class combination 3.

	Barcode_R1	Barcode_R2	Barcode_R3	Barcode_R4	Barcode_R5	Barcode_R6	Barcode_R7	Barcode_R8
Barcode_F12	CS_09	CS_10	CS_11	CS_12	CS_13	CS_14	CS_15	CS_16
Barcode_F11	CS_01	CS_02	CS_03	CS_04	CS_05	CS_06	CS_07	CS_08
Barcode_F10	CS_09	CS_10	CS_11	CS_12	CS_13	CS_14	CS_15	CS_16
Barcode_F9	CS_01	CS_02	CS_03	CS_04	CS_05	CS_06	CS_07	CS_08
Barcode_F8	WI_25	WI_26	WI_27	WI_28	WI_29	WI_30	WI_31	WI_32
Barcode_F7	WI_17	WI_18	WI_19	WI_20	WI_21	WI_22	WI_23	WI_24
Barcode_F6	WI_09	WI_10	WI_11	WI_12	WI_13	WI_14	WI_15	WI_16
Barcode_F5	WI_01	WI_02	WI_03	WI_04	WI_05	WI_06	WI_07	WI_08
Barcode_F4	CS_25	CS_26	CS_27	CS_28	CS_29	CS_30	CS_31	CS_32
Barcode_F3	CS_17	CS_18	CS_19	CS_20	CS_21	CS_22	CS_23	CS_24
Barcode_F2	CS_09	CS_10	CS_11	CS_12	CS_13	CS_14	CS_15	CS_16
Barcode_F1	CS_01	CS_02	CS_03	CS_04	CS_05	CS_06	CS_07	CS_08

3 PCR

6 PCR

18 PCR

5.2.5. Library preparation

Multiplex PCRs (cycling condition as in Vartia et al. 2014) were performed in 10 µl reactions (Size classes I and II) and 20 µl reactions (Size class III, in 3-PCR and 18-PCR size-class-combinations) with 1 µl template DNA, 1× Multiplex PCR Master Mix (Qiagen, Hilden, Germany), 2 pmol of each locus specific primer and 8 pmol of each barcoded universal primer per locus specific primer with the corresponding universal primer sequence. For each multiplex and PCR size-class-combination equal amounts of PCR product from each individual were pooled and separated by gel electrophoresis on a 2 % agarose gel at 5V/cm. PCR products were visualised using Safe Imager® (Invitrogen, Life Technologies, Carlsbad, CA, USA). Products corresponding to expected amplicon sizes were cut out from the gel and placed into 2 ml centrifuge tubes. DNA was extracted using the QIAquick® Gel Extraction Kit (Qiagen, Hilden, Germany). Amplicon DNA was quantified using Qubit® 2.0 Fluorometer with a dsDNA HS Assay kit (Life Technologies, Carlsbad, CA, USA). Amplicons from the 3-PCR and 18-PCR size-class-combinations were normalised to the optimal size-class amplicons-ratio (1:1:8). The amplicons were concentrated using the Amicon® Ultra 30K procedure (Merck Millipore Ltd., Merck KGaA, Darmstadt, Germany). An aliquot of the final library was diluted to a concentration of 500 ng in 16 µl for sequencing on a 454 Roche® platform (Margulies et al. 2005; 454 Life Sciences Corp., a Roche Company, Branford, CT, USA).

5.2.6. Sequencing

Roche adaptors were added to the amplicon library using the Rapid library preparation kit (Roche 2011). Amplicon concentrations were estimated using quantitative PCR to assess differential amplification success depending on amplicon size. Amplicons were sequenced on two ¼ regions of a PicoTiterPlate using a 454 GS-FLX Sequencer with titanium reagents at Teagasc Food Research Centre, Moorepark, Ireland. The Shotgun Data Processing Pipeline was used for signal processing to increase sequence yield (Roche 2012).

5.2.7. Genotyping

A python script was developed by a bioinformatician to process raw sequence data by identifying sequence reads containing the forward and reverse (combinatorial) barcodes and the locus specific primers and grouping them accordingly. The script was based on the Levenshtein distance metric which measures the distance between two sequences of characters. The error rate of Roche 454 GS-FLX amplicon sequences when including both sequencing and PCR errors has been estimated in *G. morhua* at 6% (Balzer et al. 2011). It was therefore necessary to allow for sequencing and PCR errors in the reads in order to avoid a significant loss of reads during the identification and grouping process. The python script allowed for up to 2 and 3 sequencing errors (both substitutions and indels) in the combinatorial barcodes and primers, respectively, to reflect a potential sequencing and PCR error rate of ~6%. All the scripts used are available on github (https://github.com/egenomics/micomba).

Grouped and classified sequences were imported into Geneious® 7 as fasta files and organised into folders per locus per individual. Loci were manually genotyped by viewing all of the reads of a particular individual at a specific locus, as a read-length histogram and verified by read alignment (Geneious Alignment - default settings) and manually edited (Kearse et al. 2012). Only individuals with five or more reads for a given locus were genotyped.

Read alignments of a subset of 16 individuals, with 10 or more sequence reads, from the Celtic Sea data set (6 PCR size-class-combinations) were screened for homoplasy. The aligned sequences were scrutinised for the presence of SNPs and indels that were not part of the microsatellite repeat structure and would not change the amplicon size. The SNP or indel had to occur in at least 20% of the reads to be considered as homoplasy.

5.2.8. Statistical analysis

Correspondence between GBS and ABI capillary-based genotyping data (see Vartia et al. 2014) was fit to a binomial model. The sufficient statistic for the log-likelihood of a binomial is given by:

Eq. 1: $LogL \propto a{*}Ln[p] + (n-a){*}Ln[1-p]$(Edwards 1992),

where ***p*** is the binomial rate parameter or proportional correspondence, ***a*** is the number of correct correspondences between 454 GBS and ABI capillary-based genotyping data, and ***n*** is the total number of potential correspondences. In the present study, ***n*** = 3196.

The impact of several factors on GBS-ABI correspondence was evaluated. These factors included microsatellite type (e.g., mono-, di-), PCR size-class-combination (3-, 6- or 18-PCR), and read depth used for genotyping (broken into 2 and 3 read depth categories). In all cases, the simplest model would be a "no effects" model, with a single binomial rate characterizing all of the data. More complex models would then propose different values of ***p*** based on different variable factors, for example a model could potentially estimate a separate value of ***p*** for each of the three PCR size-classes. More complex models will tend to have better fit to the data, as they can be "tuned" to explain specific variations in the proportional correspondence. However, they do this at the expense of increase in parameter estimation, which can lead to large variance in parameter estimates (Burnham and Anderson 2002, 2004). To select among models we employed the finite-sample Akaike Information Criterion (AICc) (Akaike 1973, 1974; Hurvich and Tsai 1989), which measures model fit to the observed data, penalized for increased model complexity;

Eq. 2: $AICc = -2LogL + 2K + \frac{2K(K+1)}{n-K-1}$ (Burnham and Anderson 2002).

In **equation 2**, ***LogL*** is the log-likelihood fit of the model to the data, ***K*** is the number of estimated parameters for the model, and ***n*** is the sample size. ***AICc*** is proportional to the parameter-corrected model likelihood by:

Eq. 3: $ModelL \propto e^{-1/2\Delta AICc}$(Burnham and Anderson 2002),

where ***ΔAICc*** is the difference in ***AICc*** score between any given model and the optimal model score. This can then be converted into a model posterior probability for the set of models under consideration by dividing each model's likelihood by the sum over all models' likelihoods.

Binomial proportions of the correspondence data were modelled for a restricted subset of the data, 16 individuals from the Celtic Sea ("Celtic Sea" data set), for which data existed across all PCR size-class-combinations. Differences in correspondence proportions among PCR size-class-combinations were evaluated using Bonferroni-corrected z-tests (Sokal & Rohlf 1995).

The effectiveness of PCR size-class-combinations on the total number of reads produced was evaluated for the Celtic Sea data. Total reads were modelled as a function of PCR size-class-combination, locus-specific primer, and universal primer type, in addition to potential intrinsic confounding variables, such as inter-locus or inter-individual variability. A multinomial probability likelihood model was employed:

Eq. 4: $LogL \propto \sum_1^R a_R * Ln(p_R)$ (Edwards 1992).

In **equation 4**, ***R*** is the number of rate classes in the proposed model structure, ***a*** is the observed number of reads in that class, and ***p*** = a/n, where n is total reads in the Celtic Sea data (n = 90194), that is, the observed proportion of reads in that class. Read data were modelled to compare the effectiveness of the different variables, as well as potential interactive effects between the variables, on total read count. As with the correspondence data, model fit was evaluated using ***AICc***.

Bonferroni-corrected Mann-Whitney tests (Sokal & Rohlf 1995) were performed on total, median and maximum number of reads per individual between the PCR size-class-combinations to assess significant differences in the performance of PCR size-class-combinations.

5.3. Results

Sequencing of the two ¼ regions of a PicoTiterPlate resulted in 228 246 reads for region 1 and 226 848 for region 2. Read length ranged from 54 bp to 1200 bp with an average read length of 275 bp (SD 84 bp). A total of 95.5% of reads had a quality score higher than Q20 and the average quality score was Q35.9. The proportions of A, C, G and T nucleotides were 27.0%, 22.7%, 26.7% and 23.6%, respectively, with a GC content of 49.4%.

A total of 180 054 reads were successfully assigned to specific individual and locus combinations (**Table 4**). The recovered reads per individual (combined across all loci) ranged from of 57 to 9192 (median: 999) (**Table 5**), and the number of reads per locus (combined over all individuals) ranged from 33 to 11 197 (median: 2294). The numbers of reads per genotype ranged from 0 to 1097 (median: 12). Analyses of PCR size-class-combinations resulted in 3325 successful genotypes from 5088 possible calls, with 3196 genotypes used for correspondence checks with the ABI – capillary genotyping data. Of the 53 loci analysed, 10 loci had low numbers of sequence reads preventing genotyping of more than 50% of the individuals (**Table 6**).

Table 4 Number of barcodes and locus-specific primers identified in the reads.

Identified primers and barcodes	No. of reads
Two barcodes and two matching primers	180054
Two barcodes and two non-matching primers	2175
Two barcodes and one primer	4843
Two barcodes and no primers	801
One barcode	205593
No barcodes	61628
Total	**455094**

Table 5 The barcode combinations used to identify individual amplicons. Reads recovered per individual after read sorting are shown under the individual code. The total number of recovered reads was 180 054. CS = Celtic Sea; WI = West of Ireland; 3-PCR is size-class combination 1; 6-PCR is size-class combination 2; and 18-PCR is size-class combination 3.

	Barcode _R1	Barcode _R2	Barcode _R3	Barcode _R4	Barcode _R5	Barcode _R6	Barcode _R7	Barcode _R8
Barcode _F12	CS_09 772	CS_10 4236	CS_11 1022	CS_12 4366	CS_13 415	CS_14 1116	CS_15 277	CS_16 4381
Barcode _F11	CS_01 499	CS_02 3424	CS_03 648	CS_04 2562	CS_05 328	CS_06 764	CS_07 216	CS_08 2993
Barcode _F10	CS_09 1695	CS_10 9192	CS_11 3010	CS_12 8988	CS_13 1261	CS_14 2690	CS_15 667	CS_16 8656
Barcode _F9	CS_01 511	CS_02 3627	CS_03 921	CS_04 3169	CS_05 462	CS_06 1038	CS_07 234	CS_08 3158
Barcode _F8	WI_25 136	WI_26 762	WI_27 285	WI_28 779	WI_29 90	WI_30 181	WI_31 57	WI_32 605
Barcode _F7	WI_17 475	WI_18 2637	WI_19 988	WI_20 2370	WI_21 371	WI_22 804	WI_23 179	WI_24 2181
Barcode _F6	WI_09 876	WI_10 5108	WI_11 1907	WI_12 4970	WI_13 641	WI_14 1361	WI_15 302	WI_16 3539
Barcode _F5	WI_01 900	WI_02 6054	WI_03 1978	WI_04 5518	WI_05 704	WI_06 1559	WI_07 339	WI_08 4204
Barcode _F4	CS_25 826	CS_26 4996	CS_27 1646	CS_28 4719	CS_29 605	CS_30 1332	CS_31 284	CS_32 3626
Barcode _F3	CS_17 891	CS_18 5163	CS_19 1724	CS_20 5028	CS_21 622	CS_22 1515	CS_23 283	CS_24 3740
Barcode _F2	CS_09 279	CS_10 1478	CS_11 472	CS_12 1476	CS_13 184	CS_14 367	CS_15 92	CS_16 1009
Barcode _F1	CS_01 384	CS_02 2179	CS_03 716	CS_04 1996	CS_05 195	CS_06 547	CS_07 114	CS_08 1408

Table 6 Loci with less than 50% of individuals genotyped.

Loci name	Percentage of individuals genotyped
A03_Hill	1.0%
C35_M13	7.3%
D10_Neo	14.6%
A30_T3	25.0%
B19_T3	39.6%
B12_T3	41.7%
D37_T3	41.7%
A19_T3	42.7%
C31_Neo	45.8%
B36_Hill	46.9%

In total, 529 genotypes were screened for homoplasy, which was detected in 32% of the genotypes and was present in 38 loci. SNPs represented 80% of the homoplasy with 20% represented by indels.

Correspondence of the GBS genotype calls and ABI genotype calls was positively correlated with read depth (**Figure 2**). AICc model selection for the Celtic Sea data demonstrated that PCR size-class-combination was the only important variable for predicting read yield. There was no support for other modelled variables (e.g., forward or reverse primers) or confounding variables (such as inter-individual and among-locus variation) (**Table 7**). Correspondence between the full GBS and ABI data sets was best explained by a model incorporating read depth and microsatellite motif type (**Table 8**). PCR size-class-combination had an effect on the GBS-ABI correspondence in the Celtic Sea data (**Table 9**), with significantly higher correspondence for both 3-PCR and 6-PCR when either is compared to 18-PCR. However, there was no significant difference between 3- and 6-PCR (**Table 9**). In addition, significantly more reads per individual were observed for 3- and 6-PCR when compared to 18-PCR, although again, there was no significant difference between the two (**Table 10**). No significant differences were noted for median number of reads among size-class-combinations (**Table 11**). Maximum number of reads per individual showed a significant difference between 6-PCR and 18-PCR only (**Table 12**).

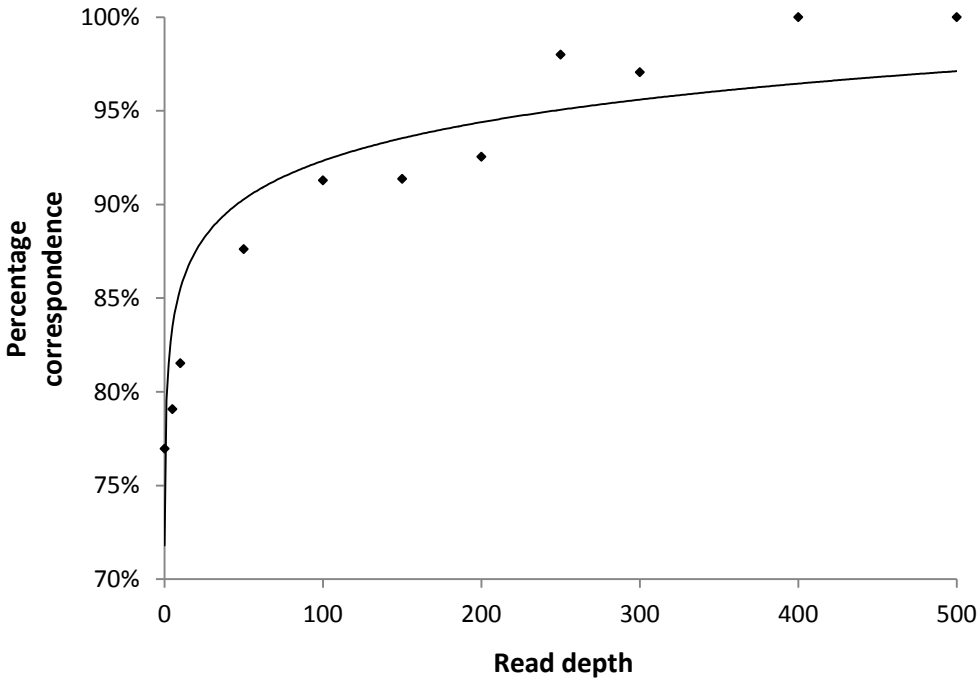

Figure 2 Correspondence of the GBS microsatellite data with ABI data for the full data set. Y-axis represents the percentage of corresponding genotype calls of all genotype calls. The X-axis represents the increasing minimum threshold of read depth required for making a genotype call. The applied thresholds were 5, 10, 50, 100, 150, 200, 250, 300, 400 and 500 reads.

Table 7 AICc model selection for the reduced ("Celtic Sea") data set on the read yield. Sample size is 90194.

Model Description	K	LogL	AICc	dAICc	Post. Prob.
PCR	3	-87,627	175,260	0	1.000
Tails	4	-123,574	247,157	71,896	0
Forward	6	-143,882	287,776	112,516	0
Reverse	8	-157,688	315,391	140,131	0
PCR * Tails	12	-210,436	420,896	245,635	0
Individuals	16	-216,345	432,722	257,461	0
Forward * Tails	24	-266,358	532,764	357,504	0
PCR * Reverse	24	-245,065	490,177	314,917	0
Reverse * Tails	32	-281,170	562,403	387,143	0
Locus	53	-318,574	637,253	461,993	0
PCR * Reverse * Tails	96	-367,637	735,467	560,206	0

K is the number of parameters estimated for a given model structure, LogL is the log-likelihood of the model, AICc is the finite sample AIC score for the model, dAICc is the difference in AICc score between the given model and the optimal model score, and Post. Prob is the model posterior probability.

Table 8 Model selection using AICc on the correspondence between 454 microsatellites and ABI microsatellites, using the full data set.

Model	Model Comments	K	LogL	AICc	dAICc	Post. Prob.
No Effects	Average over all data	1	-1724.64	3451.28	131.96	2.22 E -29
# Reads (2)	0-5 vs. 5+ reads	2	-1693.93	3391.87	72.55	1.77 E -16
# Reads (3)	0-5 vs. 5-10 vs. 10+ reads	3	-1676.04	3358.10	38.78	3.80 E -09
MST Type		5	-1693.54	3397.11	77.79	1.29 E -17
Reads (2) by MST Type	0-5 vs. 5+ reads	10	-1663.76	3347.59	28.27	7.26 E -07
Reads (3) by MST Type	0-5 vs. 5-10 vs. 10+ reads	15	-1644.58	3319.32	0.00	0.999
PCR (3,6) v. 18		2	-1706.09	3416.18	96.86	9.27 E -22
PCR 3,6,18		3	-1706.08	3418.16	98.84	3.45 E -22
Reads (2) by PCR 3,6,18	0-5 vs. 5+ reads	6	-1675.28	3362.58	43.26	4.03 E -10
Reads (3) by PCR 3,6,18	0-5 vs. 5-10 vs. 10+ reads	9	-1658.37	3334.79	15.47	0.0004

K is the number of parameters estimated for a given model structure, LogL is the log-likelihood of the model, AICc is the finite sample AIC score for the model, dAICc is the difference in AICc score between the given model and the optimal model score, and Post. Prob is the model posterior probability.

Table 9 Bonferroni corrected 2-tail p-values for z-tests for the proportion of correspondence between 454 and ABI performed on the reduced ("Celtic Sea") data set.

	3 PCR	6 PCR	18 PCR	Proportion
3 PCR				0.8298
6 PCR	1.0000			0.8333
18 PCR	0.0006	0.0005		0.7512

Table 10 Bonferroni corrected 2-tail p-values for Mann-Whitney tests performed on the reduced ("Celtic Sea") data set for differences in the total number of reads per individual by PCR condition type.

	18 PCR	6 PCR	3 PCR	Average total reads per individual
18 PCR				806
6 PCR	0.0002			3080
3 PCR	0.0406	0.0907		1751

Table 11 Bonferroni corrected 2-tail p-values for Mann-Whitney tests performed on the reduced ("Celtic Sea") data set for differences in the median number of reads per individual by PCR condition type.

	18 PCR	6 PCR	3 PCR	Average median reads per individual
18 PCR				11
6 PCR	1.0000			20
3 PCR	0.5590	1.0000		19

Table 12 Bonferroni corrected 2-tail p-values for Mann-Whitney tests performed on the reduced ("Celtic Sea") data set for differences in the maximum number of reads per individual by PCR condition type.

	18 PCR	6 PCR	3 PCR	Average maximum reads per individual
18 PCR				50
6 PCR	0.0168			402
3 PCR	0.2390	0.6619		165

5.4. Discussion

This study demonstrates the potential for NGS-based GBS as a method for microsatellite genotyping using *de novo* and existing capillary/gel electrophoresis-based multiplex marker panels. It also illustrates the potential for a rapid and cost effective method for microsatellite GBS, that can take advantage of modern NGS

platforms, using combinatorial barcoding for implementation in large-scale population genetics studies. This method provides access to the underlying sequence data, providing an additional advantage over traditional fragment length genotyping by resolving issues of size homoplasy and revealing potentially hidden genetic variation in the amplicons.

Analysis of the Celtic Sea data indicates that size-class-combination is the controlling factor for both sequence yield and GBS-ABI correspondence. Both the 3- and 6-PCR size class conditions outperformed 18-PCR, with no significant differences between the 3- and the 6-PCR conditions observed. It is possible that 6-PCR would not have performed as well as 3-PCR in the absence of pre-optimisation (see Vartia et al. 2014). However, the increased number of PCR reactions and DNA quantification steps in the 18-PCR method may have introduced quantification errors that were amplified at the pooling stage, lowering read yield, and hence correspondence. As such, it is proposed that future studies minimise the number of steps to reduce variability among PCRs. Binladen et al. (2007) reported biases in sequence recovery with 454 NGS as a result of different base pair composition of barcodes. In contrast to this, our analyses indicated no bias in sequence recovery.

Read depth had a significant effect on the genotyping correspondence between the GBS and ABI based datasets, with correspondence rapidly increasing with sequence depth (**Figure 2**). It is unlikely that complete agreement between these two approaches is achievable, due to inherent errors in capillary electrophoresis genotype calls, and GBS platform-specific sequencing errors. ABI based genotype calling from electropherograms can be obscured by spectral bleeding, cross-talk between capillaries and fluctuations in instrument parameters (Van Neste et al. 2012). Similarly, for 454 sequence amplicon reads, error rates have been reported as high as 6% for our study species, *G. morhua* (Balzer et al. 2011). While the error rate for 454 sequencing can be relatively high, amplicon sequencing, as used here, will be further affected by PCR induced errors. It is therefore important to allow for some errors in barcodes and primers. When employing other NGS platforms this requirement may be alleviated by the increased read yield of these platforms.

This study represents an example of GBS utilising established microsatellite multiplex panels that were developed for analysis on conventional capillary/gel based systems and contain loci with fragment sizes commonly reported in microsatellite studies (~ 100 - 500 bp) (Skirnisdottir et al. 2008; Higgins et al. 2009; Fordyce et al. 2011; Vartia et al. 2014). Due to the large size range of alleles in the established microsatellite multiplex panel, the choice of NGS platform was restricted to 454, as it produces up to 1 million reads of up to 700 bp (http://454.com/products/gs-flx-system/index.asp, 2015). While the 454 NGS platform has the capacity to produce long sequence reads, the sequence yield is significantly lower than for other platforms: for example, the Ion Torrent platform (Rothberg et al. 2011; Life Technologies, Carlsbad, CA, USA) produces up to 6 million 400 bp reads on their 318® chip (Life Technologies 2013) and the Illumina® Miseq platform (Bentley et al. 2008; Illumina Inc. San Diego, CA, USA) currently produces 20-30 million paired end reads of 2 x 300 bp. The short read length of these platforms limits application to relatively short amplicons. However, this limitation may be mitigated by *de novo* development of microsatellite markers, the use of partial existing microsatellite panels (i.e. loci with amplicon sizes within the read length limitation of the chosen NGS platform) or the redesign of primers for existing microsatellite markers to produce shorter amplicons.

The current study was not optimised for population genetics scale genotyping as a larger number of loci and a lower number of individuals were analysed than is often used in population genetic studies (*cf.* Vartia et al. 2014). However, in an effort to address multiple questions about the feasibility of GBS of microsatellites, it was judged advantageous to include a larger number of loci than needed for many population genetic studies. In addition, if this method was used on *de novo* developed loci, the success rate from microsatellite containing sequences to genotyped loci is only known *post hoc* and hence a larger number of initial loci would increase the likelihood of generating a sufficient number of informative loci. For deployment of the approach described here, in a population genetics setting, it may be more beneficial to interrogate fewer loci, but more individuals. This could be facilitated by increasing the number of forward and reverse combinatorial barcodes used to tag an increased number of individuals.

Manual scoring of conventional capillary and gel-based electrophoresis fragment length polymorphism is time consuming, and therefore carries a significant financial cost to genetic studies. This study describes an approach that does not require manual correction of internal size standards or genotype calls due to spectral bleed-through, thus reducing the genotyping time. In addition, the availability of the underlying sequence data lends itself to the development of automated genotyping (see Suez et al. 2015). The numbers of loci that can be multiplexed in capillary or gel based electrophoresis in conventional studies are limited by the availability of fluorescent labels/detection channels and fragment size overlaps (rarely more than 12 loci per multiplex) (Guichoux et al. 2011). The GBS approach described in this study using PCR-incorporated combinatorial barcoding has no limitations on the number of markers (other than the number of sequences produced by the chosen NGS platform), as size overlap does not affect sequence yield.

The high proportion of homoplasy (32%) observed is in concordance with other studies addressing the prevalence of homoplasy (Garza and Freimer 1996; Angers et al. 2000; Anmarkrud et al. 2008; Barkley et al. 2009). This homoplasy would have been undetectable using traditional ABI based microsatellite genotyping as the amplicon size would be unaffected. Previous detection methods include single-strand conformation polymorphism analysis and direct sequencing, with or without cloning, however, these detection methods can be laborious (Garza and Freimer 1996; Angers et al. 2000; Anmarkrud et al. 2008; Barkley et al. 2009). It should be noted that low read depth may have prevented distinguishing homoplasy from true mutations and sequencing/PCR error. However, the purpose of the analysis in the current study was to explore the capacity to detect and quantify prevalence of homoplasy using GBS based methods. The prevalence of homoplasy implies that allelic diversity in fragment-size-based studies is likely substantially underestimated. Consequently, inferred population structures in these studies may also underestimate true levels of genetic variability. Aside from homoplasy, determining microsatellite repeat numbers could improve genetic diversity comparisons (Petit et al 2005). Essentially, accessing the molecular structure of the microsatellite markers will increase our understanding of the mutation model of the studied loci and thus improve the quality of the information retrieved from the data. In addition, access

to actual sequence lengths and sequence information will greatly facilitate inter-laboratory calibration and data storage in repositories as sequence files offering a significant advantage.

This study presents a novel method for microsatellite GBS using individual combinatorial barcoding that can be faster and cheaper than current approaches while offering better and more data.

5.5. References

454.com. 2015. Products—GS FLX+ System. 454 Life Sciences, a Roche Company. Accessed 4 June 2015. Life Sciences Corp., a Roche Company, Branford, CT, USA. http://454.com/products/gs-flx-system/index.asp.

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267–281 *in* B. N. Petrov and F. Csaki, editors. Second International Symposium on Information Theory. Akademiai Kiado, Budapest.

Akaike, H. 1974. A new look at the statistical model identification. Institute of Electrical and Electronics Engineers Transactions on Automatic Control AC 19:716–723.

Anderson, E. C., and Garza, J. C. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567–2582.

Angers, B., Estoup, A., and Jarne, P. 2000. Microsatellite size homoplasy, SSCP, and population structure: a case study in the freshwater snail *Bulinus truncatus*. Molecular biology and evolution 17:1926–1932.

Anmarkrud, J. A., Kleven, O., Bachmann, L., and Lifjeld, J. T. 2008. Microsatellite evolution: Mutations, sequence variation, and homoplasy in the hypervariable avian microsatellite locus HrU10. BMC evolutionary biology 8:138.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:1–7.

Balzer, S., Malde, K., and Jonassen, I. 2011. Systematic exploration of error sources in pyrosequencing flowgram data. Bioinformatics 27:304–309.

Barkley, N. A., Krueger, R. R., Federici, C. T., and Roose, M. L. 2009. What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. Plant Systematics and Evolution 282:71–86.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Cheetham, K. R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M. J., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V, Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M. D., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Catenazzi, M. C. E., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D.,

Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Furey, W. S., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A. M., Hoschler, K., Hurwitz, S., Ivanov, D. V, Johnson, M. Q., James, T., Huw Jones, T. A., Kang, G.-D., Kerelska, T. H., Kersey, A. D., Khrebtukova, I., Kindwall, A. P., Kingsbury, Z., Kokko-Gonzales, P. I., Kumar, A., Laurent, M. A., Lawley, C. T., Lee, S. E., Lee, X., Liao, A. K., Loch, J. A., Lok, M., Luo, S., Mammen, R. M., Martin, J. W., McCauley, P. G., McNitt, P., Mehta, P., Moon, K. W., Mullens, J. W., Newington, T., Ning, Z., Ling Ng, B., Novo, S. M., O'Neill, M. J., Osborne, M. A., Osnowski, A., Ostadan, O., Paraschos, L. L., Pickering, L., Pike, A. C., Pike, A. C., Chris Pinkard, D., Pliskin, D. P., Podhasky, J., Quijano, V. J., Raczy, C., Rae, V. H., Rawlings, S. R., Chiva Rodriguez, A., Roe, P. M., Rogers, J., Rogert Bacigalupo, M. C., Romanov, N., Romieu, A., Roth, R. K., Rourke, N. J., Ruediger, S. T., Rusman, E., Sanches-Kuiper, R. M., Schenker, M. R., Seoane, J. M., Shaw, R. J., Shiver, M. K., Short, S. W., Sizto, N. L., Sluis, J. P., Smith, M. A., Ernest Sohna Sohna, J., Spence, E. J., Stevens, K., Sutton, N., Szajkowski, L., Tregidgo, C. L., Turcatti, G., Vandevondele, S., Verhovsky, Y., Virk, S. M., Wakelin, S., Walcott, G. C., Wang, J., Worsley, G. J., Yan, J., Yau, L., Zuerlein, M., Rogers, J., Mullikin, J. C., Hurles, M. E., McCooke, N. J., West, J. S., Oaks, F. L., Lundberg, P. L., Klenerman, D., Durbin, R., and Smith, A. J. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59.

Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., and Willerslev, E. 2007. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS ONE 2:1–9.

Burnham, K. P., and Anderson, D. R. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Burnham, K. P., and Anderson, D. R. 2004. Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods and Research 33:261–304.

Chowdhury, J., Kagiala, G. V, Pushpakom, S., Lauzon, J., Makin, A., Atrazhev, A., Stickel, A., Newman, W. G., Backhouse, C. J., and Pilarski, L. M. 2007. Microfluidic platform for single nucleotide polymorphism genotyping of the thiopurine S-methyltransferase gene to evaluate risk for adverse drug events. The Journal of Molecular Diagnostics 9:521–529.

Clayton, T. M., Whitaker, J. P., Sparkes, R., and Gill, P. 1998. Analysis and interpretation of mixed forensic stains using DNA STR profiling. Forensic Science International 91:55–70.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews. Genetics 12:499–510.

Delmotte, F., Leterme, N., and Simon, J. 2001. Microsatellite allele sizing: difference between automated capillary electrophoresis and manual technique. BioTechniques 31:810, 814–6, 818.

Edwards, A. W. F. 1992. Likelihood: Expanded edition. The Johns Hopkins University Press, Baltimore.

Estoup, A., Jarne, P., and Cornuet, J. M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology 11:1591–1604.

Fordyce, S. L., Ávila-Arcos, M. C., Rockenbauer, E., Børsting, C., Frank-Hansen, R., Petersen, F. T., Willerslev, E., Hansen, A. J., Morling, N., and Gilbert, T. P. 2011. High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform. BioTechniques 51:127–133.

Garza, J. C., and Freimer, N. B. 1996. Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Research 6:211–217.

Gill, P. 2001. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. International Journal of Legal Medicine 114:204–210.

Glaubitz, J. C., Rhodes, O. E., and Dewoody, J. A. 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Molecular Ecology 12:1039–1047.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., and Petit, R. J. 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources 11:591–611.

Ha, N.-T., Freytag, S., and Bickeboeller, H. 2014. Coverage and efficiency in current SNP chips. European Journal of Human Genetics 22:1124–1130.

Haasl, R. J., and Payseur, B. A. 2011. Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. Heredity 106:158–171.

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., Cariani, A., Maes, G. E., Diopere, E., Carvalho, G. R., and Nielsen, E. E. 2011. Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. Molecular Ecology Resources 11:123–136.

Higgins, B., Hubert, S., Simpson, G., Stone, C., and Bowman, S. 2009. Characterization of 155 EST-derived microsatellites from Atlantic cod (*Gadus morhua*) and validation for linkage mapping. Molecular Ecology Resources 9:733–737.

Hofreuter, D., Tsai, J., Watson, R. O., Novik, V., Altman, B., Benitez, M., Clark, C., Perbost, C., Jarvie, T., Du, L., and Galán, J. E. 2006. Unique features of a highly pathogenic *Campylobacter jejuni* strain. Infection and Immunity 74:4694–4707.

Hurvich, C. M., and Tsai, C.-L. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.

Kan, C. W., Fredlake, C. P., Doherty, E. A. S., and Barron, A. A. 2004. DNA sequencing genotyping in miniaturized electrophoresis systems. Electrophoresis 25:3564–3588.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kuhner, M. K., Beerli, P., Yamato, J., and Felsenstein, J. 2000. Usefulness of single nucleotide polymorphism data for estimating population parameters. Genetics 156:439–447.

LaFramboise, T. 2009. Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. Nucleic Acids Research 37:4181–4193.

Mardis, E. R. 2013. Next-generation sequencing platforms. Annual Review of Analytical Chemistry 6:287–303.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V, Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F., and Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–80.

Mirimin, L., O'Keeffe, D., Ruggiero, A., Bolton-Warberg, M., Vartia, S., and Fitzgerald, R. 2011. A quick, least-invasive, inexpensive and reliable method for sampling *Gadus morhua* postlarvae for genetic analysis. Journal of Fish Biology 79:801–805.

Morin, P. A., Luikart, G., and Wayne, R. K. 2004. SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution 19:208–216.

Morin, P. A., Martien, K. K., and Taylor, B. L. 2009. Assessing statistical power of SNPs for population structure and conservation studies. Molecular Ecology Resources 9:66–73.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., and Hohenlohe, P. A. 2013. Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology 22:2841–2847.

Van Neste, C., Van Nieuwerburgh, F., Van Hoofstat, D., and Deforce, D. 2012. Forensic STR analysis using massive parallel sequencing. Forensic Science International: Genetics 6:810–818.

Nielsen, R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. Genetics 154:931–942.

Oh, J. D., Kling-Bäckhed, H., Giannakis, M., Xu, J., Fulton, R. S., Fulton, L. A., Cordum, H. S., Wang, C., Elliott, G., Edwards, J., Mardis, E. R., Engstrand, L. G., and Gordon, J. I. 2006. The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. Proceedings of the National Academy of Sciences of the United States of America 103:9999–10004.

Pasqualotto, A. C., Denning, D. W., and Anderson, M. J. 2007. A cautionary tale: Lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. Journal of Clinical Microbiology 45:522–528.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7: e37135.

Petit, R. J., Deguilloux, M. F., Chat, J., Grivet, D., Garnier-Géré, P., and Vendramini, G. G. 2005. Standardizing for microsatellite length in comparisons of genetic diversity. Molecular Ecology 14:885–890.

Roche. 2009. Using Multiplex Identifier (MID) adaptors for the GS FLX titanium chemistry - Extended MID Set. Technical Bulletin No. 005-2009. 454 Life Sciences Corp. A Roche Company, Branford. CT, USA.

Roche. 2011. Rapid library preparation method manual – GS FLX+ Series — XL+. May 2011 454. Life Sciences Corp. A Roche Company, Branford. CT, USA.

Roche. 2012. 454 sequencing system guidelines for amplicon experimental design. March 2012. 454 Life Sciences Corp. A Roche Company, Branford. CT, USA.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T., and Bustillo, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. Nature 475:348–352.

Scheible, M., Loreille, O., Just, R., and Irwin, J. 2011. Short tandem repeat sequencing on the 454 platform. Forensic Science International: Genetics Supplement Series 3:e357–e358.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments – A poor man's approach to genotyping for research and high-throughput diagnostics. Nature Biotechnology 18:233–234.

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., and Seeb, L. W. 2011. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Molecular Ecology Resources 11:1–8.

Skirnisdottir, S., Pampoulie, C., Hauksdottir, S., Schulte, I., Olafsson, K., Hreggvidsson, G. O., and Hjorleifsdottir, S. 2008. Characterization of 18 new microsatellite loci in Atlantic cod (*Gadus morhua* L.). Molecular Ecology Resources 8:1503–5.

Sokal, R. R., and Rohlf, F. J. 1995. Biometry. Third edition. W.H. Freeman, New York.

Suez, M., Behdenna, A., Brouillet, S., Graça, P., Higuet, D., and Achaz, G. 2015. MicNeSs: Genotyping microsatellite loci from a collection of (NGS) reads. Molecular Ecology Resources 1:1–10.

Life Technologies. 2013. The Ion PGM™ System, with 400-base read length chemistry, enables routine high-quality de novo assembly of small genomes. Application note. Life Technologies Corporation.

Vartia, S., Collins, P. C., Cross, T. F., Fitzgerald, R. D., Gauthier, D. T., McGinnity, P., Mirimin, L., and Carlsson, J. 2014. Multiplexing with three-primer PCR for rapid and economical microsatellite validation. Hereditas 151:43–54.

Wang, S., Meyer, E., McKay, J. K., and Matz, M. V. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods 9:808–810.

Wattier, R., Engel, C. R., Saumitou-Laprade, P., and Valero, M. 1998. Short allele dominance as a source of heterozygote deficiency at microsatellite loci: Experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). Molecular Ecology 7:1569–1573.

Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.-J., Makhijani, V., Roth, G. T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C. L., Irzyk, G. P., Lupski, J. R., Chinault, C., Song, X., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D. M., Margulies, M., Weinstock, G. M., Gibbs, R. A., and Rothberg, J. M. 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872–876.

Chapter 6. Concluding synthesis

The chapters of this PhD thesis fall under three themes: 1) examination of genetic changes taking place in a newly established farmed population of cod broodstock and breeding programme; 2) study of local wild cod in waters around Ireland; and 3) developing new genetic markers and genotyping methodologies.

One of the core deliverables of the PhD was to use molecular approaches in broodstock and farmed stock management to help maintain a genetically diverse and viable stock. The establishment of the broodstock and breeding programme, in turn, prompted the study of local wild stocks for several reasons. Firstly, the wild broodstock for the breeding programme was sourced from local wild cod from the Celtic Sea, and therefore it was of importance to assess the genetic status of the sampled population. Secondly, fish escape events from cages are often an unavoidable side effect of aquaculture operations, and these escapes can present serious threats to local wild populations. In addition to an immediate threat of spreading diseases, 'genetic pollution', in terms of loss of local adaptation due to introgression, can have irreversible consequences. The magnitude of this threat depends, among other things, on how genetically differentiated the escaped fish are from their wild counterparts. Therefore, resolving fine-scale population structure in the Irish waters could help mitigate possible genetic threats to local wild stocks if aquaculture became more wide-spread along the Irish coastline. Despite a large body of work on cod population structure across its range (e.g. Hutchinson et al. 2001; O'Leary et al. 2007; Pampoulie et al. 2008; Nielsen et al. 2009; Bradbury et al. 2010), previous research on cod population structure in Irish waters had been inconclusive due to low number of markers used (i.e. Hutchinson et al. 2001; Pampoulie et al. 2008). To overcome this challenge, new microsatellite markers were developed.

Applying genetic tools in a breeding programme can offer significant advantages in management of the farmed stock, but it can also be relatively expensive. This is one of the reasons that aquaculture broodstocks across species tend not to be well characterized genetically (Liu & Cordes 2004). Next generation sequencing is offering new possibilities for generating and genotyping of genetic markers, and several protocols for SNP genotyping-by-sequencing have been published (Baird et

al. 2008; Peterson et al. 2012; Wang et al. 2012). Even though SNPs have advantages, in genome coverage for example, microsatellites can still offer better performance in some cases of parentage analysis. To explore more affordable and efficient microsatellite genotyping, a new next generation sequencing based microsatellite genotyping-by-sequencing approach was developed and tested.

Cod is one of many new aquaculture species subjected to domestication efforts. Considerable efforts to domesticate several marine species have taken place in the past 50 years in order to address a growing demand for seafood products by the means of aquaculture production. Currently, as much as 15% of all animal protein in the human diet consists of wild marine fish (Allendorf et al. 2014). The human population is still growing, increasing the demand for seafood products. At the same time, 28.8% of all commercially important marine fish stocks are overexploited and 61.3% fully exploited (FAO 2014). Diverse array of marine taxa and diversity within taxa make aquatic species very promising for domestication. So far 0.13% of known marine animal species have been domesticated, compared to only 0.0002% of known land animals (Duarte et al. 2007).

Especially in recent years, the rapid spread of technical and scientific knowledge, such as use of recirculation systems in rearing, improved disease management and accumulation of biological knowledge of the farmed species, has facilitated domestication of aquatic species on a global scale (Subasinghe et al. 2003; Duarte et al. 2007). However, concerns have also been raised over potential harmful effects of large scale farming to local wild stocks such as disease transmission, deterioration of coastal ecosystems due to farming waste, overexploitation of certain species to produce fish meal, and loss of genetic diversity and local adaptations. Genetic tools have been instrumental in the study of wild fish populations, revealing significant population structuring in cod and other marine species that were previously considered large panmictic populations spanning the whole species range. These same genetic tools have contributed to the development of aquaculture practises in various species and several breeding programmes carry out parentage analysis as a routine task when monitoring their farmed stock. Despite significant advances in some species, for example Atlantic

salmon (*Salmo salar*), many newly domesticated species are severely lacking in genetic/genomic resources preventing genetic stock management.

6.1. Farmed population

The first task in the genetic characterisation of the farmed stock was to study the source population where ova were sourced from wild broodstock fish. Genetic characterisation of the breeding programme's wild broodstock from 2008 to 2012 indicated that the sampled Celtic Sea source population was temporally stable supporting the findings of Pampoulie et al. (2008). Effective population size (N_e) is a measure of the population's genetic diversity and can be more informative than the census size (N) when assessing the viability of a population. N_e estimate of the pooled sample over the five sampled years was inconclusive (infinite), yet the lower end of the confidence interval of nearly 4000, indicated that the Celtic Sea population had large enough genetic base to retain its adaptive potential in the face of environmental changes (Frankham et al. 2014). These results confirm the suitability of the chosen sampling location for sourcing ova into the breeding programme.

Another important question was how well the initially captured genetic variation was retained from fertilisation to adulthood in the breeding programme, and if the retention of genetic variation was affected by husbandry practises. To this end, family survival was studied in three separate year classes. Indeed, parentage analysis revealed drastic differences in survival between family crosses created by strip spawning from wild broodstock. The number of families was severely reduced from the time of fertilization to the time of sampling in all three sampled year classes. Similar trend of unequal reproductive capacities was observed in a mass spawning event with farmed (F1) broodstock. Unequal contributions from parents were further accentuated by differences in survival of offspring during the on-rearing. As a result, the N_e of each year class was much lower than the available number of broodstock fish. If unequal parental contributions and differences in survival of families are not taken into account when estimating N_e in breeding programmes, discrepancies between expected values and realised values can be

significant (24 *vs* 126, for combined N_e over five year classes, 2008-2012, in this study). Similar results have been recorded in other breeding programmes, for example in Atlantic halibut (*Hippoglossus hippoglossus*) (Jackson et al. 2003).

The results have serious consequences for management of breeding programmes of not only cod but other species with similar life histories and breeding strategies as well. It is evident that the commonly employed rule of thumb of having 100 breeders or 50 breeding pairs as base population is not an adequate number to protect the farmed population from inbreeding, and guarantee retaining sufficient genetic variation, as the actual N_e is likely to be much lower than 100. These findings prompted the recommendation of continued sourcing of eggs from wild broodstock until a higher N_e has been achieved. Separate rearing of families was one of the potential solutions highlighted to ameliorate the highly variable and low survival of families as it could reduce competition and cannibalism between individuals and families. It is evident that the retention of genetic variation needs to be consistently monitored to prevent bottlenecks and fast accumulation of inbreeding.

6.2. Local wild cod

The second major question of the thesis was to establish the population structure of cod in Irish waters. Aquaculture often involves translocation of fish between areas, and escapes from captivity need to be anticipated. The risks posed to the local wild stocks by escape events depend on the existence of fine scale population structure and the state of the local wild stocks among other things. In addition, existence of fine scale population structure would have implications for fisheries management in the area.

Population structure analysis showed that the cod in the Irish waters likely form one interbreeding unit that is distinct from the western Atlantic stocks which were represented by the Scotian Shelf sample. The results were in concordance with a recent SNP study that identified three distinct reproductively isolated population units of cod in the northwest European shelf, with the western Channel, Celtic Sea,

Irish Sea and Firth of Clyde forming one population unit (Heath et al. 2014). Indeed, agreement can also be recognised with Dutil and Brander (2003) who found that Celtic Sea, Irish Sea and West of Scotland stocks clustered together showing the highest levels of productivity over 15 cod stocks across the Atlantic, as well as experiencing similar temporal trends in production. Based on the present results, the offspring from the EIRCOD programme would not pose an additional threat of genetic pollution to local wild fish in case of escape events around the Irish coast.

This situation may, however, change over time as several generations in captivity are likely to affect the genetic composition of the farmed stock due to bottlenecks, founder effect and drift, as well as, selection. Therefore, the baseline of genetic variation in the wild population recorded in this study will provide a reference point that future farmed stocks can be compared to. If cod farming were to become more widely established in Ireland in the future, one important concern would be the lack of regulation in sourcing eggs from foreign producers which would expose the local stocks to introgression with foreign stocks if escapes occurred. However, the genetic tools generated in this project will be valuable means for addressing and anticipating such issues.

Even though the local wild cod appeared to be one homogenous interbreeding unit, there was some evidence of more complicated patterns existing in parallel, such as transient migratory fish. For example, evidence of another cod population contributing to the Celtic Sea 2011 sample was detected. There was some indication of weak isolation of the Irish Sea by pairwise $F_{ST,}$ measure of population differentiation, in a few analyses; however, this was not confirmed by another population differentiation measure, the clustering method. The biological significance of the potential weak structuring will remain to be evaluated by other methods, such as common garden experiments, or use of selected markers that can detect differences in selected genes that could indicate local adaptation not detectable with neutral genetic markers.

Cod aquaculture peaked in the 2000s, and has since declined globally. Despite significant research efforts, several bottlenecks still persist in modern cod

aquaculture before it can be economically profitable. In an Irish context, the current cod aquaculture effort has produced valuable resources that can be utilised when the financial environment changes in a way that enables commercialising cod aquaculture in Ireland.

6.3. Marker validation

The third aim of the thesis was technology development for more efficient and affordable deployment of genetic markers with special emphasis on microsatellite markers. Firstly, it was deemed necessary to develop new microsatellite markers for resolving fine-scale population structure in Irish waters and as a future resource for cod research. A large proportion of existing cod microsatellite markers are EST-linked, and may thus be affected by selective forces, (i.e. Higgins et al. 2009) and are therefore not optimal for resolving population structure. In addition, previous population structure investigations between the Irish and Celtic Sea had yielded conflicting results due to low number of employed loci (Hutchinson et al. 2001; Pampoulie et al. 2008).

In Chapter 3, 55 new microsatellite markers were developed for cod. The new markers were developed from Celtic Sea cod specifically to take advantage of the potential ascertainment bias to insure that the developed markers would be variable in the farmed population that originated from the Celtic Sea. The commonly used validation method employing three-primer-PCR (Schuelke 2000; Guichoux et al. 2011) was further optimised by combining the three-primer PCR with multiplexing to allow for more economical and rapid development of microsatellite markers discovered from high throughput sequencing data, as well as deployment of these markers.

The use of 55 microsatellites, in six multiplex panels, allowed for determination of pairwise F_{ST} value between two populations with high precision (Chapter 3) and for confirmation that cod in Irish waters form a single interbreeding population unit (Chapter 5). Using larger numbers of genetic markers confers considerable advantages of increased precision and statistical power when assessing intra- and

inter-population genetic parameters such as population structure and gene flow, as well as when inferring demographic parameters, such as effective population size, population expansions and bottlenecks (Nei and Tajima 1981; Ryman et al. 2006). This improved precision allows for more robust and trustworthy management advice based on genetic data.

Using such a high number of loci offered the possibility of further investigation of how the number of loci affects the estimation of genetic parameters. Specifically, it was shown that the variability of a population differentiation measure, the pairwise multilocus F_{ST} value, decreased as more loci were added, however, the rate of reduction decreased after 20 – 30 loci suggesting that this was the point where using more loci only slightly improved the precision of the estimate. This point of diminishing returns is likely to be dependent on populations, geographic scales and organisms under study. The biggest advantage of this approach is that it can be used to validate markers on any species, including those for which extensive sequence resources are not available, and will allow for large and robust population genetic studies while reducing cost and data collection time.

6.4. Genotyping-By-Sequencing

Another topic explored in the thesis was the application of the microsatellite genotyping-by-sequencing (GBS) approach (Vartia et al. 2016). Microsatellite markers are popular in aquaculture and fisheries science, and there are several established microsatellite panels for various species (e.g. haddock (*Melanogrammus aeglefinus*), Lage and Kornfield 1999; salmon (*Salmo salar*), Norris et al. 1999; rainbow trout (*Oncorhynchus mykiss*), Fishback et al. 1999; halibut (*Hippoglossus hippoglossus*), Coughlan et al. 2000; walleye Pollock (*Theraga chalcogramma*), O'Reilly et al. 2004; herring (*Clupea harengus*), Mariani et al. 2005). In order to continue to use these valuable resources it is a logical development to harness next generation sequencing technologies in genotyping these markers. The new genotyping approach, once fully optimised, can reduce the time and cost of genotyping making genetic resources accessible for more breeding programmes

and fish breeders improving the overall genetic state of farmed stocks regardless of species.

Chapter 5 demonstrates the potential of genotyping-by-sequencing as a method for microsatellite genotyping taking advantage of NGS platforms. Universal tails were used to incorporate barcode sequences into the reads representing the microsatellite marker to be genotyped. The use of a combination of two barcodes, combinatorial barcoding, permitted the identification of amplicon reads resulting from unique individuals. This allowed for independent samples to be pooled together for sequencing, which is essential in implementing GBS in population genetics studies. The study employed existing capillary/gel electrophoresis-based multiplex marker panels developed in Chapter 3, but also attempted to test the applicability of the approach for simultaneous validation and deployment of new marker panels (*de novo*) by testing an experimental condition with similar conditions to *de novo* marker development. The results suggested that the method is applicable for both existing microsatellite panels and *de novo* marker panels. For genotyping of existing multiplex panels, the number of loci should be reduced, from the relatively high number used in this study, and the number of individuals increased. It is also suggested that future applications of this method should employ markers with short amplicon sizes (100-200 bp) in order to benefit from the superior read yield of the platforms offering shorter read lengths, such as Ion Torrent and Illumina platforms and chemistries.

One of the persistent problems with microsatellite markers has been size homoplasy where alleles can differ in underlying sequence, but appear identical in size when assayed with gel/capillary electrophoresis. The current study did indeed reveal a high prevalence of homoplasy indicating that allelic diversity in fragment-size-based studies is underestimated. Consequently, inferred population structures in these studies may also underestimate true levels of genetic variability. This new genotyping approach allows access to sequence information and exact sequence lengths and offers possibilities for inter-laboratory calibration, as well as, new

opportunities for sharing data, as genotype data can be stored in repositories as sequence files.

6.5. Conclusions

The technology development has yielded resources that will be useful in the study of wild cod and domestication of other local cod populations, as well as other marine finfish species, especially other gadoids. More universally, the microsatellite validation and deployment method, and the microsatellite GBS approach developed in this thesis are transferable to research questions far beyond aquaculture or fisheries topics, and can be applied to any question interrogated using microsatellite markers. The results obtained confirm that significant changes take place in the genetic composition of aquaculture stocks throughout rearing and close monitoring is necessary to mitigate negative effects of such changes. The findings of this thesis end previous controversy by confirming that cod in the waters surrounding Ireland form a single interbreeding population.

6.6. References

Allendorf, F. W., Berry, O., and Ryman, N. 2014. So long to genetic diversity, and thanks for all the fish. Molecular Ecology 23:23–25.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:1–7.

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D. C., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., and 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 277:3725–34.

Coughlan, J., Stefansson, M., Galvin, P., FitzGerald, R., and Cross, T. F. 2000. Isolation and characterisation of 11 microsatellite loci in Atlantic halibut *Hippoglossus hippoglossus* L. Molecular Ecology 9:11–13.

Duarte, C. M., Marbá, N., and Holmer, M. 2007. Rapid domestication of marine species. Science 316:382–3.

Dutil, J.-D., and Brander, K. 2003. Comparing productivity of North Atlantic cod (*Gadus morhua*) stocks and limits to growth production. Fisheries Oceanography 12:502–512.

FAO. 2014. The state of world fisheries and aquaculture. Food and Agriculture Oraganization of the United Nations.

Fishback, A. G., Danzmann, R. G., Sakamoto, T., and Ferguson, M. M. 1999. Optimization of semi-automated microsatellite multiplex polymerase chain reaction systems for rainbow trout (*Oncorhynchus mykiss*). Aquaculture 172:247–254.

Frankham, R., Bradshaw, C. J. A., and Brook, B. W. 2014. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. Biological Conservation 170:56–63.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., and Petit, R. J. 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources 11:591–611.

Heath, M. R., Culling, M. A., Crozier, W. W., Fox, C. J., Gurney, W. S. C., Hutchinson, W. F., Nielsen, E. E., Sullivan, M. O., Preedy, K. F., Righton, A., Speirs, D. C., Taylor, M. I., Wright, P. J., and Carvalho, G. R. 2014. Combination of genetics and spatial modelling highlights the sensitivity of cod (*Gadus morhua*) population diversity in the North Sea to distributions of fishing. ICES Journal of Marine Science 71:794–807.

Higgins, B., Hubert, S., Simpson, G., Stone, C., and Bowman, S. 2009. Characterization of 155 EST-derived microsatellites from Atlantic cod (*Gadus morhua*) and validation for

linkage mapping. Molecular Ecology Resources 9:733–737.

Hutchinson, W., Carvalho, G., and 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.

Jackson, T. R., Martin-Robichaud, D. J., and Reith, M. E. 2003. Application of DNA markers to the management of Atlantic halibut (*Hippoglossus hippoglossus*) broodstock. Aquaculture 220:245–259.

Lage, C. R., and Kornfield, I. 1999. Isolation and characterization of microsatellite loci in Atlantic haddock (*Melanogrammus aeglefinus*). Molecular Ecology 8:1351–1362.

Liu, Z. J., and Cordes, J. F. 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture 238:1–37.

Mariani, S., Hutchinson, W., Hatfield, E., Ruzzante, D., Simmonds, E., Dahlgren, T., Andre, C., Brigham, J., Torstensen, E., and Carvalho, G. 2005. North Sea herring population structure revealed by microsatellite analysis. Marine Ecology Progress Series 303:245–257.

Nei, M., and Tajima, F. 1981. Genetic drift and estimation of effective population size. Genetics 98:625–640.

Nielsen, E., Wright, P., Hemmer-Hansen, J., Poulsen, N., Gibb, I. M., and Meldrup, D. 2009. Microgeographical population structure of cod *Gadus morhua* in the North Sea and west of Scotland: the role of sampling loci and individuals. Marine Ecology Progress Series 376:213–225.

Norris, A. 1999. Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. Aquaculture 180:247–264.

O'Leary, D. B., Coughlan, J., Dillane, E., McCarthy, T. V., and Cross, T. F. 2007. Microsatellite variation in cod *Gadus morhua* throughout its geographic range. Journal of Fish Biology 70:310–335.

O'Reilly, P. T., Canino, M. F., Bailey, K. M., and Bentzen, P. 2004. Inverse relationship between F_{ST} and microsatellite polymorphism in the marine fish, walleye pollock (*Theragra chalcogramma*): Implications for resolving weak population structure. Molecular Ecology 13:1799–1814.

Pampoulie, C., Stefánsson, M. Ö., Jörundsdóttir, T. D., Danilowicz, B. S., and Daníelsdóttir, A. K. 2008. Recolonization history and large-scale dispersal in the open sea: the case study of the North Atlantic cod, *Gadus morhua* L. Biological Journal of the Linnean Society 94:315–329.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. 2012. Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135

Ryman, N., Palm, S., André, C., Carvalho, G. R., Dahlgren, T. G., Jorde, P. E., Laikre, L., Larsson, L. C., Palmé, A., and Ruzzante, D. E. 2006. Power for detecting genetic divergence: Differences between statistical methods and marker loci. Molecular Ecology 15:2031–2045.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments – A poor man's approach to genotyping for research and high-throughput diagnostics. Nature Biotechnology 18:233–234.

Subasinghe, R. P., McGladdery, S. E., Curry, D., and Bartley, D. 2003. Recent technological innovations in aquaculture. Review of the state of world aquaculture, FAO Fisheries Circular 886:59–74.

Vartia, S., Villanueva-Cañas, J. L., Finarelli, J., Farrell, E. D., Collins, P. C., Hughes, G. M., Carlsson, J. E. L., Gauthier, D. T., McGinnity, P., Cross, T. F., FitzGerald, R. D., Mirimin, L., Crispie, F., Cotter, P. D., and Carlsson, J. 2016. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. Royal Society Open Science 3:150565.

Wang, S., Meyer, E., McKay, J. K., and Matz, M. V. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods 9:808–810.

Appendices

Appendix 1 BLAST hits of the 55 microsatellite loci primers searched against *Gadus morhua* in the Whole-genome shotgun contigs database in Genbank

Primer name	Name of best hit	Accession no.	% Identity	Length of align.	Query start	Query end	Subject start	Subject end	E value
A03_F	contig236159	CAEA01127894.1	100	20	1	20	450	469	0.006
A04_F	contig561296	CAEA01381837.1	100	20	1	20	93	74	0.006
A08_F	contig329249	CAEA01195481.1	100	19	1	19	906	888	0.023
A11_F	contig547549	CAEA01373975.1	100	20	1	20	1368	1349	0.006
A16_F	contig91315	CAEA01547591.1	95	20	1	20	725	706	0.29
A18_F	contig116377	CAEA01022512.1	100	20	1	20	703	684	0.006
A19_F	contig696497	CAEA01457403.1	100	20	1	20	827	808	0.006
A20_F	contig882913	CAEA01534067.1	100	20	1	20	1552	1571	0.006
A22_F	contig264103	CAEA01146189.1	100	20	1	20	1024	1005	0.006
A30_F	contig179262	CAEA01083664.1	100	20	1	20	561	580	0.006
A31_F	contig889085	CAEA01540758.1	100	20	1	20	7068	7087	0.006
A33_F	contig96246	CAEA01551785.1	100	20	1	20	2499	2480	0.006
A34_F	contig306704	CAEA01177506.1	100	20	1	20	7224	7243	0.006
A37_F	contig406243	CAEA01258542.1	100	20	1	20	1194	1213	0.006
A39_F	contig350193	CAEA01212575.1	100	20	1	20	2369	2350	0.006
A43_F	contig58807	CAEA01397255.1	100	20	1	20	1659	1640	0.006
B01_F	contig80676	CAEA01508858.1	100	20	1	20	1292	1273	0.006
B03_F	contig320571	CAEA01188503.1	100	20	1	20	260	279	0.006
B07_F	contig884319	CAEA01535585.1	100	25	1	25	2639	2663	1.00E-05
B12_F	contig353573	CAEA01215347.1	100	20	1	20	1544	1525	0.006
B15_F	contig55910	CAEA01380961.1	100	21	1	21	4688	4668	0.002
B19_F	contig884794	CAEA01536106.1	100	20	1	20	344	325	0.006

Primer name	Name of best hit	Accession no.	% Identity	Length of align.	Query start	Query end	Subject start	Subject end	E value
B28_F	contig96405	CAEA01551885.1	95.24	21	1	20	3119	3139	0.29
B29_F	contig111672	CAEA01017946.1	95.24	21	1	20	961	941	0.29
B30_F	contig372227	CAEA01230488.1	100	20	1	20	317	298	0.006
B33_F	contig182835	CAEA01087150.1	100	20	1	20	1474	1493	0.006
B36_F	contig368008	CAEA01227052.1	100	18	1	18	321	338	0.081
B38_F	contig886567	CAEA01538036.1	95	20	1	20	4860	4878	1.1
C01_F	contig889944	CAEA01541680.1	95	20	1	20	1340	1321	0.29
C08_F	contig138840	CAEA01044355.1	100	20	1	20	577	596	0.006
C13_F	contig94964	CAEA01550699.1	100	20	1	20	1313	1332	0.006
C14_F	contig890234	CAEA01541994.1	100	20	1	20	5445	5464	0.006
C15_F	contig283435	CAEA01159677.1	100	20	1	20	61	42	0.006
C17_F	contig885749	CAEA01537152.1	100	20	1	20	627	646	0.006
C20_F	contig119710	CAEA01025720.1	100	21	1	21	1105	1125	0.002
C22_F	contig749929	CAEA01486978.1	100	19	1	19	120	102	0.023
C28_F	contig884951	CAEA01536277.1	100	20	1	20	1372	1391	0.006
C30_F	contig344699	CAEA01208094.1	100	20	1	20	105	124	0.006
C31_F	contig85706	CAEA01525561.1	100	20	1	20	1396	1377	0.006
C35_F	contig549938	CAEA01375710.1	100	20	1	20	3223	3242	0.006
C36_F	contig320962	CAEA01188840.1	100	20	1	20	1940	1921	0.006
C40_F	contig888314	CAEA01539930.1	100	20	1	20	4274	4293	0.006
C42_F	contig40710	CAEA01259242.1	100	20	1	20	6252	6233	0.006
D05_F	contig881311	CAEA01532322.1	100	20	1	20	6332	6313	0.006
D10_F	contig114745	CAEA01020896.1	100	20	1	20	2444	2425	0.006
D12_F	contig882640	CAEA01533770.1	100	20	1	20	4430	4411	0.006
D14_F	contig52321	CAEA01360772.1	100	20	1	20	1098	1079	0.006

Primer name	Name of best hit	Accession no.	% Identity	Length of align.	Query start	Query end	Subject start	Subject end	E value
D15_F	contig66409	CAEA01438378.1	100	20	1	20	2496	2515	0.006
D16_F	contig880291	CAEA01531221.1	100	20	1	20	1418	1437	0.006
D21_F	contig125767	CAEA01031570.1	100	20	1	20	956	975	0.006
D30_F	contig51563	CAEA01354798.1	100	20	1	20	936	955	0.006
D35_F	contig674890	CAEA01445464.1	100	20	1	20	123	104	0.006
D37_F	contig323213	CAEA01190654.1	100	20	1	20	8387	8406	0.006
D43_F	contig889679	CAEA01541394.1	100	20	1	20	2993	2974	0.006
D46_F	contig296077	CAEA01168917.1	100	22	1	22	344	365	2.00E-04
A03_R	contig73589	CAEA01477685.1	100	16	2	17	1732	1747	1.1
A04_R	contig351852	CAEA01213927.1	100	15	5	19	25	11	3.8
A08_R	contig329249	CAEA01195481.1	100	20	1	20	766	785	0.006
A11_R	contig547549	CAEA01373975.1	100	19	1	19	1227	1245	0.023
A16_R	contig91315	CAEA01547591.1	100	22	1	22	561	582	5.00E-04
A18_R	contig116377	CAEA01022512.1	95	20	1	20	521	540	0.29
A19_R	contig881235	CAEA01532242.1	100	16	1	16	2067	2082	1.1
A20_R	contig125694	CAEA01031500.1	100	18	2	19	277	260	0.081
A22_R	contig264103	CAEA01146189.1	100	20	1	20	818	837	0.006
A30_R	contig891916	CAEA01543820.1	100	17	3	19	4927	4943	0.29
A31_R	contig889085	CAEA01540758.1	100	20	1	20	7402	7383	0.006
A33_R	contig96246	CAEA01551785.1	100	20	1	20	2265	2284	0.006
A34_R	contig306704	CAEA01177506.1	100	20	1	20	7501	7482	0.006
A37_R	contig406243	CAEA01258542.1	100	20	1	20	1472	1453	0.006
A39_R	contig350193	CAEA01212575.1	100	20	1	20	1995	2014	0.006
A43_R	contig58807	CAEA01397255.1	100	20	1	20	1347	1366	0.006
B01_R	contig569386	CAEA01385913.1	100	20	1	20	3059	3040	0.006

Primer name	Name of best hit	Accession no.	% Identity	Length of align.	Query start	Query end	Subject start	Subject end	E value
B03_R	contig320571	CAEA01188503.1	100	20	1	20	380	361	0.006
B07_R	contig884319	CAEA01535585.1	100	20	1	20	2780	2761	0.006
B12_R	contig889720	CAEA01541440.1	100	16	5	20	767	782	1.1
B15_R	contig55910	CAEA01380961.1	100	20	1	20	4540	4559	0.006
B19_R	contig884794	CAEA01536106.1	100	20	1	20	116	135	0.006
B28_R	contig96405	CAEA01551885.1	100	20	1	20	3324	3305	0.006
B29_R	contig111672	CAEA01017946.1	100	20	1	20	718	737	0.006
B30_R	contig372227	CAEA01230488.1	100	22	1	22	185	206	5.00E-04
B33_R	contig182835	CAEA01087150.1	100	20	1	20	1575	1556	0.006
B36_R	contig364049	CAEA01223793.1	95	20	1	20	151	132	0.29
B38_R	contig886567	CAEA01538036.1	95.24	21	1	21	5245	5225	0.081
C01_R	contig34310	CAEA01206763.1	95.24	21	2	22	2912	2892	0.081
C08_R	contig138840	CAEA01044355.1	100	20	1	20	698	679	0.006
C13_R	contig94964	CAEA01550699.1	100	20	1	20	1449	1430	0.006
C14_R	contig890234	CAEA01541994.1	100	20	1	20	5575	5556	0.006
C15_R	contig753484	CAEA01489240.1	100	20	1	20	86	67	0.006
C17_R	contig885749	CAEA01537152.1	100	20	1	20	771	752	0.006
C20_R	contig344009	CAEA01207549.1	100	17	5	21	2101	2117	0.29
C22_R	contig291538	CAEA01165662.1	95.24	21	1	20	154	174	0.29
C28_R	contig884951	CAEA01536277.1	100	20	1	20	1613	1594	0.006
C30_R	contig344699	CAEA01208094.1	100	20	1	20	410	391	0.006
C31_R	contig538450	CAEA01370120.1	100	15	2	16	360	346	3.8
C35_R	contig549938	CAEA01375710.1	95	20	1	20	3471	3453	1.1
C36_R	contig320962	CAEA01188840.1	95	20	1	20	1639	1658	0.29
C40_R	ontig888314	CAEA01539930.1	100	18	1	18	4523	4506	0.081

Primer name	Name of best hit	Accession no.	% Identity	Length of align.	Query start	Query end	Subject start	Subject end	E value
C42_R	contig40710	CAEA01259242.1	100	20	1	20	5933	5952	0.006
D05_R	contig881311	CAEA01532322.1	100	20	1	20	6119	6138	0.006
D10_R	contig884493	CAEA01535777.1	94.74	19	1	19	3521	3539	1.1
D12_R	contig882640	CAEA01533770.1	100	20	1	20	4193	4212	0.006
D14_R	contig52321	CAEA01360772.1	100	18	3	20	897	914	0.081
D15_R	contig66409	CAEA01438378.1	100	20	1	20	2716	2697	0.006
D16_R	contig05146	CAEA01003665.1	100	16	2	17	5870	5885	1.1
D21_R	contig88561	CAEA01536999.1	95.45	22	1	22	1515	1535	0.081
D30_R	contig51563	CAEA01354798.1	95.83	24	1	23	1235	1212	0.006
D35_R	contig391096	CAEA01246040.1	100	17	2	18	511	495	0.29
D37_R	contig323213	CAEA01190654.1	95	20	1	20	8685	8667	1.1
D43_R	contig433731	CAEA01281599.1	94.44	18	1	18	3128	3145	3.8
D46_R	contig296077	CAEA01168917.1	100	19	1	19	466	448	0.023

Appendix 2 Summary statistics for 50 microsatellite loci for five samples.
n, number of individuals; a, number of alleles; R_S, allelic richness based on 24 individuals; as, allele size range in base pairs; H_E, expected heterozygosity; H_O, observed heterozygosity; HW, probability values of concordance with Hardy-Weinberg expectations. Values in bold type are significant probability estimates after Bonferroni correction for multiple tests (initial $\alpha = 0.05$).

Celtic Sea 2011

Locus name	n	a	R_S	as	H_E	H_O	HW
A03_Hill	45	4	4.00	158-170	0.683	0.689	0.971
A04_Neo	45	18	15.62	143-203	0.915	**0.644**	**0.000**
A08_T3	45	8	6.96	163-184	0.653	0.489	0.018
A11_Hill	45	4	3.07	163-175	0.524	0.467	0.527
A16_M13	46	13	11.36	156-194	0.877	**0.674**	**0.000**
A18_M13	46	12	10.60	211-243	0.852	0.761	0.165
A19_T3	46	6	4.34	242-274	0.257	**0.152**	0.009
A20_Neo	45	2	1.53	178-182	0.022	0.022	-
A22_Neo	46	5	4.19	214-240	0.338	**0.217**	0.005
A30_T3	45	7	6.68	245-269	0.792	0.667	0.323
A31_T3	46	9	7.55	313-374	0.555	**0.304**	**0.000**
A34_M13	46	9	8.19	297-321	0.741	0.674	0.067
A37_Neo	46	20	14.35	265-489	0.844	0.717	0.071
A39_Hill	46	13	10.24	316-408	0.741	0.696	0.125
A43_T3	46	13	9.97	331-369	0.745	0.587	0.260
B01_Hill	46	13	9.75	129-181	0.735	0.717	0.539
B03_T3	46	6	4.59	147-161	0.381	0.413	1.000
B07_M13	46	9	7.34	163-181	0.602	**0.435**	0.071
B15_T3	46	5	4.77	170-178	0.514	0.609	0.108
B19_T3	46	19	14.45	242-290	0.855	0.870	0.805
B28_Neo	46	4	3.30	226-238	0.536	0.391	0.094
B29_Neo	46	9	7.53	262-281	0.738	0.783	0.876
B30_Neo	46	6	5.77	154-169	0.744	0.804	0.206
B33_M13	46	5	4.02	120-135	0.354	0.413	0.699
B36_Hill	41	2	1.83	384-393	0.048	0.000	0.013
B38_Neo	46	4	3.30	409-415	0.370	0.370	1.000
C01_M13	46	15	12.77	127-199	0.850	**0.565**	**0.000**
C08_T3	46	13	10.41	149-185	0.765	0.761	0.759
C13_Neo	46	5	4.24	158-173	0.276	0.239	0.335
C14_Hill	45	6	5.76	146-166	0.751	0.644	0.052
C15_Hill	45	18	14.37	171-205	0.902	**0.578**	**0.000**
C17_M13	46	8	7.21	154-190	0.750	**0.543**	**0.001**
C20_Neo	45	4	3.47	227-236	0.207	**0.067**	**0.000**
C22_Hill	46	12	8.90	237-274	0.793	0.739	0.506
C28_Neo	46	7	6.44	258-282	0.580	0.543	0.350
C30_Neo	46	2	2.00	337-340	0.505	0.435	0.394
C31_Neo	44	7	5.43	330-341	0.538	0.545	0.409
C35_M13	44	5	4.09	268-281	0.520	0.432	0.076
C36_Hill	46	6	4.09	320-354	0.347	0.326	0.820
C40_M13	45	2	2.00	276-277	0.425	0.467	0.725
C42_M13	46	4	3.77	354-357	0.649	0.674	0.991
D05_Hill	45	8	6.50	230-262	0.448	0.444	0.420
D12_Hill	46	10	8.27	259-279	0.778	**0.652**	0.029
D14_Hill	46	13	10.24	228-276	0.795	0.739	0.494
D15_Hill	46	4	3.03	245-252	0.164	0.152	0.037
D16_T3	44	4	3.95	424-436	0.340	0.341	0.509
D21_Neo	46	5	4.00	336-343	0.366	**0.130**	**0.000**
D30_M13	46	5	4.41	321-333	0.535	0.565	0.799
D35_M13	46	17	13.48	316-358	0.870	**0.652**	0.003
D46_Neo	46	4	3.03	141-155	0.164	0.174	1.000

Table 1 *Continued.*

				Celtic Sea 2012			
Locus name	n	a	R_S	as	H_E	H_O	HW
A03_Hill	45	4	3.96	158-170	0.624	0.711	0.214
A04_Neo	45	19	16.43	143-207	0.932	**0.622**	**0.000**
A08_T3	45	9	7.17	159-182	0.720	0.622	0.117
A11_Hill	45	6	4.87	159-175	0.585	0.578	0.213
A16_M13	42	11	10.01	156-194	0.804	0.667	0.023
A18_M13	45	13	11.62	203-243	0.893	0.867	0.179
A19_T3	45	5	3.85	254-270	0.339	0.356	1.000
A20_Neo	45	2	1.53	178-182	0.022	0.022	-
A22_Neo	44	6	4.43	214-243	0.490	0.409	0.552
A30_T3	45	7	6.77	237-265	0.810	0.711	0.065
A31_T3	40	7	6.16	314-370	0.356	0.325	0.177
A34_M13	38	11	9.37	297-325	0.737	0.842	0.937
A37_Neo	45	24	17.08	289-430	0.885	**0.756**	0.019
A39_Hill	44	11	9.51	364-420	0.657	0.523	0.031
A43_T3	45	13	8.95	329-359	0.656	0.556	0.130
B01_Hill	45	11	9.31	125-149	0.773	0.756	0.057
B03_T3	45	4	3.51	151-157	0.411	0.444	1.000
B07_M13	45	8	6.84	165-189	0.693	0.489	0.002
B15_T3	44	5	4.50	170-178	0.579	0.455	0.147
B19_T3	44	14	11.01	242-323	0.786	0.727	0.574
B28_Neo	45	5	3.60	229-238	0.486	0.356	0.079
B29_Neo	44	6	5.94	265-277	0.712	0.773	0.799
B30_Neo	45	8	7.04	151-178	0.706	0.644	0.192
B33_M13	45	4	3.44	123-132	0.353	0.378	1.000
B36_Hill	40	2	1.84	391-393	0.049	0.050	1.000
B38_Neo	45	5	4.02	389-418	0.400	0.333	0.257
C01_M13	43	13	10.93	127-191	0.829	0.628	0.006
C08_T3	44	15	12.22	144-185	0.712	0.682	0.381
C13_Neo	45	4	3.10	161-173	0.108	0.111	1.000
C14_Hill	45	6	5.07	146-166	0.716	0.622	0.304
C15_Hill	45	20	16.10	157-210	0.896	**0.622**	**0.000**
C17_M13	44	9	8.26	154-199	0.832	**0.636**	**0.001**
C20_Neo	45	6	4.38	225-236	0.227	**0.111**	**0.001**
C22_Hill	45	13	10.77	237-281	0.781	0.778	0.618
C28_Neo	44	8	6.66	258-282	0.504	0.523	0.355
C30_Neo	45	4	3.07	316-343	0.527	0.556	0.822
C31_Neo	42	5	4.79	332-343	0.601	0.643	0.319
C35_M13	45	6	4.83	268-284	0.460	0.400	0.003
C36_Hill	45	8	5.68	322-357	0.266	0.244	0.127
C40_M13	45	2	2.00	276-277	0.449	0.489	0.736
C42_M13	44	5	4.09	345-357	0.585	0.523	0.579
D05_Hill	45	8	6.70	230-258	0.467	0.511	0.943
D12_Hill	45	13	9.91	255-291	0.732	0.667	0.692
D14_Hill	45	13	10.30	224-276	0.777	0.733	0.288
D15_Hill	45	2	2.00	245-248	0.145	0.111	0.224
D16_T3	42	4	3.81	424-436	0.381	0.405	0.260
D21_Neo	45	4	2.60	333-343	0.066	0.067	1.000
D30_M13	45	4	3.53	321-329	0.459	0.489	0.439
D35_M13	45	16	13.08	316-360	0.869	**0.756**	0.019
D46_Neo	45	2	1.98	147-151	0.106	0.111	1.000

Table 1 *Continued.*

				Irish Sea 2011/2012			
Locus name	n	a	R_S	as	H_E	H_O	HW
A03_Hill	37	4	3.95	158-170	0.589	0.568	0.786
A04_Neo	38	14	12.53	143-216	0.878	**0.632**	**0.000**
A08_T3	45	9	8.17	159-183	0.757	**0.622**	0.007
A11_Hill	44	4	3.51	159-171	0.558	0.705	0.141
A16_M13	40	14	12.33	156-194	0.891	**0.575**	**0.000**
A18_M13	44	14	12.23	203-251	0.891	0.750	**0.000**
A19_T3	44	7	4.73	246-270	0.318	0.364	1.000
A20_Neo	45	2	1.53	160-178	0.022	0.022	-
A22_Neo	45	5	4.02	214-243	0.344	0.356	0.302
A30_T3	43	6	5.88	245-265	0.763	0.651	0.317
A31_T3	36	3	2.56	314-366	0.082	0.083	1.000
A34_M13	24	8	8.00	297-317	0.743	0.708	0.108
A37_Neo	40	20	15.16	285-425	0.847	**0.725**	**0.000**
A39_Hill	43	13	10.39	364-412	0.692	0.721	0.561
A43_T3	45	15	11.28	329-369	0.779	**0.644**	0.028
B01_Hill	43	11	9.21	129-159	0.747	0.698	0.357
B03_T3	45	6	4.39	145-161	0.327	0.333	1.000
B07_M13	40	8	7.01	167-189	0.585	0.600	0.828
B15_T3	45	5	4.22	170-178	0.560	0.667	0.251
B19_T3	43	14	10.91	242-308	0.793	0.814	0.566
B28_Neo	39	4	3.23	226-232	0.467	0.513	0.869
B29_Neo	43	8	6.73	265-281	0.699	0.721	0.748
B30_Neo	42	6	5.81	154-169	0.717	0.690	0.266
B33_M13	45	6	4.13	120-138	0.380	0.444	0.029
B36_Hill	43	2	1.56	393-399	0.023	0.023	-
B38_Neo	45	5	4.68	389-415	0.453	0.422	0.432
C01_M13	44	11	8.93	127-195	0.765	0.750	0.687
C08_T3	43	10	8.42	149-185	0.629	0.744	0.788
C13_Neo	45	4	3.30	161-173	0.168	0.178	1.000
C14_Hill	40	5	4.54	150-166	0.677	0.500	0.146
C15_Hill	44	17	14.70	167-189	0.907	**0.545**	**0.000**
C17_M13	45	9	7.88	154-194	0.711	0.556	0.043
C20_Neo	44	4	3.50	226-236	0.152	0.114	0.035
C22_Hill	44	9	7.62	237-266	0.777	0.727	0.068
C28_Neo	44	9	7.31	257-282	0.506	0.523	0.719
C30_Neo	43	3	2.56	320-340	0.502	0.558	0.718
C31_Neo	31	4	3.77	332-341	0.289	0.323	1.000
C35_M13	42	6	5.13	268-284	0.610	0.524	0.140
C36_Hill	44	5	4.30	328-354	0.269	0.273	0.434
C40_M13	45	2	2.00	276-277	0.310	0.289	0.637
C42_M13	33	4	3.73	354-357	0.567	0.606	0.246
D05_Hill	43	9	6.88	230-278	0.524	0.535	**0.001**
D12_Hill	44	10	8.01	257-291	0.679	0.727	0.984
D14_Hill	44	13	10.63	228-272	0.811	0.864	0.832
D15_Hill	45	2	2.00	245-248	0.217	0.244	1.000
D16_T3	44	4	3.71	424-436	0.333	0.341	0.118
D21_Neo	44	4	3.25	337-343	0.131	0.136	1.000
D30_M13	44	5	4.46	321-333	0.602	0.545	0.565
D35_M13	40	17	12.62	316-368	0.789	0.600	0.051
D46_Neo	45	2	2.00	147-151	0.200	0.222	1.000

Table 1 *Continued.*

				West of Ireland 2011			
Locus name	n	a	R_S	as	H_E	H_O	HW
A03_Hill	45	5	4.53	154-170	0.664	0.667	0.701
A04_Neo	45	20	17.26	143-211	0.942	**0.533**	**0.000**
A08_T3	45	8	6.75	163-182	0.619	0.578	0.292
A11_Hill	45	4	3.32	163-175	0.537	0.622	0.262
A16_M13	45	13	11.07	156-194	0.843	**0.600**	**0.000**
A18_M13	45	13	11.38	211-243	0.885	**0.711**	0.056
A19_T3	44	5	3.64	254-270	0.280	0.273	0.688
A20_Neo	45	1	1.00	178-178	0.000	0.000	-
A22_Neo	44	4	3.34	214-239	0.407	0.273	0.059
A30_T3	45	7	6.51	245-265	0.797	0.756	0.279
A31_T3	45	8	6.59	314-366	0.521	0.333	**0.000**
A34_M13	45	8	7.60	297-317	0.740	0.689	0.616
A37_Neo	45	26	18.25	285-449	0.884	**0.689**	**0.000**
A39_Hill	45	13	10.86	364-424	0.775	0.800	0.403
A43_T3	45	13	10.17	307-359	0.755	**0.600**	**0.000**
B01_Hill	45	11	9.60	127-177	0.807	0.822	0.900
B03_T3	45	7	5.04	138-157	0.316	0.311	0.060
B07_M13	45	12	8.92	141-189	0.618	0.556	0.130
B15_T3	45	3	2.53	174-178	0.440	0.444	1.000
B19_T3	45	14	10.77	245-311	0.824	0.867	0.967
B28_Neo	45	4	3.07	226-238	0.525	0.444	0.325
B29_Neo	44	8	7.04	259-281	0.668	0.750	0.693
B30_Neo	45	6	5.78	154-169	0.666	0.667	0.212
B33_M13	45	5	4.32	123-135	0.549	0.533	0.667
B36_Hill	43	4	3.51	384-393	0.216	**0.093**	**0.001**
B38_Neo	45	6	4.58	389-418	0.516	0.511	0.824
C01_M13	45	16	11.79	119-195	0.842	0.800	0.369
C08_T3	45	14	11.23	149-185	0.722	**0.622**	0.009
C13_Neo	45	5	3.56	161-173	0.149	0.156	1.000
C14_Hill	45	5	4.90	150-166	0.746	0.689	0.435
C15_Hill	45	21	17.45	167-204	0.905	**0.600**	**0.000**
C17_M13	45	9	7.52	154-190	0.669	0.533	0.019
C20_Neo	43	4	3.36	227-236	0.195	0.116	0.018
C22_Hill	45	10	7.45	237-269	0.711	0.778	0.961
C28_Neo	45	6	5.42	264-282	0.594	0.556	0.583
C30_Neo	45	2	2.00	337-340	0.475	0.444	0.753
C31_Neo	45	6	5.42	332-343	0.642	0.667	0.959
C35_M13	45	5	4.49	269-284	0.531	0.400	0.026
C36_Hill	45	5	4.31	322-351	0.349	0.400	1.000
C40_M13	45	2	2.00	276-277	0.396	0.400	1.000
C42_M13	45	3	3.00	354-357	0.598	0.533	0.601
D05_Hill	45	9	7.15	230-270	0.480	0.511	0.579
D12_Hill	45	9	6.92	259-291	0.694	0.667	0.358
D14_Hill	45	11	10.17	228-268	0.787	0.689	0.005
D15_Hill	45	2	2.00	245-248	0.164	0.178	1.000
D16_T3	43	6	4.64	424-444	0.328	0.372	1.000
D21_Neo	45	5	3.39	333-345	0.108	0.111	1.000
D30_M13	45	4	3.96	321-333	0.553	0.578	0.385
D35_M13	44	17	13.92	316-354	0.894	**0.682**	0.020
D46_Neo	45	5	4.02	139-151	0.263	**0.156**	**0.001**

Table 1 *Continued.*

				Scotian Shelf 1996			
Locus name	n	a	R_S	as	H_E	H_O	HW
A03_Hill	46	4	3.97	158-170	0.586	0.652	0.221
A04_Neo	46	21	17.47	143-211	0.932	**0.587**	**0.000**
A08_T3	42	7	6.40	163-187	0.494	0.476	0.399
A11_Hill	45	4	3.68	163-175	0.244	0.267	1.000
A16_M13	44	10	8.21	170-194	0.640	0.568	0.142
A18_M13	45	11	10.32	203-239	0.868	0.711	0.012
A19_T3	45	7	5.19	254-275	0.229	0.244	1.000
A20_Neo	46	4	3.67	162-182	0.335	0.261	0.179
A22_Neo	45	5	4.05	214-243	0.428	0.422	0.926
A30_T3	44	7	6.96	245-269	0.840	**0.591**	**0.000**
A31_T3	45	11	7.98	310-386	0.580	**0.200**	**0.000**
A34_M13	45	9	6.87	297-317	0.636	0.689	0.467
A37_Neo	45	27	19.32	289-441	0.933	0.889	0.324
A39_Hill	46	12	10.17	364-416	0.685	0.696	0.177
A43_T3	45	9	7.17	331-359	0.754	0.689	0.535
B01_Hill	46	15	11.30	129-211	0.839	0.826	0.475
B03_T3	46	6	4.34	149-161	0.240	0.217	0.083
B07_M13	46	8	6.29	167-187	0.457	0.326	0.078
B15_T3	46	4	3.77	170-178	0.401	0.435	0.702
B19_T3	45	30	19.52	242-320	0.848	0.867	0.020
B28_Neo	46	5	4.44	226-235	0.534	0.565	0.017
B29_Neo	44	5	4.30	265-277	0.509	0.545	0.582
B30_Neo	46	7	6.39	151-169	0.694	0.630	0.237
B33_M13	45	7	6.37	120-135	0.494	0.533	0.353
B36_Hill	40	3	2.60	390-396	0.262	**0.100**	**0.001**
B38_Neo	46	5	3.56	389-415	0.202	0.152	0.103
C01_M13	38	7	6.98	115-139	0.824	**0.421**	**0.000**
C08_T3	45	10	9.43	145-181	0.850	0.756	0.029
C13_Neo	46	6	4.66	158-176	0.224	0.239	1.000
C14_Hill	44	4	3.55	146-158	0.625	0.523	0.415
C15_Hill	45	15	12.21	169-187	0.795	0.733	0.087
C17_M13	46	8	7.89	154-182	0.803	0.739	0.530
C20_Neo	45	5	4.41	227-239	0.317	0.178	**0.000**
C22_Hill	46	10	9.30	225-273	0.800	0.848	0.433
C28_Neo	46	7	5.54	255-279	0.480	0.500	0.964
C30_Neo	46	4	3.30	329-343	0.487	0.413	0.132
C31_Neo	46	4	3.30	332-343	0.405	0.413	0.745
C35_M13	46	5	4.32	233-281	0.275	0.261	0.693
C36_Hill	46	7	4.97	325-351	0.260	0.283	1.000
C40_M13	46	2	2.00	276-277	0.161	0.174	1.000
C42_M13	46	8	6.71	345-358	0.714	0.739	0.954
D05_Hill	46	10	7.45	230-270	0.659	0.717	0.957
D12_Hill	46	10	7.05	251-291	0.382	0.348	0.036
D14_Hill	46	7	6.39	228-244	0.560	0.565	0.224
D15_Hill	45	4	3.39	245-256	0.169	0.178	1.000
D16_T3	45	4	3.67	424-436	0.207	0.222	1.000
D21_Neo	45	6	5.01	340-389	0.334	**0.244**	0.016
D30_M13	45	8	6.56	315-333	0.644	0.444	0.002
D35_M13	44	13	10.58	316-368	0.718	**0.432**	**0.000**
D46_Neo	45	6	4.60	139-151	0.423	0.444	0.295