The genetic structuring of Atlantic salmon (*Salmo salar* L.) populations in northwest Europe as revealed through nuclear microsatellite and mtDNA PCR-RFLP analysis:

From regional to catchment level analysis and its application in conservation and management of the species.

In the style of the Journal of Fish Biology.

Submitted by Anna Kathryn Finnegan to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences, June 2009.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

(signature) .....

ABSTRACT

The structuring of Atlantic salmon (*Salmo salar* L.) into discrete, genetically differentiated populations both within and between river catchments is well documented. The utilisation of this knowledge has proved valuable in a variety of evolutionary, ecological, managerial and conservation contexts. In this thesis, the genetic structuring of Atlantic salmon populations in northwest Europe was assessed in two catchments of very different sizes, using a range of molecular and associated population genetic methods; findings from the catchment level research are set in context by a broader phylogeographic study of post-glacial colonisation of the region.

A regional study into the glacial origins and post-glacial colonisation routes of Atlantic salmon in northwest Europe was explored by analysing a pre-existing microsatellite dataset and supplementing it with haplotype data from mtDNA PCR-RFLP analysis of the same samples (N=702). Evidence from allele permutation tests undertaken on the microsatellite data alongside mtDNA haplotype frequencies suggested that there was a cryptic northern refuge in northwest France, with colonisation of the British Isles and Ireland occurring from this and the long-known Iberian Peninsula refuge.

Catchment level studies were undertaken on the river Dart and river Tweed, involving 1151 fish being genotyped with 14 microsatellite loci with a subset of 211 fish being genotyped by mtDNA PCR-RFLP. In both catchments, populations were found to be weakly differentiated genetically, and were most consistent with the meta-population theory of evolution. Similarly, individual spatial autocorrelation analysis indicated that each major tributary within the catchments could be considered as a distinct management or conservation unit. In the Tweed dataset, however, limitations in the sample coverage across the catchment reduced the robustness of some findings.

Historical stocking of the river Dart with fish from Scotland and Iceland is well-documented. The long-term implications of these activities on contemporary Dart populations were assessed by genotyping 177 fish from the donor populations using scale samples taken in the 1960s and comparing them to contemporary Dart populations by undertaking admixture analysis. Overall, admixture between the

donor and recipient populations was low and appeared to reflect natural underlying levels of genetic relationships. However, increased admixture of donor stocks with one extant Dart population was apparent, indicating some potentially long-term localised success of the stocked fish through hybridisation with the native populations; nevertheless, with the population continuing to decline, this should not be viewed as a successful supplementation programme.

Two tributaries on the river Tweed, the Gala and Leader, were inaccessible to salmon for long periods due to the construction of barriers to migration. On both tributaries, fish passes were installed in the 1940s and re-colonisation of the tributaries was possible. Assignment analysis was undertaken and indicated that, contrary to findings for between catchment studies, salmon straying from the most proximate tributaries (*i.e.* the Ettrick and Caddon) did not appear to be the principal colonisers of the current Gala and Leader populations. Rather, the highest proportion of Gala samples assigned to the Teviot (42%), with the Leader populations assigning to many tributaries across the catchment (Ettrick 28%; Upper 21%; Teviot 19%). However, given the relatively weak differentiation of the baseline samples and limitations inherent in the dataset, the correct self-assignment of baseline samples was very low (average 26%; range 0-47%), hence interpretation must be undertaken with caution. Nevertheless, the findings suggest that the Gala population may have reached a temporally stable state in the 60 years since it has been accessible to salmon.

Whilst the relatively small scale of these studies is acknowledged, the application of the findings in management and conservation of the species are discussed in a wider context. These studies would support the following recommendations: to include information on the historic (refugial) origin of contemporary populations in regional management strategies; to treat each major tributary as a distinct unit as an appropriate scale for catchment level management; and, with stocking and supplementation programmes appearing to have no significant long-term success, coupled with the relative speed with which extirpated tributaries appear to be naturally re-colonised, the use of stocking and supplementation programmes should be discouraged.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Firstly, I would like to thank my PhD. supervisor Dr. Jamie Stevens for supporting me over the last 5 years. Thanks for giving me the freedom to pursue avenues of personal interest without objection. Your insightful discussions always provided a fresh perspective on my research and ensured I kept my feet firmly on the ground.

Thanks also to my unofficial second supervisor and one of the most supportive friends anyone could wish for, Dr. Andrew Griffiths. Andrew was the font of all knowledge in Lab C4; from teaching the basics in the lab, through to invaluable discussions on the merits of various statistical programmes over afternoon tea. Your comments on earlier drafts of this thesis and the associated papers have been invaluable. Thank-you for being my primary source of knowledge for the last 5 years and the first person I would always discuss new ideas with; you were the perfect bench-partner!

I was extremely lucky to work in such a fantastic laboratory with such warm and generous people, so thank you to all past and present members of Lab C4. Particularly, thanks go to Dr. Patrick Hamilton for many interesting and helpful discussions and to Dr. Katie Sumner for her early work in microsatellite analysis of Atlantic salmon in our laboratory.

I also owe a massive thank-you to everyone who ever helped in fieldwork. Firstly, thanks to the Environment Agency Fisheries and Biodiversity electric-fishing team, lead by Dave Hoskin, and all the EA staff who were part of that team. Equally as instrumental were the University electric-fishing team, lead by Jan and Phil Shears. Thank-you to Dave, Jan and Phil for your expert knowledge and your patience while we strived for enough fish! Thanks also to everyone from Hatherly who helped out, including Jamie, Andrew, Viv, Ed, Toby, Tess, Patrick, Angela, Okyan and Jan W. Also, thanks to Dylan and Toby at the WestCoutry Rivers Trust for a gloriously sunny days sampling on the Swincombe. Without all of your generous help I wouldn't have had the most important ingredient for a good thesis – samples!!!

For funding this PhD, I owe my thanks to the Natural Environment Research Council, the WestCountry Rivers Trust, the Tweed Foundation and the Atlantic

Salmon Trust. Specifically, thanks to Dr. Jamie Stevens for believing it was a good enough idea at the inception to submit the proposal to NERC, to Dr. Dylan Bright at the WestCountry Rivers Trust for providing CASE partnership and also for introducing us to Dr. Ronald Campbell who arranged the additional funding from both the Tweed Foundation and the Atlantic Salmon Trust.

I also wish to thank Dr. Ronald Campbell at the Tweed Foundation for providing a fantastic array of samples from the Tweed catchment and the level of in-depth knowledge of a study site that one can only usually dream of. These insights into the history of the catchment have made interpretation of the Tweed samples much more interesting! Thanks also to James Hunt at the Tweed Foundation for invaluable training when setting up my GIS and continued assistance throughout the course of the project, both in GIS issues and in providing additional information on the Tweed catchment. For providing historical scale samples, I wish to thank Julian MacLean at the Fisheries Research Service, Montrose, Scotland and Sigurdur Gudjonsson at the Institute of Freshwater Fisheries, Iceland.

Finally, I owe my warmest thanks to mum and dad for giving me the confidence to pursue my dreams and the support to achieve them, and to my fiancé, George; moving in together as I started to write up this thesis could have been a disaster, but your boundless humour and constant support has made it a fantastic year.

CHAPTER 1: INTRODUCTION

The exploitation of natural resources through the expansion of human populations has dramatically increased the rate of decline and extinction for many species worldwide. In 2008 the ICUN (www.icunredlist.org) estimated that almost 40% of those organisms assessed were threatened, while the FAO (www.fao.org) estimate that over 70% of the worlds' fish species are either fully or over exploited. There is an increasing international effort to try to slow down, and even reverse these trends (*e.g*. the Earth Summit, Rio de Janeiro 1992 and Johannesburg 2002) and a growing appreciation that the maintenance of genetic diversity must be at the core of these endeavours, since it is in the genetic diversity of a species that its capacity to evolve in future climatic conditions lies.

Within a species, genetic diversity is arranged hierarchically from family units through to range wide phylogeographic lineages (Avise, 1994). To maintain this intra-specific genetic diversity, management and conservation programmes must consider genetic diversity at each hierarchical level (Kenchington *et al.*, 2003), requiring detailed investigations to define geographic boundaries and identify evolutionary mechanisms acting to maintain them. This thesis addresses these issues in relation to the Atlantic salmon (*Salmo salar* L.), a species which is currently experiencing a rapid decline throughout its native range (Parrish *et al.*, 1998). The genetic characterisation of wild populations from small scale within-river structuring, up to regional phylogeographic lineages will be considered, alongside the outcomes of stocking and supplementation programmes, and the natural re-colonisation of river systems. In this way, the value of defining and understanding the genetic characterisation at a sub-species level and the role this information can play in management and conservation of the Atlantic salmon is explored.

GLOBAL DISTRIBUTION

The native, historical freshwater distribution of the Atlantic salmon spans the east coast of North America between the Hudson River in New York State to outer Ungava Bay in Quebec. In Europe, salmon rivers extend south from Iceland, the Barents Sea and south western part of the Kara Sea, along the drainages of the eastern Atlantic coast, North Sea and Baltic Sea, with a southern limit at northern

Portugal and the Bay of Biscay (Figure 1.1) (MacCrimmon and Gots, 1979; Klemetsen *et al.*, 2003). Much of this historical freshwater range has been compromised over the last 200 years due to environmental changes resulting from habitat degradation caused by anthropogenic activities, such as dam construction and pollution events. Consequently, local extinctions have been witnessed across the global distribution of the Atlantic salmon (MacCrimmon and Gots, 1979; Parrish *et al.*, 1998). Furthermore, many native populations are only maintained through supplementation programmes where artificially reared fish are introduced into compromised or extirpated populations (MacCrimmon and Gots, 1979).

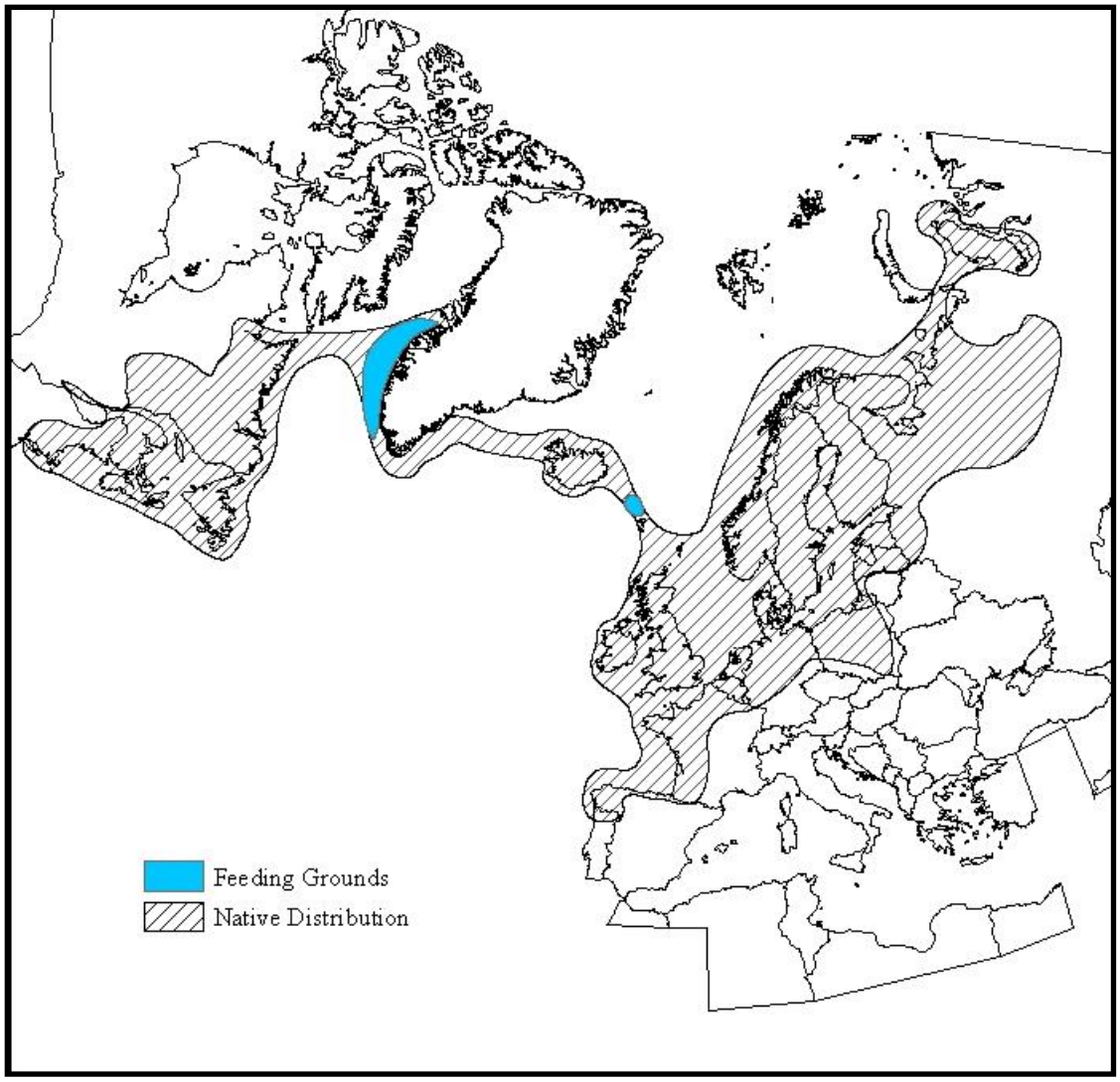


Figure 1.1 Atlantic salmon distribution. Modified from MacCrimmon and Gots (1979) and Klemetson *et al*. (2003).

The native range of the Atlantic salmon has been artificially extended through the introduction of hatchery-reared fish into non-native river systems (and

hatchery facilities) in Mexico, South America, South Africa, India, Indonesia, Australia and New Zealand. However, the majority of these introductions have failed to become naturalised and most populations are maintained through supportive breeding programmes (MacCrimmon and Gots, 1979). Consequently, despite the demise of the wild, native Atlantic salmon, the species is not threatened, and indeed, with the continued growth of the aquaculture industry, there may be more Atlantic salmon alive today than ever in history. This, however, does not detract from the fact that the state of wild populations is still a major cause for concern (Dodson *et al.*, 1998; Parrish *et al.*, 1998) and that the escape of farmed fish can potentially have negative repercussions for the native, wild populations (McGinnity *et al.*, 1997).

LIFE HISTORY

The Atlantic salmon is a semelparous, anadromous fish that migrates between freshwater and marine environments at different stages in its life history: reproduction and juvenile phases occur in freshwater, while its main feeding and growth phase occurs at sea (Mills, 1971) (Figure 1.2).

LIFE IN FRESHWATER

Sexually mature Atlantic salmon can enter the river at any time of year, and are called spring, summer and autumn fish accordingly, depending on this run time. If the salmon has spent just one year at sea before this return migration, the fish is termed a 'grilse', or for two or more years at sea they are termed multi-sea-winter salmon. Once in freshwater again, the salmon stop feeding until after spawning (Mills, 1971).

The returning salmon migrate upstream at varying speeds depending on the time of year, water temperature and stream flow. Spawning begins in November, but some late running fish will not deposit their eggs until January or February (Mills, 1971). The female salmon constructs nests in silt-free gravel substrate in their breeding territories, in the upside of riffles or gravel bars where dissolved oxygen levels are high (Mills, 1971; Fleming, 1998). To construct the nest, the female adult salmon turns on its side, pressing the caudal fin against the gravel and then rapidly lifts the caudal fin, creating a water vortex that displaces the gravel (Fleming, 1998).

Adult

Upon completion of smoltification, the salmon undertake long migrations to feeding grounds in the Ocean. British Atlantic salmon have been found on feeding grounds off west Greenland and the Faeroe Islands. Most salmon mature at sea before making the return migration to spawn in their natal rivers.

Smolt

Smoltification occurs in parr after one year or more. The morphological and physiological changes undergone in this process adapt the salmon for life in the oceans.

Parr

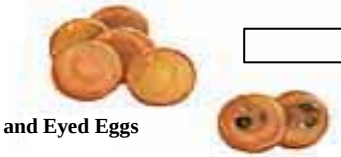
Characterised by 9-10 parr marks, these territorial fish defend established feeding territories, while feeding off aquatic invertebrates. Some males will mature as parr.

Fry

Remain in localised 'home areas' with siblings for approximately one year before developing into parr.

Spawning

Females deposit their eggs in a string of nests, called a redd. Large anadromous males compete to fertilise the eggs and small precocious parr sneak in under the larger fish at the last moment to add their milt.



Eggs and Eyed Eggs

Eggs are laid down in autumn and winter

Alevin

Emerge from the eggs in spring, feeding from their yolk-sac until it has been fully absorbed 4-5 weeks later.

Figure 1.2 Life cycle of the Atlantic Salmon. Modified from the Atlantic Salmon Federation ©.

The female deposits her eggs in the nest, and males compete to fertilise them. Once fertilised the nest is immediately covered with gravel. This protects the eggs from predators, gravel shifts by freshets and digging by other females, and desiccation during low water and freezing (Fleming, 1998). After covering the eggs, the female will slowly switch to excavating a new nest in front of the previous one. A string of nests together like this is termed a redd. The number of nests in a redd correlates with body size; small one sea winter salmon create redds with three to five nests, while large multi sea winter salmon can construct redds comprised of as many as 14 nests (Fleming, 1996; Fleming *et al.*, 1997). This nest building period lasts for about five to six days for females, but males may remain sexually active for a month or more (Fleming, 1996; Fleming *et al.*, 1997).

Eggs are laid down in streams during autumn and winter, hatching in late March, early April, with the juvenile salmon emerging as alevins. The alevins remain in the gravel, feeding from their yolk-sac until they emerge as fry four to five weeks later once the yolk-sac has been fully absorbed (Mills, 1971). The fry remain in localised areas with their siblings for some time after emergence, in areas with shallow riffles and moderate current velocities (Saunders and Gee, 1964; McCormick *et al.*, 1998). After approximately one year's feeding, the fry develop into parr, characterised by 9 to 10 black stripes down the sides of their bodies known as parr marks (Mills, 1971). Parr inhabit fast flowing riffles with rough gravel, defending established feeding territories (McCormick *et al.*, 1998). They are sit-and-wait predators, largely feeding on aquatic invertebrates being carried past in the drift (Mills, 1971; Metcalfe, 1998). During this phase the parr may move to other nursery sites away from their natal stream where there is more food and better wintering habitat, but they rarely move far (McCormick *et al.*, 1998). This characteristic nature of fry and parr remaining close to their 'home area' where they were spawned is maintained by rheotactic behaviour to olfactory cues; the juvenile fish demonstrate a positive rheotaxis to natal odorant cues, and a negative rheotaxis in the absence of it (Saunders and Gee, 1964).

PARR-SMOLT TRANSFORMATION

At a specific stage of development, parr undergo morphological and physiological transformations that pre-adapt them for survival and growth in the marine

environment. This is known as the parr-smolt transformation, or smoltification (Boeuf, 1994).

TIMING

Smolting can occur after one year, but has been observed in parr of eight years old, with single rivers often producing three or more age classes of smolts (Metcalfe and Thorpe, 1990). Smolt transformation is triggered in late summer, seven to ten months before entry to sea (Metcalfe, 1998). McCormick *et al.* (1998) described the presence of a 'smolt window', where certain physiological and environmental factors must coincide for smoltification to be triggered. An environmental window was defined by seasonal changes in environmental conditions appropriate for smolt survival, which coincided with a physiological window, which determined the physiological and behavioural readiness for smolt migration and seawater entry (McCormick *et al.*, 1998). Temperature and photoperiod are believed to be the most important environmental variables involved in triggering smolting (Morin and Doving, 1992; McCormick *et al.*, 1998). However, it appears that whether or not an individual parr will respond to these environmental and physiological cues and undergo the parr-smolt transformation is dependent on the fish reaching a genetically pre-determined size threshold (Boeuf, 1994; McCormick *et al.*, 1998; Metcalfe, 1998). By late summer, if the fish is set to exceed a projected size threshold by the following spring, then smolt transformation will be initiated. If the projected size falls below this threshold value, then the fish will remain as parr in freshwater for another year (Metcalfe, 1998). Larger individuals reach the threshold for smolting earlier than their smaller counterparts, and hence leave the rivers earlier. This results in the occurrence of two size groups of smolts with differing temporal patterns of smolt development and seaward migration (Boeuf, 1994; Arnesen *et al.*, 2003).

MORPHOLOGICAL AND PHYSIOLOGICAL CHANGES

During the smoltification period, parr undergo a combination of cytological, morphological, biochemical, physiological, endocrinological and behavioural changes in freshwater, with the active downstream migration concluding the process (Boeuf, 1994). Morphological changes include a silvering of the skin and scales caused by deposition of guanine and hypoxanthine, which is thought to aid predation avoidance in the marine environment (McCormick *et al.*, 1998). The fish also become more elongated and streamlined by growing more in length than weight, and

show a marked decrease in condition factor (McCormick *et al.*, 1998; Arnesen *et al.*, 2003). Physiologically, the fish show increased hypo-osmoregulatory capacity, an increased scope for growth, increased buoyancy, and a shift in visual pigments and change in haemoglobin (McCormick *et al.*, 1998; Arnesen *et al.*, 2003). At the end of the parr-smolt transformation the smolts are stimulated to embark on a downstream migration (Boeuf, 1994), which lasts between three to six weeks in most populations and is typically undertaken at night (McCormick *et al.*, 1998). Numerous behavioural changes are observed during this period, including increased negative rheotaxis (*i.e*. downstream orientation) and schooling, decreased agonistic and territorial behaviour, and increased salinity preference (McCormick *et al.*, 1998). The resulting transformation ensures that the salmon smolts are not just adapted to survive in seawater, but to thrive there (McCormick *et al.*, 1998).

Olfactory imprinting

During the smolting period the fish develops a high sensitivity to environmental factors (Boeuf, 1994), including olfactory sensitivity and odour learning important for the imprinting process. Olfactory activity varies over the course of smoltification, with two distinct periods observed that coincide with two peaks of odour learning (Morin *et al.*, 1989; Morin and Doving, 1992). The first, and strongest, peak occurs at the first signs of parr-smolt transformation, when the fish is still close to its 'home area' in its natal river. The second peak, of much lower intensity, occurs after smoltification. It is believed that this first period of increased olfactory activity, in a sensitive period for olfactory imprinting, is the critical stage for the imprinting process which allows migratory adults to return to their natal rivers to reproduce (Stabell, 1984; Morin *et al.*, 1989; Morin and Doving, 1992).

Sexual maturation

Sexual maturation of Atlantic salmon commences approximately one year prior to spawning (Kadri *et al.*, 1996), triggered if the individual fish is projected to exceed a threshold level of resources by the following autumn (Metcalfe, 1998) in an analogous manner to the triggering of smoltification. Atlantic salmon can mature as parr, 1-3 sea winter fish and, on rare occasions, older ages, though it is rare for females to mature as parr (Fleming, 1998; Klemetsen *et al.*, 2003). Maturation of male Atlantic salmon as both parr, grilse and multi sea winter fish gives rise to two phenotypes of mature male Atlantic salmon; large anadromous males that fight for

access to females, and small precocious parr that sneak access to the females by occupying a position near to the female (Fleming, 1998), though parr can successfully mate with the anadromous females in absence of the large anadromous males (Klemetsen *et al.*, 2003). Anadromy greatly increases the fecundity of the individual, and there is a reduction in mortality of the eggs and fry in freshwater of anadromous parentage (McCormick *et al.*, 1998). However, mature male parr are capable of fertilising 11-40% of all eggs (Fleming, 1998). These mature parr can go on to smolt and reproduce again as multi sea winter fish, though the probability of smolting is reduced if the fish mature as parr (Berglund *et al.*, 1991). The presence of these two reproductive strategies has been described as a 'bet-hedging' strategy, which maximises survival and population stability (Klemetsen *et al.*, 2003). On average, about 11% of the population will survive to breed again as kelts, though in some populations this may be as high as 43% (Ducharme, 1969). Large anadromous females are the least likely contingent of the population to be repeat spawners due to the large energy expenditure incurred in their first spawning event (Fleming, 1998).

LIFE AT SEA

In contrast to the freshwater phase of the Atlantic salmon life-cycle, relatively little research has been undertaken exploring the marine phase, and accordingly, our understanding about their time at sea is limited (Hansen and Quinn, 1998). The marine phase involves extensive migrations for thousands of miles to productive feeding grounds in the North Atlantic where they feed on pelagic animals such as crustaceans, fish and squid (Hansen and Pethon, 1985). Research from scale analysis and tagging studies suggests that salmon from most areas of the European range spend some time feeding north of the Faeroe Islands in the Norwegian Sea (Hansen *et al.*, 1993; Hansen and Quinn, 1998), and many progress on to feeding areas west of Greenland, which are also frequented by salmon of North American origin (Reddin and Friedland, 1999) (Figure 1.1).

In a review on this subject, Jonsson and Jonsson (2004) estimated marine survival to be less than 10% for most rivers, with the highest mortality occurring in the first year at sea. Furthermore, it is a widely held opinion that increased mortality during the marine phase has significantly contributed to the continued decline of Atlantic salmon over the last few decades (Dodson *et al.*, 1998; Parrish *et al.*, 1998; Jonsson and Jonsson, 2004). Potential factors involved in the success of the marine

migration are many and varied (Jonsson and Jonsson, 2004). The importance of environmental conditions encountered during the critical period as post-smolts first enter the sea has been demonstrated by Jonsson and Jonsson (2004), who revealed a positive correlation between the North Atlantic Oscillation Index (NAOI) for May and one sea winter fish survival. Friedland *et al.* (2005) also stressed the importance of thermal conditions during this period by demonstrating relationships between sea surface temperature (SST) and early marine growth. These effects, however appeared complex and stock specific, for example, in one Scottish population, increased SST as the smolts first entered the sea in May negatively affected salmon growth. This was attributed to the direct effect of temperature on feeding activity, whereby the increased temperatures exceeded the optimal range for feeding. However, although similar correlations were observed in a Canadian population, the temperatures were outside ecologically relevant ranges (*i.e*. the temperatures of 3-4ºC observed are outside the range experienced by smolts entering the marine environment and cooler than temperatures post-smolts are believed to utilize during early marine residence), and thus it was suggested that indirect measures were acting on the population (Friedland *et al.*, 2005).

The effects of this early marine growth will also propagate through the marine life of salmon. For example, growth at sea may influence susceptibility to predation, as larger fish are less likely to be preyed upon (Friedland *et al.*, 2000). Growth rate has also been linked to age at maturity, which determines the length of time the salmon stay at sea (Hansen *et al.*, 1993; Hutchings and Jones, 1998). Beaugrand and Reid (2003) demonstrated correlations between salmon catches and zooplankton data, suggesting that production at lower trophic levels can influence salmon post-smolts via prey availability (Friedland *et al.*, 2005) and again highlights the link between environmental conditions, growth and success.

Atlantic salmon remain at sea for, on average, between one and four years before maturing and making the return migration to their home rivers (Hansen *et al.*, 1993). Many factors have been presented as contributing to the timing of this event. Freshwater life history such as smolt size, smolt age and timing of seaward migration, have all been proposed as contributors (reviewed in Jonsson and Jonsson, 2004). Furthermore, Metcalfe (1998) proposed that a pre-determined resource

threshold must be exceeded for maturation to occur, again highlighting the important role of growth at sea and environmental conditions.

MIGRATION AND HOMING

The mechanisms involved in driving migration are largely unknown and several hypotheses are currently being debated, including: navigation, facilitated by the sensitivity Atlantic salmon show to magnetic fields; compass orientation using the sun's positioning and magnetic fields; and piloting, using information learned on the outward journey (reviewed in Hansen and Quinn, 1998). It has also been proposed that genetically determined, population specific, migration patterns may exist for Atlantic salmon, as has been suggested for Pacific salmon (Hansen and Quinn, 1998). However, no single hypothesis has taken precedence over the others as yet.

In contrast, once close to the natal river, it has been established that odours in the form of pheromones and physical stimulants, which are learnt during the imprinting period of smoltification, guide the salmon in the final stage of homing to ensure reproduction ensues in their natal rivers (Stabell, 1984; Morin *et al.*, 1989; Morin and Doving, 1992; further details can be found previously in this chapter in OLFACTORY IMPRINTING). Homing accuracy in wild Atlantic salmon is believed to be in the range of 94-98%, as estimated in a review by Garcia de Leaniz *et al*. (2007). A degree of straying is likely to occur in all populations (Jonsson *et al*., 2003), since effective migration can be an important mechanism for maintaining small populations, restoring extirpated populations after localised extinction events (Saura *et al.,* 2008) and colonising virgin habitat, *i.e.* after glaciation events (further details can be found later in this chapter in GENETIC DIFFERENCES BETWEEN CATCHMENTS). Straying, however, may not always result in gene flow since migrants may have a lower fitness in a particular environment; a concept referred to as 'selection against immigration' (Hendry, 2004; Nosil, 2005; Bolnick & Nosil, 2007). Hence, the precise homing of wild Atlantic salmon acts to create reproductive isolation and the formation of distinct populations, whilst the low frequency of straying maintains the potential to re-colonise extirpated populations (Saura *et al.,* 2008) or virgin habitat.

ANTHROPOGENIC INFLUENCES ON ATLANTIC SALMON POPULATIONS

Many populations over the global range of the Atlantic salmon are in decline (Crozier *et al.*, 2004; ICES, 2006) (Figure 1.3). Areas with high human population densities, *i.e*. in the southern parts of the species' range, appear to be the worst affected, whilst the northern populations, which are less impacted upon by human activities due to lower population densities (*i.e*. Russia and Iceland), are maintaining relatively stable populations (Parrish *et al.*, 1998). Factors contributing to these declines are many and varied, and whilst cause and effect are difficult to infer overfishing practices, habitat destruction (MacCrimmon and Gots, 1979), escape of farmed fish (McGinnity *et al.*, 1997; McGinnity *et al.*, 2003), supplementation programmes (Youngson and Verspoor, 1998), pollution (MacCrimmon and Gots, 1979), pathogens and disease (Bakke *et al.*, 1990) have all been cited as contributing factors to the decline either globally or locally. However, in most instances it is impossible to isolate one causal factor, and multiple factors are likely to be acting in concert (Parrish *et al.*, 1998).

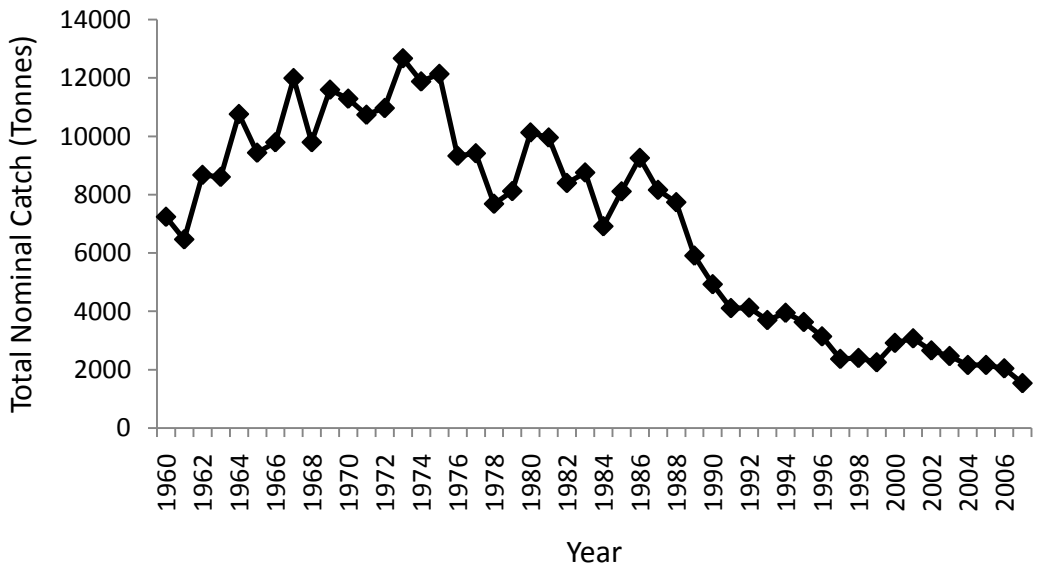


Figure 1.3 Total nominal catches (Tonnes) of Atlantic salmon from 1960-2007 (ICES, 2008). NB CPUE data unavailable.

Within rivers the construction of barriers to migration, for example the construction of dams, weirs and hydropower developments, has been the single most detrimental human activity to riverine populations of Atlantic salmon (MacCrimmon and Gots, 1979; Parrish *et al.*, 1998). Their construction extirpates all populations upstream of the barrier unless a fish pass is installed, in which case passage is allowed with a slight delay (Smith *et al.*, 1997) and then movement in the upstream loch, reservoir or river appears to be achieved with little difficulty (Gowans *et al.*, 2003). As these barriers are removed the upstream habitat may be re-colonised either naturally or through stocking programmes, though the efficacy and outcomes of these processes are rarely explored.

Other factors that have increased the incidence of local extinctions or have potentially exacerbated the decline observed in rivers include: the introduction of the parasite *Gyrodactylus salaries* to susceptible populations (Bakke *et al.*, 1990; Dalgaard *et al.*, 2003); intensive agricultural and industrial practices leading to habitat degradation and reduced survival of fish through such effects as siltation (Soulsby *et al.*, 2001); pesticide run-off (Moore and Waring, 1996; 2001); pollution events (Hendry *et al.*, 2003); and, deposition of acid rain (Clair *et al.*, 2004). This is by no means an exhaustive account, but evidently many factors may exhort negative impacts on riverine populations of Atlantic salmon and can potentially act in unison, further contributing to the continuing decline we see throughout the species range.

In order to counteract this decline, a common management practice has been to implement supplementation programmes whereby fish reared in captivity, or taken from a different population, or both, are used to augment the native populations or restore extirpated populations (MacCrimmon and Gots, 1979; Dodson *et al.*, 1998; Aprahamian *et al.*, 2003).

The transplantation of fish from exogenous sources introduces novel genetically based characteristics into the native population. This practice has found applications for severely inbred populations of other wild species (Beebee and Rowe, 2004) and historical stocking programmes for Atlantic salmon did champion this as improving the ‘hybrid-vigour’ of wild populations. However, the interbreeding of the introduced and native fish can potentially lead to outbreeding depression, whereby the mating of distantly related fish reduces the fitness of the offspring

(Taylor, 1991; Einum and Fleming, 2001), thus potentially exacerbating the decline rather than abating it. This is likely due to adaptations fish show for their native rivers, which has been cited as a common reason for the widespread failures of transplanted stocks (reviewed in Garcia de Leaniz *et al.*, 2007). As our awareness of the importance of population structuring has increased, there has been a move away from transplanting exogenous fish into recipient rivers, and today the common practice, where possible, is to use locally caught fish as broodstock and raise their progeny in a local hatchery before planting out into the rivers at various life stages. Whilst this may theoretically avoid the pitfalls of outbreeding depression, it is not without risk as hatchery fish commonly differ from their wild conspecifics in a multitude of ways. In the first instance, hatchery bred fish often have reduced genetic diversity relative to wild fish (Verspoor, 1988; Wang and Ryman, 2001; Koljonen *et al.*, 2002), probably as a result of the limited number of adults used as broodstock and stochastic processes induced in the hatchery environment (Verspoor, 1988; Youngson and Verspoor, 1997; Wang and Ryman, 2001; Wang *et al.*, 2001). Equally, the extreme environmental and ecological differences between hatchery and wild conditions (*e.g.* differing flow regimes and prey availability) exhort differing selection pressures on the fish (Youngson and Verspoor, 1998). Hatchery rearing substantially increases survival during egg and juvenile stages, thus bringing through genotypes that are potentially eradicated in the wild, and selection pressures during hatchery life may also act to select for behavioural and physiological traits that are disadvantageous in nature (Einum and Fleming, 2001).

Once transplanted out into the recipient rivers, the introduced salmon can interact with wild fish in a multitude of ways. Einum *et al* (2001) used a meta analytical approach, combining data from multiple ecological based studies and reported that hatchery fish consistently exhibit increased levels of aggression relative to wild fish and that they showed a reduced response to predation risk. Differences in the feeding behaviour, habitat use and morphology were also reported (Einum and Fleming, 2001). This unnatural behaviour of introduced fish in the wild appears to negatively impact the native population and long-term sustainability of the river population. For example, the increased levels of aggression of hatchery and farmed fish have been shown to displace native populations (McGinnity *et al.*, 1997; McGinnity *et al.*, 2003). The interbreeding of stocked and wild fish has also been

reported, thus leading to introgression of foreign and native genes and the erosion of locally adapted gene complexes (Martinez *et al*., 2001a). Indeed, Araki *et al.* (2007) showed that the genetic effects of captive breeding programmes can reduce subsequent reproductive capabilities by up to 40%, further highlighting the potential long term repercussions of these programmes (Araki *et al.*, 2007). Hence, the lower reproductive success of the domesticated individuals can culminate in a lower overall productivity of the river (McGinnity *et al*., 1997; Martinez *et al*., 2001a; McGinnity *et al*., 2003; Saisa *et al*., 2003; Saltveit, 2006). Furthermore, the introduced fish may undergo extensive migrations within the river system, thus the repercussions of introductions are not necessarily limited to the locations of input (Vasemagi *et al*., 2005).

In conclusion, human activities on or near our waterways are evidently a contributing factor to the decline of the Atlantic salmon. The measures taken to offset this decline have a real potential to further exacerbate the situation and accelerate the decline. Many current management practices attempt to improve fishery returns to rivers, *i.e*. supplementation programmes, but in reality these practices may exacerbate the decline further rather than counteracting it. Furthermore, the limited number of pre- and post- assessments undertaken as part of the stocking programmes is alarming and should be addressed in order to achieve management objectives and minimise detrimental side effects.

CURRENT GENETIC CHARACTERIZATION

It has long been appreciated that Atlantic salmon populations are highly structured across their native range, from large scale phylogenetic divisions to within river population differentiation.

INTER-SPECIES LEVEL

The Atlantic salmon (*Salmo salar* L.) is a teleost fish of the family Salmonidae. Salmonids are unusual in that they are tetrapoid, with four sets of chromosomes rather than the usual two in the diploid state (Allendorf and Thorgaard, 1984). The Salmonidae are thought to have arisen from a single tetraploid event in a common ancestor, between 25 and 100 million years ago. Sometime after the initial tetraploidy event, the chromosomes arranged themselves into two pairs, giving rise

to duplicate pairs of the original chromosomes. These duplicate pairs are being gradually lost over time through diploidization, and hence are returning to disomic inheritance (Allendorf and Thorgaard, 1984). This is only one of two cases where a polyploidy event has given rise to a whole family and the Salmonidae are still in the process of diploidization.

The family Salmonidae is comprised of three subfamilies: Coregoninae (whitefish and ciscoes), Thymalline (grayling) and Salmoninae (char, trout and salmon) (Allendorf and Thorgaard, 1984; Crespi and Fulton, 2004). Of the three, the Salmoninae is the most specious with at least 5 genera and 30 species, one of which being *Salmo salar* (Figure 1.4). The evolutionary history of the Salmoninae has been disputed for many years (Allendorf and Thorgaard, 1984; Oakley and Phillips, 1999; Crespi and Fulton, 2004). In a phylogenetic tree produced from analysis of 49 morphological characters, a stepwise transition from exclusively freshwater forms (*Thymallus, Brachymystax*) to intermediate anadromy (*Salvelinus, Salmo*) to increasing loss of dependence on freshwater stages (*Oncorhynchus*) was observed (Elliott, 1994). However, recent developments have produced robust phylogenies from combined DNA analysis in answer to some of the main uncertainties (Crespi and Fulton, 2004). Most notably, Crespi and Fulton (2004) and Oakley and Phillips (1999) have reclassified the Salmoninae, asserting that it is *Oncorhynchus* and *Salvelinus* that are sister taxa, rather than the more traditional view that *Oncorhynchus* and *Salmo* are sister taxa. This gives rise to an interesting evolutionary situation, where anadromy must have evolved at least twice within the subfamily: once in *Oncorhynchus* and once in *Salmo* (Crespi and Fulton, 2004). The tetraploid nature of salmonids has given rise to duplicate copies of many genes. It has been proposed that the expression of these different duplicates during the freshwater and marine stages of salmonid life history may have, in part, contributed to the unparalleled success of anadromy within the Salmonidae (Allendorf and Thorgaard, 1984).

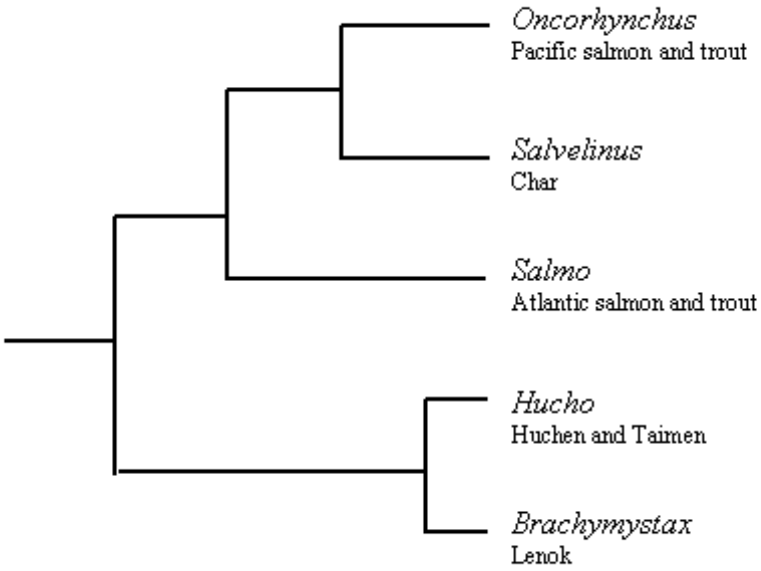


Figure 1.4 Phylogeny of Salmoninae (Family: Salmonidae). Modified from Crespi and Fulton (2004).

INTRA-SPECIES LEVEL

Atlantic salmon populations are highly structured throughout their native range (Verspoor, 1997); strong genetic differentiation is observed between regions and among rivers (Ståhl, 1987; Nielsen *et al.*, 1999; King *et al.*, 2001; Consuegra *et al.*, 2002), with weaker, yet significant, differentiation of populations being observed between locations within river systems (Nielsen *et al.*, 1999; Garant *et al.*, 2000; Primmer *et al.*, 2006; Vaha *et al.*, 2007).

REGIONAL STRUCTURING

Ståhl (1987) was among the first to describe genetically differentiated populations both within and between major drainages of the species native range. He described the three distinct regions of the Western Atlantic, Eastern Atlantic and Baltic drainages, characterised by high, intermediate and low genetic diversity respectively (Ståhl, 1987) (Figure 1.5; Figure 1.6). The greatest genetic differences were found between populations from the two continents of North America and Europe with differentiation between the two European clusters (*i.e.* between Eastern Atlantic and Baltic populations) estimated at less than half of this (Ståhl, 1987; Saisa *et al.*, 2005). It is estimated that the two main lineages of North America and Europe diverged

more than one million years ago (Nilsson *et al.*, 2001), while the shallower differentiation of the European lineage into Atlantic and Baltic sub-lineages is expected to have occurred more recently but probably before the last glacial maximum around 22,000 years before present (YBP) (Verspoor *et al.*, 1999). The division of Atlantic salmon into these three main lineages has since been supported by many studies using multiple techniques including allozymes (Bourke *et al.*, 1997), mitochondrial DNA (King *et al.*, 2000; Consuegra *et al.*, 2002), microsatellites (McConnell *et al.*, 1995; Nielsen *et al.*, 1999; King *et al.*, 2001; Saisa *et al.*, 2005), ribosomal RNA genes (Cutler *et al.*, 1991) and chromosome number (Phillips and Hartley, 1988).

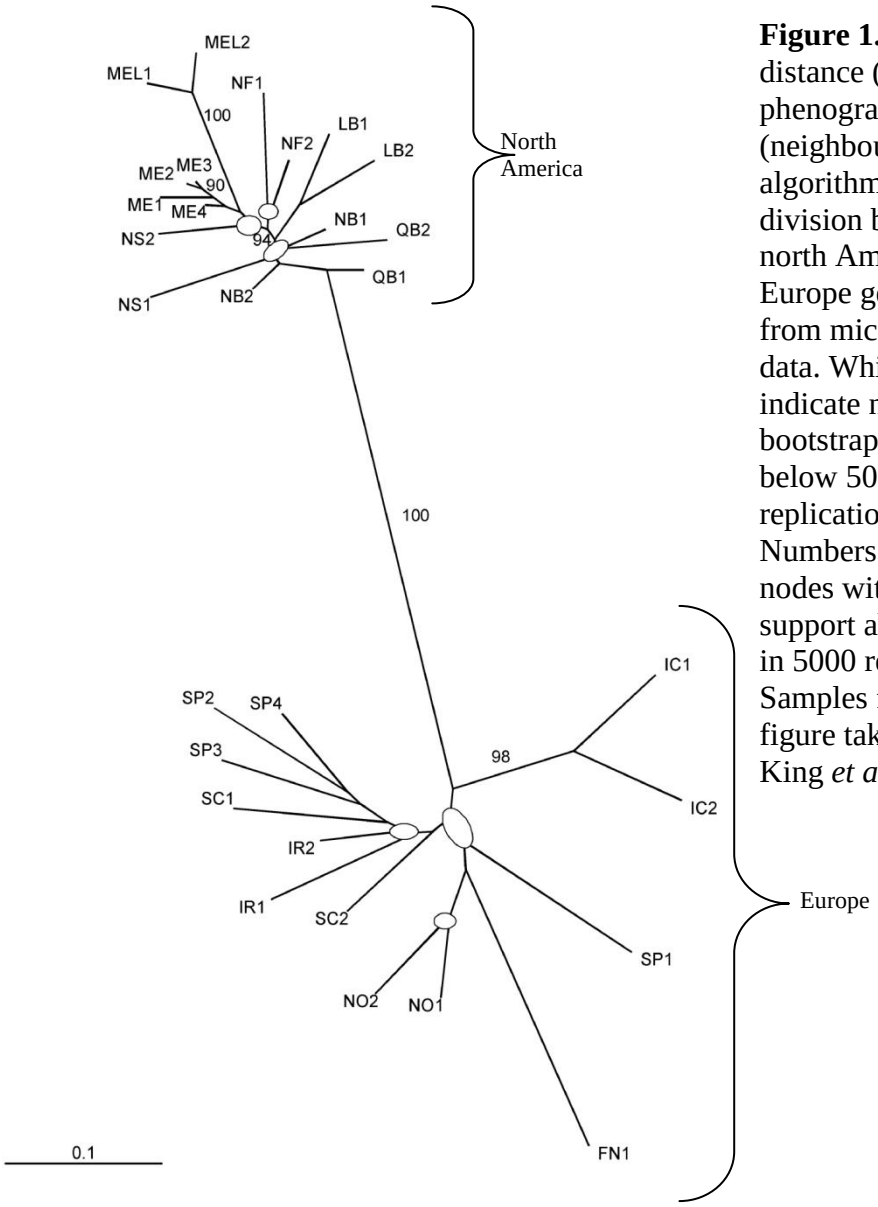


Figure 1.5 Genetic distance (*Da*) phenogram (neighbour-joining algorithm) depicting division between north America and Europe generated from microsatellite data. White ellipses indicate nodes with bootstrap support below 50% in 5000 replications. Numbers indicate nodes with bootstrap support above 90% in 5000 replications. Samples follow, and figure taken from King *et al.* (2001).

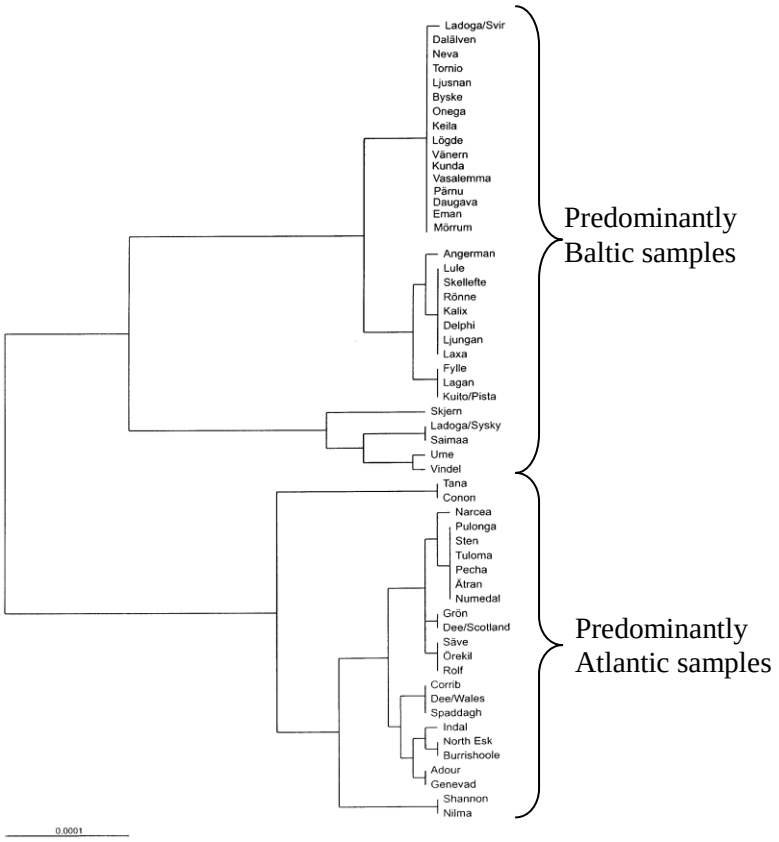


Figure 1.6 UPGMA clustering of Atlantic salmon populations based on the number of net nucleotide substitutions per site (*dA*) at the mitochondrial *ND1* gene showing divergence between salmon from the Atlantic and Baltic drainages. Samples follow, and figure taken from, Nilsson *et al.* (2001)

The differentiation of these lineages and the structuring observed within each region is thought to reflect differences in the quaternary history of the three regions. The most significant historical event to have occurred during the evolutionary lifespan of most extant species is arguably the Pleistocene glaciations (Bernatchez and Wilson, 1998). The Pleistocene period lasted approximately two million years, terminating around 10,000 years ago with a period of global warming. This termination marked the beginning of the Holocene, which continues to the present time (Andersen and Borns, 1994). During the major glaciations of the Pleistocene polar ice sheets expanded considerably, extirpating many populations and compressing species ranges to their southerly limits, thereby restricting populations to glacial refugial locations (Hewitt, 1999; Beebee and Rowe, 2004). Refugial populations remained relatively small during this period, which lasted for 1000s to 10,000's years. During this time, isolated and fragmented populations differentiated genetically primarily as the result of genetic drift, culminating in eventual lineage sorting in the speciation process. Following the onset of glacial retreat, populations at the northern limits of

the glacial range would have expanded into areas of suitable habitat (Hewitt, 1999; Beebee and Rowe, 2004).

The study of this worldwide phenomenon of range compression, population fragmentation and differentiation, and subsequent post-glacial population expansion from northern leading edge populations to the current geographic range by the exploration of gene genealogies over a defined geographic range is termed phylogeography (Avise *et al.*, 1987). Many such studies have been undertaken for Atlantic salmon populations in order to understand the differentiation of the three main lineages, and also regional structuring within each lineage. Most of these studies attempt to retrace the movements of the fish in relation to Pleistocene glaciation events, primarily since the last glacial maximum around 22,000 YBP, and explore geographic correlations in genealogical findings to reveal the evolutionary relationships between contemporary populations.

Evidently, freshwater habitats of Atlantic salmon would have been drastically different from those observed today; the advance of the glaciers from the north would have destroyed old systems resulting in displacement and extirpation of local populations (Bernatchez and Wilson, 1998). At the ice edge major river valleys would have been dammed (Gibbard, 1988; Gibbard and Lewin, 2003) and some river systems would have had much extended lower valleys creating land-bridges between currently separated land masses, such as the land-bridge connecting the British Isles and mainland Europe (Andersen and Borns, 1994; Bridgeland and D'Olier, 1995). Displaced animals would have had to survive for thousands of years in refugia and fringe habitats along glacial margins before re-colonisation at the onset of de-galciation (Bernatchez and Wilson, 1998).

THE PHYLOGEOGRAPHIC HISTORY OF EUROPEAN ATLANTIC SALMON

In Europe, the British and Irish ice sheet extended to around 52º north at the time of the last glacial maximum (Wingfield, 1995; Bowen *et al.*, 2002). This was not believed to be confluent with the Scandinavian ice sheet (Bowen *et al.*, 2002), which extended over the Baltic region and western Russia. Displaced populations of Atlantic salmon could potentially have existed in glacial lakes along the ice margins and ice-dammed lakes throughout the region south of the glaciers. It is also possible that their range extended further south than their current distribution, whilst keeping

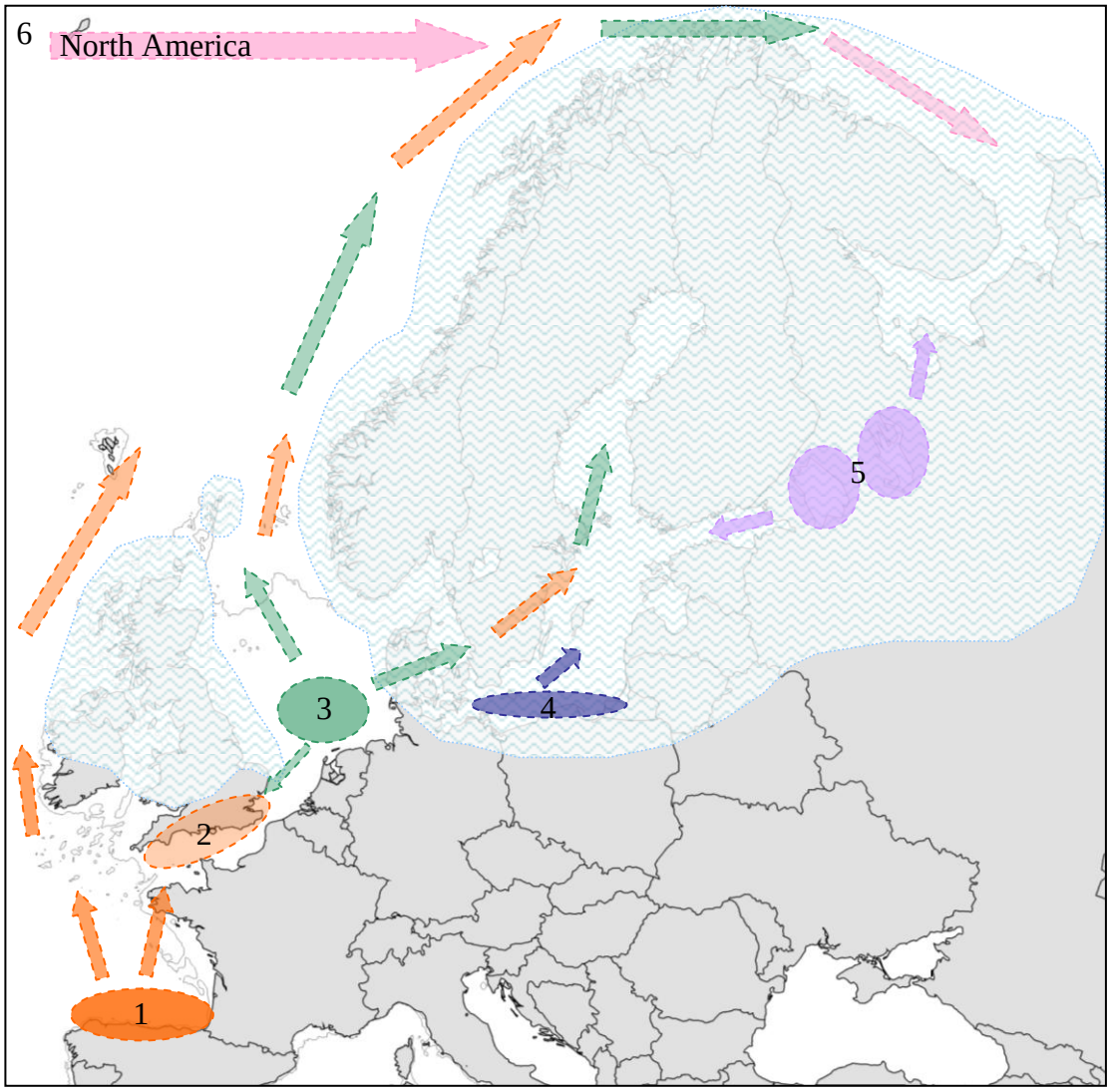
within their thermal limits. The only sound evidence of a refugial location comes in the form of bone relics recovered from a cave in the Iberian Peninsula, which demonstrates the presence of Atlantic salmon in this region for at least the last 40,000 years, *i.e.* during the last glacial period (Consuegra *et al.*, 2002). Speculation on the location of other refugia comes from extensive phylogeographic studies of the region and includes ice lakes present in Russia (Kazakov and Titov, 1991; Nilsson *et al.*, 2001; Saisa *et al.*, 2005), the southern North Sea (Koljonen *et al.*, 1999; Verspoor *et al.*, 1999; Saisa *et al.*, 2005) and the southern Baltic region (Saisa *et al.*, 2005), alongside unglaciated regions of the southern British Isles (Payne *et al.*, 1971; Child *et al.*, 1976; Verspoor, 1986b) and colonisers from North America (Asplund *et al.*, 2004; Wennevik *et al.*, 2004; Makhrov *et al.*, 2005) (Figure 1.7).

Eastern Atlantic and Baltic sub-lineages

The eastern Atlantic drainages, extending south from the Kola Peninsula and Iceland, around the British Isles to the Iberian Peninsula were separated from the Baltic Sea during the Pleistocene until the Närke straight connected the Baltic Ice Lake to the North Sea across southern Sweden in the Yolida stage (approximately 10,000 YBP), and then the Danish Strait finally opened approximately 2000 years later. This would have precluded the mixing of eastern Atlantic and Baltic populations and explains some of the differentiation observed between the two regions. The dichotomy of eastern Atlantic and Baltic lineages has been observed in numerous studies utilising an array of different markers, for example allozymes (Ståhl, 1987; Bourke *et al.*, 1997), mitochondrial DNA (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001; Consuegra *et al.*, 2002) and microsatellites (King *et al.*, 2001; Saisa *et al.*, 2005).

Environmental, behavioural and immunological observations add weight to the proposed genetic differentiation of eastern Atlantic and Baltic salmon populations. The Baltic Sea is the largest brackish water reservoir in the world and as such the environmental conditions of the eastern Atlantic and Baltic regions differ markedly. The migratory behaviour of present day Baltic salmon differs from that of eastern Atlantic salmon in that almost no Baltic salmon migrate into the Atlantic Ocean. Furthermore, the eastern Atlantic and Baltic populations show differential susceptibility to the parasite *Gyrodactylus salaris* with Baltic populations immune

from any negative effect but eastern Atlantic populations highly susceptible with whole populations devastated by its invasion, which is considered to be due to genetic control of susceptibility, or resistance towards *Gyrodactylus* (Bakke *et al.*, 1990). Hence, these differences in life history traits are commonly used to corroborate theories of sub-lineages and differential colonization of the eastern Atlantic and Baltic drainages.



Extent of ice at Last Glacial Maximum (LGM) ~22,000 years BP

Extent of land at LGM when sea level was ~ 100m below present

Proposed refugial locations:

1: Iberian Peninsula;

2: Southwest England/Northern France;

3: Southern North Sea;

4: Southern Baltic Sea;

5: Eastern Baltic/Russian Ice Lakes;

Proposed colonisation routes

Figure 1.7 Proposed refugial locations for Atlantic salmon during the last glacial period and colonisation routes subsequent to de-glaciation.

Northern Europe (Russian, White and Barents Seas drainages)

At the northeast limit of the species range in northern Russia, investigations focused on this region appear to show a degree of structuring not observed in other studies into the phylogeographic history of the eastern Atlantic lineage. Colonisation of this region from three directions appears to have taken place following de-glaciation. Early studies showed populations from this region clustering together at the edge of the Atlantic lineage (Bourke *et al.*, 1997) and colonisation from the Atlantic, possibly from the speculated refugia in the southern North Sea, was proposed as a likely refuge (Verspoor *et al.*, 1999; Asplund *et al.*, 2004; Tonteri *et al.*, 2005). Since the early studies of Kazakov and Titov (1991), the occurrence of predominantly North American alleles or haplotypes has been observed in a number of Russian populations (Kazakov and Titov, 1991; 1993; Bourke *et al.*, 1997; Asplund *et al.*, 2004; Wennevik *et al.*, 2004; Makhrov *et al.*, 2005), but never in more southerly eastern Atlantic populations, thus suggesting the colonisation of Russia from north America. Colonisation from an eastern refugium has also received credible attention. The White Sea coastal populations have mtDNA haplotypes typical of Baltic salmon (Asplund *et al.*, 2004; Makhrov *et al.*, 2005), suggesting colonisation of these populations from a common refuge. The Ice Lake present near Lake Onega could potentially have served as a refuge from which Atlantic salmon colonised both Eastern Europe and the White Sea, given that historically the White Sea and Lake Onega were connected for a period (Ingerslev *et al.*, 2006). Thus, it can be surmised that the western populations of Russia are likely descendents of the eastern Atlantic refuge, possibly the southern North Sea, and colonisers from North America, whereas the eastern most populations are likely descendents of colonisers from an eastern ice lake near Lake Onega, which also served as a refuge for Baltic populations.

Western Europe (Eastern Atlantic and North Sea drainages)

The phylogeographic origins of the eastern Atlantic and North Sea drainages of Western Europe are less well defined relative to the Baltic regions and Russia. Since the uncovering of bone relics in the Iberian Peninsula (Consuegra *et al.*, 2002), this area is unequivocally designated as a refugial location for Atlantic salmon of northwest Europe and is a highly likely source of colonising salmon for the region. However, due to the high diversity found in the region, particularly in the southern

British Isles, this region has been proposed as a contact zone for salmon colonising from multiple refugial locations (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001; Consuegra *et al.*, 2002). Alternatively, the lower genetic diversity in the Baltic region could be attributed to a population bottleneck event, either historically or in recent times, due to the loss of natural spawning habitat in the region (Verspoor *et al.*, 1999). Similarly, the lower genetic diversity of Spanish populations may be due to the continued decline of populations in the region (Parrish *et al.*, 1998). Regardless, support for the polyphyletic origin of this region has been increasing. The existence of two lineages in northwest Europe was initially proposed by Payne *et al.* (1971) based on polymorphisms at the transferrin (Tf) locus in the first phylogeographic study focusing attention on this area. The so-called 'Boreal' race was proposed to occupy most of northern Europe, including the Baltic, whilst the 'Celtic' race was confined to the southwest of England. Payne *et al.* (1971) speculated that the 'Boreal' race originated from refugia present in an ice lake located in the southern North Sea, while the 'Celtic' race had remained in un-glaciated parts of the southwest of England throughout the glacial period. This theory was supported by Child *et al.* (1976) who added sampling locations to Payne's *et al.* (1971) original data set, and then by (Verspoor, 1986b) in work with allozymes, who further proposed the extension of the 'Celtic' race into northern France and Scandinavia, with a more northern group including Iceland and northern Norway.

It is possible that Atlantic salmon persisted north of the Iberian Peninsula, potentially in southwest England, during the last glacial period, which would be in line with Payne's *et al.* (1971) 'Celtic' hypothesis and Verspoors (1986b) findings. However, the 'Boreal-Celtic' hypothesis has also faced major opposition. Ståhl (1987), while exploring the phylogeography of Atlantic salmon across its whole geographic range in an allozyme study found no evidence of the Boreal-Celtic divide. Similarly, Blanco *et al.* (1992) and Bourke *et al.* (1997), also with allozyme studies, submitted results that were not consistent with Payne's *et al.* (1971) hypothesis. More recently, in work focusing on Europe, neither Verspoor *et al.* (1999) nor Consuegra *et al.* (2002) found any evidence to support the Boreal-Celtic theory when exploring matrilineal lineages using RFLP haplotypes of mtDNA.

Nevertheless, there has been growing support for the polyphyletic origin of Atlantic salmon in Europe (Payne *et al.*, 1971; Child *et al.*, 1976; Koljonen *et al.*,

1999; Verspoor *et al.*, 1999; Nilsson *et al.*, 2001; Consuegra *et al.*, 2002; Asplund *et al.*, 2004). European wide mtDNA analysis revealed the British Isles to have the highest nucleotide diversity within the area covered, which led Consuegra *et al.* (2002) to suggest the British Isles may have been a contact zone for colonising salmon expanding from multiple refugia. Furthermore, on closer inspection of the literature, all cited papers in the Boreal-Celtic debate, except those of Payne *et al.* (1971) and Child *et al.* (1976), had very few samples from the British Isles (usually 3 with a maximum of 4). Allozyme studies of Bourke *et al.* (1997), Ståhl (1987) and Verspoor (1986b) are further bought into question by the revelation that a commonly used allozyme (ME-2, mMEP-2*) is likely to be acting under selection pressures, as there appears to be a correlation between allele frequency at mMEP-2* and temperature as well as a change in allele frequency with directional selection in the hatchery traits such as growth rate or late maturity (Verspoor and Jordan, 1989; Bourke *et al.*, 1997; Jordan *et al.*, 1997). As such it is not a suitable marker for phylogeographic studies as it violates the assumption of neutrality and therefore has implications for estimates of genetic divergence, such as F_{ST}. However, this may also be the case for transferrin (Verspoor and Jordan, 1989) as used by Payne *et al.* (1971) and Child *et al.* (1976) in the initial investigation into phylogeography of Atlantic salmon of the British Isles.

More recently attention has shifted away from a 'Celtic' race in southwest England, and an ice lake present in the southern North Sea has been proposed as a potential refuge (Verspoor *et al.*, 1999). Specifically, this area has been proposed as a second refugial location for salmon of the British Isles, supporting the polyphyletic origin of the region and explaining the high genetic diversity observed here. Furthermore, colonisation of the Baltic and Iceland from this refuge has been used to explain the similarities observed between these two distant regions (Verspoor *et al.*, 1999). In addition, links between the most northern populations of Europe in the White and Barents Seas with the Atlantic populations have been explained by colonising salmon moving northwards to these areas from the Atlantic, possibly from a refugium in the southern North Sea (Verspoor *et al.*, 1999; Tonteri *et al.*, 2005).

Evidently, the phylogeographic origins of Atlantic salmon in northwest Europe are far from concluded. Colonisation from a refugium in the Iberian

Peninsula is undisputed, yet the polyphyletic origin of salmon in this region seems likely. An Ice Lake situated in the present day southern North Sea is the most widely accepted second refugium from which Atlantic salmon colonised northwest Europe, though there is insufficient evidence to categorically rule out the possibility of another refugium being present north of the Iberian Peninsula, possibly in the southern British Isles. The competing hypotheses regarding the location of a second refugium remain largely speculative and a well-focused investigation centred on this region would be beneficial.

GENETIC DIFFERENTIATION WITHIN CATCHMENTS

It has long been appreciated that genetically distinct populations of Atlantic salmon occur in neighbouring catchments, which has largely been attributed to the precise homing ability of the species (Stabell, 1984). However, the occurrence of a number of differentiated populations inhabiting a single river system is now widely documented in numerous studies over the species range (Ståhl, 1987; Elo *et al.*, 1994; Beacham and Dempson, 1998; Garant *et al.*, 2000; Primmer *et al.*, 2006; Vaha *et al.*, 2007; Dillane *et al.*, 2008). It is thought that populations became differentiated in this way due to the discontinuous nature of spawning habitat, thus isolating groups of breeders, and the subsequent adaptations these groups made to their local environment in order to maximise survival and recruitment in that river or tributary (Youngson *et al.*, 1991). The growing realisation of within river structuring has prompted further investigations to determine the evolutionary mechanisms acting to drive and maintain this differentiation in order to manage within river stocks effectively (Garant *et al.*, 2000; Primmer *et al.*, 2006). Garant *et al.* (2000) and Primmer *et al.* (2006), both using microsatellite analysis, explored intra-river variation in relation to the 'member-vagrant' (Figure 1.8) and 'meta-population' (Figure 1.9) theories for population structuring. Both models are appropriate for Atlantic salmon populations inhabiting rivers with discontinuous stretches of available habitat.

The member-vagrant hypothesis assumes that physical (*e.g.* hydrodynamic) and biological (*e.g.* rheotaxis and homing behaviour) forces retain young fish and attract returning adults to their natal stretch of river, thus maintaining a stable population (Iles and Sinclair, 1982). These individuals are 'members' in this model and make up the majority of the population, whereas individuals that do not achieve

this are termed 'vagrants', and potentially contribute as migrants to other populations. Under this scenario, a high level of genetic differentiation is expected between populations inhabiting discontinuous stretches of suitable habitat within the river catchment. The populations would be temporally stable and an isolation-by-distance model of genetic relatedness should be observed as the vagrants are most likely to migrate and breed in other local populations (Figure 1.8).

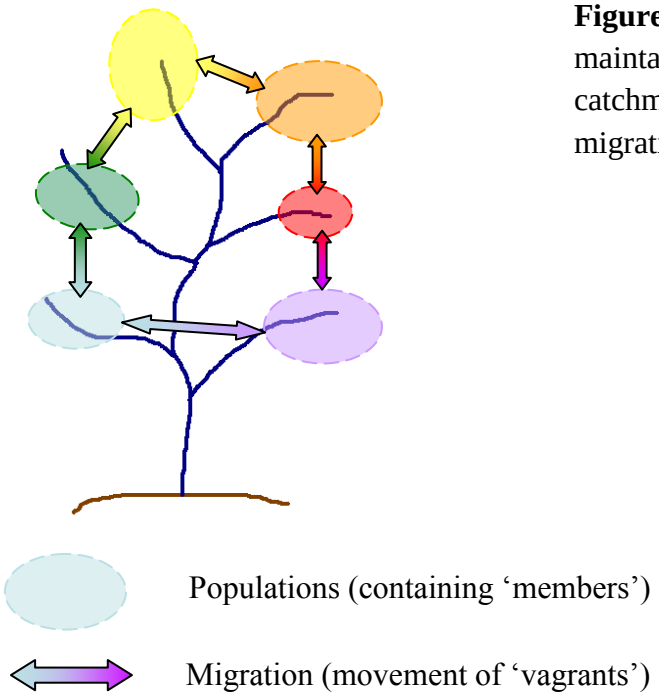


Figure 1.8 Member-Vagrant model maintaining population structuring within catchments, characterised by low migration between populations.

In contrast, although based on similar attributes of retaining young fish and returning adult fish, the meta-population theory assumes that these populations suffer intermittent local extinction events and subsequent re-colonisation from other populations (McQuinn, 1997; Rieman and Dunham, 2000). This can be a highly realistic scenario for Atlantic salmon populations as their riverine habitats may be prone to flash floods, drought, freeze/thawing and/or pollution events, to name a few determinants of extinction events, thus maintaining meta-population structuring within the catchment and precluding the progression onto a member-vagrant system,

as would likely be achieved in the absence of any extinction events. In catchments where meta-population models of population structuring prevail, vagrants, or migrants, are an intrinsic part of maintaining each population. Hence, in these situations, genetic divergence among populations is likely to be weak, yet still significant, implying a degree of straying/migration and relatively high geneflow. Similarly, in contrast with the member-vagrant model, no significant correlation between genetic and geographic distance would be expected (Figure 1.9).

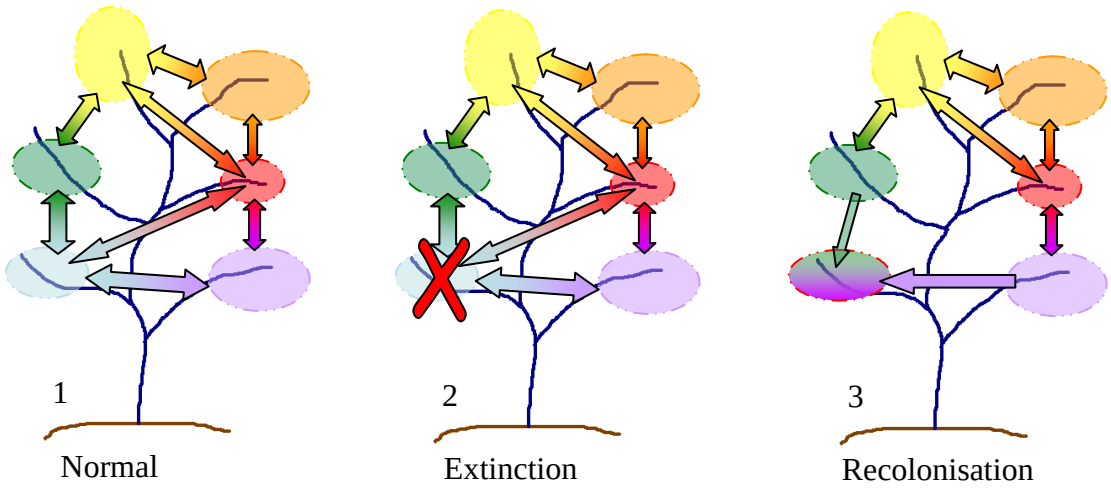


Figure 1.9 Meta-population model maintaining population structuring within catchments, characterised by extinction-recolonisation events and therefore increased migration.

To date, there have been two well focused investigations designed to clarify which evolutionary model best describes the genetic structuring of Atlantic populations observed within rivers. Garant *et al.* (2000) genotyped Atlantic salmon from seven locations in the Sainte-Marguerite River system (Canada) using five microsatellite loci. Sampling locations were separated by 20-55km of river. They found weak, but significant, genetic differentiation between samples but no isolation-by-distance, thus their findings were most consistent with the meta-population model. Conversely, Primmer *et al.*(2006) genotyped Atlantic salmon from 11 locations of the Varzuga River (Russia) with 17 microsatellite loci. On average,

sampling locations were situated 60 km apart (range: 5-165km) and a strong isolation-by-distance signal was revealed, leading the authors to conclude that the Atlantic salmon population of the Varzuga River best fitted the member-vagrant model.

The discrepancies between the outcome of these two investigations may be due to experimental design; Primmer *et al.* (2006) used 17 microsatellite loci to Garants *et al.* (2000) five, and their samples were taken over a much larger geographic scale. Both factors may have increased the likelihood of determining a member-vagrant model relative to Garants *et al.* (2000) study. Indeed, it was only when waterway distances increased to 69km or more between sampling locations that genetic differentiation between all populations became significant (Primmer *et al.*, 2006), whereas the maximum waterway distance in Garants *et al.* (2000) study was 55km. Thus, it is possible that the different outcomes of these investigations are due to sampling strategy and experimental design. However, it is also possible that different evolutionary mechanisms prevail in rivers of differing sizes as seems to be the case for the closely related brown trout (*Salmo trutta*) (Hansen *et al.*, 2002; Ostergaard *et al.*, 2003), and indeed the Sainte-Marguerite River has a total catchment area of approximately a quarter of the Varzuga River. In order to clarify this aspect of the population structuring of Atlantic salmon, an investigation centred on rivers of differing sizes, using the same molecular techniques is required.

CHARACTERISATION METHODS

Until the advent of molecular markers the resolution of phylogenies was predominantly attempted through the exploration and comparison of morphological variation, which has since been shown often to be environmentally rather than genetically driven. Thus, the widespread and increasing use of molecular markers over the last half century, allowing direct comparisons of individuals' genomes and quantification of genetic diversity, has revolutionised ecological and phylogenetic studies. Currently, there is a wide range of molecular markers from which to choose; the most appropriate candidates are discussed below.

Allozymes

Allozymes were the first mainstream molecular markers to have found widespread application. These enzyme variants differed in size and electrical charge and so could be separated by electrophoresis on starch gels and many early population genetic studies of Atlantic salmon were conducted using allozyme analysis (Ståhl, 1987; Verspoor *et al.*, 2005). However, by today's standards, allozymes are indirect and insensitive since they do not look directly at an individuals' DNA. Furthermore, a number of markers have been shown to be acting under selective pressures, such as the MEP locus in Atlantic salmon (Verspoor and Jordan, 1989). Subsequently, molecular markers which look directly at variation in DNA itself have been developed, thus circumventing the problems encountered with allozymes.

Restriction Fragment Length Polymorphisms (RFLPs)

Restriction enzymes cleave DNA after a specific sequence, producing an RFLP banding pattern for an individual. Changes in the DNA sequence due to indels, base substitutions or rearrangements involving the restriction site can result in the gain, loss or relocation of a restriction site, thus producing different banding patterns among individuals. A number of restriction enzymes are usually used in tandem to create composite haplotypes; the frequencies of which are compared between samples of interest. The potential power of RFLP markers in revealing genetic variation is relatively low compared to more recently developed markers. However, the RFLP of mitochondrial DNA (mtDNA) has found successful application in phylogeographic studies (Avise *et al.*, 1987). Mitochondrial DNA has several useful aspects for phylogeographic studies. It is considered to be selectively neutral and the effective population size of mtDNA is a quarter that of nuclear genes and so haplotype frequencies can drift rapidly, thereby creating genetic drift among populations in relatively short time periods. As mtDNA is maternally inherited there is just one ancestor in the previous generation and it does not normally undergo recombination. This allows the genetic relationships among haplotypes to be inferred and hence relative haplotype frequencies in different populations can be used to infer historical relationships between populations (Beebee and Rowe, 2004). Indeed, matrilineal phylogeographical histories of Atlantic salmon in the Baltic region have successfully been inferred using RFLP of mtDNA techniques (Nilsson *et al.*, 2001).

MINISATELLITES

Minisatellites can also involve the digestion of geneomic DNA with restriction enzymes. However, they differ from RFLP analysis as they consist of tandem repeats, 10-100 base pairs (bp) in length, which frequently show length polymorphisms to form repeat regions 100-2000bp in length, thus producing a bar-code-like banding pattern. These markers have found their niche in forensics and paternity testing, being termed 'DNA fingerprinting', but are inappropriate for population genetics studies as often the complex banding pattern precludes the assignment of alleles to a given locus, hence standard population genetic analysis cannot be applied (Schlotterer, 2004).

MICROSATELLITES

Microsatellites are a tandem repeat motif, similar to a minisatellite but much smaller; the repeat unit is typically no more than 6bp in length, repeated from a few to a few hundred times. This smaller size allows easy amplification using polymerase chain reaction (PCR) and, hence was the first widespread marker to take full advantage of the advent of PCR. The repetitive nature of microsatellites makes their replication accident prone, leading to changes in length. This type of mutation is called DNA slippage, the most common mutations being changes of a single repeat unit, following a stepwise mutation model (SMM) (Goldstein and Pollock, 1997). Mutation rate is estimated at 10^{-2} to 10^{-6} per locus per generation (Ellegren, 2000), which is several orders of magnitude greater than that of non-repetitive DNA (Chistiakov *et al.*, 2006). These different sized products are termed alleles, and the high number of different alleles possible provides high levels of polymorphism; thus microsatellites have found broad application in population biology. Microsatellites are inherited in a Mendelian fashion as co-dominant and are largely considered to be selectively neutral, though they can occasionally represent functionally relevant polymorphisms (Chistiakov *et al.*, 2006). In population biology they have become the marker of choice over the last few decades (Goldstein and Schlotterer, 2001) and have been widely used in the study of Atlantic salmon population structuring (Garant *et al.*, 2000; Primmer *et al.*, 2006; Vaha *et al.*, 2007; Dillane *et al.*, 2008).

RAPDS, ISSRS, IRAPS, AFLPS

These are a group of markers, which require no prior knowledge of primer sequences and use PCR amplification to yield multiple bands showing presence/absence

variation in individuals. They include: RAPDs (Randomly Amplified Polymorphic DNAs, with short PCR primers); ISSRs (Inter-Simple-Sequence-Repeats, with primers complementary to repetitive elements such as microsatellites); IRAPs (Inter-Retrotransposon Amplified Polymorphisms, with primers complementary to repetitive elements such as retrotransposons); and, AFLPs (Amplified Fragment Length Polymorphisms, which are restriction fragments amplified by adding linkers and subsequent selective amplification). However, often the reproducibility of these techniques, in particular for RAPDs, has been brought into question, and analysis is difficult, thus reducing their application in population genetic studies (Liu and Cordes, 2004; Schlotterer, 2004).

SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS)

Single nucleotide polymorphisms detect variation at a single nucleotide position in the genome. Rapidly detectable at a moderate cost using genechip technology, SNPs are the most abundant polymorphism in any organism and reveal hidden polymorphisms not detected with other markers. They are co-dominant and bi-allelic and mutate at a slower rate than microsatellites ($10^{-8} – 10^{-9}$), making the probability of homoplasy events less of a concern for SNPs than for microsatellites (Brumfield *et al.*, 2003). Furthermore, due to their biallelic nature, statistical analysis should be more straightforward for SNPs than for microsatellites; hence, the use of SNPs in population biology is increasing. However, SNP identification can be time consuming and costly (Brumfield *et al.*, 2003; Liu and Cordes, 2004). This thesis is principally hypotheses driven, hence, the methodological development of appropriate, reliable SNPs was not appropriate here under the given time and economic constraints. Nevertheless, with the ongoing GRASP project (Genetic Research on All Salmon Project; http://web.uvic.ca/grasp) and the ever improving technologies in this area, it is likely to be only a matter of time until SNP technology becomes widely and readily available for Atlantic salmon.

DNA SEQUENCING

DNA sequencing provides complete information for the region analysed, making it the most fine-scale genetic information possible. However, invariant sites are needlessly sequenced and, given the number of individuals required to be genotyped for population genetics, for this study DNA sequencing is an uneconomical option.

Overall Objectives of the Thesis

The overall objective of this thesis is to advance understanding of the genetic population structuring of Atlantic salmon in northwest Europe, the processes involved in creating and maintaining this phenomenon, the implications of certain human activities on the status of wild Atlantic salmon populations within rivers, and how information of this kind, *i.e*. derived from population genetics investigations, can be used to drive management and conservation initiatives at a local and regional scale.

To achieve this end a series of investigations will be undertaken to address the different aspects here raised. Specifically, these are as follows:

Unravelling the phylogeographic history of Atlantic salmon in northwest Europe (Chapter 3).

During the last ice age Atlantic salmon populations were displaced by the advancing ice sheets to refugial locations in the south before re-colonising the region following the retreat of the glaciers. Whilst the locations of glacial refugia and colonisation routes of Atlantic salmon for the Baltic region and Russia are relatively well understood, colonisation processes for northwest Europe are largely speculative with no clear consensus achieved in the literature to date. Furthermore, there has not been a focused investigation on this region since the 1970's (Payne *et al.*, 1971). Thus, an expansive investigation spanning from northern Spain and France, through the British Isles and Ireland will be undertaken.

Colonisation of this region from a refugial location in the Iberian Peninsula is widely accepted, but the polyphyletic origin of the region has repeatedly been proposed. Hence, the following general hypotheses will be address in this chapter:

H_0: Northwest Europe was colonised by Atlantic salmon expanding out of a single refugial location in the Iberian Peninsula;

H_1: Northwest Europe was colonised by Atlantic salmon expanding out of multiple refugial locations in the region (particular attention will focus on potential refugial locations in, and colonisation routes from, the southern North Sea and southwest England.).

CHOSEN MARKERS: MTDNA RFLP

For the phylogeographic analysis, RFLP of mtDNA will be undertaken. This marker has proved successful in determining phylogeographic lineages of Atlantic salmon, particularly in the Baltic region (Nilsson *et al.*, 2001), and comparisons with previously published material is relatively straightforward. However, despite the apparent advantages of using mtDNA for phylogeographic studies, there are some limitations and some of the underlying assumptions enabling these analyses have been brought into question. In the first instance, it must be remembered that the mitochondrial genome is approximately 0.00055% of the total genome, and RFLP analysis assess only a very small fraction of this. Furthermore, it is evident that selective pressures can act on mtDNA (Gemmell *et al*., 2004; Rand, 2001). Moreover, some species show clear recombination, whilst in other species recombination seems unlikely. Partial biparental inheritance has also been shown to occur in a number of taxa (see Ballard and Whitlock, 2004 and references therein). Potential recombination rates and the occurrence of biparental inheritance of mtDNA in Atlantic salmon are currently unknown, which makes it impossible to account for, but these inherent limitations must be acknowledged. Regardless, this technique has proved successful in prior studies (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001; Consuegra *et al.*, 2002), hence RFLP of mtDNA can still have a role in unravelling phylogeographic histories, therefore, this technique will be used in tandem with microsatellite analysis (see below) to try to unravel the phylogeographic histories of Atlantic salmon in northwest Europe.

WITHIN RIVER STRUCTURING OF ATLANTIC SALMON POPULATIONS (CHAPTERS 4 AND 6)

It is well documented that genetically differentiated populations can exist within river systems; however, it is unclear as to how these are maintained. Two theories prevail; the member-vagrant model is prevalent in more environmentally stable conditions and is characterised by high divergence among populations, limited geneflow and a strong isolation-by-distance signal, conversely the meta-population model is characterised by intermittent extinction and re-colonisation events resulting in weak but still significant genetic divergence, increased geneflow between populations and therefore no significant isolation-by-distance signal. In the two investigations focussed on this issue to date, one concluded that the member-vagrant

model was operating within the river system, while the other determined the meta-population model was more realistic. However, in the study which favoured the meta-population model, fewer molecular loci were used and the investigation was conducted over a much smaller geographic scale. This therefore begs the question: are these discrepancies due to experimental design, are they due to the size of the river catchment, or are other processes at work here?

This thesis aims to resolve this question by undertaking a thorough assessment of within river structuring of two catchments of varying sizes and natures, namely the rivers Tweed and Dart (see Chapter 2 for details). By performing identical analysis of these two rivers we should be able to determine whether the same evolutionary mechanisms are indeed acting within rivers of differing sizes, or whether catchment size or other factors determine the prevalent evolutionary model acting to maintain within river structuring. Knowledge of this kind, regarding evolutionary mechanisms operating within rivers of varying sizes can aid management at the regional and catchment level.

The specific hypotheses to be tested for each of the Tweed and Dart catchments in turn are as follows:

H_0: Atlantic salmon within the catchment form a single panmictic population, *i.e.* no significant genetic differentiation is observed between samples;

H_1: There are multiple populations of Atlantic salmon within the catchment, *i.e*. significant genetic differentiation between samples is observed;

H_{1A}: The samples/populations follow a meta-population evolutionary mechanism, *i.e*. the samples show weak, yet significant genetic differentiation, but there is no isolation-by-distance signal;

H_{1B}: The samples/populations follow a member-vagrant evolutionary mechanism, *i.e*. the samples show highly significant genetic differentiation there is a strong isolation-by-distance signal.

ARTIFICIAL PROPAGATION OF WILD STOCKS: DO THE STOCKED FISH PERSIST AND CONTRIBUTE TO OVERALL PRODUCTIVITY? (CHAPTER 4)

To abate the decline of Atlantic salmon stocks a common management practice has been to augment natural populations with fish from other rivers or with fish raised in hatcheries. Due to the source location of broodstock or the selection pressures acting within hatchery environments, stocked and native fish often differ in their genetic constitution. Given the highly structured nature of the species, the prospect of these differentiated stocks interacting and potentially interbreeding has raised concerns. Here we will investigate the long-term impacts of supplementation programmes using the river Dart as a case study. This river was stocked historically during the 1960s with fish from Scotland and Iceland. By genotyping these donor stocks alongside contemporary Dart populations we hope to determine if these fish and their progeny have persisted in the Dart for this time. This investigation assess the efficacy of these historical supplementation programmes. Based on the outcomes, modifications to the existing stocking policies may be possible.

The following hypotheses will be explored:

H_0: There has been no long-term impact of historical stocking activities on the contemporary Dart salmon populations, *i.e*. no genetic signal of the Scottish or Icelandic fish can be detected anywhere in the current river Dart populations;

H_1: There has been a continued impact of the historical activities on the contemporary Dart salmon populations, *i.e*. a genetic signal of the Scottish and/or Icelandic populations can be detected in the current Dart population;

H_{1A}: The continued impact of the historical stocking activities is exhibited uniformly across the catchment, *i.e.* a genetic signal of the Scottish and Icelandic fish is observed with equal intensity in all current Dart populations;

H_{1B}: The continued impact of the historical stocking activities is not uniformly exhibited across the catchment, *i.e*. a genetic signal of the Scottish and Icelandic fish is observed with unequal intensity in the current Dart populations.

NATURAL RE-COLONISATION OF PREVIOUSLY EXTIRPATED TRIBUTARIES (CHAPTER 7)

Atlantic salmon populations are experiencing severe declines across their species range. One of the principle causes of local extinctions within rivers is due to the construction of unsurpassable barriers to migrations, such as hydro-power developments. Unless an effective fish pass is installed, this wipes out all Atlantic salmon populations upstream of the developments for the period of their existence. In cases where historically constructed man-made barriers are no longer in use, a number of rivers have been re-opened to salmon runs. This has occurred in two major tributaries of the river Tweed and it is the aim here to assess the status of the re-colonised stocks, and ideally determine which stocks the tributaries were re-colonised from and if a fully self-sustaining and self-recruiting population has been reached since the tributary was re-opened.

On a between catchment scale it has been reported that fish straying from the most proximate neighbouring rivers have initially been involved in the re-colonisation process. Hence the initial hypotheses to test are:

H_0: The recently re-opened tributary was re-colonised by individuals straying from tributaries across the catchment;

H_1: The recently re-opened tributary was re-colonised by individuals straying from the most proximate tributaries.

H_0: The recently re-colonised tributary has not reached a stable state since it has been re-opened;

H_1: The recently re-colonised tributary has reached a stable state since it has been re-opened.

CHOSEN MARKERS: MICROSATELLITES

For this thesis, microsatellite markers are considered to be the most suitable marker to undertake the main body of genotyping in order to address the hypotheses set out regarding phylogeographic history of northwest Europe (Chapter 3) within river structuring (Chapters 4 and 6) the outcome of stocking practices (Chapter 5) and the re-colonisation of extirpated populations (Chapter 7). The high variability of these markers, coupled with the wealth of statistical packages with which to analyse the

data are fundamental to this choice. However, the limitations of the marker system must also be considered. Due to the high mutation rate of microsatellite regions it is possible that homoplasy events may occur unnoticed, hence, not revealing the true evolutionary history of the microsatellite. Unfortunately, this cannot be controlled for and so, while it must be acknowledged, there are no measures which can be readily undertaken to account for this. Another consideration for the analysis of microsatellite data is the complicated mutation process which microsatellites undergo. While the majority of mutations seem to involve a single repeat motif, mutations to alleles of much larger sizes have been observed (Goldstein and Pollock, 1997). Hence, an appreciation of the underlying mutation model used in each statistical package during analysis must be observed. These models usually follow either the stepwise mutation model (SMM), involving a single repeat motif, or the infinite allele model (IAM), which suggests that every mutation results in the creation of a new allele, though other models, such as the Two Phased Model (TPM) do exist. Nevertheless, successful application of microsatellite markers in a multitude of previous studies exploring population genetics (Garant *et al.*, 2000; Primmer *et al.*, 2006; Vaha *et al.*, 2007; Dillane *et al.*, 2008) installs confidence that the expressed hypotheses can be addressed using these techniques.

CHAPTER 2: METHODS

STUDY AREA CATCHMENT DETAILS

Two river catchments with differing natures will be used as study sites for work considered in this thesis, namely the river Dart and river Tweed.

RIVER DART

The river Dart (N50:20:35 W3:33:53) in southwest England drains the granite massif of Dartmoor National Park, 545m above sea level in a southerly direction, with a total catchment area of 475km^2 over 45 km. The upper tributaries fall into Dartmoor National Park, most of which is owned by the Duchy of Cornwall. The upper moorlands are very acid, with high rainfall compared to the lower lying area of the South Hams, where the Wash, Hems, Bidwell Brook and Am Brook originate. These lower reaches tend to have slightly poorer water quality compared to the moorlands and more incidences of physical degraded habitat.

The region is sparsely populated and there is very little major industry within the catchment aside from agriculture and tourism. Agricultural land use is predominantly grazing animals on the moorlands. Due to the extensive peat deposition on the moor, the river Dart is highly acidic, nevertheless, it still conforms to the RE1 Rivers Ecosystem Classification (water of good to very good quality suitable for all species), and as such is capable of providing valuable spawning grounds for salmonid fish. There are no lakes or reservoirs within the catchment, largely because Devon was un-glaciated during the last glacial period. Furthermore, there are no major barriers to migratory fish within the sampling area covered in this thesis; any man-made barriers have fish passes installed which are accessible at most flows (Environment Agency, 2003).

Densities of Atlantic salmon are higher in the moorlands compared to the lower reaches. Atlantic salmon in the Dart catchment, particularly the Spring salmon element, appear to be declining, as evidenced through rod and net catches (Figure 2.1). It could be argued that fishing effort has also decreased over this period, and indeed, Peress (1999; 2002) reported that exploitation from net fisheries had decreased from 45% to 9% and rod exploitation had decreased from 39% to 3% during the 10 year period from 1992 to 2002. Nevertheless, spawning targets and

conservation limits over this same time period, which define the numbers of spawners and eggs required to maintain the Dart populations, have not been met. Given the reduction of within river exploitation, low marine survival and/or low juvenile freshwater survival have been raised as potential causes for the continued decline and failure to meet the conservation limits (Environment Agency, 2003).

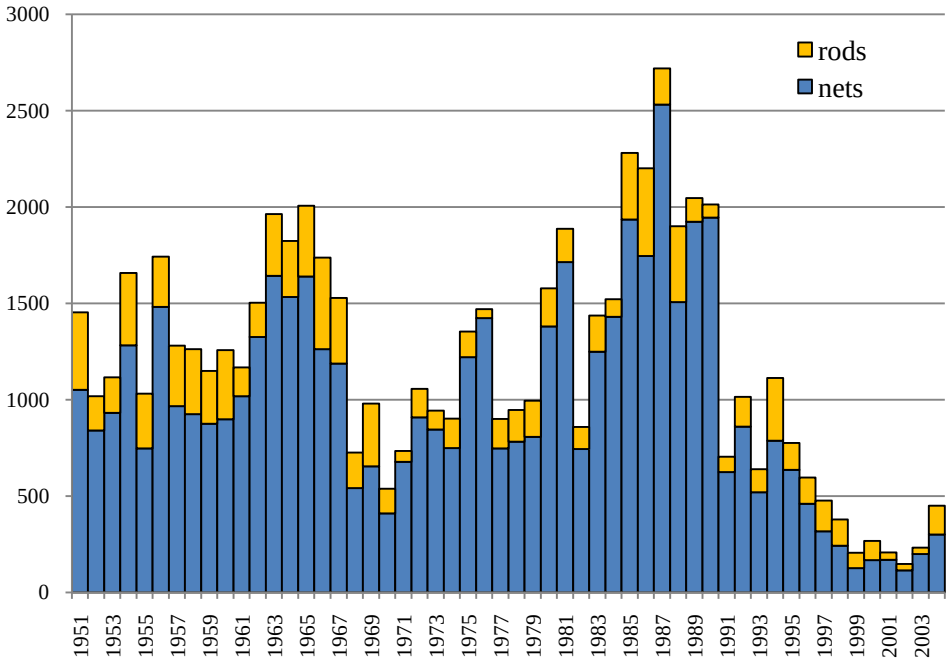


Figure 2.1 Net and rod catches for Atlantic salmon on the river Dart, 1951-2004. NB no effort data was available.

During the 1950s and 60s, as many as 300,000 eyed ova from hatcheries in Scotland or Iceland were transplanted into the river Dart each year. The rationale behind these stocking initiatives appears to be a desire to bring large fish to the river Dart for recreational sport fishing; such fish are commonly found in the Scottish and Icelandic source rivers. Our study of historical stocking records indicated the East Dart to be the only recipient tributary in the river Dart, and was duly prepared by the prior removal of all likely competitors *e.g*. brown trout, *Salmo trutta* L. and the native Atlantic salmon. Historical National River Authority reports infer that this practice occurred frequently over at least a 15 year period, but had certainly curtailed by the 1980s. However, supplementation of Atlantic salmon populations in the Dart

still occurs. Currently these practices involve the striping of broodstock taken from salmon returning to the Dart. The progeny are then raised in hatchery facilities within the catchment and planted out at the smolt stage into the main river channel.

River Tweed

The river Tweed (N55:45:50 W1:59:35) is the second largest river in Scotland with a catchment area of 5000km^2 over 160km of river. It straddles the English - Scottish borders and is bounded to the northwest by the Lammermoor and Moorfoot Hills and to the south by the Cheviot Hills, which form part of the Northumberland National Park. The steep valley sides of these hills give way to more open rolling hills in the east, with flatter, more fertile land. The majority of the river is designated a Site of Special Scientific Interest and a Special Area of Conservation under the EU Habitats Directive. There is very little industrial development within the Tweed catchment and few large settlements. Land use is predominantly agricultural, with some forestry and woodlands. The water quality of the river Tweed is deemed good to excellent under the Scottish Environment Protection Agency (SEPA) water quality guidelines and there is abundant habitat for adult spawning salmon and all freshwater life history stages. The catchment has 13 reservoirs and is an important source of water for the Borders and Edinburgh, but this is not thought to negatively impact upon the river (Tweed Foundation, 2005). Historically, there have been numerous barriers to migratory fish throughout the catchment, though fish passes are installed on most of these, allowing migration through most of the catchment (Tweed Foundation, 2005). No known supplementation or stocking programmes have been implemented within the Tweed catchment (Ronald Campbell, *personal communication*).

Salmon fishing here is worth £18m a year and 500 jobs. As many as 12,000 salmon are caught each year, with the majority returned to the river in line with the voluntary catch and release scheme in place locally. In contrast to the majority of salmon rivers, the salmon populations of the Tweed are relatively healthy, and not currently experiencing the common declines experienced in many other rivers (Figure 2.2).

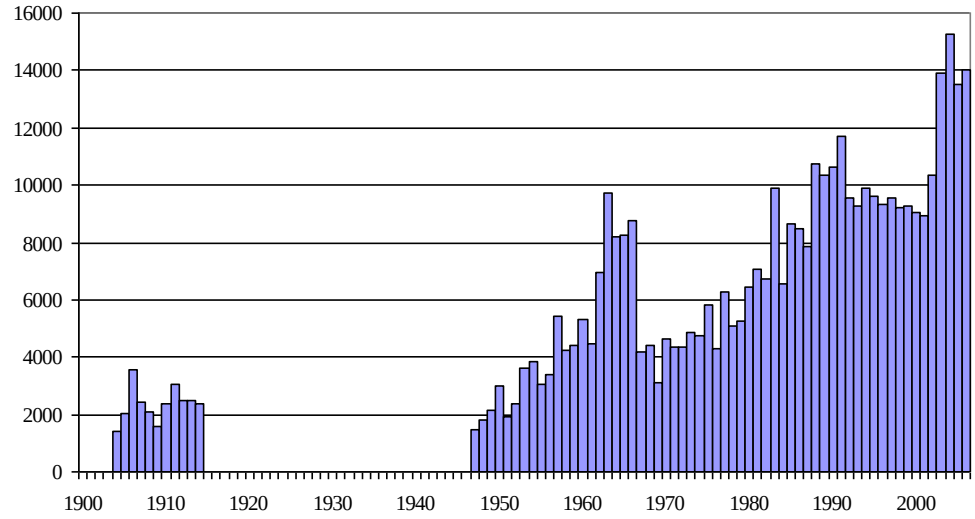


Figure 2.2 Rod catches of Atlantic salmon on the river Tweed, 1914-2006. NB no effort data was available.

However, there does appear to be a long-term decline of large fish over 13lbs, with a corresponding increase in the numbers of smaller fish. This may be a cyclical phenomenon. Historical records indicate that between 1887 and 1915 smaller fish were comparatively abundant, and then underwent a decline, with a corresponding rise in larger fish, until the 1950's. During the 1950s and 60s, there was a return to the situation found at the beginning at the century, with more small fish and fewer large fish. However, the decline of the larger fish has continued, such that there are now more small fish compared to large fish; a situation not previously observed in the Tweed records. The tropicalisation of exploited fish stocks, *i.e*. the reduction in the average size of fish and the onset of early maturation induced through the preferential exploitation of larger individuals has been noted previously (Stergiou, 2002). However, whether this shift to smaller individuals in the Tweed is in any way reflective of this process has not been investigated, but it is a possibility.

There is already some evidence to suggest that there is stock structure, possibly with a genetic component within the Tweed catchment (Campbell, 1995; Verspoor *et al.*, 1997). Scale reading analyses, used to identify returning salmon as either early running "spring" fish or later running "autumn" salmon was implemented. Results identified the Ettrick to be mostly populated by spring fish,

rather than autumn fish, with the Upper Tweed reaches indicating the reverse. These results were then supported in a radio-tracking study. In this study, returning salmon were tagged at the river mouth throughout the year and tracked through their upriver migration to their spawning grounds. Again, these results indicated that the majority of spring salmon, and a few early running grilse, returned to just two tributaries; the Whiteadder and Ettrick. A few were tracked to other tributaries but none returned to the Leader or Upper Tweed. In contrast, autumn fish, mainly grilse, had the reverse pattern of the spring fish, with large numbers entering the Teviot, Upper Tweed, Till and Gala, but few entering the Ettrick and none entering the Whiteadder. Hence, it appears that the Ettrick is largely populated by spring and summer multi sea winter (MSW) salmon, with the other tributaries mainly populated by 1SW, later running grilse (Campbell, 1995). Furthermore, a preliminary study undertaking genetic analysis on samples from the Ettrick, Gala and Upper Tweed tributaries determined that there was some genetic differentiation between salmon populations inhabiting these tributaries (Verspoor *et al.*, 1997).

SAMPLE COLLECTION

RIVER DART

Twelve locations throughout the river Dart were sampled in electrofishing surveys between May and September of 2005, 2006 and 2007; some sites were sampled in multiple years (Figure 2.3; Table 2.1; Figure 2.5; Table 2.3). Several of these were undertaken with the help of the Environment Agency, Devon, while the remainder were undertaken with the University of Exeter electric-fishing team. For all samples, adipose fin clips from parr in multiple year classes were taken and stored in 95% ethanol.

Figure 2.3 Sampling locations on the river Dart. Details follow Table 2.1.

Table 2.1 Sample details for the river Dart

MapID	Tributary	Site	Year	Number
1	Swincombe	Wydemeet	2007	26
2	Strane*	Whiteworks	2006*	30
3	West Dart*	Prince Hall	2006*	31
4	Blackbrook*	Blackbrook Bridge	2006*	27
5	West Dart	Crockern Tor	2007	34
6	Cherry Brook*	Higher Cherry Brook Bridge	2005*/06*/07	15/26/21
7	East Dart*	Postbridge	2005*/06*	45/40
8	Walla Brook*	Babeny	2006*	33
9	West Webburn	Lower Cator Bridge	2007	35
10	West Webburn*	Ponsworthy Bridge	2005*/06*/07	25/36/30
11	East Webburn	Cockingford Bridge	2007	31
12	Hems*	Tally Ho! Bridge	2006*	30

Used in Chaper 5*

RIVER TWEED

Sample collection took place during routine electric-fishing surveys carried out by the Tweed Foundation from 2004 to 2007. Fin clips from either the adipose or caudal fin were taken from Atlantic salmon parr and preserved in ethanol. These samples span the whole of the Tweed catchment, except the Whiteadder (Figure 2.4; Table 2.2; Figure 2.5; Table 2.3). A limited number of scale samples collected from the Gala Water in 1997, 2001 and 2003 were also available for analysis.

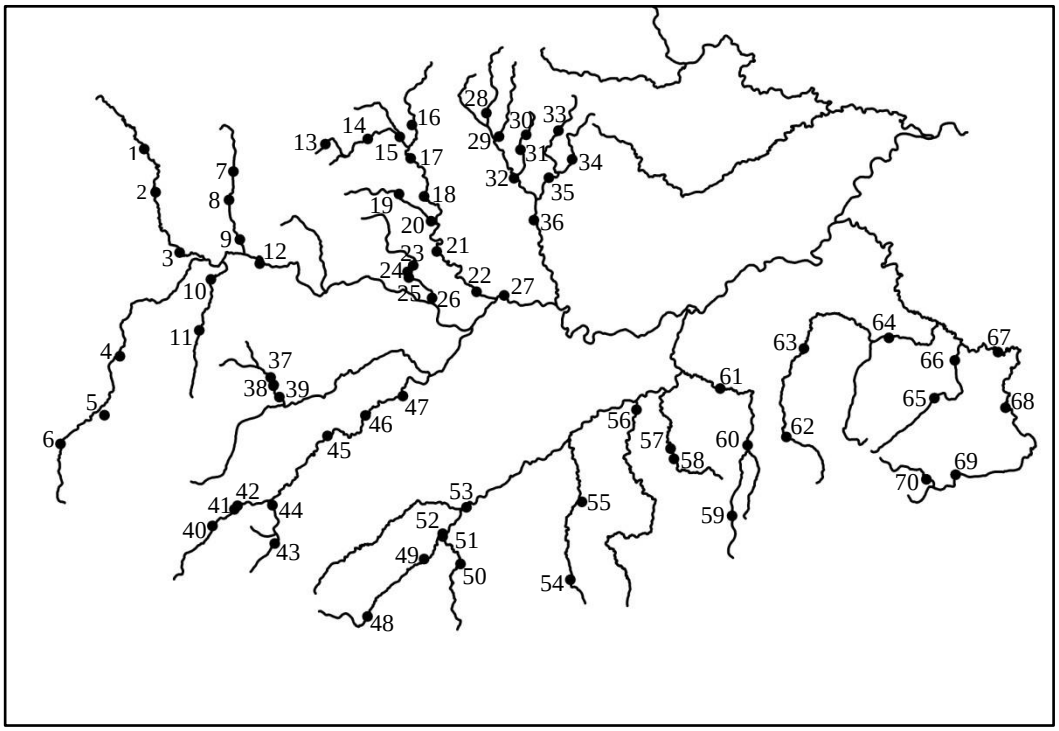


Figure 2.4 Sampling locations on river Tweed. Details follow Table 2.2.

Table 2.2 Sample details for the river Tweed

Map	Tributary	Year	Number	Map ID	Tributary	Year	Number
1	Lyn	2003	3	36	Leader	2003	8
2	Lyn	2003	2	37	Ettrick	2005	6
3	Lyn	2003/2004	3/15	38	Ettrick	2005	31
4	Upper	2003	8	39	Ettrick	2005	19
5	Upper	2003	7	40	Ettrick	2005	7
6	Upper	2004	9	41	Ettrick	2005	12
7	Eddleston	2003	6	42	Ettrick	2005	5
8	Eddleston	2003	3	43	Ettrick	2005	12
9	Eddleston	2003	1	44	Ettrick	2005	3
10	Manor	2003	21	45	Ettrick	2005	14
11	Manor	2003	19	46	Ettrick	2005	8
12	Upper	2003	5	47	Ettrick	2005	23
13	Gala	1997/2003	3/3	48	Teviot	2004	5
14	Gala	2001/2003	2/21	49	Teviot	2004/2007	15/7
15	Gala	1997/2001/2003	7/2/12	50	Teviot	2004	7
16	Gala	1997/2001/2003	2/4/4	51	Teviot	2004	1
17	Gala	2001/2003	3/25	52	Teviot	2004	10
18	Gala	1997/2003	3/6	53	Teviot	2005	10
19	Gala	1997/2001	4/1	54	Teviot	2004	5
20	Gala	1997/2001/2003	3/2/13	55	Teviot	2004	11
21	Gala	1997/2001/2003	3/3/5	56	Teviot	2004	8
22	Gala	1997/2001/2003	1/2/18	57	Teviot	2004	5
23	Caddon	2007	4	58	Teviot	2004	5
24	Caddon	2007	2	59	Teviot	2004	5
25	Caddon	2007	4	60	Teviot	2004	11
26	Caddon	2003/2007	7/22	61	Teviot	2004	4
27	Middle	2007	23	62	Till	2004	8
28	Leader	2003	6	63	Till	2004	3
29	Leader	2003	2	64	Till	2004	13
30	Leader	2003	4	65	Till	2004	3
31	Leader	2003	2	66	Till	2004	6
32	Leader	2003	5	67	Till	2004	4
33	Leader	2003	1	68	Till	2004	5
34	Leader	2003	10	69	Till	2004	8
35	Leader	2003	1	70	Till	2004	12

HISTORICAL SCALE SAMPLES

Historical scale samples from rivers involved in stocking programmes on the river Dart, namely the rivers Tweed, Tay in Scotand and the river Ellidaar in Iceland were located and collected (Figure 2.5; Table 2.3). The Freshwater Fisheries Laboratories in Montrose donated scale samples of returning smolts from 1968 from the rivers Tay and Tweed. The last running fish of the season were chosen to be genotyped since broodstock were typically taken at this time of year. The exact location of the Icelandic source population was not documented. However, the river Ellidaar was used in stocking of other European rivers during this period and so was deemed the most likely source of the Icelandic fish (Sigurdur Gudjonsson, *personal communication*). The Institute of Freshwater Fisheries in Iceland donated scale samples of returning salmon to the river Ellidaar from July in 1962 and 1988 for analysis.

Figure 2.5 Sampling locations for stocking analysis (Chapter 5). For details refer to Table 2.3.

Table 2.3 River locations for stocking analysis

MapID	River
1	Ellidaar
2	Tay
3	Tweed
4	Dart

PHYLOGEOGRAPHY

Samples collected as part of the ASAP project, undertaken jointly at the Universities of Exeter and Oviedo (Spain), from across northwest Europe were available for phylogeographic analysis, alongside within river samples collected for this thesis (Figure 2.6; Table 2.4). For details see Herreo *et al.* (2008).

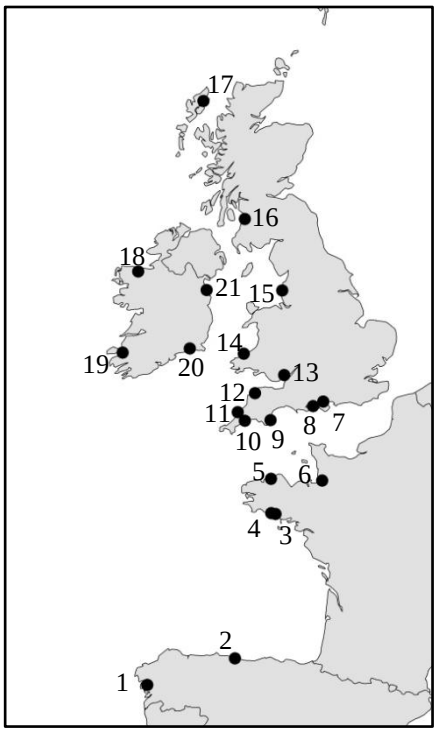


Figure 2.6 Samples available from the ASAP project for phylogeographic analysis. Details follow Table 2.4.

ID	River	N	ID	River	N
1	Ulla	40	12	Taw	32
2	Sella	41	13	Usk	30
3	Scourff	30	14	Teifi	27
4	Elle	46	15	Ribble	25
5	Leuger	28	16	Ayr	32
6	See	31	17	Creed	35
7	Itchen	29	18	Moy	29
8	Avon	23	19	Laune	47
9	Dart	38	20	Barrow	37
10	Fowey	36	21	Boyne	36
11	Camel	30			

Table 2.4 Samples available from the ASAP project for phylogeographic analysis.

METHODOLOGICAL DEVELOPMENT

DNA EXTRACTIONS

The initial stage for any genetic analysis is to first successfully extract DNA from the samples. The Chelex extraction method (Estoup *et al.*, 1996) is routinely used in the laboratory at Exeter by other researchers' also using microsatellite analysis to genotype Atlantic salmon populations from fin and scale samples.

CHELEX

In this protocol, 500µl of 10% Chelex is added to the sample and 6µl Proteinase K in a 1.5ml Eppendorf tube. The samples are heated at 55ºC for 1 hour in a waterbath and vortexed every 15 minutes. The samples are then transferred to a waterbath at 100ºC for 15 minutes.

The heating breaks open the cells, releasing the DNA. The Chelex binds to magnesium during the boiling phase to prevent DNA breakdown by DNAases. The extracted DNA is left in the supernatant and is ready for immediate use in PCR reactions with no need for dilution. Extractions can be stored in the fridge for a few days, or long term at -20ºC or -80ºC.

The Chelex extraction method is quick and simple, but the purity and long-term stability of the extracted DNA is thought to be inferior to other more robust extraction protocols. Some samples are 50-year-old scale samples, which will evidently have more degraded DNA than fresh tissue or scale samples. It was therefore decided to test the Chelex protocol against the more robust salt extraction method.

SALT EXTRACTION

In this protocol the sample is incubated with 250µl Digisol buffer (50mM Tris, 20M EDTA, 117mM NaCl and 1% SDS) and 10µl Proteinase K (10mg/ml) at 37ºC overnight in a rotary oven. Once digested, 300µl of 4M ammonium acetate is added, which precipitates the denatured protein. Vortexing takes place several times over a two to three hour period whilst samples are left at room temperature.

After centrifugation at 13,000rpm in a microcentrifuge, cell debris is pelleted and the clear liquid is aspirated into clean 1.5ml microcentrifuge tubes. 1ml of

ethanol precipitates the DNA and centrifugation and washing with 70% ethanol follows. Finally, samples are suspended in double distilled water.

In order to assess the relative success and reliability of these two DNA extraction methods, DNA was extracted from 5 fresh adipose fin clips and 5 historical scale samples using both the Chelex and salt extraction protocols. Since DNA extracted using the Chelex method cannot be visualised directly on agarose gel, following extractions the DNA was PCR amplified using 5Sa primers (Pendas *et al.*, 1995), before being run on a 2% agarose gel, stained with ethidium bromide (EtBr) for UV transillumination.

The salt extraction and Chelex extration protocols yielded similar results. Weaker products were initially observed for Chelex extractions with the historical scale samples, so the volume of Chelex added to each sample was reduced to 300µl. This increased the strength of product obtained. Thus, the Chelex extraction protocol was used for all subsequent DNA extractions as it is a much faster protocol to execute and had the additional advantage that other laboratory members commonly use it.

However, both of these protocols were unable to extract DNA of a high enough quality from the degraded historical Icelandic samples. For these samples, the phenol:chloroform (Taggart *et al.*, 1992) and the Qiagen DNEasy kits were tested.

PHENOL:CHLOROFORM

Tissue or scale samples are digested with 500µl of Digisol buffer (50mM Tris, 20mM EDTA, 117mM NaCl, 1% SDS) and 7µl proteinase K overnight in a rotary oven at 37ºC. 500µl of phenol:chloroform:isoamyl alcohol (25:24:1) is added and gently mixed for 10 minutes before being centrifuged at 12,000g for 3 minutes. This separates the liquid into two layers; a top aqueous layer, containing the digested DNA, and a bottom layer containing the phenol mixture. The top aqueous layer is transferred into a clean Eppendorf tube and the bottom layer discarded. The phenol:chloroform:isoamyl process is then repeated. Following this, ice cold 100% ethanol is added in approximately twice the volume to the DNA containing solution. This is mixed thoroughly, and then left on ice for 10 minutes to precipitate out the

DNA. The liquid is then centrifuged at 10,000g for 10 minutes then the ethanol is poured off leaving a small pellet of DNA. This is washed in ice cold 70% ethanol; the ethanol is poured off and the pellet left to dry. Once dry, the pellet is re-suspended in 100-200µl ddH_2O.

Both the phenol:chloroform and the Qiagen DNEasy method (undertaken according to the manufacturers manual) produced comparable results, hence the Qiagen DNEasy kits were used to extract DNA from the 1966 sample, while the phenol:chloroform method was used to extract DNA from the 1988 sample.

MICROSATELLITES

MULTIPLEX DESIGN

After consulting the wider literature and other laboratory members, it was decided that two multiplexes, containing as many primer sets as will fit together, would be the minimum required to answer the questions set out in this investigation.

A table of potential microsatellites detailing the published size ranges of their alleles, from relevant literature and from those markers used by other laboratory members was initially set up. Primer sets were graded according to their use in the Exeter laboratory and their use and success in the wider scientific community. In this way, two potential multiplexes for initial screening were designed (Table 2.5).

PRIMER SCREENING

Of the 14 microsatellites chosen for initial screening, only two were not routinely used in the Exeter laboratory. Despite this, all 14 microsatellites were optimised through varying the magnesium concentrations from 0.3mM to 2.5mM, and annealing temperatures from 50ºC to 64ºC. All microsatellites performed well at a magnesium concentration of 1.5mM, and so this is used in all PCR reactions. Optimum annealing temperatures are detailed in Table 2.5. The primer to be labelled was chosen by undertaking the same PCR amplification three times, once containing only forward primers, once containing only reverse primers and once containing both; the primer to be labelled is the one which produces the least stutter.

Table 2.5 Details of microsatellites used

Loci	Source	Multiplex	Annealing Temp	Allele Range
*Ssa*289	(McConnell *et al.*, 1995)	1	53˚C	104-146
Sssp2213	(Paterson *et al.*, 2004)	1	58˚C	142-250
Sssp2201	(Paterson *et al.*, 2004)	1	58˚C	236-372
*Ssa*171	(O'Reilly *et al.*, 1996)	1	58˚C	198-258
*Ssa*14	(McConnell *et al.*, 1995)	1	53˚C	135-155
SSOSL438	(Slettan *et al.*, 1996)	1	58˚C	90-142
SSOSL85	(Slettan *et al.*, 1996)	1	55˚C	175-227
SsaD157a	(King *et al.*, 2001)	1	58˚C	262-418
*Ssa*85	(O'Reilly *et al.*, 1996)	2	58˚C	107-177
Sssp1605	(Paterson *et al.*, 2004)	2	58˚C	209-269
*Ssa*412	(Cairney *et al.*, 2000)	2	58˚C	221-327
*Ssa*197	(O'Reilly *et al.*, 1996)	2	58˚C	127-297
SSOSL311	(Slettan *et al.*, 1996)	2	58˚C	116-190
*Ssa*202	(O'Reilly *et al.*, 1996)	2	58˚C	231-279

PCR REACTIONS

The microsatellite region is isolated using specific oligonucleotide priming sequences for the immediate region flanking either side of the microsatellite, and amplified in a PCR reaction.

The polymerase chain reaction (PCR) begins with a small amount of DNA and copies it exponentially to generate much larger quantities. Purified DNA is incubated with a thermostable DNA polymerase (*Taq*), the four deoxyribonucleotide substrates for DNA synthesis, priming oligonucleotides and a mixture of salts necessary to sustain polymerase activity.

PCR cycles include a denaturation step, in which the DNA strands are separated by heat; an annealing step, in which the primers bind to complementary sequences on each strand; and a synthesis step during which the polymerase copies each strand starting from the primer ends. The procedure is exponential and so 30-40 cycles are usually sufficient to generate ample product (Beebee and Rowe, 2004). Amplification of all microsatellites were undertaken in single 10µl reactions containing: 1xPCR buffer, 0.2mM dNTPs, 1.0µM of the unlabelled primer, 0.5µM of the fluorescently labelled primer and 0.5 units of *Taq* polymerase. The PCR cycle consisted of 94ºC for 3 minutes, followed by 35 cycles of 94ºC for 30 seconds,

specific annealing temperature (Table 2.5) for 30 seconds and 72ºC for 30 seconds, concluding with a final extension period of 72ºC for 10 minutes.

Sizing Alleles

To accurately size the alleles, one primer is labelled with a fluorescent dye (fluorophore), and the products are run on an automated sequencer system with an internal size standard.

A Beckman Coulter CEQ 8000 sequencer is used at Exeter University. This has an 8 capillary system, which allows products to be loaded in a 96 well plate system. The alleles are separated by size on a capillary gel. When the PCR product passes a laser at a fixed migration distance, the fluorophore is excited and the resulting flourescence measured automatically for comparison with an internal size standard which is run concurrently. A computer generated sequence readout is generated from which alleles can be sized.

The Beckman Coulter system has four dyes available to users. One dye is required to label the size standard, leaving three dyes that can be used in combination with differently sized microsatellites, allowing many primer sets to be pooled in a single lane of the gel (Table 2.5). The multiplexes in this thesis were designed to make optimum use of this system. The Beckman Coulter machine is sensitive to the amount of salts loaded. Hence, a maximum of ~6.5µl of PCR product can be loaded into each well. The relative amounts of PCR product for each microsatellite PCR product therefore needed to be adjusted in order for all microsatellites to be visualized correctly and simultaneously. This ranged from 0.5µl of the stronger, more stable products to 1.5µl of the weaker products, and usually needed to be adjusted for each run on the sequencing machine.

mtDNA RFLP

Composite haplotypes from RFLP of mtDNA are easily comparable between studies undertaken in different laboratories. Hence, in order to enable comparisons with published data the most commonly used restriction enzymes in published literature were chosen for use in this thesis. This involves the restriction enzymes *Hinf I, Rsa I, Hae III* and *Ava II* digesting the ND1 region of the control loop of mtDNA. Initially, the ND1 region was amplified in 30µl reactions using the primers designed by Hall

and Nawrocki (1995) and modified by Nilsson *et al.* (2001). The PCR cycle consisted of 95ºC for 3 minutes, followed by 35 cycles of 95ºC for 30 seconds, 58ºC for 45 seconds and 72ºC for 1 minute, concluding with a final extension period of 5 minutes. Restriction digests were undertaken in 10µl reactions containing 3.2µl H_2O, 0.3µl enzyme, 0.5µl corresponding buffer and 6µl of ND1 PCR product. These were digested overnight in a rotary oven at 37ºC. 8µl of digested product were run out on a 2% agarose gel stained with EtBr and subsequently visualised under ultraviolet light. Classification of banding patters follows Nielsen *et al.* (1996) (Figure 2.7), as does composite haplotype designation.

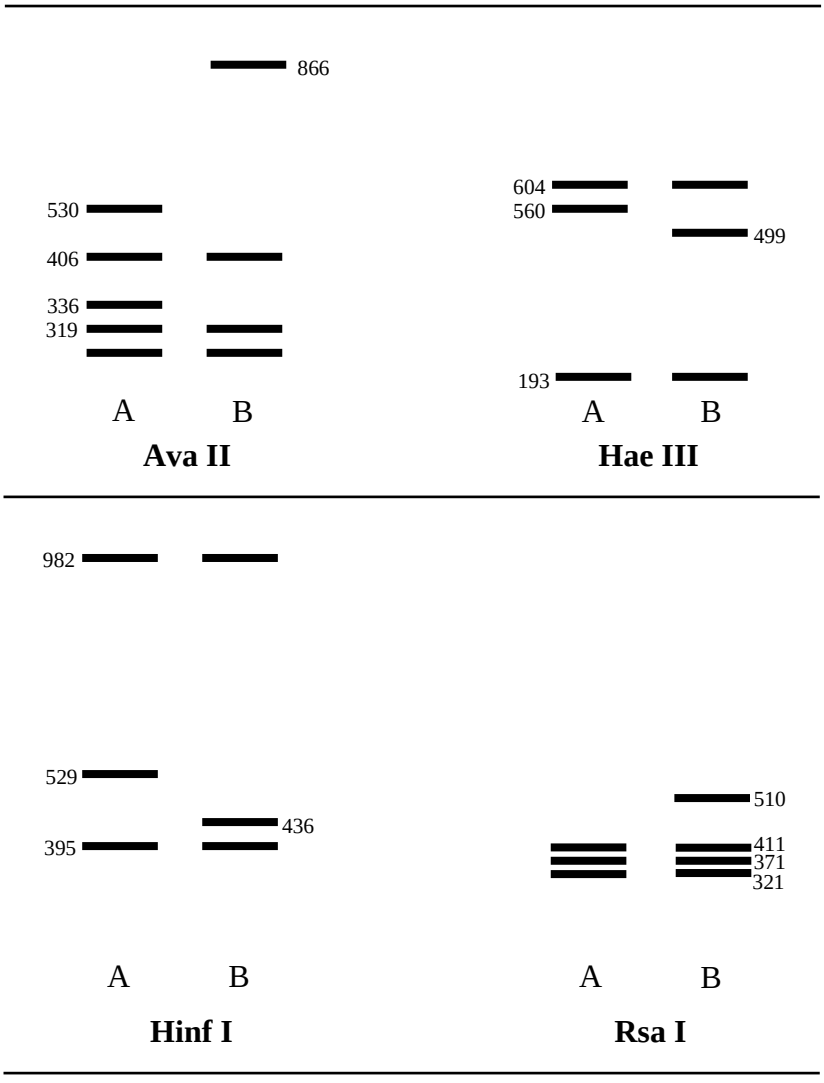


Figure 2.7 RFLP haplotypes of the ND1 region of mtDNA following (Nielsen *et al.*, 1996).

ASSESSING THE BEHAVIOUR AND POWER OF MOLECULAR MARKERS

All data was checked for abnormalities and the prescence of null alleles prior to any analysis using Microchecker (Van Oosterhout *et al*., 2004) and the Microsatellite Analysis Toolkit available in Excel.

LINKAGE AND SELECTION

Two of the basic assumptions in the use of microsatellite markers, and inherent in the majority of statistical analysis of microsatellite data, is that the markers used are neutral and unlinked. Hence, it is important to explore this for the markers used in the current study before the commencement of thorough statistical analysis. This initially involved searching the Genomic Research on Atlantic Salmon Project (GRASP) database on their website (http://web.uvic.ca/grasp/), alongside current literature, before undertaking some additional analysis on the data gathered in this thesis.

The physical linkage of the markers Sssp2213 and Sssp2201 on linkage group AS-5f, at a distance of between 59.7cM (Gilbey *et al.*, 2004) to 112.5cM (GRASP; http://web.uvic.ca/grasp/) has been observed. Also, the markers Sssp1605 and Ssa85 have been found to be located on different subgroups of linkage group As-18f (Sssp1605 is on subgroup AS-18f+2f), at a distance of 63cM (GRASP; http://web.uvic.ca/grasp/) . For each marker detailed in the GRASP project, a BLAST search is implemented across the medaka, zebrafish, stickleback and human genomes; Ssa85 had 112 contigs, many of which were associated with proteins, and SSa197 had two contigs with sequences of unknown function. If the region is conserved this may suggest, particularly for marker Ssa85, that the marker may be physically linked to areas under selection. In other studies, Ssa202 has been shown to be linked to the sex determining region (Gilbey et al 2003; Woram et al 2003), and Ssa14 and Ssa171 have been identified as potential outlier markers and, as such, may also be under the influence of selection (Vasemagi *et al.*, 2005b).

In order to investigate the behaviour of markers in this study, linkage dis-equilibrium tests were undertaken in GenePop *v3.4* (Raymond and Rousset, 1995) using the default settings. Of the 3003 tests undertaken, significant linkage dis-equlibrium was observed between 85 pairs of loci ($p<0.05$) after sequential

Bonferroni corrections were applied at the population level according to Rice (1989). Significant linkage dis-equlibrium between the loci Sssp2213 and Sssp2201 was observed in 5 populations; these loci are known to be physically linked (Gilbey *et al.*, 2004; GRASP; http://web.uvic.ca/grasp/). Of the other 80 significant results, 26 were only observed in one population, 18 were observed in two populations, and six were observed in three populations, hence, the only consistent significant result across populations appears to be the linkage of markers Sssp2213 and Sssp2201. Hence, locus Sssp2213 was removed from all analyses, except those undertaken in Chapter 5.

ASSESSING ANOMALOUS FINDINGS

*SSA*171

The microsatellite marker *Ssa*171 is a tetra-nucleotide with a range of 193-272bp. However, a small number of alleles (21 from the river Dart and 32 from the river Tweed) appeared to be di-nucleotides at loci across the whole range reported in this thesis (198-258bp). A subset of these individuals was re-genotyped and in all cases they were deemed true alleles, hence they were deemed true alleles for all individuals. This phenomenon of the occurrence of half alleles has also been seen in other studies using this microsatellite (Saisa *et al.*, 2005; Tonteri *et al.*, 2005).

SALMON-TROUT HYBRID TESTS

For a number of markers, the allele range reported in this thesis are outside the range previously reported, principally due to individuals from the river Dart which has not been genotyped before. In individuals where this was apparent, the salmon-trout hybrid test (Pendas *et al.*, 1995) was undertaken to ascertain if trout or salmon-trout hybrid individuals had inadvertently been genotyped. Of the 50 fish tested, eight were found to be trout and six were found to be salmon-trout hybrids; these 14 fish were therefore removed from all statistical analyses. The remaining 36 were Atlantic salmon.

CHAPTER 3: USE OF MULTIPLE MARKERS DEMONSTRATES CRYPTIC NORTHERN REFUGIA AND POSTGLACIAL COLONISATION ROUTES OF ATLANTIC SALMON (*SALMO SALAR* L.) IN NORTHWEST EUROPE.

Submitted to *Heredity* on 12th March 2009 with authors as follows:

A.K. Finnegan[†], A.M. Griffiths[†], G. Machado-Schiaffion[‡], E. Vazquez[‡], D. Bright[§], J.R. Stevens[†]

† University of Exeter, Hatherly Laboratories, Prince of Wales Road, Exeter, Devon, EX4 4PS.

‡Universidad de Oviedo, Facultad de Medicina, Dpto. Biología Funcional, Area de Genética, C/ Julián Clavería s/n, 33006 Oviedo, Spain.

§ WestCountry Rivers Trust, Rain-Charm House, Kyl Cober Parc, Stoke Climsland, Callington, Cornwall, PL17 8PH.

A.K. Finnegan undertook mtDNA genotyping, all statistical analysis and all drafts of this paper.

A.M. Griffiths undertook microsatellite and mtDNA genotyping as part of the ASAP project, provided guidance for statistical approaches and commented on earlier manuscripts.

G. Machado-Schiaffion undertook microsatellite genotyping as part of the ASAP project.

E. Vazquez and D. Bright are the principal project partners for the ASAP project.

J.R. Stevens provided guidance throughout and commented on earlier versions of the manuscript.

NB: All sample collection and microsatellite genotyping was undertaken as part of the Atlantic Salmon Arc Project (ASAP).

INTRODUCTION

Genetic differentiation within a species is governed by both contemporary and historical processes. The Pleistocene glaciations, which lasted approximately two million years and terminated around 10,000 years ago with a period of global warming (Andersen and Borns, 1994), are arguably one of the most important historical events involved in shaping the large-scale population structuring observed in many species today (Bernatchez and Wilson, 1998; Taberlet *et al.*, 1998). During the major glaciations of the Pleistocene, polar ice sheets expanded considerably, extirpating many populations and compressing species ranges to their southerly limits, thereby restricting populations to refugial locations (Hewitt, 1999). These populations remained relatively small during this period, which lasted for 1000s to 10,000s years. They were isolated and fragmented and became genetically differentiated primarily as the result of genetic drift. Following the onset of glacial retreat, populations at the northern edge of the glacial range would have expanded into areas of suitable habitat (Hewitt, 1999). Due to the genetic differentiation that occurred between refugial populations (due to drift and differential selection), it is possible through phylogeographic exploration to trace the movement of species from refugial areas into the previously glaciated regions which now constitute their native range.

For most European species, ancestry from the Pleistocene period can be traced back to one or more of the three main refugia in the Iberian, Italian or Balkan Peninsulas (Taberlet *et al.*, 1998; Hewitt, 1999). However, in the case of the Atlantic salmon, their current range only extends to the historical Iberian Peninsula refuge (MacCrimmon and Gots, 1979), which is widely accepted as a glacial refugium for Atlantic salmon during the Pleistocene glaciations (Consuegra *et al.*, 2002), but there is no evidence for their extension into the Italian or Baltic regions at any time.

Within Europe, Atlantic salmon appear to have differentiated into Atlantic and Baltic sub-lineages prior to the last glacial period (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001), hence it is expected that the two lineages would have differing phylogeographic histories. Extensive phylogeographic exploration in the Baltic region and around Russia to the Whites and Barents Seas has proved successful in elucidating likely glacial refugia for Atlantic salmon in freshwater lakes present in

the Baltic region and their subsequent colonisation routes across the region (Kazakov and Titov, 1991; Koljonen *et al.*, 1999; Nilsson *et al.*,2001; Asplund *et al.*, 2004; Makhrov *et al.*, 2005; Saisa *et al.*, 2005; Tonteri *et al.*, 2005; see Figure 1.7). In contrast, the phylogeographic history of Atlantic salmon in northwest Europe (northern Spain, France, Ireland and the UK) remains speculative (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001). Hence, the aim of this study was to elucidate the phylogeographic history of Atlantic salmon in this region.

In most Europe-wide studies conducted to date, samples of wild populations from the northwest, *i.e.* Spain, France, Britain and Ireland, have been largely lacking, particularly for France and Britain (a maximum of five samples from across the region were included in studies by Verspoor *et al.*, 1999 and Bourke *et al.*, 1997). In the only studies to directly focus on this region, Payne *et al.* (1971) and Child *et al.* (1976) described two lineages based on polymorphisms at the transferrin locus. The so-called 'Boreal' lineage radiated from a glacial lake in the southern North Sea and was proposed to occupy most of northern Europe, including the Baltic, whilst the 'Celtic' race was confined to the south west of England and southern Ireland. Much has been achieved in other areas of Europe since these early studies, and whilst no evidence for a Boreal-Celtic divide has since been observed (Ståhl, 1987; Blanco *et al.*, 1992; Bourke *et al.*, 1997; Verspoor *et al.*, 1999; Consuegra *et al.*, 2002), this region has been proposed as a contact zone for salmon expanding out of multiple refugium due to the high genetic diversity consistently reported for Britain and Ireland (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001; Consuegra *et al.*, 2002).

Sound evidence of a refuge for Atlantic salmon in this region comes in the form of bone relics recovered from a cave in the Iberian Peninsula, which demonstrates the presence of Atlantic salmon in this area for at least the last 40,000 years, *i.e.* during the last glacial period (Consuegra *et al.*, 2002). As proposed for other freshwater fish (Hanfling *et al.*, 2002), it is possible that Atlantic salmon existed north of this refuge, in northern France or southern England (Verspoor, 1986a), in line with Paynes *et al.* (1971) 'Celtic' hypothesis. Additionally, a glacial lake present in the southern North Sea has been proposed as a potential refuge for Atlantic salmon colonising west into the Atlantic and North Sea drainages, east into the Baltic (Koljonen *et al.*, 1999; Verspoor *et al.*, 1999; Saisa *et al.*, 2005) and north into the White and Barents Seas (Verspoor *et al.*, 1999; Asplund *et al.*, 2004; Tonteri

et al., 2005), in line with the 'Boreal' hypothesis of Payne *et al.* (1971). However, the allozyme studies of Ståhl (1987), Bourke *et al.* (1997) and Verspoor (1986a) have been bought into question by the revelation that a commonly used allozyme (ME-2, mMEP-2*) is likely to be acting under selection pressures (Verspoor and Jordan, 1989; Bourke *et al.*, 1997; Jordan *et al.*, 1997) and as such is not a suitable marker for phylogeographic studies. This may also be the case for transferrin (Verspoor and Jordan, 1989) as used in the Payne *et al.* (1971) and Child *et al.* (1976) initial investigations.

Evidently, the phylogeographic origins of Atlantic salmon in northwest Europe are far from resolved. Colonisation from a refuge in the Iberian Peninsula is widely accepted, yet a polyphyletic origin of salmon in this region seems probable. An Ice Lake situated in the present day southern North Sea is the most widely accepted second refuge from which Atlantic salmon colonised northwest Europe, though there is insufficient evidence to categorically rule out the possibility of other refugia being present north of the Iberian Peninsula, possibly in northern France or the southern British Isles, as seems to be the case for the bullhead, *Cottus gobio* (Hanfling *et al.*, 2002).

In recent years, the use of multiple classes of molecular markers has proved highly effective in elucidating phylogeographic histories of freshwater fish (Koskinen *et al.*, 2002; Hanfling *et al.*, 2002). Specifically for Atlantic salmon, the differing yet complimentary properties of microsatellite and PCR-RFLP of mtDNA have led to robust inferences of phylogeographic histories in the Baltic Seas and northern Europe (Saisa *et al.*, 2005; Tonteri *et al.*, 2005). The use of mitochondrial DNA in phylogeographic studies has a long history due to several useful properties (Avise *et al.*, 1987; Avise, 1994); it is considered to be selectively neutral (but see Gemmell *et al.*, 2004 and Rand, 2001); the effective population size of mtDNA is one quarter that of nuclear genes, hence, haplotype frequencies drift rapidly, therefore contributing to genetic differentiation in relatively short time periods. Moreover, as mtDNA is maternally inherited there is just one ancestor in the previous generation and, ordinarily, it does not undergo recombination (but see Ballard and Whitlock, 2004, and references therein). This allows the genetic relationships among haplotypes to be inferred and, hence, the relative haplotype frequencies in different populations can be used to infer historical relationships

among populations. Indeed PCR-RFLP of mtDNA has proved highly effective in helping to resolve the overarching phylogeographic history of Atlantic salmon across Europe (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001). Thus, by utilising the same techniques, Europe-wide comparisons can be drawn with this study.

More recently microsatellites have been advocated as a useful marker for phylogeographic studies, specifically to estimate the number of refugia that contributed to the colonisation of a region (Hanfling *et al.*, 2002; Tonteri *et al.*, 2005). This is possible due to the timing of the colonisation events, as microsatellite mutations are only expected to contribute to the genetic divergence of populations after 2500 generations which, coincidentally, coincides with the last glacial retreat around 10,000 years BP. Hence, microsatellite mutations should only have contributed to the divergence of populations which were colonised from different refugia, thus, this analysis can be useful for estimating the number of refugia involved in colonising a particular region (Hardy *et al.*, 2003; Tonteri *et al.*, 2005). Our aim therefore was to utilise these markers to resolve the phylogeographic history of Atlantic salmon in northwest Europe, by identifying refugial locations and determining colonisation pathways. Specifically, the null hypothesis (H_0) to be tested was that salmon colonising northwest Europe expanded out of a single refuge located in the Iberian Peninsula, with alternate hypotheses of salmon residing in additional glacial refugia located north of the Iberian peninsula (in northern France and/or southern Britain; H_{1A}) and/or in the southern North Sea (H_{1B}). Subsequently, we aim to reconstruct the colonisation pathways of Atlantic salmon into Britain and Ireland from these source location(s).

METHODS

SAMPLE COLLECTION

In this study 723 specimens of Atlantic salmon were collected as part of the ASAP project (see www.AtlanticSalmon.org.uk and Hurreo *et al.* 2008) from 21 rivers that drain into the east coast of the Atlantic Ocean, English Channel and Irish Sea (Figure 2.6; Table 2.4). The majority of sampling was carried out in 2004 and 2005 during routine juvenile salmon abundance surveys and targeted 1+ parr. For French specimens, scales from rod caught adult salmon were utilized.

DNA ANALYSIS: MICROSATELLITE

The genotyping of microsatellite loci was undertaken by Dr Andrew Griffiths of the University of Exeter (UK) and Gonzalo Machado-Schiaffion of the Universidad de Oviedo (Spain) as part of the ASAP project, as follows. Genomic DNA was extracted from scales or fin clips of Atlantic salmon according to a chelex resin protocol (Estoup *et al.*, 1996). Variation was determined at 12 microsatellite loci: *Ssa*157a, SsaD144b (King *et al.*, 2005), *Ssa*197, *Ssa*202 *Ssa*171 (O'Reilly *et al.*, 1996), SSsp2210, SSsp1605 SSsp 2201, SSsp G7 (Paterson *et al.*, 2004), *Ssa* 289 (McConnell *et al.*, 1995) and SSOSL 417, SSOSL85 (Slettan *et al.*, 1996). PCR amplifications were performed on reaction mixtures containing approximately 50 ng of extracted Atlantic salmon DNA template, 10 nM Tris-HCl pH 8.8, 1.5-2.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton x-100, 0.35 μM of fluorescently labeled primers, 0.5 Units of DNA *Taq* Polymerase (Promega, Madison, WI) and 250 μM of each dNTP in a final volume of 20 μL. Microsatellite analysis was carried out in two laboratories (see Horreo *et al.*, 2008 for further details); each used a different set of six loci to keep identification of alleles as consistent as possible, which also avoided the need for cross-laboratory standardisation. Size determination of labelled PCR products was performed using a Beckman-Coulter CEQ8000 automatic DNA sequencer with the associated fragment analysis software (Beckman-Coulter) at the University of Exeter and an ABI 3100 with GENEMAPPER V.3.5 (Applied Biosystems) software at the DNA Sequencing Unit of the University of Oviedo.

DNA ANALYSIS: MTDNA RFLP

The ND1 region of mtDNA was amplified using the primers given by Hall and Nawrocki (1995) and modified by Nilsson *et al.* (2001). The 30μL reaction volume

contained 0.25 units of *Taq* DNA polymerase (Bioline), 1X reaction buffer, 0.2mM dNTP, 2mM $MgCl_2$ and 0.5µM of each primer. PCR amplification consisted of an initial denaturation step of 95ºC for 3 mins, followed by 35 cycles of 95ºC for 30 secs, 58ºC for 45 secs and 72ºC for 60 secs, concluding with a final extension phase of 5 mins at 72ºC.

Amplified DNA was digested by the restriction enzymes *Hae III, Hinf I, Ava II* and *Rsa I* in individual reactions. 4µL of enzyme and recommended buffer (Promega) mix were added to 6µL of amplified DNA and incubated at 37ºC overnight. Fragments were separated by electrophoresis and variant fragment patterns, and thus composite haplotypes, were determined following Nielsen *et al.* (1996); see Figure 2.7.

MICROSATELLITE DATA ANALYSES

GENETIC DIVERSITY, HARDY-WEINBERG EQUILIBRIUM AND GENOTYPIC LINKAGE EQUILIBRIUM

All data was checked for errors and the presence of null alleles using Microchecker (Van Oosterhout *et al.*, 2004) and the Microsatellite Analysis Toolkit in Excel prior to any statistical analysis. Allele number and allelic richness (allele number corrected for sample size) were calculated for all loci within populations using FSTAT *v2.9.3* (Goudet, 1995). Deviations from Hardy Weinberg expectations across samples (within loci) and across loci (within samples) were estimated using the method of Guo and Thompson (1992) as implemented in Genepop *v3.4* (Raymond and Rousset, 1995), with sequential Bonferroni corrections applied for multiple tests across populations according to Rice (1989). The occurrence of linkage disequilibrium between loci was tested using the default parameters in Genepop *v3.4* with sequential Bonferroni corrections applied for multiple tests across populations.

GENETIC DIFFERENTIATION

FSTAT *v2.9.3* was used to calculate F_{ST} values between populations and their significance. Genetic distances between populations were estimated according to Nei's (1983) DA distance, from which a phylogenetic tree was constructed using the Neighbour Joining method as implemented in POWERMARKER *v3.25* (Liu and Muse, 2005). Strength of support for each node was assessed by undertaking 1000

bootstrap replicates in POWERMARKER, and a consensus tree was obtained using the CONSENSE programme in PHYLIP *v3.67* (Felsenstein, 1993).

INFERENCE OF REGIONAL POPULATION STRUCTURING

The optimal number of regional groups was estimated using STRUCTURE *v2.2* (Pritchard *et al.*, 2000). This Bayesian Markov Chain Monte Carlo (MCMC) based approach uses a clustering based algorithm to partition individuals into groups based on criteria to minimise Hardy Weinberg disequilibrium within populations and linkage disequilibrium between loci within populations. A burn-in period of 30,000 steps was set, followed by 1,000,000 MCMC replicates; allele frequencies were set as 'weakly correlated' as recommended by Pritchard *et al.* (2000). The number of groups to be simulated (K) was set as 1 to 10, with 10 replicates for each group. The optimal number of groups was estimated using the method of Evanno *et al.* (2005).

Hierarchical diversity of Atlantic salmon in northwest Europe was assessed in an AMOVA, as implemented in ARLEQUIN *v3.1* (Schneider *et al.*, 2000). Three hierarchical levels were defined as: (*i*) regional groups, (*ii*) populations within groups, (*iii*) within populations. Regional groups were defined based on the outcome of the STRUCTURE analysis and phylogenetic tree, which resolved five broad regions (Spain, France, UKSouth, UKSouthwest, UKNorthwest/Ireland; see Results). Differences in the average values of allelic richness, observed heterozygosity and expected heterozygosity among these regions were also assessed statistically using FSTAT *v2.9.3*.

An allele size permutation test (Hardy *et al.*, 2003), implemented in SPAGeDi *v2.1* (Hardy and Vekemans, 2002) was undertaken to assess the relative contribution of stepwise mutations in the genetic differentiation of populations within and between the five regions described above. If R_{ST} is significantly greater than F_{ST} this is an indication that the populations diverged a sufficiently long time ago for stepwise mutations of microsatellite regions of DNA to have contributed to this differentiation. Due to the mutation rate of microsatellites, this would have occurred more than 2500 salmon generations ago, approximating to 10,000 years ago, *i.e*. before the last glacial maximum. Hence, a significant result is indicative of the region being colonised by Atlantic salmon radiating from multiple refugia (Hardy *et al.*, 2003). Confidence intervals were calculated by undertaking 10,000

permutations of allele sizes among alleles within each locus. The locus *Ssa171* was excluded from this analysis as both di- and tetra-nucleotide repeats are observed at this locus; therefore it could not have mutated through a step-wise process alone.

PCR-RFLP MTDNA DATA ANALYSES

The sizes of restriction fragments, as estimated by Nielsen *et al.* (1996) were used to generate a restriction site binary matrix of presence/absence of fragment bands for each enzyme, and a binary representation of the composite haplotypes of each individual using REAP *v4.0* (McElroy *et al.*, 1992). These were used in subsequent programmes in the REAP *v4.0* package to calculate haplotype and nucleotide diversity within populations and a matrix of pairwise distances between populations (Nei, 1987) based on population haplotype frequencies. Subsequently, a UPGMA phylogenetic tree was constructed in PHYLIP *v3.67*. Mann-Whitney U-tests were used to statistically investigate the difference of haplotype and nucleotide diversity between the three broad geographic regions depicted in the phylogenetic tree (Spain, France, UK/Ireland; see Results). Hierarchical diversity was assessed with an AMOVA implemented in ARLEQUIN as for the microsatellite analysis (see above), except with the regions defined as Spain, France, Britain/Ireland (see above).

RESULTS

MICROSATELLITE DATA ANALYSES

GENETIC DIVERSITY, HARDY WEINBERG EQUILIBRIUM AND GENOTYPIC LINKAGE EQUILIBRIUM

A total of 289 alleles were observed across the 12 loci here used, ranging from six (*Ssa*289) to 33 (Sssp2201; SsaD144). Within populations, the number of alleles ranged from 2 (*Ssa*289; Moy) to 27 (SaaD144; Laune), the average number of alleles ranged from 7.00 (Avon) to 15.56 (Barrow) and allelic richness ranged from 7.000 (Avon) to 13.384 (Teifi) (Table 3.1; 3.2). After sequential Bonferroni corrections were applied, significant deviations from Hardy Weinberg expectations were observed on 11 occasions, involving 9 populations and 6 loci; Sella (Sssp1605), Elle (Sssp2201, g7), Leuger (*Ssa*197, g7), See (*Ssa*417), Camel (*Ssa*202), Ribble (g7), Ayr (*Ssa*202), Creed (*Ssa*202) and Barrow (g7) (Table 3.1). Of the 1386 tests of linkage disequlibrium between loci, only 26 (*i.e.* <2%) were found to be significant after sequential Bonferroni corrections were applied at the population level. Of these 26, 20 were single observations only occurring in one population and three loci pairs were found to be significant in 2 populations (144 and 157, 171 and 197, 144 and 289). Hence, linkage disequilibrium of loci was deemed to be negligible.

Table 3.1 Genetic diversity indices for microsatellite data, mtDNA RFLP data. mtDNA RFLP haplotype frequencies.

			Microsatellite diversity			mtDNA RFLP diversity		mtDNA RFLP haplotypes			
River	**ID**	**N**	**A_r**	**H_o**	**H_e**	**Haplotype Diversity**	**Nucleotide diversity**	**AAAA**	**AABA**	**BBBB**	**BBBA**
Ulla	1	40	8.43	0.70	0.72	0.000	0.00000	0.00	0.00	1.00	0.00
Sella	2	41	11.87	0.81	0.81	0.337	0.00883	0.05	0.00	0.80	0.15
Scourff	3	30	11.26	0.80	0.79	0.646	0.01774	0.50	0.27	0.00	0.23
Elle	4	46	11.64	0.82	0.78	0.684	0.01949	0.22	0.48	0.09	0.22
Leuger	5	28	12.62	0.81	0.79	0.577	0.01531	0.61	0.25	0.04	0.11
See	6	31	11.12	0.81	0.80	0.475	0.01368	0.71	0.16	0.03	0.10
Itchen	7	29	7.79	0.74	0.75	0.616	0.01336	0.00	0.14	0.34	0.52
Avon	8	23	7.00	0.73	0.81	0.557	0.02841	0.61	0.04	0.30	0.04
Dart	9	38	10.96	0.83	0.88	0.495	0.01016	0.03	0.03	0.66	0.29
Fowey	10	36	11.48	0.81	0.82	0.708	0.02697	0.19	0.31	0.42	0.08
Camel	11	30	12.68	0.83	0.84	0.690	0.02170	0.03	0.43	0.30	0.23
Taw	12	32	12.57	0.84	0.87	0.692	0.02186	0.13	0.13	0.47	0.28
Usk	13	30	12.58	0.84	0.86	0.632	0.01869	0.13	0.10	0.20	0.57
Teifi	14	27	13.38	0.84	0.85	0.732	0.02600	0.22	0.11	0.41	0.26
Ribble	15	25	11.29	0.82	0.84	0.530	0.01498	0.04	0.36	0.00	0.60
Ayr	16	32	11.36	0.83	0.83	0.627	0.02054	0.00	0.34	0.50	0.16
Creed	17	35	11.90	0.83	0.83	0.570	0.01402	0.00	0.14	0.60	0.26
Moy	18	29	11.74	0.82	0.85	0.680	0.02004	0.14	0.07	0.34	0.45
Laune	19	47	12.11	0.83	0.82	0.531	0.01144	0.02	0.09	0.26	0.64
Barrow	20	37	13.33	0.86	0.86	0.739	0.02261	0.14	0.22	0.27	0.38
Boyne	21	36	11.57	0.82	0.80	0.386	0.00526	0.00	0.00	0.75	0.25

ID: Population ID for Figure 2.6; N: Number of salmon; A_r: Allelic richness; H_o: Observed heterozygosity; H_e: Expected heterozygosity; mtDNA RFLP composite haplotypes derived from restriction enzymes *Ava II*, *Hae III*, *Hinf I*, *Rsa I*, according to Nilsson *et al*. (2001).

Table 3.2 Microsatellite diversity indices by population.

		Sssp2201	Ssa85	SsaD144	Ssa289	Ssa171	Ssa157	Ssa197	Ssa202	Sssp1605	Sssp2210	Ssa417	g7
Ulla	**Allele No.**	15	9	17	5	10	14	11	7	7	7	6	10
	A_R	12.566	8.033	14.479	3.972	8.422	12.181	8.664	6.144	6.136	6.714	5.077	8.796
	H_E	0.863	0.610	0.893	0.489	0.592	0.863	0.755	0.709	0.628	0.666	0.540	0.732
	H_O	0.875	0.600	0.925	0.550	0.600	0.825	0.800	0.750	0.625	0.700	0.625	0.750
	HWE	0.947	0.548	0.443	0.093	0.253	0.540	0.737	0.370	0.056	0.311	0.285	0.094
Sella	**Allele No.**	20	13	23	5	12	22	15	7	7	11	18	16
	A_R	18.168	10.311	19.203	4.817	10.161	18.324	12.150	6.662	6.208	9.383	15.218	11.832
	H_E	0.923	0.799	0.918	0.647	0.812	0.911	0.850	0.724	0.646	0.805	0.872	0.786
	H_O	0.900	0.775	0.925	0.600	0.875	0.923	0.878	0.756	0.610	0.829	0.805	0.780
	HWE	0.223	0.188	0.427	0.906	0.407	0.929	0.634	0.276	**0.000**	0.336	0.048	0.185
Scourff	**Allele No.**	18	11	23	4	13	18	13	9	8	7	13	11
	A_R	17.088	9.937	19.567	3.937	11.582	16.636	11.711	8.480	7.866	6.480	12.197	9.643
	H_E	0.911	0.806	0.895	0.588	0.844	0.893	0.823	0.809	0.771	0.681	0.841	0.695
	H_O	0.903	0.871	0.903	0.613	0.839	0.806	0.774	0.839	0.871	0.548	0.871	0.581
	HWE	0.274	0.741	0.528	0.342	0.445	0.072	0.117	0.575	0.959	0.077	0.522	0.005
Elle	**Allele No.**	22	11	23	4	14	22	14	12	10	11	17	11
	A_R	17.826	9.740	17.626	3.919	11.175	17.767	11.502	9.495	9.187	9.012	13.583	8.809
	H_E	0.917	0.857	0.919	0.512	0.854	0.921	0.839	0.808	0.768	0.747	0.887	0.748
	H_O	0.980	0.840	0.940	0.440	0.780	1.000	0.900	0.800	0.720	0.560	0.760	0.620
	HWE	0.446	0.116	0.375	0.071	0.035	0.931	0.593	0.627	0.062	**0.000**	0.005	**0.003**
Leuger	**Allele No.**	21	12	24	4	14	17	12	11	8	13	14	10
	A_R	19.428	11.554	21.956	3.996	13.164	16.097	11.343	10.724	7.792	12.318	13.670	9.451
	H_E	0.897	0.861	0.927	0.472	0.854	0.882	0.808	0.833	0.773	0.815	0.886	0.720
	H_O	0.964	0.714	0.964	0.500	0.821	0.893	0.607	0.893	0.786	0.929	0.857	0.500
	HWE	0.537	0.246	0.671	0.457	0.170	0.373	**0.002**	0.812	0.090	0.645	0.438	**0.000**
See	**Allele No.**	19	14	21	4	11	14	9	11	10	8	13	12
	A_R	16.880	12.416	19.165	4.000	9.853	12.958	8.467	10.612	9.265	7.413	11.754	10.684
	H_E	0.904	0.847	0.921	0.655	0.795	0.873	0.768	0.867	0.747	0.709	0.844	0.808
	H_O	0.906	0.844	0.906	0.625	0.813	0.813	0.900	0.938	0.844	0.625	0.719	0.625
	HWE	0.356	0.562	0.763	0.660	0.935	0.035	0.072	0.850	0.150	0.341	**0.001**	0.007

		Sssp2201	Ssa85	SsaD144	Ssa289	Ssa171	Ssa157	Ssa197	Ssa202	Sssp1605	Sssp2210	Ssa417	g7
Itchen	**Allele No.**	14	5	11	4	5	12	9	4	7	10	8	11
	A_R	12.689	5.000	10.531	3.993	4.793	10.926	8.545	3.793	6.712	8.924	7.880	9.718
	H_E	0.787	0.743	0.826	0.508	0.723	0.842	0.819	0.655	0.663	0.681	0.800	0.769
	H_O	0.862	0.931	0.931	0.448	0.724	0.862	0.828	0.690	0.586	0.655	0.655	0.862
	HWE	0.797	0.271	0.080	0.005	0.773	0.508	0.869	0.955	0.110	0.137	0.074	0.263
Avon	**Allele No.**	13	5	11	4	6	9	10	5	4	5	5	7
	A_R	13.000	5.000	11.000	4.000	6.000	9.000	10.000	5.000	4.000	5.000	5.000	7.000
	H_E	0.850	0.756	0.850	0.554	0.708	0.812	0.808	0.674	0.626	0.593	0.748	0.777
	H_O	0.957	0.913	1.000	0.565	0.783	0.826	0.913	0.739	0.652	0.609	0.913	0.826
	HWE	0.587	0.837	0.177	0.746	0.664	0.163	0.440	0.022	0.096	0.364	0.299	0.065
Dart	**Allele No.**	17	11	17	4	10	18	15	10	7	11	13	15
	A_R	13.821	10.890	15.033	3.590	8.784	15.581	13.026	8.820	6.949	9.430	11.742	13.888
	H_E	0.885	0.885	0.896	0.633	0.814	0.899	0.844	0.789	0.801	0.791	0.870	0.895
	H_O	0.949	0.949	0.974	0.641	0.769	0.974	0.795	0.923	0.821	0.949	0.923	0.846
	HWE	0.050	0.502	0.601	0.483	0.133	0.181	0.030	0.412	0.416	0.315	0.512	0.119
Fowey	**Allele No.**	21	13	20	4	12	17	16	9	8	11	11	13
	A_R	17.565	11.517	16.885	3.956	10.662	15.609	13.865	8.591	7.589	9.731	10.614	11.229
	H_E	0.900	0.810	0.885	0.535	0.865	0.903	0.846	0.827	0.776	0.749	0.880	0.797
	H_O	0.889	0.833	0.889	0.556	0.889	0.917	0.778	0.833	0.750	0.806	0.833	0.833
	HWE	0.039	0.850	0.248	1.000	0.450	0.171	0.109	0.375	0.155	0.247	0.027	0.695
Camel	**Allele No.**	22	11	18	4	13	19	16	13	7	13	12	15
	A_R	20.022	10.143	16.882	4.000	12.358	17.928	15.287	12.264	6.884	11.362	11.510	13.492
	H_E	0.920	0.779	0.912	0.628	0.875	0.891	0.891	0.838	0.766	0.761	0.882	0.851
	H_O	1.000	0.733	0.900	0.667	0.909	0.867	0.867	0.733	0.769	0.767	0.967	0.933
	HWE	0.221	0.102	0.037	0.794	0.551	0.678	0.678	**0.002**	0.320	0.122	0.417	0.058
Taw	**Allele No.**	18	9	25	4	12	20	17	10	9	10	15	14
	A_R	16.557	8.356	22.464	3.999	11.133	18.231	15.195	9.955	8.341	9.264	13.882	13.459
	H_E	0.882	0.791	0.930	0.633	0.865	0.907	0.893	0.867	0.755	0.774	0.884	0.870
	H_O	0.938	0.750	0.938	0.750	0.875	0.968	0.906	0.936	0.750	0.844	0.938	0.875
	HWE	0.805	0.057	0.115	0.871	0.240	0.602	0.019	0.871	0.430	0.944	0.344	0.571

		Sssp2201	Ssa85	SsaD144	Ssa289	Ssa171	Ssa157	Ssa197	Ssa202	Sssp1605	Sssp2210	Ssa417	g7
Usk	**Allele No.**	17	15	20	4	12	20	15	10	8	13	11	17
	A_R	15.874	13.756	18.833	3.989	11.516	18.364	13.705	9.469	7.762	11.548	10.196	15.947
	H_E	0.900	0.875	0.923	0.623	0.881	0.913	0.861	0.842	0.810	0.804	0.812	0.874
	H_O	0.967	0.967	0.967	0.467	0.933	0.867	0.833	0.867	0.900	0.700	0.867	0.933
	HWE	0.997	0.887	0.367	0.036	0.745	0.028	0.760	0.879	0.326	0.071	0.629	0.136
Teifi	**Allele No.**	21	14	20	4	13	17	15	11	10	12	15	17
	A_R	19.871	13.215	19.239	3.980	12.403	16.189	14.052	10.388	9.555	11.495	14.067	16.156
	H_E	0.916	0.855	0.908	0.616	0.877	0.884	0.857	0.821	0.819	0.821	0.864	0.877
	H_O	0.926	0.963	0.815	0.704	1.000	0.889	0.815	0.815	0.815	0.778	0.778	0.852
	HWE	0.560	0.220	0.008	0.229	0.943	0.123	0.122	0.805	0.203	0.104	0.093	0.104
Ribble	**Allele No.**	21	9	14	4	13	18	15	10	9	12	11	11
	A_R	18.764	8.694	12.807	3.997	11.868	16.613	13.795	9.674	7.968	11.014	10.402	9.901
	H_E	0.903	0.807	0.861	0.600	0.845	0.905	0.877	0.855	0.753	0.763	0.854	0.820
	H_O	0.936	0.742	0.871	0.677	0.871	1.000	0.936	0.936	0.807	0.742	0.935	0.581
	HWE	0.865	0.348	0.784	0.617	0.862	0.745	0.795	0.024	0.454	0.428	0.240	**0.002**
Ayr	**Allele No.**	18	11	19	4	11	18	16	10	9	10	11	14
	A_R	15.724	10.171	15.722	3.999	9.892	16.057	13.813	9.631	8.242	9.110	10.807	13.131
	H_E	0.898	0.826	0.873	0.622	0.824	0.901	0.868	0.842	0.786	0.789	0.847	0.881
	H_O	0.912	0.882	0.882	0.706	0.853	0.882	0.912	0.706	0.735	0.735	0.882	0.882
	HWE	0.526	0.266	0.951	0.914	0.157	0.192	0.407	**0.002**	0.044	0.012	0.310	0.731
Creed	**Allele No.**	18	10	24	4	12	17	21	9	8	11	13	15
	A_R	16.423	9.088	19.972	3.873	10.956	15.022	17.743	8.651	7.107	10.381	11.053	12.489
	H_E	0.911	0.818	0.916	0.663	0.844	0.892	0.910	0.851	0.747	0.772	0.843	0.834
	H_O	1.000	0.861	0.889	0.722	0.833	0.917	0.861	0.714	0.778	0.833	0.806	0.750
	HWE	0.530	0.764	0.299	0.620	0.191	0.305	0.022	**0.003**	0.713	0.904	0.260	0.010
Moy	**Allele No.**	18	13	16	2	10	19	15	12	7	11	14	13
	A_R	16.625	12.257	15.055	2.000	9.642	17.749	14.249	11.252	6.986	10.390	13.336	11.323
	H_E	0.906	0.860	0.897	0.491	0.798	0.915	0.887	0.840	0.802	0.821	0.868	0.795
	H_O	0.968	0.871	0.903	0.548	0.871	0.935	0.759	0.966	0.897	0.742	0.933	0.806
	HWE	0.960	0.920	0.044	0.725	0.884	0.218	0.030	0.665	0.649	0.066	0.041	0.192

		Sssp2201	Ssa85	SsaD144	Ssa289	Ssa171	Ssa157	Ssa197	Ssa202	Sssp1605	Sssp2210	Ssa417	g7
Laune	**Allele No.**	21	12	27	4	15	24	15	10	9	10	14	17
	A_R	17.621	9.724	20.777	3.969	11.695	18.903	12.618	9.095	7.869	8.114	11.722	13.238
	H_E	0.918	0.792	0.932	0.673	0.854	0.923	0.883	0.845	0.744	0.695	0.867	0.853
	H_O	0.957	0.745	0.936	0.638	0.936	0.936	0.872	0.787	0.681	0.660	0.787	0.809
	HWE	0.236	0.330	0.674	0.455	0.171	0.832	0.525	0.155	0.013	0.267	0.164	0.099
Barrow	**Allele No.**	21	13	20	4	20	18	18	14	9	11	18	22
	A_R	18.106	11.349	16.153	4.000	17.232	14.596	15.819	12.391	7.848	9.779	15.603	17.111
	H_E	0.923	0.847	0.901	0.697	0.903	0.889	0.906	0.882	0.775	0.778	0.899	0.884
	H_O	0.975	0.900	0.950	0.600	0.950	0.897	0.925	0.795	0.795	0.775	0.925	0.821
	HWE	0.792	0.812	0.443	0.104	0.878	0.977	0.628	0.007	0.635	0.066	0.368	**0.001**
Boyne	**Allele No.**	19	15	19	4	12	17	16	10	7	10	14	12
	A_R	16.841	12.172	17.303	3.956	11.390	15.260	14.520	9.294	6.623	9.076	11.449	10.999
	H_E	0.906	0.816	0.903	0.625	0.867	0.897	0.879	0.784	0.789	0.772	0.750	0.865
	H_O	0.972	0.722	0.917	0.583	0.917	0.917	0.806	0.639	0.806	0.667	0.778	0.833
	HWE	0.980	0.130	0.635	0.354	0.929	0.057	0.013	0.031	0.327	0.320	0.510	0.809
Total	**Allele No.**	33	25	33	6	28	32	28	18	11	17	29	29
	A_R	19.716	12.521	21.320	4.281	13.551	18.584	15.012	10.366	8.603	11.164	14.737	13.996
	H_E	33.000	25.000	33.000	6.000	28.000	32.000	28.000	18.000	11.000	17.000	29.000	29.000
	H_O	0.950	0.871	0.955	0.658	0.905	0.944	0.904	0.874	0.807	0.815	0.915	0.881
	HWE	0.940	0.823	0.921	0.598	0.849	0.907	0.844	0.811	0.758	0.734	0.831	0.774

Given are allele number (Allele N.), Allelic richness (Ar), Expected heterozygosity (He), Observed heterozygosity (Ho) and Hardy-Weinberg expectations (HWE; significant deviations indicated in bold).

Table 3.3 Genetic divergence of populations.

	Avon	Ayr	Barrow	Boyne	Camel	Creed	Dart	Elle	Fowey	Itchen	Laune	Leuger	Moy	Ribble	Scourff	See	Sella	Taw	Teifi	Ulla	Usk
Avon	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ayr	0.089	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Barrow	0.073	0.021	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Boyne	0.105	0.035	0.021	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Camel	0.074	0.033	0.017	0.037	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.035	0.000	0.000
Creed	0.071	0.027	0.015	0.032	0.020	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Dart	0.096	0.042	0.026	0.042	0.022	0.033	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Elle	0.073	0.056	0.036	0.059	0.025	0.045	0.040	-	0.000	0.000	0.000	**0.477**	0.000	0.000	**0.494**	0.000	0.000	0.000	0.000	0.000	0.000
Fowey	0.068	0.041	0.028	0.042	0.021	0.031	0.028	0.024	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Itchen	0.037	0.088	0.079	0.097	0.073	0.077	0.090	0.076	0.077	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Laune	0.070	0.032	0.014	0.028	0.011	0.019	0.027	0.027	0.027	0.064	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Leuger	0.076	0.056	0.031	0.052	0.021	0.035	0.037	**0.001**	0.021	0.080	0.025	-	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000
Moy	0.076	0.023	0.017	0.040	0.023	0.022	0.037	0.043	0.036	0.074	0.022	0.039	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ribble	0.081	0.031	0.015	0.034	0.024	0.027	0.034	0.046	0.033	0.088	0.025	0.043	0.029	-	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Scourff	0.079	0.056	0.036	0.061	0.030	0.044	0.048	**0.001**	0.029	0.079	0.030	0.006	0.048	0.051	-	0.000	0.000	0.000	0.000	0.000	0.000
See	0.069	0.046	0.028	0.050	0.020	0.033	0.043	0.022	0.032	0.074	0.026	0.020	0.040	0.037	0.020	-	0.000	0.000	0.000	0.000	0.000
Sella	0.075	0.051	0.037	0.054	0.033	0.043	0.052	0.048	0.045	0.090	0.037	0.044	0.044	0.049	0.048	0.037	-	0.000	0.000	0.000	0.000
Taw	0.074	0.024	0.021	0.033	0.007	0.024	0.020	0.033	0.014	0.070	0.017	0.029	0.027	0.020	0.037	0.023	0.039	-	0.000	0.000	0.000
Teifi	0.073	0.024	0.014	0.037	0.006	0.020	0.020	0.020	0.015	0.074	0.015	0.020	0.015	0.026	0.029	0.026	0.036	0.010	-	0.000	0.003
Ulla	0.136	0.118	0.100	0.123	0.107	0.112	0.116	0.124	0.114	0.156	0.102	0.117	0.110	0.115	0.133	0.114	0.056	0.111	0.104	-	0.000
Usk	0.070	0.027	0.015	0.036	0.015	0.024	0.020	0.022	0.020	0.075	0.016	0.020	0.020	0.023	0.025	0.026	0.040	0.015	0.009	0.106	-

Pairwise F_{ST} estimates (below diagonal); associated *p* values (above diagonal); Bold indicates non-significant values at p=0.05.

GENETIC DIFFERENTIATION

Global F_{ST} was estimated at 0.0436, with pairwise estimates ranging from 0.001 (Elle-Leuger; Elle-Scourff) to 0.156 (Ulla-Itchen). Although generally low, all were significant at $p<0.05$ except the two lowest values involving Elle-Leuger and Elle-Scourff (Table 3.3). These three populations are all in northwest France; the rivers Elle and Scourff share the same river mouth and comprised samples of adults ascending the rivers. The Neighbour Joining (NJ) population tree constructed from Nei's DA distance clearly depicted five broad groups of populations reflecting geographical regions (Figure 3.1). Bootstrap analysis indicated very strong support for the group of Spanish populations (96.2%) and reasonably strong support for the group of French populations (75.2%), with increased bootstrap support at branches within this group (94.4% and 95.4%). Within Britain and Ireland, the clustering of populations into three geographic regions was clearly depicted; specifically UKSouth, UKSouthwest and UKNorthwest/Ireland. Bootstrap support is very strong for the two UKSouth populations branching together (100%), but less so for the grouping of the UKSouthwest (39.4%) and UKNorthwest/Ireland (26.5%) populations, though within these groups there is stronger support at some branches (Figure 3.1).

INFERENCE OF REGIONAL POPULATION STRUCTURING

The clustering analysis, as implemented in STRUCTURE, defined an optimum of 5 regions (*i.e.* k=5), corresponding well with geographic origins (Figure 3.2). Complementing the NJ population tree, individuals clustered into the five regions of Spain, France, UKSouth, UKSouthwest and UKNorthwest/Ireland. Similarly, Spain, France and UKSouth all formed very distinct clusters, whilst the UKSouthwest and UKNorthwest clusters were less distinct and showed considerable overlap.

Hierarchical analysis demonstrated that 3% of variation lay among these 5 regional groups, 2.2% lay among populations within these regions, and the remaining 94.8% lay within populations. Genetic diversity, expressed as average allelic richness and heterozygosity was lowest in the UKSouth regional group; this was significantly lower ($p<0.05$) compared to all other regional groups, except Spain (which was close to being significant; Ar: $p=0.057$; He: $p=0.038$) (Table 3.3).

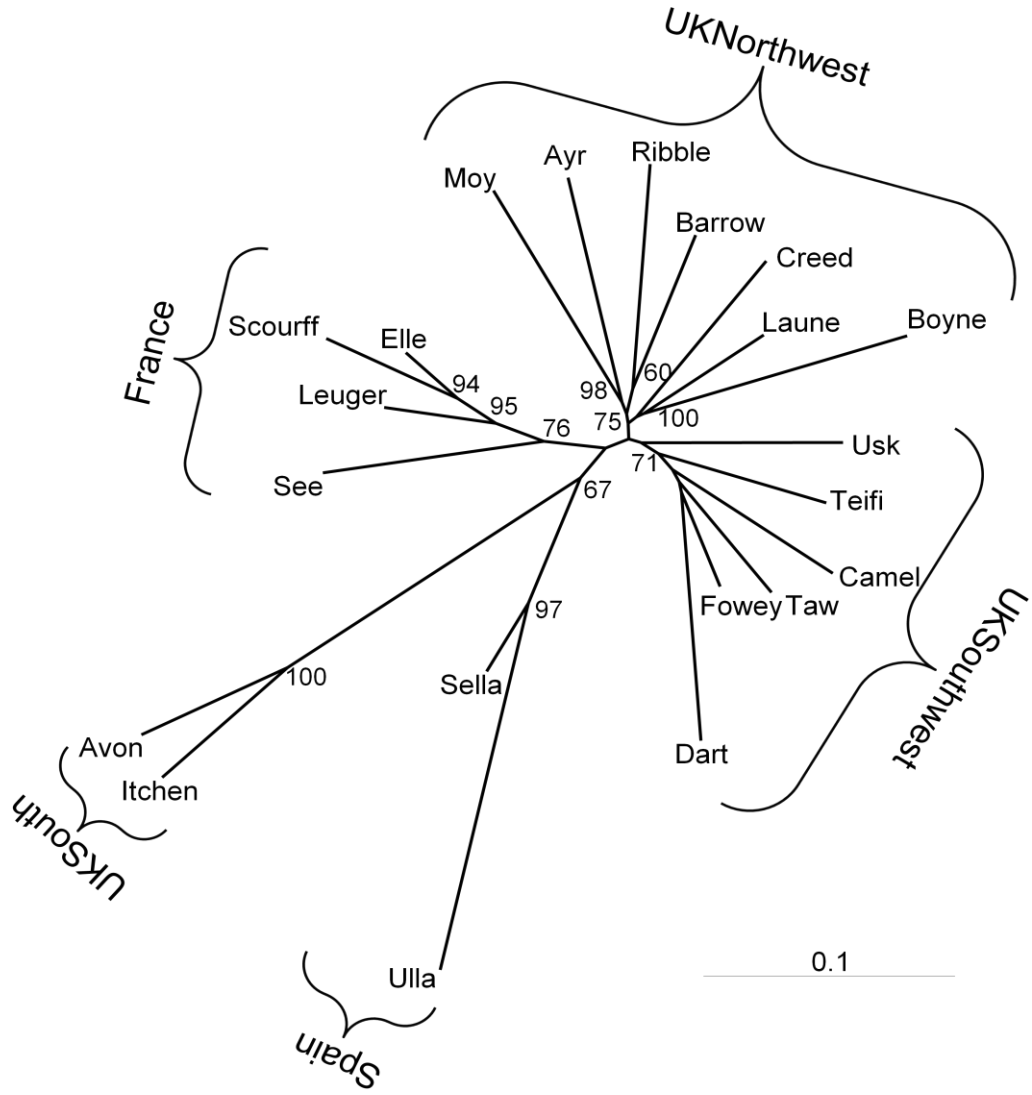


Fig. 3.1 Neighbour Joining phylogenetic tree generated from microsatellite data. Scale depicts the number of nucleotide substitutions per site.

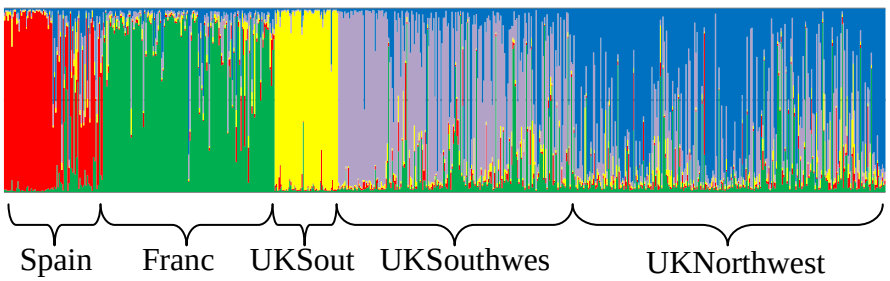


Figure 3.2 STRUCTURE output depicting 5 regional groups. Each bar represents one individual

Table 3.4 Regional genetic diversity indices

	Microsatellite diversity		mtDNA RFLP diversity	
	A_r	H_e	Haplotype diversity	Nucleotide diversity
Spain	10.151	0.771	0.166	0.0044
France	11.661	0.834	0.586	0.0166
UKSouth	7.396	0.767	0.575	0.0209
UKSouthwest	12.276	0.86	0.647	0.0209
UKNorthwest	11.9	0.856	0.571	0.0156

A_r: Allelic richness; H_o: Observed heterozygosity; H_e: Expected heterozygosity.

The allele size permutation test revealed that, on a large scale, multilocus R_{ST} values were significantly higher than F_{ST} values (F_{ST}=0.0436, R_{ST} =0.0635, p<0.01; Figure 3.3). This demonstrates that stepwise mutations of microsatellite repeat regions have contributed to the genetic differentiation of the five broad regions and thus is consistent with the hypothesis that northwest Europe was colonised by Atlantic salmon from different refugia. To further test the hypothesis of distinct refugia being present in Spain (Iberian Peninsula) and northern France, these two regions were compared and a significant contribution of stepwise mutations to population differentiation was revealed (F_{ST} =0.0511, R_{ST} =0.0985, p<0.05), indicating that these two regions were colonised from, or acted as, different refugia. To test the competing hypotheses of colonisation of the UK and Ireland, populations from the three clusters comprising this region (UKSouth, UKSouthwest, UKNorthwest/Ireland) were compared. No significant contribution of stepwise mutations to genetic differentiation was observed; this indicates that the three regions were colonised from the same refugial population(s). However, when the British and Irish populations were then compared to the potential source/refugial populations of Spain and France, both comparisons were found to be significant (France *vs.* UK: F_{ST} =0.0352, R_{ST} =0.0457, p<0.05; Spain *vs.* UK: F_{ST} =0.0448, R_{ST} =0.0613, p<0.05). This could be interpreted in two ways; it could suggest that there was a third refuge in southern England, from which salmon colonised the whole of the UK and Ireland, alternatively this could be evidence that the UK and Ireland was a contact zone for Atlantic salmon colonising out of the different refugia in Spain

and France, the test being significant because the UK and Ireland populations are comprised of salmon descended from both refugia. Viewed in light of the mtDNA RFLP analysis, the second interpretation seems the most likely (See results below).

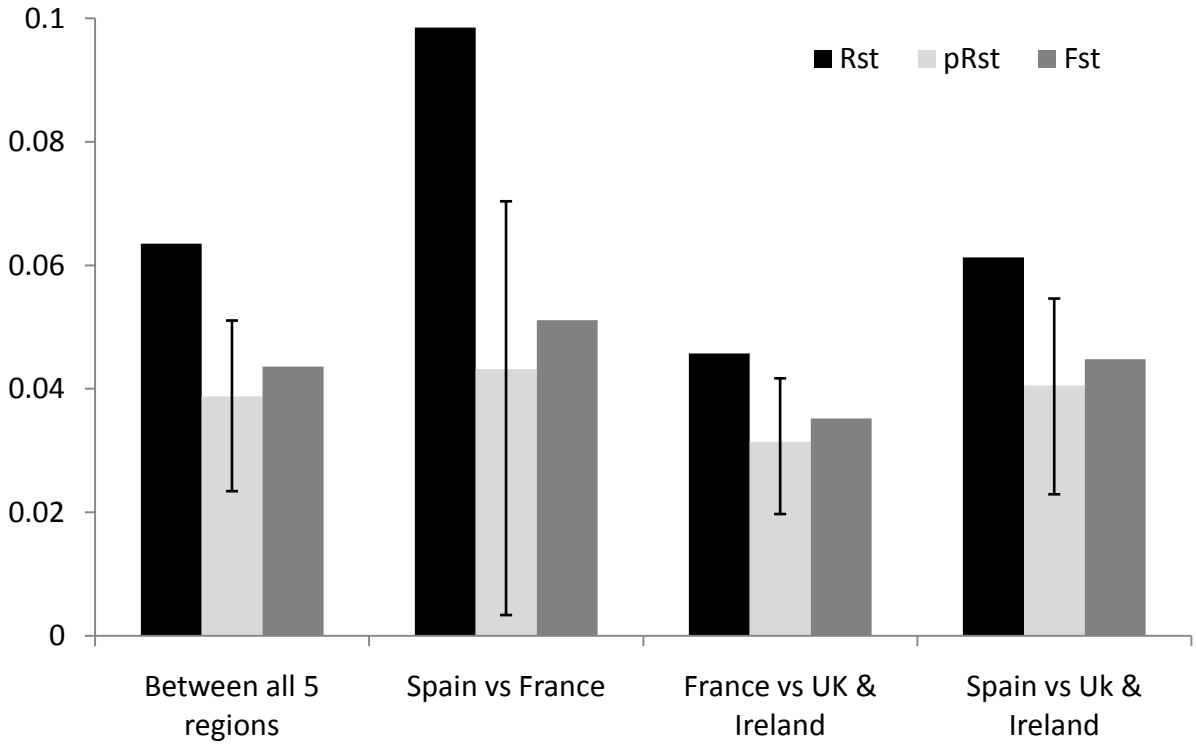


Figure 3.3 Allele size permutation test results, testing whether stepwise mutations of microsatellite regions have contributed to genetic differentiation among the 5 regions of northwest Europe depicted in the STRUCTURE analysis and phylogenetic tree (see text). Shown are global R_{ST} estimates that are significantly larger that the null distribution of the permuted R_{ST} values (pR_{ST}; 10,000 permutations; $p<0.05$), which is analogous to global F_{ST}.

mtDNA RFLP Data analysis

RFLP analysis of the ND1 region of mtDNA revealed a total of four composite haplotypes (AAAA, AABA, BBBA, BBBB), all of which have been observed previously in populations from northwest Europe (Nielsen *et al.*, 1996; Verspoor *et al.*, 1999; Nilsson *et al.*, 2001). The majority of populations exhibited varying frequencies of all four haplotypes (Figure 3.4; Table 3.1), with only one population fixed for a single haplotype (Ulla; BBBB). All four haplotypes are thought to have been widely distributed before the last glacial period (Nilsson *et al.*, 2001), but due to their frequency distributions, the AAAA and AABA haplotypes have previously been termed the 'Baltic' haplotypes while the BBBB and BBBA haplotypes were termed 'Atlantic' haplotypes (Nilsson *et al.*, 2001); for continuation we will also refer to this nomenclature in this study, though it is important to remember that this does not reflect their true origins, rather the nomenclature simply reflects their comparative dominance within the two regions. The Spanish populations were dominated by 'Atlantic' haplotypes (BBBB; BBBA; 97.5%), whilst the French populations were dominated by 'Baltic' haplotypes (AAAA; AABA; 80%; Table 3.4). Overall, the UK and Ireland were largely dominated by the 'Atlantic' haplotypes, however the southern UK populations exhibit a higher proportion of 'Baltic' haplotypes compared to the northern UK populations. Genetic diversity, expressed as haplotype and nucleotide diversity, was highest in the UK and Ireland, specifically the UKSouthwest group and lowest in Spain (Table 3.4); a significant reduction in genetic diversity was observed in Spain compared to the UK and Ireland ($p<0.05$). All other comparisons were non-significant.

The UPGMA population tree constructed from Nei's D_A distance, revealed the division of populations into three clusters; one dominated by French populations exhibiting high frequencies of the 'Baltic' haplotypes, which also contained the geographically proximate Avon; one largely Spanish clade dominated by high frequencies of the 'Atlantic' haplotypes, with two rivers from other regions (Dart and Boyne) which show similarly high frequencies of the 'Atlantic' haplotypes; and a third clade containing the majority of British and Irish samples which have a more equal distribution of both the 'Atlantic' and 'Baltic' haplotypes (Figure 3.5). Hierarchical analysis for these three broad regions revealed that 34.37% of genetic

diversity lay among the three regions, 8.74% lay among populations within each region and 56.89% resided within populations.

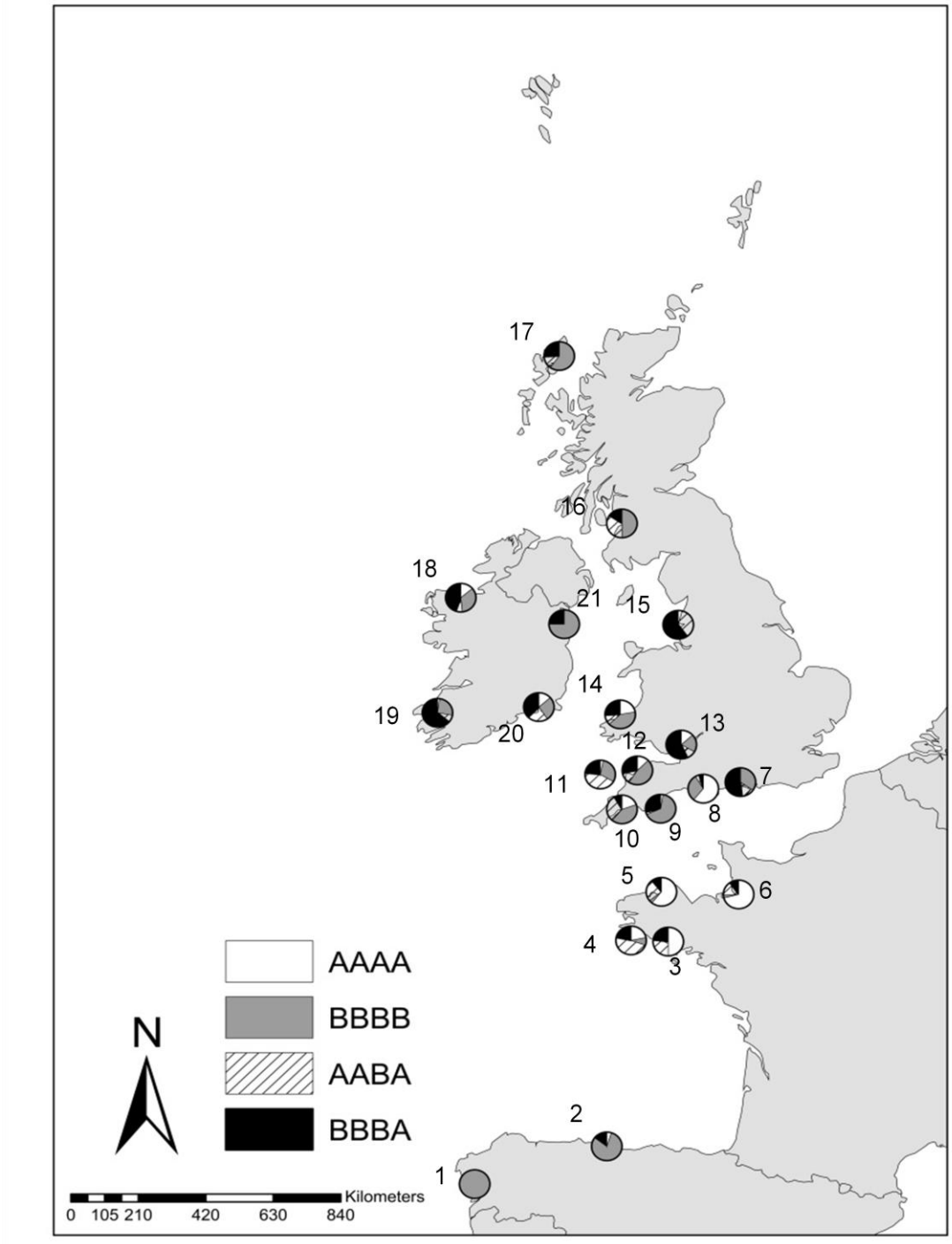


Figure 3.4 Geographical distribution of mtDNA haplotype frequencies for Atlantic salmon across northwest Europe. Population ID follows Table 3.1; Composite haplotypes derived from the restriction enzymes *HeaIII, HinfI, Ava II* and *Rsa I* as described in Nielsen *et al.* (1996).

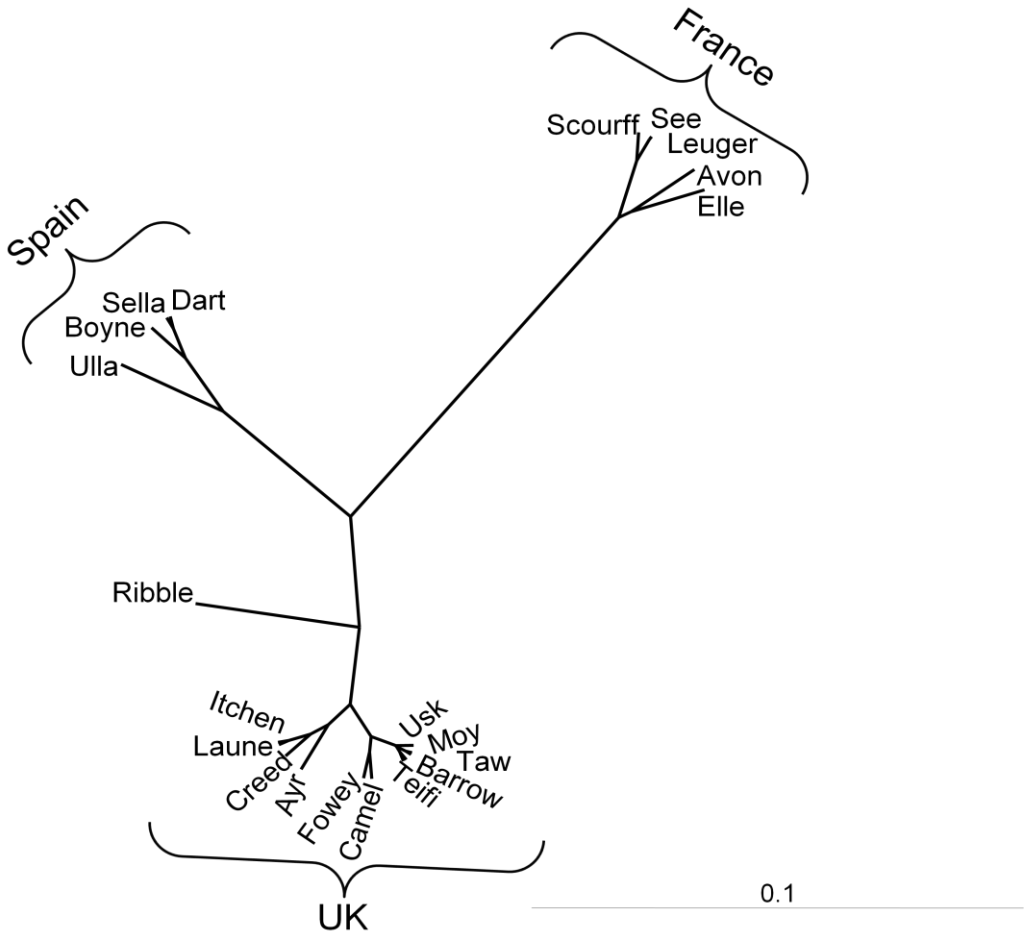


Figure 3.5 UPGMA phylogenetic tree generated from mtDNA RFLP data depicting relationships between matrilineal Atlantic salmon populations across northwest Europe. Scale depicts the number of nucleotide substitutions per site.

DISCUSSION

The high genetic diversity observed in Britain and Ireland, particularly in the southwest of England, indicated that Atlantic salmon did not colonise this region following the last glacial retreat from salmon expanding out of a single Iberian refuge in northern Spain (*i.e.* H_0 is rejected). Rather, the results are consistent with there being a second, cryptic northern refuge for Atlantic salmon in northern France (*i.e.* H_{1A} is accepted) with colonisation of Britain and Ireland from both refugia. Thus, the use of these two different classes of molecular markers has proven highly effective in elucidating the post-glacial history of Atlantic salmon in northwest Europe.

GENETIC DIVERSITY AND DIFFERENTIATION

The genetic diversity of 21 populations of Atlantic salmon in northwest Europe as revealed by microsatellite analysis (12 loci, 289 alleles) was found to be comparable to that reported for other regions (Global:12 loci, 266 alleles, King *et al.*, 2001; Northern Europe: 14 loci, 242 alleles, Tonteri *et al.*, 2005; Baltic: 9 loci, 203 alleles, Saisa *et al.*, 2005). Expected heterozygosity (He=0.808) was consistent with that reported for populations from Britain and Ireland by Saisa *et al.* (2005; He = 0.812) and higher than that reported for other regions (Baltic: He=0.73, Saisa *et al.*, 2005); Northern Europe: He=0.56, Tonteri *et al.*, 2005; North America: He=0.60, King *et al.*, 2001). Within northwest Europe, genetic diversity was found to be highest in southwest England (UKSouthwest: Ar=12.276, He=0.86) and lowest in southern England (UKSouth: Ar=7.396, He=0.767), though this latter estimate was still higher than previously reported for other regions (King *et al.*, 2001; Saisa *et al.*, 2005; Tonteri *et al.*, 2005). Genetic divergence was highest between southern England (UKSouth) and Spain (F_{ST} =0.114) and lowest between the southwest of England (UKSouthwest) and northwest England/Ireland (UKNorthwest/Ireland) (F_{ST} =0.026).

The genetic diversity of matrilineal Atlantic salmon populations in northwest Europe as revealed by PCR-RFLP mtDNA and measured using average haplotype diversity estimates (Global H=0.509) was found to be comparable to that observed in previous studies (British Isles and Spain H=0.507, Nilsson *et al.*, 2001), higher than that reported for the Atlantic as a whole or in the Baltic (H=0.478; H=0.217 respectively, Nilsson *et al.*, 2001), but lower than that reported for northern Europe

(H=0.543 Asplund *et al.*, 2004). In terms of mtDNA variability, within northwest Europe, genetic diversity was found to be highest in southwest England (UKSouthwest: H=0.647) and lowest in Spain (H=0.166). Lower diversity in Spanish populations compared to those from the British Isles in PCR-RFLP mtDNA studies has been reported previously by Verspoor *et al.* (1999) and Consuegra *et al.* (2002).

All four composite haplotypes observed in this study have been described previously (Nielsen *et al.*, 1996) and are known to be widespread across Europe (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001; Asplund *et al.*, 2004). The consensus is that the evolution of these haplotypes predates the Pleistocene glaciations and hence all haplotypes were probably widespread prior to the last glacial maximum (Nilsson *et al.*, 2001; Asplund *et al.*, 2004; Saisa *et al.*, 2005). Thus, the effect of the last glacial period would primarily be observed by lineage sorting through random drift, initially in the differing frequency distributions of these haplotypes in different refugial populations, and subsequently in the mixing of these haplotypes in postglacial processes, such as processes of secondary contact.

The dominance of the 'Atlantic' (BBBB and BBBA) haplotypes in Spain (97.5%) has been noted previously by Verspoor *et al.* (1999) (81% 'Atlantic') and Consuegra *et al.* (2002) (97% 'Atlantic'), though Consuegra *et al.* (2002) used only three of the four restriction enzymes used here. Similarly, the relatively high frequencies of the 'Atlantic' haplotypes in Britain and Ireland (72%) have been noted by Verspoor *et al.* (1999) (73% 'Atlantic') and Consuegra *et al.* (2002) (77% 'Atlantic' with three of four restriction enzymes). However, the dominance of the 'Baltic' (AAAA and AABA; 80%) haplotypes in northwest France has not been observed previously, though none of the previous studies had sampled from this region.

Refugial locations for Atlantic salmon in northwest Europe

The British/Irish Ice sheet extended to around 52ºN at the time of the last glacial maximum, with permafrost extending over most of Britain and Ireland except the far southwest corner of England (Murton and Lautridou, 2003). South of the ice sheet, the Iberian Peninsula (northern Spain and Portugal) has been widely accepted as a refuge for Atlantic salmon since bone relics dating back 40,000 years were

uncovered in a cave in this region (Consuegra *et al.*, 2002). The initial hypothesis set out in this study was to ascertain whether this was the sole source of contemporary Atlantic salmon populations in northwest Europe; results from this study suggest that this is highly unlikely, and there is evidence to propose that there was a distinct second refuge for Atlantic salmon in northern France. Supporting evidence comes from both microsatellite and mtDNA data.

Separation of the French populations from the Spanish populations is depicted in the clustering and phylogenetic analysis, where the French populations formed a distinct cluster in the STRUCTURE analysis and distinct groupings on both the microsatellite and mtDNA population trees. Furthermore, the allele size permutation test of microsatellite data revealed that stepwise mutations had contributed to the differentiation of the Spanish and French populations, suggesting these groups have been separated for 10,000 years or more, *i.e.* since before the last glacial maximum. Hence, these regions must have been distinct during the last glacial period. The discontinuity of mtDNA haplotype frequencies also corroborates this theory. In line with previous studies (Verspoor *et al.*, 1999; Consuegra *et al.*, 2002), Spanish populations were dominated by 'Atlantic' haplotypes, however, French populations were dominated by 'Baltic' haplotypes. This again indicates that French populations are unlikely to be descendants of the Spanish populations. Northwest France has not been stocked with fish from the Baltic region, and due to the current migratory behaviour of Baltic salmon, whereby very few fish pass the Danish strait into the North Sea and Atlantic Ocean, it seems improbable that northern France would have been colonised naturally by Baltic salmon. Therefore, it seems likely that the high frequencies of these 'Baltic' haplotypes exhibited by salmon populations in northwest France are due to lineage sorting through random drift. Thus, corroborating the results of the allele size permutation test, it seems probable that the French populations were isolated from the Spanish populations during the last glacial period while this process took place.

The data shows that at least two refugia for Atlantic salmon occupied northwest Europe during the last glacial period; one in northern Spain (Iberian Peninsula) and one in northern France. There is increasing evidence that many species existed further north of the traditionally accepted southern Peninsula refuge, in sheltered areas experiencing stable microclimates (Stewart and Lister, 2001);

northern France may have been once such location. In a recent review (Bhagwat and Willis, 2008) it was proposed that species whose current northerly range extends to 60ºN or further would have been more able to persist in northerly refugia; this seemingly reflects the ecological and physiological adaptations these animals have in order to survive these conditions (Bhagwat and Willis, 2008). In the case of the Atlantic salmon, their current distribution extends to over 70ºN in the Barents and Kara Seas (MacCrimmon and Gots, 1979) and they exhibit high tolerance to extreme cold conditions, hence it is plausible that Atlantic salmon could have persisted in the conditions in northern France during the late glacial period. In an analogous situation, the cold adapted freshwater fish, the bullhead *Cottus gobio*, seems to have persisted in a refuge in the southern British Isles (Hanfling *et al.*, 2002). However, there is no evidence here that the French refuge extended into southern England, as previously proposed by Payne *et al.* (1971) and Child *et al.* (1976).

In earlier studies, the southern North Sea has been proposed as the most likely location of a second refuge for Atlantic salmon in northwest Europe (Verspoor *et al.*, 1999; Asplund *et al.*, 2004; Saisa *et al.*, 2005). No indication of this was observed in this study but our research focussed on western Britain and Ireland. An investigation centred on drainages of the North Sea would be necessary to explore this theory, though this would be difficult since many of the North Sea drainages are now extinct of salmon.

Colonisation of Britain and Ireland

A second objective for this study was to reconstruct the colonisation pathway of Atlantic salmon from these source populations into Britain and Ireland. Results are consistent with the hypothesis that Britain and Ireland is a secondary contact zone for Atlantic salmon expanding out from the Spanish and French refugial populations. Furthermore, it seems likely that there was no differential colonisation of regions within Britain and Ireland by the French and Spanish colonisers. Evidence of this came from the allele size permutation test comparing populations within Britain and Ireland, which revealed that stepwise mutations had not contributed to the differentiation of populations within the region, suggesting that these populations are likely to be descended from the same refugial population(s). Moreover, when populations from Britain and Ireland were compared to the potential source populations of France and Spain, the allele size permutation tests revealed a

significant contribution of stepwise mutations to the observed differentiation. This presumably reflects the fact that both the French and Spanish colonisers were successful in colonising all of Britain and Ireland, with little differential colonisation occurring.

Despite the supposition that colonisers from France and Spain were equally successful across Britain and Ireland, it is possible to speculate on the colonisation route of female Atlantic salmon out of these refugial populations through the exploration of matrilineal haplotype frequencies. The French populations were largely dominated by the 'Baltic' haplotypes, whereas the Spanish populations were dominated by the 'Atlantic' haplotypes. Within Britain, the highest frequency of 'Baltic' haplotypes was in the region geographically most proximate region to France, the UKSouth. Indeed, contemporary populations of salmon colonising uninhabited rivers appears to initially be individuals straying from the closest neighbouring catchments (Jonsson *et al.*, 2003; Saura *et al.*, 2008). The proportion of 'Baltic' haplotypes then decreased in the UKSouthwest region, where the frequency distribution of the 'Baltic' and 'Atlantic' haplotypes were most closely balanced, and decreased further in UKNorthwest/Ireland, where the 'Atlantic' haplotypes began to dominate once more.

This would support the following colonisation process: when the ice retreated and Atlantic salmon could begin to colonise the previously glaciated Britain and Ireland, salmon expanded out of the French refuge into southern England (UKSouth and UKSouthwest) before heading north (UKNorthwest/Ireland) with less success. Simultaneously, salmon expanded out of the Spanish refuge into southwest England (UKSouthwest), where they initially came into contact with French colonisers, before heading east (UKSouth) with limited success and north (UKNorthwest/Ireland) with more success. Correspondingly, we see the highest proportion of Baltic/French haplotypes in UKSouth, decreasing into the west and north; the highest genetic diversity is observed in UKSouthwest, where the two colonising groups initially made contact; and the highest proportion of Atlantic/Spanish haplotypes in UKNorthwest, where they were the more successful colonisers.

CONCLUSIONS

The use of multiple classes of molecular markers has proved highly effective in resolving the phylogeographic history of Atlantic salmon in a region which to date has received little focussed attention. In addition to the well established refugia in the Iberian Peninsula, a cryptic northern refugium present in northern France also seems to be likely, with colonisation of Britain and Ireland occurring from both refugia.

The clear distinction of the populations of northern France from those in Spain has not been observed in previous studies since samples from France have previously been lacking. To effectively protect these two ancient lineages it would be necessary to locate the boundary between the two refugia; an obvious avenue for future research. The ability to reliably identify such variation is important in order to safeguard the genetic diversity of the species, especially in these small populations existing in somewhat atypical warmer environments at the extreme southern edge of the species.

CHAPTER 4: TRIBUTARY LEVEL STRUCTURING OF ATLANTIC SALMON (*SALMO SALAR* L.) POPULATIONS WITHIN A SMALL RIVER CATCHMENT: IMPLICATIONS FOR CONSERVATION AND MANAGEMENT.

Due to be submitted to *Ecology of Freshwater Fish,* with authors as follows:

A.K. Finnegan[†], A.M. Griffiths[†], D. Bright[§] & J.R. Stevens[†]

[†]University of Exeter, Hatherly Laboratories, School of Biosciences, Prince of Wales Road, Exeter, Devon, EX4 4PS.

[§] Westcountry Rivers Trust, Rain-Charm House, Kyl Cober Parc, Stoke Climsland, Callington, Cornwall, PL17 8PH.

A.K. Finnegan undertook all laboratory and statistical analysis and prepared all drafts of the manuscript.

A.M. Griffiths provided guidance for statistical analysis and commented on earlier manuscripts.

D. Bright, at the Westcountry Rivers Trust, is a Case partner for the studentship.

J.R. Stevens provided guidance throughout and commented on earlier manuscripts.

INTRODUCTION

While the decline and extinction of species is an intrinsic part of evolution, there can be no denying that the expansion of human populations and the subsequent over-exploitation of the planet's natural resources have dramatically increased the rate of these processes. In 2006 the ICUN estimated that 40% of all organisms were threatened, while the FAO (www.fao.org) estimate that over 70% of the world's fish species are either fully or over-exploited. As numbers decline and populations are lost, the overall genetic diversity of a species can become compromised (O' Brien, 1994). This makes it extremely important to understand the distribution of genetic diversity within a species in order to best preserve it. Within a species, genetic diversity is arranged hierarchically from family units through to range wide phylogeographic lineages (Avise, 1994). To maintain this intra-specific genetic diversity, management and conservation programmes must consider genetic diversity at each hierarchical level (Kenchington *et al.*, 2003), requiring detailed investigations to define geographic boundaries and to identify the evolutionary mechanisms acting to maintain them.

The Atlantic salmon (*Salmo salar* L.) is an anadromous teleost fish which migrates between the freshwater environment for breeding and early life history stages and the marine environment for feeding and growth stages (Mills, 1971). Over the last 200 years there has been a marked decline of salmon populations and an increase in localised extinction events, particularly since the 1970s (Parrish *et al.*, 1998). It has long been appreciated that genetically distinct populations of Atlantic salmon occur in neighbouring catchments (Ståhl, 1987), however, the occurrence of a number of differentiated populations inhabiting a single river system is now widely documented in numerous studies across the species range (Ståhl, 1987; Elo *et al.*, 1994; Beacham and Dempson, 1998; Garant *et al.*, 2000; Primmer *et al.*, 2006; Vaha *et al.*, 2007). When tasked with maintaining the genetic diversity of this species in a managerial or conservation context, understanding the extent of within-river structuring and the processes involved in driving and maintaining this structuring are essential considerations. Indeed, to be consistent with the North Atlantic Salmon Conservation Organisation's (NASCOs) Precautionary Approach, catchment level management of salmon is aspired to across the species range. Hence, if it is possible to determine the extent of genetic differentiation between salmon populations at the

catchment level and the evolutionary mechanism acting to drive and maintain this differentiation, then this information can be fed into catchment level management. With knowledge of this kind, recommendations relating to the movement or stocking of fish can be given; similarly, estimates for the area over which conservation measures need to be taken in order to maintain sufficient genetic diversity within the catchment can be provided. Furthermore, if commonalities between rivers can be found, then general management and conservation strategies can be adopted across rivers.

It is thought that salmon populations become differentiated within rivers in part due to the discontinuous nature of spawning habitat, which acts to isolate groups of breeders such that, over time, genetic drift acts to alter allele frequencies of neutral loci between isolated groups, while the development of adaptations to the local environment acts to maximise survival and recruitment in a particular river or tributary. Subsequently, such differentiation reinforces the precise homing behaviour that Atlantic salmon exhibit, whereby after their extensive feeding migrations they return to their tributary of origin to reproduce (Stabell, 1984; Vaha *et al.*, 2007); this cycle is one of the principle mechanisms driving and maintaining within-river population structuring. However, external factors, such as environmental conditions, can also act to influence this phenomenon. The interplay between homing behaviour and environmental perturbations can give rise to differing levels of within-river genetic structuring and two differing evolutionary theories have been advocated as good models to explain these processes (Garant *et al.*, 2000), namely the 'member-vagrant' model and the 'metapopulation' model. The member-vagrant hypothesis (Figure 1.8) assumes that physical (*e.g.* hydrodynamic forces) and biological (*e.g.* rheotaxis and homing behaviour) processes retain young fish and attract returning adults to their natal stretch of river, thus maintaining a stable population (Iles and Sinclair, 1982). These individuals are 'members' in this model and make up the majority of the population, whereas individuals that do not accomplish this are termed 'vagrants', and potentially contribute as migrants to other populations. Under this scenario, a high level of genetic differentiation is expected between populations inhabiting discontinuous stretches of suitable habitat within the river catchment. The populations would be temporally stable and an isolation-by-distance model of

genetic relatedness should be observed as the vagrants are most likely to migrate and breed in other local populations.

In contrast, although based on similar attributes of retaining young fish and returning adult fish, the meta-population theory (Figure 1.9) assumes that these populations suffer intermittent local extinction events and subsequent re-colonisation from other populations (McQuinn, 1997; Rieman and Dunham, 2000). This can be a highly realistic scenario for Atlantic salmon populations as their riverine habitats may be prone to flash floods, drought, freeze/thawing and/or pollution events, thus maintaining a meta-population structure within the catchment and precluding the progression onto a member-vagrant system, as would likely be achieved in the absence of any extinction events (Garant *et al.*, 2000). In catchments where meta-population models of population structuring prevail, vagrants, or migrants, are an intrinsic part of maintaining each population. Hence, in these situations, genetic divergence among populations is likely to be weak, yet still significant, implying a degree of straying/migration and relatively high geneflow. Similarly, in contrast with the member-vagrant model, no correlation between genetic and geographic distance would be expected.

An additional consideration, which to date appears to have been overlooked, is the possibility that some genetic differentiation of populations within river catchments may be due to differential post-glacial colonisation of tributaries within river catchments. This scenario may be envisaged for previously glaciated areas where colonisation is known to have occurred from salmon expanding out of more than one glacial refuge. If found to be the case, this could potentially raise the conservation status for some within-river populations of Atlantic salmon.

Here, we use genetic data from Atlantic salmon from the river Dart to explore these fundamental issues in relation to catchment level management. The river Dart (N50:20:35 W3:33:53) in southwest England drains the granite massif of Dartmoor National Park, 545m above sea level, in a southerly direction, with a total catchment area of 475km^2 over 45 km. By genotyping Atlantic salmon populations from across the catchment using nuclear microsatellite loci and mtDNA PCR-RFLP analysis, we have been able to test the hypothesis that Atlantic salmon within the catchment are a single panmictic population, *i.e.* no significant genetic differentiation is observed

between samples (H_0). Subsequently, we have been able to evaluate a series of alternate hypotheses concerning the applicability of the member-vagrant and/or meta-population evolutionary models, *i.e.* H_1: there are multiple populations of Atlantic salmon within the catchment and significant genetic differentiation between samples is observed; H_{1A}: the differentiation detected suggests that samples follow a meta-population evolutionary mechanism, where the samples show weak, yet significant genetic differentiation, but there is no isolation-by-distance signal; H_{1B}: the differentiation detected suggests that samples follow a member-vagrant evolutionary mechanism, where the samples show highly significant genetic differentiation and there is a strong isolation-by-distance signal; H_2: differential colonisation of samples/populations by post-glacial colonisers is observed, *i.e.* genetic differentiation of populations occurred before colonisation of the Dart while in different glacial refugia.

METHODS

SAMPLE COLLECTION

Twelve locations throughout the river Dart were sampled in electric-fishing surveys between May and September of 2005, 2006 and 2007 with some sites sampled in multiple years (Figure 2.3; Table 2.1; Figure 2.5; Table 2.3). The average (±SD) waterway distance between sampling sites was 28 ± 18 km (range 5 -74 km). For all samples, adipose fin clips from parr in multiple year classes were taken and stored in 95% ethanol.

DNA ANALYSIS: MICROSATELLITE

Genomic DNA was extracted from all adipose fin clips using the Chelex method of Estoup *et al.* (1996). Variation was determined at 13 microsatellite loci: Ssa157a (King *et al.*, 2005), Ssa197, Ssa202, Ssa171, Ssa85 (O'Reilly *et al.*, 1996), SSsp1605, SSsp 2201 (Paterson *et al.*, 2004), Ssa 289, Ssa14 (McConnell *et al.*, 1995), SSOSL438, SSOSL85, SSOSL311 (Slettan *et al.*, 1996) and Ssa412 (Cairney *et al.*, 2000). Microsatellite loci were amplified in single 10µL PCR reactions containing 0.25 units Taq DNA polymerase (Bioline), 1X reaction buffer, 0.2mM dNTPs, 1.5mM $MgCl_2$, 0.5µM labelled primer and 1µM unlabelled primer. Amplification was undertaken in a thermal cycler block (Thermo) where samples were initially heated to 94˚C for 2 min for denaturation then exposed to 35 cycles of 94˚C for 2 min, annealing temperature 58˚C for 30 secs (except Ssa289, Ssa14 at 53ºC; SSOSL85 at 55ºC), 72˚C for 30 secs, with a final extension step of 72˚C for 10 mins. The 13 microsatellite loci were then pooled into two multiplexes and subsequently genotyped on a Beckman-Coulter CEQ 8000 automated sequencer.

DNA ANALYSIS: MTDNA RFLP

The ND1 region of mtDNA was amplified using the primers given by (Hall and Nawrocki) and modified by Nilsson et al. (2001). The 30µL reaction volume contained 0.25 units of Taq DNA polymerase (Bioline), 1X reaction buffer, 0.2mM dNTP, 2mM $MgCl_2$ and 0.5µM of each primer. PCR amplification consisted of an initial denaturation step of 95ºC for 3 mins, followed by 35 cycles of 95ºC for 30 secs, 58ºC for 45 secs and 72ºC for 60 secs, concluding with a final extension phase of 5 mins at 72ºC.

Amplified DNA was digested by the restriction enzymes *Hae III, Hinf I, Ava II* and *Rsa I* in individual reactions. 4µL of enzyme and recommended buffer (Promega) mix were added to 6µL of amplified DNA and incubated at 37ºC overnight. Fragments were separated by electrophoresis and variant fragment patterns, and thus composite haplotypes, were determined following Nielsen *et al.* (1996).

STATISTICAL ANALYSIS: ASSESSING THE IMPACT OF POST-GLACIAL COLONISATION PROCESSES IN DETERMINING POPULATION STRUCTURING

MTDNA RFLP ANALYSIS

The sizes of restriction fragments, as estimated by Nielsen et al. (1996) were used to generate a restriction site binary matrix of presence/absence of fragment bands for each enzyme, and a binary representation of the composite haplotypes of each individual using REAP *v4.0* (McElroy *et al.*, 1992). These were used in subsequent programmes in the REAP *v4.0* package. Haplotype and nucleotide diversity within populations was calculated according to Nei (1987) in the DA programme of REAP *v4.0*. A matrix of pairwise distances (Nei, 1987) based on population haplotype frequencies was generated in REAP and used to construct a UPGMA phylogenetic tree in PHYLIP *v3.67* (Felsenstein, 1993).

MICROSATELLITE ANALYSIS

An allele size permutation test (Hardy *et al.*, 2003), implemented in SPAGeDi *v2.1* (Hardy and Vekemans, 2002) was undertaken to assess the relative contribution of stepwise mutations in the genetic differentiation of populations within the Dart catchment. If R_{ST} is significantly greater than F_{ST} this is an indication that the populations diverged a sufficiently long time ago for stepwise mutations of microsatellite regions of DNA to have contributed to this differentiation. Due to the mutation rate of microsatellites, this would have had to occur more than 2500 Atlantic salmon generations ago, approximating to 10,000 years ago, *i.e.* before the last glacial maximum. Hence, a significant result is indicative of the region being colonised by Atlantic salmon radiating from multiple refugia (Hardy *et al.*, 2003). Confidence intervals were calculated by undertaking 10,000 permutations of allele sizes among alleles within each locus. The locus Ssa171 was excluded from this

analysis as both di- and tetra-nucleotide repeats are observed at this locus and could therefore not have mutated through step-wise processes alone.

STATISTICAL ANALYSIS: ASSESSING THE ROLE OF METAPOPULATION AND MEMBER-VAGRANT EVOLUTIONARY MODELS IN DETERMINING POPULATION STRUCTURING (MICROSATELLITE ANALYSIS)

GENETIC DIVERSITY, HARDY WEINBERG EQUILIBRIUM AND GENOTYPIC LINKAGE EQUILIBRIUM

Allele number and allelic richness (allele number corrected for sample size) were calculated for all loci within populations using FSTAT *v2.9.3* (Goudet, 1995). Deviations from Hardy Weinberg expectations across samples (within loci) and across loci (within samples) were estimated using the method of Guo and Thompson (1992) as implemented in GENEPOP *v3.4* (Raymond and Rousset, 1995), with sequential Bonferroni corrections applied for multiple tests across populations according to Rice (1989). The occurrence of linkage disequilibrium between loci was tested using the default parameters in GENEPOP *v3.4* with sequential Bonferroni corrections applied for multiple tests across populations (Rice, 1989).

GENETIC DIFFERENTIATION AND RELATIONSHIPS AMONG POPULATIONS

Genetic differentiation was measured by calculating F_{ST} values between populations and their significance using FSTAT *v2.9.3* (Goudet, 1995). Genetic distances between populations were estimated according to Nei's (1987) DA distance, from which a phylogenetic tree was constructed using the Neighbour Joining method as implemented in POWERMARKER *v3.25* (Liu and Muse, 2005). Strength of support for each node was assessed by undertaking 1000 bootstrap replicates in POWERMARKER, and a consensus tree was obtained using the CONSENSE programme in PHYLIP *v3.67* (Felsenstein, 1993). The extent of genetic differentiation among populations was also measured with assignment tests (Cornuet *et al.*, 1999). Individuals were assigned to their most likely population of origin based on their microsatellite multi locus genotype using GENECLASS2 (Piry *et al.*, 2004). The Bayesian method using the 'leave one out' procedure was implemented (Rannala and Mountain, 1997).

ASSESSING SPATIAL GENETIC STRUCTURE

The underlying population structure within the river was inferred using STRUCTURE *v2.2* (Pritchard *et al.*, 2000). This Bayesian Markov Chain Monte Carlo (MCMC)

based approach uses a clustering based algorithm to partition individuals into groups based on criteria to minimise Hardy Weinberg disequilibrium within populations and linkage disequilibrium between loci within populations. A burn-in period of 30,000 steps was set, followed by 1,000,000 MCMC replicates; allele frequencies were set as 'weakly correlated' as recommended by Pritchard *et al.* (2000). The number of groups to be simulated (K) was set as 1 to 15, with 20 replicates for each group. The optimal number of clusters was estimated using the ΔK method of Evanno *et al.* (2005) and the *ad hoc* method of Pritchard *et al.* (2000).

Isolation by distance of populations was tested by assessing the association between population genetic differentiation ($F_{ST}/(1- F_{ST})$) and geographic distance between sampling sites (waterway distances in km) by Mantel tests implemented in GENALEX *v6.1* (Peakall and Smouse, 2006).

Spatial genetic structure at the individual level was assessed following the approach of Primmer *et al*. (1996) as implemented in GENALEX *v6.1* (Peakall and Smouse, 2006). The multivariate approach as implemented in GENALEX *v6.1* strengthens the spatial signal by simultaneously assessing the signal generated by multiple genetic loci, thereby reducing stochastic noise introduced through more classical allele-to-allele or locus-to-locus comparisons. Genetic distances between individuals were estimated using the squared distance method (Peakall *et al.*, 2003) and compared to waterway distances (km). Individuals from the same pair of sampling sites were assigned the same geographic distance and individuals from the same population were assigned a geographic distance of zero. An autocorrelation coefficient (r) is generated. This coefficient provides a measure of the genetic similarity between pairs of individuals, where geographic separation falls within a spatial distance class; this can be summarised in a correlelogram. In the correlelogram, the autocorrelation coefficient (r) was plotted as a function of eight discrete classes (5, 10, 15, 20, 30, 40, 60, 76 km), with individuals separated by 0-5km included in the first class, individuals separated by 6-10km included in the second class, *etc*. This analysis provides an estimate of the extent of non-random genetic structure across the whole study site. A second genetic autocorrelation analysis allowed analysis of multiple distance class sizes in a way that is designed to investigate the interplay between sample size and distance class. The same distance classes were used. however, individuals from more distant groups were added to the

previous groups, *i.e*. individuals separated by 0-5km were included in the first class, individuals separated by 0-10km were included in the second class, *etc*. This second analysis enabled a more accurate estimate of the scale across which spatial genetic structure can be detected.

Statistical significance for these tests was assessed in two ways. First, 1000 permutations were undertaken by shuffling individuals among geographic locations to calculate 95% confidence intervals (CIs). If *r* fell outside these 95% CIs then significant spatial structure was inferred. Additionally, 1000 bootstrap replicates were undertaken within the set of pairwise comparisons for each distance class, from which 95% CIs were calculated. When the CIs did not straddle $r=0$, again significant spatial structuring was inferred.

ASSESSING TEMPORAL GENETIC STRUCTURE

To quantify the relative contribution of spatial variation in gene diversity among sites versus the temporal variance in gene diversity within sites, two approaches were taken. Initially, hierarchical gene diversity was assessed in an AMOVA, as implemented in ARLEQUIN *v3.1* (Schneider *et al.*, 2000). Three hierarchical levels were defined in order to assess the component of genetic diversity attributable to (i) variance among sample sites (geographic component), (ii) variance between temporal samples within sites (temporal component), and (iii) variance among individuals within sites. Additionally, the statistical significance in the differences observed between spatial and temporal components were assessed statistically by performing Mann Whitney U tests in MINITAB. This involved comparing spatial and temporal data for (i) the number of loci showing significant differences in allele frequency distributions, (ii) DA values, and (iii) pairwise F_{ST} values. These analyses were performed on the subset of sample sites for which there were temporal replicates (3 sites, over 2-3 years; Table 2.1).

RESULTS

ASSESSING THE IMPACT OF POST-GLACIAL COLONISATION PROCESSES IN DETERMINING POPULATION STRUCTURING

MTDNA RFLP ANALYSIS

Four composite haplotypes (AAAA, AABA, BBBB, BBBA), all of which are widespread across northwest Europe (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001; Finnegan *et al.*, Submitted), were generated from the RFLP analysis of the ND1 region of mitochondrial DNA using four restriction enzymes (Table 4.1). No population was fixed for one haplotype and most (5 of 7) exhibited varying frequencies of all four haplotypes. In line with previous studies (Nilsson *et al.*, 2001; Chapter 3), the BBBB and BBBA haplotypes are referred to as 'Atlantic' haplotypes, while the AAAA and AABA haplotypes are referred to as 'Baltic' haplotypes. This does not reflect their evolutionary origin, as all four haplotypes are believed to have been widespread before the last glacial maximum (Verspoor *et al.*, 1999; Nilsson *et al.*, 2001), but is a useful nomenclature in order to make comparisons with previous studies. Within the river Dart, the 'Atlantic' haplotypes dominated in most populations, apart from the Walla sample which was dominated by 'Baltic' haplotypes (Table 4.1). Haplotype diversity ranged from 0.3265 (Cherry06) to 0.7375 (Web/Pons06), with an average of 0.6066. The overall average genetic distance, as estimated with NeiS (1987) DA distance was 0.023452, ranging from 0.015927 between the Cherry06 and East06, to 0.031489 between the Cherry06 and Walla samples.

Table 4.1 mtDNA RFLP results and diversity indices

	N	BBBB	BBBA	AAAA	AABA	Haplotype diversity	Nucleotide diversity
West/Prince	33	51.5	21.2	18.2	9.1	0.6583	0.023992
Walla	32	21.9	15.6	21.9	40.6	0.7262	0.024056
East06	38	65.8	28.9	2.6	2.6	0.4884	0.010158
Cherry06	25	80.0	0.0	20.0	0.0	0.3265	0.020303
Web/Cock	23	17.4	47.8	0.0	34.8	0.6338	0.017418
Web/Pons06	34	32.4	20.6	14.7	32.4	0.7375	0.024954
Hems	26	46.2	30.8	15.4	7.7	0.6757	0.022142

Detailed are number of individuals (N); mtDNA RFLP composite haplotypes derived from restriction enzymes *Ava II, Hae III, Hinf I, Rsa I*, according to Nielsen et al (1996).

In the population tree constructed from Neis DA distance, the grouping of samples bears little relation to geographic proximity. When the Dart samples are included in the population tree with samples from across northwest Europe (see Chapter 3), they do not cluster together within the broad UK group, rather one population (Walla) is included in the predominantly French group; two populations (Cherry06, East06) are included in the predominantly Spanish group, whilst the remaining five populations join the UK group, but do not cluster together within it (Figure 4.2).

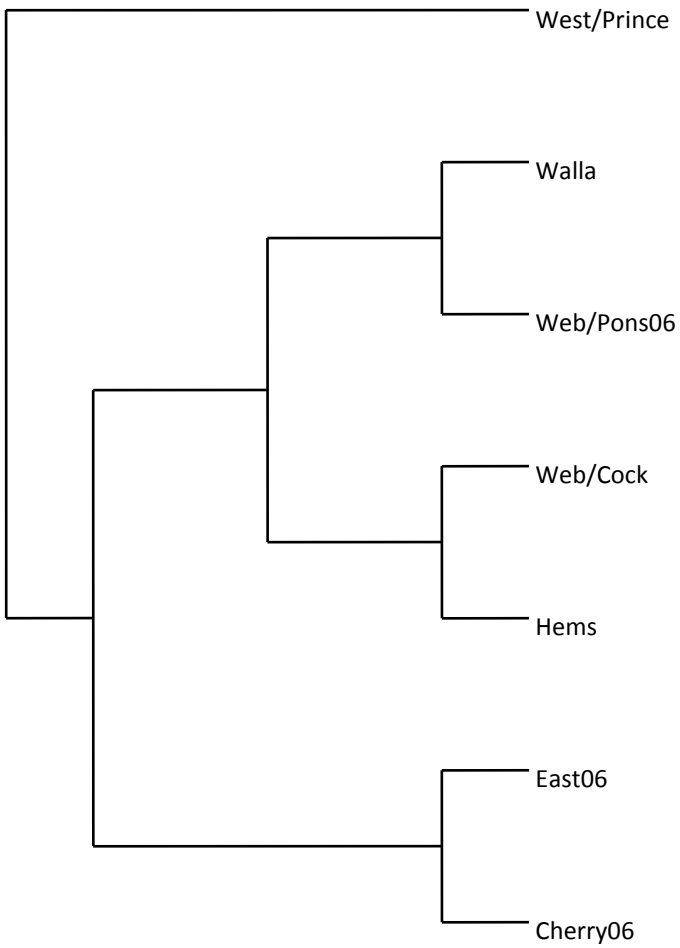


Figure 4.1 UPGMA phylogenetic tree from mtDNA RFLP analysis of river Dart populations.

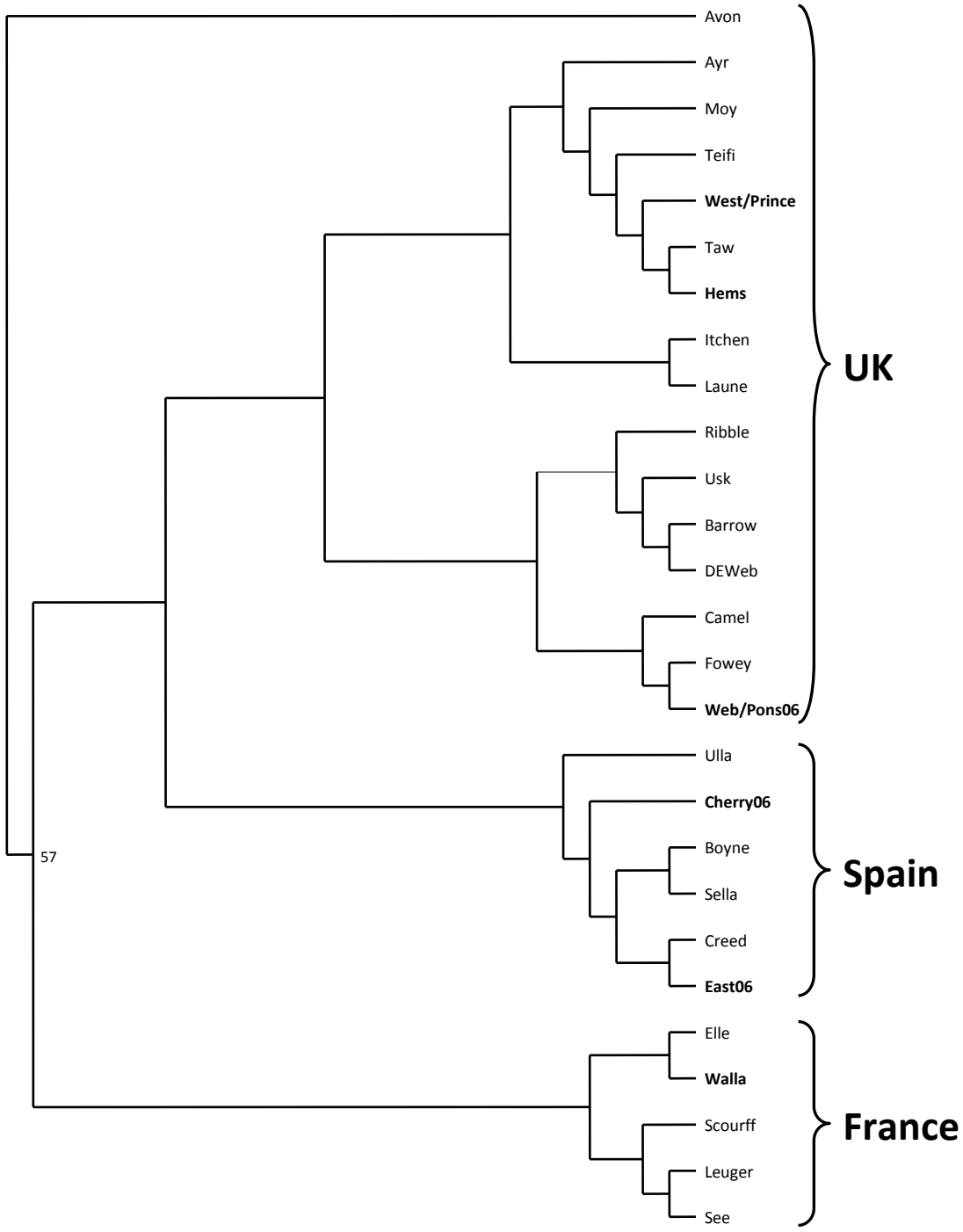


Figure 4.2 UPGMA phylogenetic tree from mtDNA RFLP analysis.

Dart populations (in bold) and other samples from across northwest Europe, see Chapter 3.

MICROSATELLITE ANALYSIS

The allele size permutation test revealed that multilocus R_{ST} values were not significantly higher than F_{ST} values (F_{ST}=0.0295, R_{ST}=0.0303, p=0.58). This demonstrates that stepwise mutations of microsatellite repeat regions have not contributed to the genetic differentiation populations within the river Dart. Such a result indicates that populations within the catchment were colonised by Atlantic salmon from the same glacial refugium and the differences in haplotype frequencies observed must be due to postglacial processes.

ASSESSING THE ROLE OF METAPOPULATION AND MEMBER-VAGRANT EVOLUTIONARY MODELS IN DETERMINING POPULATION STRUCTURING

GENETIC DIVERSITY, HARDY WEINBERG EQUILIBRIUM AND GENOTYPIC LINKAGE EQUILIBRIUM

A total of 275 alleles were observed across the 13 microsatellite loci used in this study, ranging from 9 (Ssa412) to 30 (Sssp2201) at a single locus. Within populations, the number of alleles per locus ranged from 2 (Ssa412; Ssa14) to 22 (Sssp2201), the average number of alleles per population ranged from 8.07 (Cherry055, Cherry066) to 13.53 (West/Prince) and population level allelic richness ranged from 7.06 (Cherry06) to 10.60 (West/Prince) (Table 4.2). After sequential Bonferroni corrections were applied, significant deviations from Hardy Weinberg expectations were observed on 33 occasions, involving 13 of 17 samples and 12 of 13 loci (Table 4.2). Linkage disequilibrium tests revealed 44 (3%) of the 1326 tests to be significant after sequential Bonferroni corrections were applied. Of these 44 significant associations, 22 were observed in only one population, while 11 were observed in two populations. Hence, given the number and spread of significant associations, linkage disequilibrium was deemed negligible.

Table 4.2 Microsatellite diversity indices for the river Dart by population

		SSSP2201	SSOSL85	SSOSL438	SSA289	SSA171	SSA157	SSA14	SSSP1605	SSOSL311	SSA85	SSA412	SSA202	SSA197
Swincombe	**Allele N.**	17	11	8	5	11	16	2	8	17	15	2	11	13
	Ar	13.852	9.017	7.311	4.920	8.837	13.284	2.000	7.324	13.250	12.699	1.998	9.135	10.902
	He	0.925	0.828	0.746	0.762	0.842	0.922	0.372	0.830	0.905	0.925	0.207	0.850	0.891
	Ho	0.885	0.885	0.346	0.654	0.808	0.692	0.346	0.692	0.885	0.962	0.154	0.808	0.923
	HWE	0.072	0.893	**0.000**	0.023	0.568	**0.000**	1.000	0.114	0.090	0.039	0.287	0.528	0.175
Strane	**Allele N.**	14	13	5	6	8	13	4	6	12	7	3	8	13
	Ar	11.487	10.158	4.686	5.482	7.021	11.219	3.508	5.835	9.519	6.692	2.754	7.560	11.008
	He	0.908	0.799	0.648	0.742	0.762	0.908	0.558	0.736	0.800	0.832	0.484	0.837	0.900
	Ho	0.833	0.867	0.433	0.633	0.633	0.800	0.433	0.633	0.633	0.933	0.633	0.833	0.833
	HWE	0.010	0.719	**0.002**	**0.001**	0.020	**0.001**	**0.002**	0.280	0.004	0.005	0.009	0.004	0.004
West/Prince	**Allele N.**	20	16	9	6	12	20	4	10	20	20	7	12	20
	Ar	15.271	12.685	6.662	5.157	9.782	15.732	3.862	8.785	15.505	14.924	5.552	9.065	14.857
	He	0.850	0.895	0.698	0.740	0.862	0.948	0.618	0.799	0.914	0.845	0.475	0.805	0.905
	Ho	0.806	0.903	0.742	0.548	0.742	0.871	0.516	0.742	0.935	0.806	0.194	0.710	0.871
	HWE	0.191	0.476	0.781	0.018	0.031	0.198	0.000	0.127	0.235	0.123	0.000	0.357	0.025
Blackbrook	**Allele N.**	12	14	7	5	8	15	2	7	11	10	3	7	12
	Ar	10.900	11.224	6.022	4.740	6.830	12.189	2.000	6.015	8.885	8.404	2.966	6.126	10.984
	He	0.797	0.768	0.651	0.576	0.652	0.781	0.444	0.599	0.701	0.741	0.365	0.671	0.785
	Ho	0.806	0.742	0.710	0.516	0.581	0.742	0.484	0.742	0.645	0.742	0.065	0.774	0.710
	HWE	0.009	**0.001**	0.164	0.193	0.099	0.004	0.710	0.768	0.014	0.097	**0.000**	**0.002**	0.014
West/Tor	**Allele N.**	17	13	8	6	9	13	2	8	18	13	2	8	10
	Ar	12.725	8.836	6.461	5.319	6.980	10.905	2.000	7.127	12.751	10.807	1.979	6.315	8.811
	He	0.890	0.823	0.738	0.701	0.777	0.874	0.448	0.810	0.917	0.896	0.163	0.804	0.856
	Ho	0.824	0.912	0.559	0.588	0.706	0.824	0.471	0.618	1.000	0.971	0.176	0.912	0.941
	HWE	**0.000**	0.114	**0.000**	0.270	0.052	0.030	1.000	0.004	**0.000**	**0.001**	1.000	0.272	0.010

		SSSP2201	SSOSL85	SSOSL438	SSA289	SSA171	SSA157	SSA14	SSSP1605	SSOSL311	SSA85	SSA412	SSA202	SSA197
Cherry05	**Allele N.**	12	8	4	5	5	15	2	8	12	13	3	9	9
	Ar	12.000	8.000	4.000	5.000	5.000	15.000	2.000	8.000	12.000	13.000	3.000	9.000	9.000
	He	0.913	0.784	0.646	0.683	0.763	0.947	0.508	0.853	0.903	0.878	0.297	0.825	0.862
	Ho	1.000	0.800	0.800	0.667	0.867	0.867	0.467	0.867	1.000	0.800	0.200	0.733	0.800
	HWE	0.980	0.259	0.075	0.272	0.626	0.113	1.000	0.305	0.322	0.016	0.321	0.125	0.597
Cherry06	**Allele N.**	12	9	6	5	9	9	3	7	14	12	3	8	9
	Ar	9.705	6.847	5.366	4.928	7.633	8.085	2.577	6.631	10.765	10.640	2.973	7.296	8.352
	He	0.842	0.693	0.640	0.733	0.835	0.836	0.461	0.776	0.878	0.903	0.410	0.819	0.830
	Ho	0.923	0.692	0.577	0.769	0.885	0.885	0.654	0.731	0.885	1.000	0.346	0.962	0.846
	HWE	0.700	0.606	0.049	0.253	0.163	0.031	0.052	0.442	0.300	0.142	0.147	0.137	0.467
Cherry07	**Allele N.**	13	11	5	5	7	14	2	6	12	12	2	6	10
	Ar	11.792	10.083	4.561	4.428	6.614	12.164	2.000	5.638	10.665	9.903	1.999	5.429	8.666
	He	0.900	0.865	0.621	0.656	0.803	0.912	0.251	0.800	0.892	0.793	0.215	0.770	0.778
	Ho	0.952	0.857	0.333	0.667	0.905	1.000	0.286	0.619	0.905	0.857	0.238	0.857	0.857
	HWE	0.008	0.403	**0.001**	0.270	0.533	0.052	1.000	0.111	**0.001**	0.686	1.000	0.396	0.382
East05	**Allele N.**	21	15	13	11	10	18	7	8	16	18	3	11	17
	Ar	13.825	11.073	9.215	8.412	8.020	12.471	4.974	7.000	11.267	13.506	2.558	9.734	12.178
	He	0.925	0.878	0.855	0.817	0.849	0.908	0.643	0.792	0.872	0.931	0.427	0.878	0.891
	Ho	0.933	0.822	0.778	0.711	0.800	0.822	0.600	0.689	0.911	0.800	0.400	0.800	0.867
	HWE	**0.000**	0.065	**0.000**	**0.000**	0.534	0.029	**0.000**	0.211	0.066	**0.000**	0.017	0.310	0.711
East06	**Allele N.**	17	12	5	4	9	17	5	8	14	16	3	9	14
	Ar	11.602	10.704	4.757	3.761	7.437	12.725	3.837	7.054	10.299	11.580	2.375	7.273	10.832
	He	0.904	0.910	0.716	0.666	0.836	0.916	0.546	0.822	0.883	0.892	0.386	0.811	0.872
	Ho	0.925	0.950	0.725	0.625	0.700	0.950	0.425	0.825	0.825	0.850	0.400	0.725	0.875
	HWE	0.040	0.548	0.082	0.139	0.021	0.082	**0.000**	0.534	0.205	0.018	1.000	0.155	0.023

		SSSP2201	SSOSL85	SSOSL438	SSA289	SSA171	SSA157	SSA14	SSSP1605	SSOSL311	SSA85	SSA412	SSA202	SSA197
Walla	**Allele N.**	18	16	8	7	9	16	6	8	19	14	3	8	14
	Ar	13.281	12.526	6.261	5.205	7.912	12.463	4.208	7.044	13.152	11.444	2.706	7.837	11.327
	He	0.922	0.872	0.701	0.671	0.784	0.919	0.563	0.798	0.879	0.913	0.369	0.869	0.903
	Ho	0.758	0.788	0.576	0.576	0.606	0.788	0.545	0.879	0.818	0.818	0.333	0.879	0.909
	HWE	0.008	0.003	0.013	0.006	0.009	**0.000**	0.008	0.195	0.023	**0.000**	0.741	0.141	0.571
Web/Cator	**Allele N.**	18	14	7	6	14	15	5	8	19	19	3	9	14
	Ar	14.346	10.939	5.530	4.286	10.839	11.818	4.272	6.817	13.601	12.733	2.429	7.477	11.107
	He	0.933	0.861	0.746	0.682	0.865	0.909	0.467	0.793	0.914	0.857	0.380	0.834	0.904
	Ho	0.971	0.857	0.686	0.600	0.857	0.943	0.400	0.829	0.857	0.857	0.314	0.914	0.714
	HWE	0.031	0.024	0.190	0.017	0.172	0.281	0.126	0.104	0.150	0.120	0.482	0.358	**0.000**
Web/Pons05	**Allele N.**	17	10	7	4	10	13	5	6	16	12	3	9	12
	Ar	15.033	9.092	5.984	3.593	8.849	10.913	4.821	5.142	13.105	10.587	2.845	8.306	10.961
	He	0.949	0.885	0.718	0.607	0.830	0.890	0.604	0.718	0.866	0.876	0.528	0.847	0.912
	Ho	0.920	0.960	0.720	0.760	0.960	0.840	0.320	0.800	0.840	1.000	0.440	0.880	0.960
	HWE	0.497	0.935	0.324	0.083	0.437	0.632	**0.000**	0.446	0.232	0.575	0.032	0.304	0.954
WebPons06	**Allele N.**	22	18	8	7	11	19	3	7	17	15	2	10	15
	Ar	15.442	12.761	7.059	5.303	8.965	13.567	2.417	6.688	13.582	12.407	2.000	8.329	12.231
	He	0.936	0.917	0.779	0.699	0.823	0.917	0.387	0.806	0.926	0.905	0.430	0.851	0.891
	Ho	0.917	0.889	0.778	0.722	0.833	1.000	0.444	0.833	0.917	0.889	0.389	0.917	0.944
	HWE	0.704	0.531	0.018	0.612	0.049	0.557	0.736	0.535	**0.000**	0.301	0.697	0.216	0.453
Web/Pons07	**Allele N.**	18	12	11	6	11	18	4	7	16	16	3	10	17
	Ar	13.929	10.428	8.021	4.749	9.312	14.311	3.474	6.839	13.159	12.660	2.754	9.045	12.798
	He	0.896	0.891	0.747	0.631	0.833	0.934	0.445	0.820	0.889	0.922	0.472	0.885	0.920
	Ho	0.767	0.900	0.700	0.667	0.800	0.967	0.500	0.967	0.900	0.900	0.400	0.833	0.733
	HWE	**0.001**	0.222	0.016	0.789	0.272	0.453	0.117	0.741	0.599	0.060	0.026	0.248	**0.000**

		SSSP2201	SSOSL85	SSOSL438	SSA289	SSA171	SSA157	SSA14	SSSP1605	SSOSL311	SSA85	SSA412	SSA202	SSA197
Web/Cock	**Allele N.**	16	12	10	4	12	17	5	7	17	16	2	10	14
	Ar	12.735	9.735	8.167	3.967	10.570	12.906	3.837	6.657	13.888	13.233	2.000	8.919	11.873
	He	0.881	0.876	0.807	0.608	0.815	0.914	0.514	0.778	0.930	0.929	0.474	0.851	0.912
	Ho	0.903	0.871	0.806	0.516	0.677	0.903	0.484	0.742	1.000	0.935	0.484	0.839	0.839
	HWE	0.212	0.849	0.150	0.031	0.023	0.354	0.091	0.322	0.348	0.415	1.000	0.356	0.044
Hems	**Allele N.**	19	13	8	5	9	15	5	7	14	15	4	10	16
	Ar	14.242	10.554	6.517	4.198	7.203	11.988	4.390	6.616	11.275	12.525	3.742	8.824	12.222
	He	0.920	0.885	0.735	0.590	0.711	0.913	0.592	0.802	0.894	0.928	0.567	0.869	0.907
	Ho	0.800	0.767	0.700	0.300	0.533	0.900	0.500	0.733	0.867	0.833	0.333	0.833	0.867
	HWE	0.103	0.043	0.051	**0.001**	**0.001**	0.176	**0.001**	0.571	0.202	0.027	**0.000**	0.512	0.230
All	**Allele N.**	30	24	20	16	18	30	10	14	35	29	9	13	27
	Ar	15.723	11.802	8.686	5.409	9.266	14.271	3.983	7.110	14.270	12.856	2.955	9.008	12.656
	He	0.917	0.857	0.726	0.686	0.816	0.912	0.506	0.795	0.890	0.892	0.399	0.837	0.889
	Ho	0.895	0.859	0.653	0.623	0.769	0.878	0.474	0.772	0.881	0.891	0.326	0.846	0.863

Given are allele number (Allele N.), Allelic richness (Ar), Expected heterozygosity (He), Observed heterozygosity (Ho) and Hardy-Weinberg expectations (HWE; significant deviations indicated in bold).

Table 4.3 Genetic differentiation (pairwise F_{ST}) and geographic separation (km) of Dart populations

	Swinc	Strane	Prince	Black	Tor	Cher05	Cher06	Cher07	East05	East06	Walla	Cator	Pons05	Pons06	Pons07	Cock	Hems
Swinc	-	7	6	14	11	11	11	11	19	19	11	27	22	22	22	23	62
Strane	0.037	-	13	21	18	18	18	18	26	26	18	34	29	29	29	30	69
Prince	0.018	0.019	-	8	5	6	6	6	24	24	15	32	27	27	27	27	66
Black	0.039	0.035	0.019	-	9	14	14	14	31	31	23	39	34	34	34	34	74
Tor	0.007	0.041	0.021	0.046	-	11	11	11	28	28	20	36	31	31	31	32	71
Cher05	0.032	0.026	0.015	0.039	0.045	-	0	0	28	28	19	36	31	31	31	31	70
Cher06	0.037	0.051	0.033	0.060	0.040	0.053	-	0	28	28	19	36	31	31	31	31	70
Cher07	0.032	0.064	0.046	0.077	0.049	0.052	**0.078**	-	28	28	19	36	31	31	31	31	70
East05	0.022	0.023	0.017	0.038	0.028	0.026	0.039	0.066	-	0	10	32	27	27	27	28	67
East06	0.024	0.028	0.011	0.027	0.029	0.030	0.036	0.058	0.019	-	10	32	27	27	27	28	67
Walla	0.025	0.016	0.009	0.031	0.026	0.015	0.040	0.063	0.013	0.019	-	24	19	19	19	19	58
Cator	0.023	0.061	0.035	0.075	0.042	0.057	0.070	0.040	0.042	0.050	0.041	-	5	5	5	10	49
Pons05	0.038	0.032	0.019	0.045	0.043	0.036	0.049	0.066	0.028	0.029	0.022	0.037	-	0	0	5	44
Pons06	0.023	0.027	0.005	0.031	0.030	0.020	0.041	0.042	0.022	0.012	0.011	0.029	0.010	-	0	5	44
Pons07	0.016	0.042	0.017	0.047	0.024	0.048	0.053	0.050	0.024	0.025	0.026	0.016	0.024	0.012	-	5	44
Cock	0.027	0.050	0.031	0.060	0.031	0.056	0.060	0.043	0.034	0.040	0.037	0.017	0.030	0.021	0.008	-	44
Hems	0.026	0.017	0.006	**0.020**	0.027	0.026	0.036	0.062	0.020	0.017	0.007	0.048	0.018	0.011	0.025	0.030	-

Geographic waterway distances (km; above diagonal) and pairwise F_{ST} values (below diagonal); non-significant values indicated in bold.

Table 4.4 Self-assignment of Dart populations.

	Swinc	Strane	Prince	Black	Tor	Cher05	Cher06	Cher07	East05	East06	Walla	Cator	Pons05	Pons06	Pons07	Cock	Hems	Regional
Swinc	**6 (23)**		1 (4)	1 (4)	6 (23)	1 (4)		2 (8)		2 (8)	2 (8)	1 (4)		1 (4)	1 (4)		2 (8)	66
Strane		**19 (63)**	2 (7)	1 (3)		1 (3)			3 (10)		1 (3)				1 (3)	1 (3)	1 (3)	76
Prince	1 (3)	2 (6)	**9 (29)**			1 (3)			3 (10)		4 (13)		2 (6)	4 (13)		1 (3)	3 (10)	41
Black			3 (11)	**17 (63)**					1 (4)	2 (7)				3 (11)			1 (4)	74
Tor	4 (12)	1 (3)	2 (6)	1 (3)	**18 (53)**			1 (3)	3 (9)			1 (3)	1 (3)			2 (6)		80
Cher05	1 (7)	2 (13)				**6 (40)**	1 (7)				1 (7)			3 (20)			1 (7)	67
Cher06					3 (12)		**18 (69)**		2 (8)		1 (4)						2 (8)	81
Cher07	3 (14)							**14 (67)**		1 (5)	1 (5)	1 (5)				1 (5)		81
East05	1 (2)		1 (2)		1 (2)				**31 (69)**	3 (7)	2 (4)	1 (2)		2 (4)	2 (4)		1 (2)	80
East06	1 (2)	2 (5)	1 (2)	5 (13)					3 (8)	**22 (55)**	4 (10)			1 (2)			1 (2)	73
Walla		2 (6)	3 (9)				1 (3)	1 (3)	3 (9)	1 (3)	**13 (40)**			3 (9)	1 (3)		5 (15)	52
Cator	1 (3)		1 (3)		1(3)			1 (3)				**23 (74)**		2 (6)	3 (10)	3 (10)		100
Pons05			1 (4)			1 (4)							**17 (68)**	6 (24)				92
Pons/06			2 (6)	3 (8)		3 (8)		1 (3)		2 (6)	2 (6)		3 (8)	**17 (47**)	4 (11)			66
Pons07	1 (3)		1 (3)						1 (3)		1 (3)	6 (20)		3 (10)	**14 (47)**	3 (10)		80
Cock												2 (6)	1 (3)	1 (3)	4 (13)	**22 (71)**	1 (3)	96
Hems	1 (3)	3 (10)	2 (6)	1 (3)	1 (3)				2 (6)	2 (6)	3 (10)			2 (6)	1 (3)	2 (6)	**10 (33)**	

Given are the numbers (and percentage) of individuals probabilistically assigned to each population using the 'leave one out' method of Rannala and Mountain (1997) GeneClass *v.3*.2 (Raymond and Rousset, 1995; Piry *et al.*, 2004). Numbers in bold indicate assignment to correct population of origin, shading indicates region of origin. Correct assignment to regional of origin is given as a percentage but rounded to complete figure.

GENETIC DIFFERENTIATION AND RELATIONSHIPS AMONG POPULATIONS

Global F_{ST} was estimated at 0.0333, with pairwise estimates ranging from 0.006 (Blackbrook-Hems) to 0.078 (Cherry06-Cherry07). Although generally low, all tests were significant at $p<0.05$ after sequential Bonferroni corrections, except the two lowest values involving Blackbrook-Hems and Walla-Hems (Table 4.3). The Hems is the only sample taken from the lower reaches of the river Dart, and the close association it shows with many of the samples from the upper tributaries may be an indication that returning salmon which originated in the upper Dart tributaries do not make it upstream to their natal tributary, but remain in the lower tributaries of the Dart. However, this has not been fully assessed in this study. The Neighbour Joining population tree showed moderate support for some groupings of populations (Figure 4.3); however, often these populations were not geographically close to each other. The only exception were the two samples from the Swincome tributary (Swincome and Strane samples), however bootstrap support for this was poor, being <50%. Furthermore, temporal replicates from the same sample sites were not routinely grouped together, with the exception of Web/Pons samples from 2006 and 2007 (61%). Overall, only 54% of individuals were assigned to the correct population of origin (range: 23% Swincombe to 74% Web/Cator). When assigned to region of origin this rose to an average of 75% correct assignment (range: 41% West/Prince to 100% Web/Cator) (Table 4.4).

ASSESSING THE TEMPORAL COMPONENT IN GENETIC DIVERSITY

Hierarchical analysis demonstrated that a significant component (3.16%, $p<0.001$) of genetic variation was attributable to variance among sampling sites (geographic component), whereas only a small, non-significant component (0.24%, $p=0.12023$) of genetic variance was attributable to variance among temporal replicates within sampling sites. The remaining 96.60% was attributable to genetic variance within samples. Further evidence for the greater contribution of spatial variance in genetic structuring than temporal variance is provided by the comparisons of the number of loci showing significant differences in allele frequency distribution, D_A values and pairwise F_{ST} values. Temporal replicates within sites showed significantly fewer differences in allele frequency distributions when compared to spatial samples within years (Mann Whitney U test: $W=58.5$; $p<0.01$); similarly genetic distances (D_A) between temporal replicates were significantly lower than between spatial sites

(Mann Whitney U test: W = 64.0; P<0.05), however, there was no significant difference in the average F_{ST} values between spatial samples within years than for temporal samples within sites (Mann Whitney U test: W = 114.5; P>0.05). This indicates that, whilst significant variation can be observed between samples taken from the same site in different years, spatial variation accounts for most of the genetic diversity exhibited between populations within the river catchment.

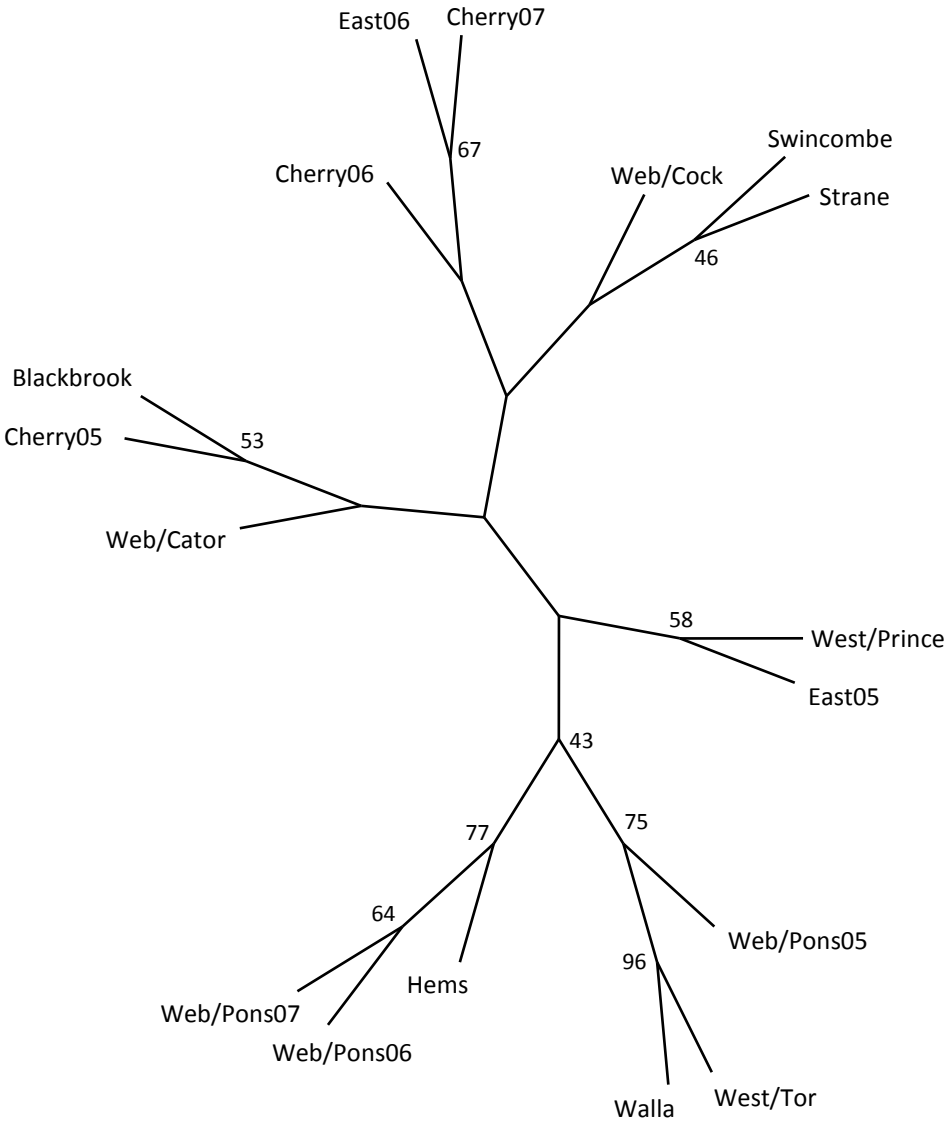


Figure 4.3 Neighbour Joining phylogenetic tree of salmon populations from the river Dart using microsatellite analysis.

ASSESSING SPATIAL GENETIC STRUCTURE

The clustering methods implemented in the program STRUCTURE did not partition individuals into distinct groups. Neither the ΔK method of Evanno *et al.* (2005) nor the *ad hoc* methods of Pritchard *et al.* (2000) produced robust estimates for the optimum number of clusters; rather, individuals were probabilistically assigned to numerous clusters with more or less equal probabilities, indicating no overarching structure of populations within the Dart catchment.

Tests for associations between genetic divergence ($F_{ST}/(1- F_{ST})$) and geographic distance revealed no significant association and therefore no isolation by distance signal (r_{xy}=-0.027, p=0.506, r^2=0.0007). However, it was noted that the Hems sample from the lower Dart showed close affinity to many of the upper Dart samples, as evidenced in F_{ST} and DA estimates, despite the large geographic separation, which may have affected the Mantel analysis. Accordingly, when the Hems sample was removed, highly significant isolation by distance between samples in the upper Dart was revealed (r_{xy}=0.272, p=0.004, r^2=0.074). To remove the effect of temporal variation, if only samples from 2006 were included for sites with temporal replicates this association increased further (r_{xy}=0.427, p=0.004, r^2=0.182) (Fig. 4.4). This therefore indicates that the process of isolation by distance is operating on the upper Dart tributaries.

Tests for spatial genetic structure at the individual level revealed that the genetic autocorrelation coefficient, *r*, was significantly positive for the 0-5 and 0-10 distance classes, intercepting the *y* axis at 17km. Significant negative genetic autocorrelation was observed for the 20 to 40km size classes (Fig. 4.5). The genetic correlation for increasing waterway distance classes indicated that *r* was significantly positive for waterway distance classes up to 30 km, above this positive genetic autocorrelation was no longer detectable (Fig. 4.6).

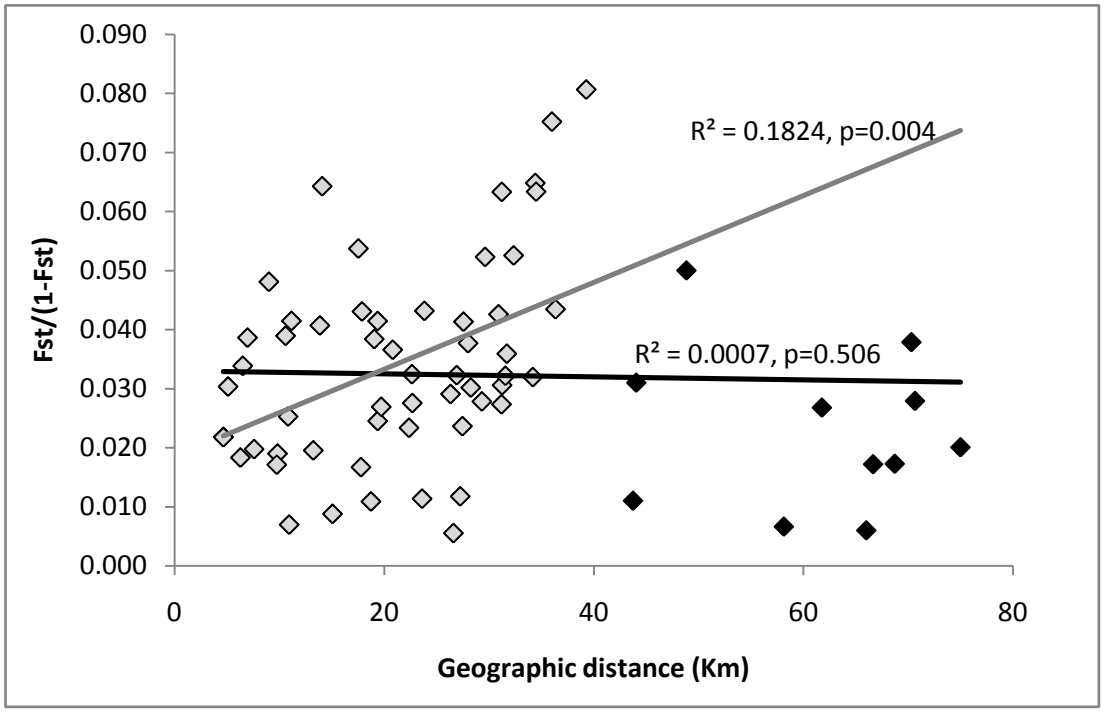


Figure 4.4 Isolation by distance of all Dart populations (all diamonds; black regression line) and of upper Dart 2006 populations only (no Hems, no temporal replicates; grey diamonds only, grey regression line).

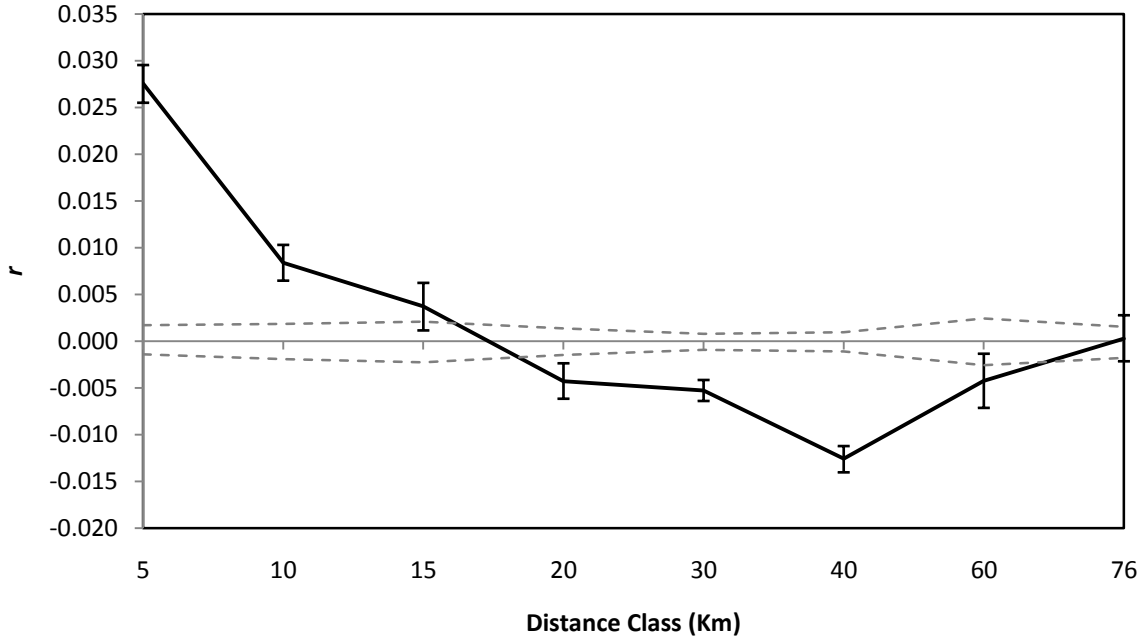


Figure 4.5 Genetic autocorrelation analysis of individual Dart salmon for different distance classes. Correlelogram showing the genetic correlation, *r*, as a function of distinct distance classes. Dotted grey lines indicate 95% CL about the null hypothesis of no genetic structure and error bars about *r* indicate 95% CI as determined by bootstrapping (see text for details).

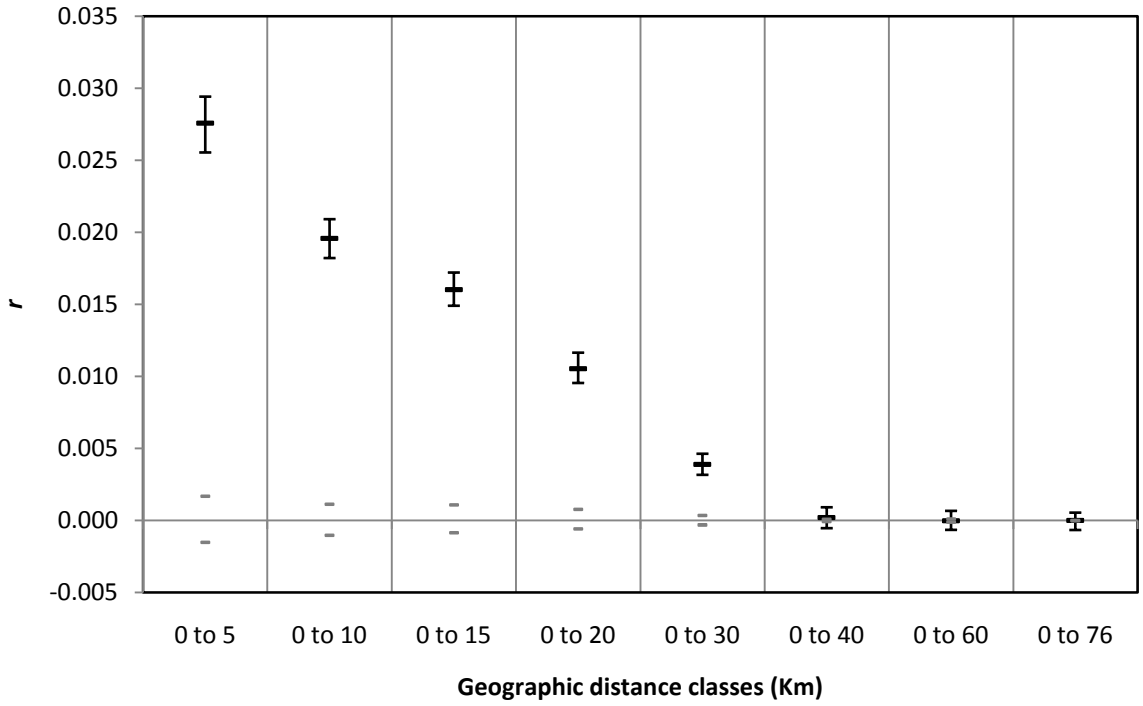


Figure 4.6 Genetic autocorrelation analysis of individual Dart salmon. Genetic correlation *r*, for increasing distance classes. Grey bars (-) indicate 95% CI about the null hypothesis of no genetic structure and error bars about *r* (-) indicate 95% CI as determined by bootstrapping.

DISCUSSION

Atlantic salmon samples within the river Dart catchment show weak but significant genetic differentiation. Temporal stability was not observed over short timescales but genetic diversity attributable to temporal variability was significantly less than genetic diversity attributable to spatial variation, however, there was a greater spread of spatial to temporal samples. A significant isolation-by-distance signal was detected among samples of the upper Dart tributaries and there is sufficient evidence to suggest that populations of Atlantic salmon in the river Dart follow the meta-population evolutionary model, and that treating each tributary as a distinct population or Operational Conservation Unit would be a prudent approach to catchment level management.

GENETIC DIVERSITY AND DIVERGENCE WITHIN THE RIVER DART

The high levels of polymorphism observed with the 13 microsatellite loci used in this study (275 alleles, ranging from 9 to 30 alleles at a single locus within a single sample) are consistent with those observed in other studies of within-river populations of Atlantic salmon (Garant *et al.*, 2000: 4-33 alleles at a single locus in a single population; Primmer *et al.*, 2006: 2-30 alleles at a single locus within a population). Genetic diversity, as measured using estimates of allelic richness (allele number corrected for sample size) and expected heterozygosity was higher in this study (Ar=7.06-10.60; He=0.68-0.82) than for previous within river studies undertaken in northern Europe (Primmer *et al.*, 2006: Ar=6.2-7.2, He=0.62-0.68; Vaha *et al.*, 2007: Ar=4.24-7.02, He=0.58-0.70), which is consistent with regional studies which have indicated greater genetic diversity in populations of the British Isles compared to other European regions (Verspoor *et al.*, 1999; Consuegra *et al.*, 2002; Saisa *et al.*, 2005; Chapter 3). Genetic divergence, as measured with pairwise F_{ST} estimates between samples within the river Dart (global F_{ST} =0.033; range: 0.00593-0.078) was comparable to the estimate of Garant *et al.* (2000) in their study of populations of the Sainte Marguerite river in Canada (Global F_{ST} =0.034; range: 0.0077-0.0874), which is higher than that observed by Primmer *et al.* (2006) for the Varzuga river on the Kola Peninsula, NW Russia (Global F_{ST} =0.014; range: 0.006-0.07) but lower than reported by Vaha *et al.* (2007) on the river Teno of Norway and Finland (F_{ST}: 0.001-0.138).

INFLUENCE OF PHYLOGEOGRAPHIC HISTORY ON CONTEMPORARY POPULATION STRUCTURING

Genetic diversity within the river Dart as revealed by PCR-RFLP analysis of mtDNA was comparable to that observed in the British Isles previously and higher that that observed in other regions of Europe (Nilsson *et al.*, 2001; Chapter 3). When analysed alongside data from northwest Europe in Chapter 3, the five samples from within the river Dart did not group together (Figure 4.1; 4.2). Rather, two samples grouped with a Spanish clade, one sample grouped with a French clade, while the remaining three samples grouped with other samples analysed previously from the UK. Data from Chapter 3 provided evidence to suggest that, alongside the Iberian Peninsula refuge in northern Spain, there was also a cryptic northern refuge for Atlantic salmon in northern France, with the UK being colonised by salmon expanding out from both of these refugia. Given the results of the phylogenetic tree, this then raises the theory of differential colonisation of tributaries within the river Dart from the two refugia, *i.e.* the Walla population is predominantly descended from French colonisers and the East06 and Cherry06 populations are predominantly descendants from Spanish colonisers. However, the allele size permutation test indicates that this is unlikely; pairwise R_{ST} values are not significantly greater than pairwise F_{ST} values, therefore stepwise mutations appear not to have contributed to the differentiation of these populations. Hence, differentiation is likely to have occurred in the last 10,000 years, *i.e.* after the retreat of the ice sheets and after the colonisation of the British Isles. The differences in RFLP haplotype frequencies must therefore be due to post-glacial processes, thereby allowing us to reject the theory that genetic differentiation of populations occurred before colonisation of the Dart while in different glacial refugia (H_2).

TEMPORAL VARIATION WITHIN THE RIVER DART

Significant genetic differentiation between temporal replicates taken from the same sampling location were observed for all three locations where temporal replicates were available (Table 4.3), moreover, temporal replicates from the same location did not always group together in the phylogenetic tree (Figure 4.3). On first consideration, such a result indicates that populations within the Dart catchment are not temporally stable over short time scales of two to three years. However, the relative genetic variation attributable to temporal variation within sampling sites is

lower (0.24%) than can be attributable to spatial variation among sampling sites (3.4%); this 0.24% was deemed a non-significant contribution to genetic variation in the AMOVA analysis. The greater variation attributable to spatial variation compared to temporal variation was also reflected in other analyses; genetic distances were found to be significantly shorter between temporal replicates within sites when compared to spatial samples between sites and there were fewer significant differences in allele frequency distributions between temporal replicates compared to among spatial samples. Hence, while significant genetic differentiation is observed between temporal replicates in the pairwise F_{ST} estimates, overall, genetic diversity within the catchment is predominantly attributable to spatial variation. However, it must be acknowledged that there are more spatial replicates than temporal, which may have some bearing on these results.

In previous studies which have explored temporal stability of Atlantic salmon populations within river catchments, results have been inconsistent. Vaha *et al.* (2008) found populations of the river Teno to be temporally stable over periods of up to 24 years and attributed 5.2% of the overall genetic diversity to variation among sampling locations with only a non-significant 0.7% of the genetic variation attributable to temporal replicates taken from the same location. In contrast, Garant *et al.* (2000) found four of seven Canadian salmon populations to be stable over a two year period, and found that genetic variability attributable to temporal replicates accounted for almost three times the genetic variability that spatial genetic variation did (AMOVA analysis: 0.9% spatial, 2.5% temporal). It may be the case that inter-annual temporal variability is common in all populations and it is only when systems are viewed on a much longer time frame, as was undertaken on the river Teno, that temporal stability can be observed. Furthermore, it has been suggested (Vaha *et al.*, 2008) that large salmon rivers, such as the Teno, may support temporally stable populations of salmon (Nielsen *et al.*, 1997; 1999; Hansen *et al.*, 2002; Vaha *et al.*, 2008), whereas populations in smaller rivers may be temporally unstable, as seems to be the case for brown trout (Ostergaard *et al.*, 2003). The river Dart is certainly a small salmon river and these results accord with this theory.

Spatial variation within the river Dart

Analysis at both the individual and population level indicated that genetic similarity in the headwater tributaries of the river Dart decreased with increasing geographic

separation. This was evidenced at the population level by a strong isolation by distance signal detected among the headwater populations (Figure 4.4) and was corroborated in the spatial autocorrelation analysis for individual samples. Here, a significant positive association between individuals separated by waterway distances of up to 30km was observed, indicating that migration and geneflow is likely up to this distance (Figure 4.6); for individuals separated by waterway distances of over 30km there is a negative association between genetic similarity and geographic separation indicating that migration and geneflow is limited over distances of 30km or more. Spatial autocorrelation analysis has been advocated as a useful tool in defining operational conservation units (OCUs) (Diniz-Filho and Teves, 2002). Specifically, the intercept of the correlelogram on the *x* axis has been used as an indication of the minimum distance between samples for which genetic diversity can be assessed and conserved effectively at low costs (Diniz-Filho and Teves, 2002); for the upper tributaries of the river Dart this would equate to 17km (Figure 4.5). The mainstem of the tributaries in the upper Dart are on average (S.D.) 13(±4)km in length (range: 10-23km), hence, under these circumstances each main tributary of the upper Dart would make a suitable OCU in order to maintain genetic diversity within the catchment. Further corroborating this are the results of the assignment tests, whereby correct assignment to population was low (average 54%, range: 23-74%) but rose substantially when assigned to tributary of origin (average 75%, range 41-100%; Table 4.4). Collectively this would suggest that there is substantial migration and geneflow within each major headwater tributary, which then decreases as you move out and away from the tributary. However, neither the clustering analysis nor the phylogenetic tree demonstrated these tributary level associations.

ROLE OF EVOLUTIONARY MODELS ACTING TO MAINTAIN WITHIN RIVER GENETIC STRUCTURING

The river Dart population is clearly not a single panmictic unit and therefore should not be managed as such. Clear genetic structuring and temporal stability were not observed; hence the member-vagrant model seems unlikely. Rather the weak but significant differentiation of populations seems more in accordance with the meta-population model. However, a significant isolation-by-distance signal was detected between the upper Dart tributaries, as would be consistent with the member-vagrant theory, hence an intermittent state between the two evolutionary theories could be

acting within the catchment, as has previously been suggested for the Sainte-Marguerite river in Canada (Garant *et al.*, 2000).

Garant *et al.* (2000) genotyped Atlantic salmon fry from seven locations in the Sainte-Marguerite River system (Canada) using five microsatellite loci; sampling locations were separated by waterway distances of 20-55km. Weak, but significant, genetic differentiation between samples but no isolation-by-distance was observed; similarly to this study, findings were most consistent with the meta-population model. Conversely, Primmer *et al.* (2006) genotyped Atlantic salmon parr from 11 locations of the Varzuga River (Russia) with 17 microsatellite loci. On average, sampling locations were situated 60 km apart (range: 5-165km) and a strong isolation-by-distance signal was revealed, leading the authors to conclude that the Atlantic salmon population of the Varzuga River best fitted the member-vagrant model. Similarly, Vaha *et al.* (2007) genotyped returning adults from 12 locations on the river Teno (Norway/Finland) using 32 microsatellite loci; strong genetic structuring was observed which related well to geographic location, and although a significant isolation-by-distance signal was not detected, the member-vagrant model rather than the meta-population model best explained the data.

Overall, a number of factors appear to have played a part in the different findings of these studies and the differences observed appear to be attributable to the number of microsatellite loci used, the life-stage sampled and the geographical scale of the study. Indeed, the catchment area of a tributary has recently been proposed to be one of the most important factors determining genetic diversity within catchments (Vaha *et al.*, 2008). Hence, it may the case that smaller rivers, such as the Dart and the Sainte-Marguerite are maintained through meta-population evolutionary processes (H_{1A}), which due to frequent extinction events precludes them from progressing onto the member-vagrant state (H_{1B}), while larger rivers, such as the Varzuga and Teno, have the capacity to reach the member-vagrant state and have the buffering capacity to deal with local extinction events within populations. This has previously been advocated by Primmer *et al.* (2006) and also seems to be the case for brown trout (Hansen *et al.*, 2002; Ostergaard *et al.*, 2003; Griffiths *et al.*, In press). Alternatively, in order to resolve clear genetic structuring using clustering methods and/or phylogenetic trees, as revealed by Vaha *et al.* (2008), additional microsatellite loci might be required.

MANAGEMENT AND CONSERVATION CONSIDERATIONS

In line with recommendations made for salmon in the river Teno (Vaha *et al.*, 2007; 2008), results from the spatial autocorrelation analysis would advocate that, for catchment level management, each main tributary should be regarded as a separate distinct population. Hence, this may be an indication that, regardless of catchment size, this might be a general phenomenon for Atlantic salmon populations. Under these circumstances it is to be recommended that movement of fish between tributaries due to human activities, *e.g.* supportive breeding programmes, should be avoided and that where supplementation programmes are used, adults used for broodstock should be taken from the same tributary and the progeny should be released in that same tributary in order to maximise their chances of survival. From a conservation perspective, multiple tributaries from across the catchment would need to be preserved in order to maintain the genetic diversity of salmon populations within the catchment.

CHAPTER 5 ASSESSING THE LONG-TERM GENETIC IMPACT OF HISTORICAL STOCKING EVENTS ON CONTEMPORARY POPULATIONS OF ATLANTIC SALMON (*SALMO SALAR* L.).

Published in *Fisheries Management and Ecology* (2008) **15**, 315-326.

Authors: A.K. Finnegan and J.R. Stevens.

A.K. Finnegan undertook all laboratory and statistical analysis and prepared all drafts of the paper.

J.R. Stevens provided guidance throughout and commented on earlier drafts of the paper.

INTRODUCTION

The supplementation of declining wild populations with individuals from a distant location or with those reared in captivity is a common management practice for many fish species (MacCrimmon and Gots, 1979; Hindar *et al.*, 1991; Dodson *et al.*, 1998; Aprahamian *et al.*, 2003; Araki *et al.*, 2007). The rationale behind these activities is generally case specific, but the majority aim to enhance natural productivity or maintain an artificial fishery and, ultimately, to provide long term benefits to the local riverine populations (Fleming and Petersson, 2001; Aprahamian *et al.*, 2003). However, the possibility of potential detrimental interactions between introduced and native fish have raised concerns as to the long term implications of such activities on native stocks (Youngson and Verspoor, 1998; Einum and Fleming, 2001; Wang *et al.*, 2001; Wang *et al.*, 2002; Youngson *et al.*, 2003; Araki *et al.*, 2007). These practices are particularly common with salmonid species (Hindar *et al.*, 1991) and often historical activities are well documented and archived scales are available. Here we attempt to determine the long-term impact of historical stocking activities on current populations of Atlantic salmon (*Salmo salar* L.).

Atlantic salmon populations are highly structured throughout their native range (Verspoor, 1997); strong genetic differentiation is observed between regions and among rivers (Ståhl, 1987; Nielsen *et al.*, 1996; King *et al.*, 2001; Consuegra *et al.*, 2002), with weaker, yet significant, differentiation of populations being observed between locations within river systems (Garant *et al.*, 2000; Nilsson *et al.*, 2001; Primmer *et al.*, 2006; Vaha *et al.*, 2007; Dillane *et al.*, 2008). This differentiation is in part the result of the high fidelity these salmonids show to their natal rivers (Stabell, 1984) and is further facilitated through the discontinuous nature of habitats within catchments, which acts to segregate spawning habitat. This observed genetic diversity is in most cases associated with differences in biological performance relevant to local survival and recruitment, which highlights an important concept; that individual populations possess only a fraction of the overall diversity of the species in which the potential ability to adapt to future environmental change lies (Boyce, 1992).

Hence, any activities that result in erosion of the natural genetic hierarchies observed in Atlantic salmon, even at the population level, have the potential to

negatively impact the genetic heritage of the species as a whole. This includes supplementation programmes, whereby the transfer of non-native fish may introduce novel genetically based characteristics into local wild populations leading to a reduction in fitness, possibly resulting from a loss of local adaptation (Taylor, 1991; Einum and Fleming, 2001; Araki *et al.*, 2007). Currently, common practice is to raise locally sourced broodstock in hatcheries before stocking out into recipient rivers, but this is also not without associated risks. Hatchery life may act to select for behavioural and physiological traits that are disadvantageous in nature (Einum and Fleming, 2001) the extreme environmental and ecological differences between hatchery and wild conditions (*e.g.* differing flow regimes and prey availability) exhort differing selection pressures on the fish (Youngson and Verspoor, 1998). Hatchery rearing substantially increases survival during egg and juvenile stages, thus bringing through genotypes that are potentially eradicated in the wild, and selection pressures during hatchery life may also act to select for behavioural and physiological traits that are disadvantageous in nature (Einum and Fleming, 2001) (See main Introduction pg 29 for more details).

Once transplanted out into the recipient rivers, the introduced salmon can interact with wild fish in a multitude of ways. Einum *et al* (2001) used a meta analytical approach, combining data from multiple ecological based studies and reported that hatchery fish consistently exhibit increased levels of aggression relative to wild fish and that they showed a reduced response to predation risk. Differences in the feeding behaviour, habitat use and morphology were also reported (Einum and Fleming, 2001). This unnatural behaviour of introduced fish in the wild appears to negatively impact the native population and long-term sustainability of the river population. For example, the increased levels of aggression of hatchery and farmed fish have been shown to displace native populations (McGinnity *et al*., 1997; McGinnity *et al*., 2003). The interbreeding of stocked and wild fish has also been reported, thus leading to introgression of foreign and native genes and the erosion of locally adapted gene complexes (Martinez *et al*., 2001a). Indeed, Araki *et al.* (2007) showed that the genetic effects of captive breeding programmes can reduce subsequent reproductive capabilities by up to 40%, further highlighting the potential long term repercussions of these programmes (Araki *et al.*, 2007). Hence, the lower reproductive success of the domesticated individuals can culminate in a lower

overall productivity of the river (McGinnity *et al*., 1997; Martinez *et al*., 2001a; McGinnity *et al*., 2003; Saisa *et al*., 2003; Saltveit, 2006). Furthermore, the introduced fish may undergo extensive migrations within the river system, thus the repercussions of introductions are not necessarily limited to the locations of input (Vasemagi *et al*., 2005).

Moreover, hatchery populations have been shown to have a lower genetic diversity, even in first generation hatchery populations, and reduced fitness when reproducing in the wild (Verspoor, 1988; Wang and Ryman, 2001; Koljonen *et al.*, 2002; Araki *et al.*, 2007). Once transplanted out into the recipient rivers, the introduced salmon can interact with wild fish in a multitude of ways, which in some cases has been shown to negatively impact the native individuals and the long term sustainability of the river population (Einum and Fleming, 2001; McGinnity *et al.*, 2003). However, the vast majority of management programmes do not carry out pre- or post-stocking assessments to monitor the outcome of their activities.

This investigation assessed the long-term impact of non-native fish introduced into the river Dart during the 1960s through the genetic analysis of contemporary Dart populations and historical source populations of the stocked fish. The river Dart (N50:20:35 W3:33:53) in southwest England drains the granite massif of Dartmoor National Park in a southerly direction, with a total catchment area of 475km^2. Dartmoor is characterised by high rainfall and acidic, nutrient poor peaty soil, and represents the largest expanse of unglaciated upland area in the British Isles. During the 1950s and 60s, as many as 300,000 eyed ova from hatcheries in Scotland or Iceland were transplanted into the River Dart each year. The rationale behind these stocking initiatives appears to be a desire to bring large spring-running fish to the river Dart for recreational sport fishing; such fish are commonly found in the Scottish and Icelandic source rivers. Our study of historical stocking records indicated the East Dart at Postbridge to be the only recipient tributary in the River Dart, and was duly prepared by the prior removal of all likely competitors *e.g*. brown trout, *Salmo trutta* L. and the native Atlantic salmon. During the 1950s and 1960s, the average number of returning adults to the Dart was 1390 (±361) and, at present, it is estimated that 1400 returning adults to the Dart would produce approximately 4 million eggs (Environment Agency, 2003). The 300,000 transplanted eggs during the 1950s and 60s would therefore make up 7.5% of this approximate figure, not an

inconsequential amount given that this practice occurred frequently over at least a 15 year period, but had certainly curtailed been by the 1980s at the latest.

Over 40 to 50 years have elapsed since the time of these stocking programmes and it is possible that if the stocked fish have persisted then some interbreeding with the native Dart fish may have occurred, creating a hybrid population. This investigation therefore aims to assess the potential of artificially increased genetic relationships between donor and recipient stocks of Atlantic salmon, if indeed this has occurred and is still detectable in current populations. In this way the long-term outcomes of these stocking programmes can be assessed.

METHODS

SAMPLE COLLECTION

Eight locations throughout the River Dart were sampled by electric fishing between May and September of 2005 and 2006 (Figure 2.3; Table 2.1). Adipose fin clips from 1+ parr or older were taken and stored in 95% ethanol. The Freshwater Fisheries Laboratories in Montrose donated scale samples of returning adults from 1968 from the rivers Tay and Tweed, both documented donor populations of the fish stocked into the river Dart. The last running fish of the season were chosen to be genotyped since broodstock were typically taken at this time of year (November) and it is thought that run timing may have a genetic component. The River Ellidaar was used in stocking of many European rivers during this period and appears the most likely source of the Icelandic fish. The Institute of Freshwater Fisheries in Iceland donated scale samples of returning adults to the River Ellidaar from July in 1962 and 1988 for analysis.

GENETIC ANALYSIS

DNA EXTRACTION

Genomic DNA was extracted from all adipose fin clips and from the dried, historical scale samples from the rivers Tay and Tweed using the Chelex method (Estoup *et al.*, 1996). DNA was extracted from river Ellidaar scale samples using either the Qiagen DNEasy kits (1962 samples) or using phenol:chloroform extractions (1988 samples) (Taggart *et al.*, 1992). These extraction methods maximized the quality of DNA from the more degraded Icelandic samples.

MICROSATELLITE AMPLIFICATION

Fourteen fluorescently labelled microsatellite loci (Table 2.5) were amplified in single 10μL PCR reactions containing 0.25 units *Taq* DNA polymerase (Bioline), 1X reaction buffer, 0.2mM dNTPs, 1.5mM $MgCl_2$, 0.5μM labelled primer and 1μM unlabelled primer. Amplification was undertaken in a thermal cycler block (Thermo) where samples were initially heated to 94˚C for 2 mins for denaturation then exposed to 35 cycles of 94˚C for 2 mins, annealing temperature X˚C for 30 secs (Table 2.5), 72˚C for 30 secs, with a final extension step of 72˚C for 10 mins. The 14 microsatellite loci were then pooled (Table 2.5) and subsequently genotyped on a Beckman-Coulter CEQ 8000 automated sequencer.

DATA ANALYSIS

POPULATION GENETIC PARAMETERS

Allele number and allelic richness (allele number corrected for sample size) were calculated for all loci within populations using FSTAT version 2.9.3 (Goudet, 1995). Deviations from Hardy Weinberg equilibrium across samples (within loci) and across loci (within samples) were estimated using the method of Guo and Thompson (1992) as implemented in GENEPOP *v3.4* (Raymond and Rousset, 1995), with sequential Bonferroni corrections applied for multiple tests (Rice, 1989). ARLEQUIN *v2.000* (Schneider *et al.*, 2000) was used to calculate genetic differentiation among samples by calculating pairwise F_{ST} values and their significance values.

ASSESSING GENETIC RELATIONSHIPS BETWEEN DONOR AND RECIPIENT STOCKS

INFERENCE OF POPULATION STRUCTURE

The underlying population structuring of the donor and recipient populations was inferred using STRUCTURE *v.2.0* (Pritchard *et al.*, 2000). This Bayesian Markov Chain Monte Carlo (MCMC) based approach uses a clustering based algorithm to partition individuals into groups based on criteria to minimise Hardy Weinberg disequilibrium within populations and linkage disequilibrium between loci within populations. For this, and all subsequent analyses, a burn-in period of 30,000 steps was set, followed by 1,000,000 MCMC replicates; allele frequencies were set as 'weakly correlated' as recommended by Pritchard *et al.* (2000). The inferred number of populations (k) was estimated using the *ad hoc* methods of Pritchard *et al.* (2000) and that of Evanno *et al.* (2005), and applied in subsequent analyses (k=11).

MODEL I: ESTIMATES OF CURRENT GENETIC RELATIONSHIPS BETWEEN DONOR AND RECIPIENT STOCKS

STRUCTURE was subsequently used to calculate individual admixture coefficients (Q) within the samples to provide an estimate of population admixture between the river Dart and the source locations of the stocked salmon. In the first analysis (which we term Model I) the source locations of the stocked fish were defined geographically (*i.e.* the Tweed, Tay and Ellidaar) and STRUCTURE partitioned the Dart fish in order to optimise the criteria of minimised Hardy Weinberg disequilibrium and linkage disequilibrium, with k=11 as estimated previously (see above). Hence, individual

fish clustered either with the three pre-defined source locations of the stocked fish, or formed new clusters of Dart fish, if this allowed for a more genetically realistic population. In this way, the genetic relationships of each individual fish with each population can be calculated with admixture coefficients (Q). Model I therefore estimates the maximum level of shared ancestry apparent in the genetic relationships between the donor and recipient stocks as calculated with individual and population level admixture coefficients. Critically, since no movement of fish is believed to have occurred between the source locations of the stocked fish, this analysis also provides a comparative estimate of natural genetic relationships among stocks within northwest Europe in situations where no stocking has taken place.

MODEL II: ESTIMATES OF HISTORICAL GENETIC RELATIONSHIPS BETWEEN DONOR AND RECIPIENT STOCKS

For comparative purposes, in Model II the process was reversed with the Dart populations defined geographically and the source locations partitioned according to the criteria of STRUCTURE. This approach allows a further assessment of natural genetic relationships between the donor and recipient stocks, albeit a maximum estimation based on the likelihood that, for example, the Icelandic fish are more likely to cluster together, rather than with Icelandic-Dart hybrids introgressed over a 40-50 year period. This is not intended be an accurate measure of genetic distance, but merely a comparative measure, relative to Model I, in order to account for the natural underlying genetic relationships. By calculating this relative estimate for the natural underlying genetic relationships (Model II) and subtracting this from the overall population level admixture coefficients (Model I), it is possible to assess the role of natural (historic) versus current (introgression) admixture.

RESULTS

POPULATION GENETIC PARAMETERS

After initial data exploration, the locus Ssa14 was shown to be out of Hardy Weinberg equilibrium (HWE) in nine of the 16 samples genotyped. This locus was therefore excluded from further analysis.

The allele range for some loci in this study is outside of that previously reported. Hence, for these individuals, hybrid tests (Pendas *et al.*, 1995) were undertaken to ascertain if they were brown trout or salmon/trout hybrids. These fish were found to be Atlantic salmon.

DNA extracted from the 1962 Icelandic sample (n=40) proved to be so degraded that it was only possible to reliably amplify 6 microsatellite loci; however, a sample of preserved scales from 1988 from the same river was also available for analysis. All 13 loci were successfully amplified from the 1988 sample and the pairwise F_{ST} value and its significance (calculated based on the six shared loci for the two Icelandic samples in Arlequin *v2.000*) showed no significant differentiation between the two time points (F_{ST}=0.00207; p>0.05). Thus, the 1988 sample from the river Ellidaar in Iceland was used in all subsequent analysis.

Less than 4% of all pairwise comparisons in the linkage disequilibrium tests were significant after sequential Bonferroni corrections. No more than eight (and generally less than five) of the pairwise tests were significant for any population and rarely were the same loci linked in differing populations, thus, overall, linkage disequilibrium was deemed to be negligible.

Successful multilocus genotypes based on 13 microsatellite loci were produced for 472 fish taken from a total of 11 sites across the four locations. Most samples were within Hardy Weinberg (HW) expectations, with no more than two loci in any population being out of HW equilibrium after sequential Bonferroni corrections (Table 5.1). However, six loci were out of HW equilibrium in the river Tay sample suggesting the possibility of disequilibrium. To verify that this result was not due to large allele drop-out due to poor DNA quality 48 of the 53 individuals from this sample were re-genotyped and no additional heterozygotes were observed. The sample was comprised of returning adult salmon and so it is

likely that, if there is population structuring within the Tay catchment, as might be expected in such a system, this sample is not representative of a single panmictic population, which could explain why HW expectations are not met for this sample.

The Scottish samples (rivers Tay and Tweed) had the greatest number of alleles and highest allelic richness of all samples (river Tay: average number of alleles: 15.6, average allelic richness: 10.587), whilst the river Ellidaar in Iceland had the lowest values (average number of alleles: 7.9; average allelic richness: 6.521), with all Dart samples falling within this range (Table 5.1).

Genetic differentiation between samples, as estimated by F_{ST}, ranged between 0.00494 (West/Prince-Hems) to 0.17244 (Eli-Taw). Of the 120 pairwise comparisons, only two were not significant (Walla-Hems: 0.00599; West/Prince-Hems: 0.00494) (Table 5.2). This, therefore, excludes random mating between the majority of samples and suggests population structuring within the Dart catchment.

Table 5.1 Microsatellite diversity indices by population.

		Sssp2213	sssp2201	Ssosl85	Ssosl438	Ssa289	Ssa171	Ssa157	Sssp1605	Ssosl311	Ssa85	Ssa412	Ssa202	Ssa197	All loci
East05	**Allele No.**	13	19	15	13	11	10	18	8	16	18	3	11	17	13
	Ar	10.570	13.286	11.029	9.075	8.430	8.061	12.423	6.975	11.260	13.591	2.568	9.642	12.178	9.930
	Ho	0.773	0.932	0.818	0.773	0.727	0.818	0.818	0.682	0.909	0.795	0.409	0.795	0.886	0.780
	He	0.894	0.922	0.874	0.850	0.812	0.850	0.906	0.791	0.869	0.932	0.434	0.877	0.911	0.840
	HWE	0.005	0.026	0.207	0***	0.001**	0.464	0.011	0.208	0.090	0.000	0.017	0.378	0.712	
East06	**Allele No.**	14	17	12	5	4	9	17	8	14	16	3	9	14	11
	Ar	11.031	11.602	10.704	4.757	3.761	7.437	12.725	7.054	10.299	11.580	2.375	7.273	10.832	8.572
	Ho	0.900	0.925	0.950	0.725	0.625	0.700	0.950	0.825	0.825	0.850	0.400	0.725	0.875	0.790
	He	0.891	0.904	0.910	0.716	0.666	0.836	0.916	0.822	0.883	0.892	0.386	0.811	0.872	0.808
	HWE	0.071	0.031	0.464	0.077	0.133	0.038	0.062	0.562	0.210	0.025	1.000	0.155	0.008	
Walla	**Allele No.**	14	18	16	8	7	9	17	8	19	14	3	8	14	12
	Ar	11.127	13.207	12.526	6.261	5.205	7.849	12.898	7.044	13.152	11.444	2.706	7.837	11.327	9.429
	Ho	0.879	0.909	0.788	0.576	0.576	0.667	0.848	0.879	0.818	0.818	0.333	0.879	0.909	0.760
	He	0.903	0.917	0.872	0.701	0.671	0.789	0.923	0.798	0.879	0.913	0.369	0.869	0.903	0.808
	HWE	0.410	0.673	0.0028*	0.009	0.006	0.008	0***	0.216	0.030	0.001	0.744	0.112	0.563	
West/Prince	**Allele No.**	16	23	16	11	7	12	21	8	20	22	6	12	18	15
	Ar	12.219	16.256	12.712	7.505	5.839	10.062	16.223	7.329	14.708	14.743	4.614	9.279	13.367	11.143
	Ho	0.844	0.906	0.875	0.750	0.531	0.719	0.844	0.813	0.938	0.875	0.188	0.750	0.844	0.760
	He	0.911	0.946	0.899	0.717	0.755	0.869	0.950	0.833	0.932	0.929	0.425	0.850	0.924	0.842
	HWE	0.221	0.298	0.415	0.560	0.0008*	0.022	0.089	0.295	0.297	0.225	0***	0.115	0.053	
Blackbrook	**Allele No.**	9	12	15	7	5	8	15	9	11	10	3	7	12	9
	Ar	8.387	10.951	11.709	6.077	4.772	6.801	12.312	7.209	9.027	8.458	2.973	6.126	11.056	8.143
	Ho	0.923	0.923	0.846	0.808	0.577	0.654	0.846	0.846	0.769	0.846	0.077	0.923	0.808	0.757
	He	0.849	0.915	0.882	0.742	0.667	0.748	0.894	0.711	0.817	0.847	0.431	0.800	0.901	0.785
	HWE	0.138	0.022	0.008	0.226	0.155	0.099	0.038	0.546	0.058	0.066	0***	0.0008*	0.047	

		Sssp2213	sssp2201	Ssosl85	Ssosl438	Ssa289	Ssa171	Ssa157	Sssp1605	Ssosl311	Ssa85	Ssa412	Ssa202	Ssa197	All loci
Cherry05	**Allele No.**	11	12	8	4	5	5	16	8	12	13	3	9	9	9
	Ar	11.000	12.000	8.000	4.000	5.000	5.000	16.000	8.000	12.000	13.000	3.000	9.000	9.000	8.846
	Ho	0.867	1.000	0.800	0.800	0.667	0.867	0.867	0.867	1.000	0.800	0.200	0.733	0.800	0.790
	He	0.903	0.913	0.784	0.646	0.683	0.763	0.952	0.853	0.903	0.878	0.297	0.825	0.862	0.789
	HWE	0.350	0.968	0.224	0.070	0.257	0.627	0.086	0.335	0.372	0.113	0.327	0.133	0.573	
Cherry06	**Allele No.**	10	12	9	6	5	9	9	8	14	12	3	8	9	9
	Ar	8.861	9.953	6.847	5.366	4.928	7.633	8.085	7.200	10.765	10.640	2.973	7.296	8.352	7.608
	Ho	0.846	0.885	0.692	0.577	0.769	0.885	0.885	0.769	0.885	1.000	0.346	0.962	0.846	0.796
	He	0.861	0.847	0.693	0.640	0.733	0.835	0.836	0.819	0.878	0.903	0.410	0.819	0.830	0.777
	HWE	0.208	0.356	0.491	0.062	0.256	0.158	0.026	0.332	0.279	0.116	0.140	0.165	0.446	
Strane	**Allele No.**	10	14	13	6	6	8	13	6	12	7	3	8	13	9
	Ar	9.114	11.487	10.158	4.869	5.482	7.021	11.219	5.805	9.519	6.692	2.754	7.560	11.008	7.899
	Ho	0.900	0.833	0.867	0.500	0.633	0.633	0.800	0.600	0.633	0.933	0.633	0.833	0.833	0.741
	He	0.885	0.908	0.799	0.641	0.742	0.762	0.908	0.724	0.800	0.832	0.484	0.837	0.900	0.786
	HWE	0.026	0***	0.564	0.012	0.0011*	0.021	0.0027*	0.178	0.013	0.005	0.009	0.010	0.0023*	
Web/Pons05	**Allele No.**	14	18	12	8	5	10	14	7	17	14	5	9	14	11
	Ar	12.497	15.408	10.173	6.732	4.144	8.743	11.572	5.910	13.536	11.604	3.980	8.257	12.013	9.582
	Ho	0.769	0.923	0.962	0.692	0.769	0.923	0.808	0.769	0.846	1.000	0.462	0.846	0.962	0.825
	He	0.921	0.950	0.894	0.739	0.618	0.830	0.897	0.738	0.875	0.885	0.564	0.851	0.919	0.822
	HWE	0.056	0.525	0.838	0.079	0.124	0.491	0.445	0.136	0.091	0.476	0.0021*	0.231	0.920	
Web/Pons06	**Allele No.**	14	22	18	8	7	11	19	7	17	15	2	10	16	13
	Ar	11.344	15.442	12.761	7.059	5.303	8.965	13.567	6.688	13.582	12.407	2.000	8.329	12.772	10.017
	Ho	0.806	0.917	0.889	0.778	0.722	0.833	1.000	0.833	0.917	0.889	0.389	0.917	0.944	0.833
	He	0.905	0.936	0.917	0.779	0.699	0.823	0.917	0.806	0.926	0.905	0.430	0.851	0.921	0.832
	HWE	0.037	0.766	0.491	0.014	0.621	0.033	0.538	0.537	0***	0.292	0.696	0.212	0.247	
Hems	**Allele No.**	14	20	13	8	6	10	15	7	14	15	4	10	16	12
	Ar	11.153	14.885	10.554	6.517	4.890	7.671	11.988	6.616	11.275	12.525	3.742	8.824	12.222	9.451
	Ho	0.667	0.900	0.767	0.700	0.433	0.800	0.900	0.733	0.867	0.833	0.333	0.833	0.867	0.741
	He	0.893	0.925	0.885	0.735	0.602	0.777	0.913	0.802	0.894	0.928	0.567	0.869	0.907	0.823
	HWE	0.0037*	0.427	0.043	0.052	0.278	0.658	0.152	0.573	0.135	0.023	0.0003**	0.535	0.268	

		Sssp2213	sssp2201	Ssosl85	Ssosl438	Ssa289	Ssa171	Ssa157	Sssp1605	Ssosl311	Ssa85	Ssa412	Ssa202	Ssa197	All loci
Tweed	**Allele No.**	13	23	12	14	4	17	19	10	24	18	4	12	21	15
	Ar	10.441	15.606	8.518	8.703	3.957	11.494	13.938	7.896	15.029	12.265	2.638	10.084	14.942	10.424
	Ho	0.660	0.830	0.787	0.702	0.745	0.894	0.830	0.681	0.809	0.957	0.298	0.872	0.809	0.759
	He	0.891	0.945	0.847	0.808	0.679	0.890	0.935	0.855	0.924	0.910	0.405	0.889	0.937	0.840
	HWE	0.0004**	0.008	0.011	0.030	0.339	0.255	0.014	0.010	0***	0.381	0.066	0.010	0.025	
Tay	**Allele No.**	17	23	14	11	7	15	26	10	24	19	4	11	22	16
	Ar	12.331	15.440	9.260	7.302	5.123	10.675	16.626	7.458	15.707	12.570	2.976	8.520	13.648	10.587
	Ho	0.642	0.774	0.774	0.566	0.698	0.736	0.698	0.736	0.868	0.906	0.396	0.660	0.925	0.721
	He	0.915	0.942	0.842	0.733	0.702	0.871	0.953	0.808	0.945	0.915	0.516	0.867	0.923	0.841
	HWE	0***	0***	0.066	0.0029*	0.0013*	0.318	0***	0.162	0.112	0.405	0.007	0.0016*	0.244	
Ellidaar	**Allele No.**	7	15	8	6	4	11	13	6	9	6	2	6	10	8
	Ar	6.573	12.366	5.655	4.311	3.132	8.239	10.775	5.767	7.068	4.868	2.000	5.084	8.937	6.521
	Ho	0.588	0.941	0.353	0.529	0.353	0.853	0.853	0.706	0.824	0.706	0.471	0.618	0.765	0.658
	He	0.814	0.917	0.504	0.623	0.383	0.843	0.909	0.717	0.766	0.679	0.479	0.705	0.869	0.708
	HWE	0.053	0.874	0.0028*	0.229	0.035	0.683	0.445	0.411	0.868	0.399	1.000	0.667	0.007	
Means	**Allele No.**	13	17	13	8	6	11	16	8	16	14	3	9	14	11
	Ar	10.464	13.106	9.767	6.152	4.854	8.328	12.673	6.886	11.711	11.024	2.870	8.043	11.367	9.019
	Ho	0.803	0.891	0.793	0.671	0.609	0.782	0.868	0.780	0.857	0.878	0.359	0.820	0.866	0.767
	He	0.888	0.915	0.827	0.708	0.653	0.813	0.912	0.783	0.876	0.880	0.430	0.838	0.895	0.802
	HWE	0.139	0.414	0.298	0.177	0.201	0.247	0.154	0.304	0.239	0.161	0.352	0.213	0.316	

Figures provided are Allele No. (number of alleles); AR (allelic richness); Ho (observed heterozygosity); He (expected heterozygosity); HWE (Hardy Weinberg Equilibrium).Significant deviations (after Bonferroni corrections) indicated as * at 0.05, ** at 0.01, *** at 0.001.

Table 5.2 Genetic differentiation of populations as estimated with pairwise *F*ST between Atlantic salmon samples.

	East05	East06	Walla	West/Prince	Black	Cherry05	Cherry06	Strane	Web/Pons05	Web/Pons06	Hems	Tweed	Tay	Ellidaar
East05	0													
East06	0.01825	0												
Walla	0.01237	0.01789	0											
West/Prince	0.01561	0.01036	0.00653	0										
Black	0.03754	0.02382	0.02893	0.0171	0									
Cherry05	0.02774	0.03194	0.01647	0.01395	0.04268	0								
Cherry06	0.03765	0.03641	0.03801	0.02924	0.05772	0.05591	0							
Strane	0.02446	0.02887	0.01727	0.01764	0.03433	0.02946	0.0507	0						
Web/Pons05	0.02418	0.02604	0.01925	0.01676	0.03729	0.03311	0.04763	0.0302	0					
Web/Pons06	0.01898	0.01138	0.00927	0.00497	0.02529	0.02015	0.04158	0.02621	0.00685	0				
Hems	0.01889	0.01442	**0.00599**	**0.00494**	0.01864	0.02647	0.03478	0.01646	0.01566	0.00846	0			
Tweed	0.03106	0.03606	0.02624	0.02379	0.04202	0.0408	0.05258	0.05255	0.02492	0.01724	0.02642	0		
Tay	0.03001	0.03995	0.02894	0.03	0.05282	0.04651	0.05422	0.05306	0.02182	0.02293	0.02595	0.00817	0	
Ellidaar	0.1105	0.11552	0.11697	0.11203	0.13674	0.1346	0.15211	0.12868	0.08384	0.09192	0.10775	0.10157	0.10414	0

Non-sgnificant values are given in bold

INFERENCE OF POPULATION STRUCTURING

The optimal number of clusters, as estimated using both the Pritchard *et al*. (2000) and Evanno *et al.* (2005) methods, was 11; one distinct Icelandic cluster (river Ellidaar), two distinct Scottish clusters and eight Dart clusters. This value of k was used in all subsequent analysis.

MODEL I: ESTIMATES OF CURRENT GENETIC RELATIONSHIPS BETWEEN DONOR AND RECIPIENT STOCKS

For each individual fish, admixture coefficient values (Q) indicating the level of ancestry within each cluster created by STRUCTURE were calculated (Model I). An admixture coefficient of 1 indicates full ancestry within a population; conversely, an admixture coefficient of 0 indicates no ancestry in that population. The distribution of Q values of Dart fish in each of the pre-defined donor populations indicated that shared ancestry levels are generally low and that shared ancestry is highest between the Dart and the River Tweed, and lowest between the Dart and the river Ellidaar (Table 5.3). Of note, the tributary into which the salmon were stocked (East Dart at Postbridge) exhibited average levels of shared ancestry, but samples from the West Webburn at Ponsworth (2005 and 2006) had consistently high admixture coefficients with all donor populations, which may be attributable to the donor stocks (Table 5.3). Less than 1% of shared ancestry between all three source locations was observed using this model, indicating a possible base-level estimate of natural genetic relationships with which to compare this figure.

MODEL II: ESTIMATES OF HISTORIC GENETIC RELATIONSHIPS BETWEEN DONOR AND RECIPIENT STOCKS

In the second analysis, where only Dart populations were defined geographically (Model II), STRUCTURE identified the Icelandic fish as a very distinct cluster (Table 5.4). Conversely, the Scottish fish did not cluster into discrete Tweed and Tay populations, but rather they formed two largely Scottish clusters (Table 5.4). In this second analysis, the river Ellidaar fish appeared to have only 1-2% shared ancestry with the Dart, Tweed or Tay, in agreement with the output from the previous analysis (Model I). Admixture of the Scottish fish was generally attributed to the two Scottish clusters, with 1-3% ancestry from within the Ellidaar and ~5% ancestry

from within the Dart, implying that the estimates of natural levels of relatedness calculated previously may be an underestimate.

Table 5.3 Results from admixture analysis (Q) carried out using STRUCTURE; DOnor populations (Tweed, Tay, Ellidaar) pre-defined geographically.

	Admixture In:					
	Tweed		Tay		Ellidaar	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
East05	0.052	0.101	0.051	0.09	0.019	0.022
East06	0.027	0.038	0.022	0.027	0.014	0.018
Cherry05	0.052	0.056	0.038	0.036	0.021	0.023
Cherry06	0.023	0.04	0.017	0.016	0.016	0.023
Web/Pons05	0.231	0.182	0.147	0.172	0.073	0.094
Web/Pons06	0.146	0.143	0.092	0.142	0.039	0.049
Strane	0.012	0.007	0.011	0.007	0.019	0.044
West/Prince	0.069	0.094	0.056	0.08	0.033	0.067
Black	0.014	0.012	0.011	0.008	0.01	0.01
Hems	0.048	0.061	0.07	0.074	0.047	0.084
Walla	0.075	0.11	0.054	0.052	0.024	0.032
Dart Average	**0.068**	**0.111**	**0.053**	**0.091**	**0.028**	**0.052**
Tweed	—		0.008	0.035	0	0
Tay	0.002	0.004	—		0	0
Ellidaar	0.001	0.005	0.001	0.002	—	

Table 5.4 Results from admixture analysis (Q) carried out using STRUCTURE; Recipient populations (Dart) defined geographically.

	Admixture In:							
	Dart		Tweed		Tay		Ellidaar	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Dart	0.84	0.29	0.049	0.041	0.046	0.028	0.012	0.004
Scot1	<0.01	0	0.314	0.249	0.313	0.229	0.015	0.026
Scot2	<0.01	0	0.125	0.099	0.163	0.130	0.012	0.019
Ellidaar	<0.01	0	0.027	0.044	0.015	0.020	0.842	0.164

DISCUSSION

Measures of admixture between current native Atlantic salmon populations and non-native populations stocked into the recipient river 40-50 years previously generally show a minimal long-term impact of the stocked fish but high variability between tributaries of the recipient river. Between <1 to 7% population level admixture between contemporary Atlantic salmon populations of the river Dart and historical samples taken from populations used to stock this river during the 1950s and 1960s was observed. After allowances were made for the natural underlying genetic relationships between the donor and recipient populations this value decreased to 1-2% population level admixture that could potentially be attributable to the long-term persistence of the progeny of native and non-native hybrids. Given that it is not possible to satisfactorily calculate confidence intervals and the high variability, it is impossible to conclude with confidence whether this represents the persistence of the progeny of the stocked and native fish over a 40-50 year time period and, overall, this indicates a long-term dominance of native over non-native fish at a catchment level. However, these levels vary among tributaries with the effects appearing to be highest in samples from the West Webburn at Ponsworthy, which was not the original recipient tributary for the stocked fish. In this tributary, admixture coefficient values were consistently higher than elsewhere in both sampling years (3-18% after allowances were made for natural levels of genetic relationships). Again, high variability is observed here, but this does potentially suggest that the non-native stocked fish may have experienced localized success in this tributary.

ESTIMATES OF UNDERLYING HISTORICAL GENETIC RELATIONSHIPS

Movement of fish in this case was unidirectional (*i.e.* only into the Dart and not out of) and, as far as the authors are aware, there has been no man-mediated movement of fish between the source locations of the stocked fish (*i.e.* between the rivers Tweed, Tay and Ellidaar). This allowed estimates of shared ancestry between the donor populations to be an estimate of the underlying levels of shared ancestry between these populations at a regional level, which was calculated as <1%. Note, these estimates are not intended to be representative of the precise genetic distances between populations - rather, they are presented as a means of providing relative estimates of the underlying relationships between populations, undertaken within the same framework as the admixture analysis in this study. Hence, this provides an

estimate of underlying genetic relationships between populations relative to the admixture model used in the main analysis (Model I), thereby allowing for comparisons to be drawn. By undertaking the second analysis, in which the Dart fish were defined geographically and the source fish were allowed to partition themselves according to the assumptions of the model (Model II), a further assessment of how much admixture the source locations have with the recipient river populations was possible. Of course, these higher estimates may be skewed if introgression between the stocked and native fish has occurred, but given the long interval, the influence of these hybrid fish in the calculation of the admixture coefficients is likely to be minimal. These maximum estimates of natural genetic relationships between the Dart population and the Scottish and Icelandic source rivers are calculated as 5% with the rivers Tweed and Tay, and 1% with the Ellidaar.

Continued impact of Icelandic fish in the Dart

The Ellidaar fish consistently formed a distinct cluster of individuals at each level of analysis; similarly, only 1% admixture of Ellidaar fish with the two dominant Scottish groupings was consistently indicated. Conversely, there appears to be approximately 3% admixture of Dart and Ellidaar fish. This, therefore, suggests that at most 1% of the admixture between the Dart and Ellidaar fish is due to the underlying natural shared ancestry continuing from the last glacial period, and that 2% is due to admixture of Dart and Ellidaar fish resulting from interbreeding between the stocked and native fish. With the high levels of variation observed, it would suggest that the long-term persistence of Icelandic fish in the Dart was minimal or non-existent.

Continued impact of Scottish fish in the Dart

The situation with the Scottish fish is less clear-cut since they fail to form such distinct Tweed and Tay clusters. However, results indicate ~5% population admixture attributable to the underlying historical shared ancestry and 5-7% admixture when this is not taken into account. Hence, if we assume that 5% admixture between the Dart and the two Scottish rivers is a maximum estimation of natural, historic ancestry, then we can attribute at least 2% admixture with the Tweed to the persistence of stocked fish introduced during the 1950s and 1960s, but none with the River Tay. Again, with the high variability displayed, it is impossible to

conclude that there has been any significant long-term impact of the stocked fish, despite the long-running nature of the programme and the high numbers of fish used.

VARIATIONS BETWEEN TRIBUTARIES

Despite the low average levels of admixture at a catchment level, one tributary consistently exhibited considerably higher admixture levels with the donor populations. The samples from the West Webburn at Ponsworthy, taken in consecutive years from the same location, indicated as much as 11-18% admixture with the Tweed, 4-10% with the Tay and 3-6% with the Ellidaar, even when the underlying relationships between the donor and recipient populations are taken into account. In spite of the high variability exhibited, it is impossible to conclude with certainty that this is not due to the persistence of the progeny of the fish stocked into the Dart 40-50 years previously and, as such, may be evidence of the continued long-term impact of the stocked fish at a very local scale. This tributary was not the recipient tributary into which the stocked fish were originally introduced, but a large neighbouring tributary, slightly down-stream of this. The straying of stocked fish has been noted elsewhere (Vasemagi *et al.*, 2005a), as may be the case here. The observed higher admixture coefficient values observed with the Tweed, relative to the Tay or Ellidaar, is possibly due to the frequency with which the fish were stocked; it is known that the Tweed was used as a donor population on a number of occasions, but the Tay and the Ellidaar are reportedly used only once (though this may be a factor of reporting, rather than a reflection of the true situation). Quite why the persistence of the stocked fish is greater in the West Webburn (Web/Pons) tributary is unknown. This tributary is slightly down-stream from the original recipient tributary, so perhaps returning adults encountered these suitable spawning grounds sooner on the return migration; perhaps the conditions in this tributary are more suitable for the Scottish and Icelandic fish, or maybe this tributary was also used as the recipient tributary at other times, which have not been documented.

From these analyses, at a catchment-wide level it appears that the stocking practices undertaken during the 1950s and 60s have had a minimal long-term impact on the resident Dart populations. However, one tributary exhibited consistently higher admixture coefficients, indicating that analyses at the catchment-level may mask

localized effects. It is possible that the higher admixture coefficient values for this population are due to the long-term persistence of the progeny of the stocked fish over the last 40-50 years. These estimates of 3-18% are comparable to those observed by Martinez *et al.*, 2001a) when assessing the effect of stocking Scottish salmon into the Nivelle River in France. In the Nivelle study, less than 10% introgression between the stocked and native fish was reported after a 13 year period since the stocking events, but no allowances for natural underlying genetic relationships were made. In a comparative study in brown trout (*Salmo trutta*) into the effects of stocked hatchery fish on the indigenous populations of two recipient rivers, similar methodologies implemented in STRUCTURE estimated that there was approximately 8% admixture between the stocked and recipient populations in a population in which stocking had desisted, but 18% admixture between a second endemic population and the stocked fish in a river in which stocking was ongoing (Hansen *et al.*, 2001). The higher levels of admixture in the population that continued to be stocked, relative to the population where stocking had ceased, may suggest that a wild population can potentially revert back to its original, natural state over time, at least in part. It may therefore be possible that the low levels of continued hybridisation potentially observed in our study may be indicative of the wild Dart population acting to purge the non-native genotypes introduced through stocking practices and, given enough time, may eventually fully remove the genetic impact of stocking. Indeed, there is increasing evidence to suggest that non-native or hatchery reared fish do not perform as well as their wild, native counterparts after introduction (Einum and Fleming, 2001; Einum *et al.*, 2003; Araki *et al.*, 2007), and that often introduced fish undergo extensive migrations within their new catchment, thus not confining the ramifications to the location of input (Vasemagi *et al.*, 2005a), which is also consistent with the findings presented here.

IMPLICATIONS FOR MANAGEMENT

The actual consequences of any stocking activities are difficult to discern with any great confidence. However, as the adoption of the precautionary approach spreads increasingly to management agendas, these findings have already been shown to be a useful tool in responsible Atlantic salmon management. In the latest recommendations to the Freshwater Fisheries Act review (MAFF, 2001), it was implored that supplementation programmes should be undertaken only as a last

resort, while the reasons behind the declines are explored and remedied if at all possible. Moreover, if supplementation is deemed necessary, best practice guidelines should be adhered to. The MAFF (2001) recommendations included the directive that, where possible, fish to be used as broodstock should be taken from rivers or tributaries which have been minimally affected by previous supplementation or stocking activities. It is suggested from this investigation that the impacts of past stocking regimes are not consistent throughout a catchment, with levels of admixture between the native and introduced fish varying between tributaries. The river Dart currently has a supplementation programme operating whereby offspring from river Dart broodstock are raised to the smolt stage in a local hatchery before being released into the main river. Based on the results of this study the local groups undertaking the work are endeavouring to obtain their broodstock from the apparently less impacted tributaries of the river Dart. Certainly our findings highlight the importance of determining the outcome of historical stocking and supplementation programmes, as well as current programmes, in order to effectively manage riverine Atlantic salmon stocks, while minimizing potential detrimental effects.

CONCLUSIONS

With regards to the initial premise behind the historical stocking activities on the Dart, which was to introduce large spring-running fish to the catchment, these activities can be deemed failures. The current Dart population is primarily made up of late-running fish, the population has continued to decline since the 1960s and is currently at its most depleted state on record with egg deposition rates also at their lowest, falling well below their target threshold (Environment Agency, 2003). Although long-term introgression between the stocked and wild fish appears to be low, if present at all on a catchment scale, one tributary does potentially display a long-term impact of stocking and therefore genetic erosion of the Dart population. This may have resulted in reduced fitness in the Dart population, and thus have reduced the natural productivity of the population further (Taylor, 1991; Einum and Fleming, 2001; Araki *et al.*, 2007), hence exacerbating the observed decline. However, over the whole catchment our findings demonstrate the resilience of native stocks, which continue to dominate despite efforts to displace them with non-native

counterparts. This presumably reflects the fact that native fish are inherently better adapted to their home environments.

CHAPTER 6: THE GENETIC STRUCTURING OF ATLANTIC SALMON POPULATIONS IN THE RIVER TWEED: IMPLICATIONS FOR MANAGEMENT AND CONSERVATION.

The genetic analysis undertaken in this chapter was funded by the Tweed Foundation and the Atlantic Salmon Trust.

Chapters 6 and 7 have been combined with a non-technical summary and submitted to the Tweed Foundation and Atlantic Salmon Trust.

INTRODUCTION

It has long been appreciated that genetically distinct populations of Atlantic salmon (*Salmo salar* L.) occur in neighbouring catchments, which has largely been attributed to the precise homing ability of the species (Stabell, 1984). However, the occurrence of a number of differentiated populations inhabiting a single river system is now widely documented in numerous studies over the species range (Stabell, 1984; Elo *et al.*, 1994; Beacham and Dempson, 1998; Garant *et al.*, 2000; Primmer *et al.*, 2006; Vaha *et al.*, 2007). It is thought that populations became differentiated in this way due to the discontinuous nature of spawning habitat, thus isolating groups of breeders, and the subsequent adaptations these groups made to their local environment in order to maximise survival and recruitment in that river or tributary (see Garcia de Leaniz *et al.*, 2007 and references therein).

To effectively manage and conserve these within river stocks it is necessary to define the extent of population structuring within the catchment; this requires some understanding of what is driving and maintaining this phenomenon. Three evolutionary theories have been proposed as potential mechanisms driving within river population structuring of Atlantic salmon (Garant *et al.*, 2000); namely panmixia, 'meta-population' and 'member-vagrant' theories. A panmictic population is one where individuals can move about freely within their habitat and breed at random; hence all individuals are potential partners. By definition, for a panmictic population to be achieved there can be no barriers to movement and the individuals within the population must be pre-disposed to migration within the population area. Several factors indicate that Atlantic salmon populations within rivers are unlikely to be panmictic. Firstly, river catchments are often highly dendritic systems and frequently there are barriers to migration, both natural and man-made. Over time this will act to segregate individuals and populations will become differentiated due to random genetic drift. Alongside this, evidence suggests that Atlantic salmon do not undertake long migrations within catchments, rather they remain close to their natal area as both fry and parr (McCormick *et al.*, 1998). This characteristic of remaining close to their 'home area' where they were spawned is maintained by rheotactic behaviour to olfactory cues; the juvenile fish demonstrate a positive rheotaxis to odorant cues, and a negative rheotaxis in the absence of it (Saunders and Gee, 1964). Furthermore, it is a well known phenomenon that after their marine feeding

migrations, adults return to their natal rivers and tributaries to reproduce (Stabell, 1984; Vaha *et al.*, 2008). Hence, given this behaviour and the characteristics of river catchments it is unlikely that Atlantic salmon populations within rivers are panmictic.

In contrast, both the meta-population and member-vagrant models describe the life history characteristics of Atlantic salmon more accurately. The member-vagrant hypothesis assumes that physical (*e.g*. hydrodynamic forces) and biological (*e.g*. rheotaxis and homing behaviour) factors retain young fish and attract returning adults to their natal stretch of river, thus maintaining a stable population (Iles and Sinclair, 1982). These individuals are 'members' in this model and make up the majority of the population, whereas individuals that do not achieve this are termed 'vagrants', and potentially contribute as migrants to other populations. Under this scenario, a high level of genetic differentiation is expected between populations inhabiting discontinuous stretches of suitable habitat within the river catchment. The populations would be temporally stable and an isolation-by-distance model of genetic relatedness should be observed as the vagrants are most likely to migrate and breed in other local populations.

Although based on similar attributes of retaining young fish and returning adult fish, the meta-population theory assumes that these populations suffer intermittent local extinction events and subsequent re-colonisation from other populations (McQuinn, 1997; Rieman and Dunham, 2000). This can be a highly realistic scenario for Atlantic salmon populations as their riverine habitats may be prone to flash floods, drought, freeze/thawing and/or pollution events, to name a few determinants of extinction events, thus maintaining meta-population structuring within the catchment and precluding the progression onto a member-vagrant system, as would likely be achieved in the absence of any extinction events. In catchments where salmon populations appear to conform to a meta-population model of population structuring, vagrants, or migrants, are still an intrinsic part of maintaining each population. Hence, in these situations, genetic divergence among populations is likely to be weak, yet still significant, implying a degree of straying/migration and relatively high geneflow. Similarly, in contrast with the member-vagrant model, no correlation between genetic and geographic distance would be expected under a meta-population model.

For Atlantic salmon, there is often behavioural, morphological or physiological evidence for the occurrence of multiple populations within a river catchment, but without genetic characterisation it is impossible to determine whether this is simply evidence of the plasticity inherent in the species or whether it is a true indication of distinct populations exhibiting different behavioural, morphological or physiological characteristics. Atlantic salmon populations of the river Tweed are one such case.

The river Tweed (N55:45:50 W1:59:35) is the second largest river in Scotland with a catchment area of 5000km^2 over 160km of river. It straddles the English - Scottish border and is bounded to the northwest by the Lammermoor and Moorfoot Hills and to the south by the Cheviot Hills. The majority of the river is designated a Site of Special Scientific Interest and a Special Area of Conservation under the EU Habitats Directive. The water quality of the river Tweed is deemed good to excellent under the Scottish Environment Protection Agency (SEPA) water quality guidelines and there is abundant habitat for adult spawning salmon and all freshwater life history stages. The catchment has 13 reservoirs and is an important source of water for the Borders and Edinburgh, but this is not thought to negatively impact upon the river (Tweed Foundation, 2005). Historically, there have been numerous barriers to migratory fish throughout the catchment, though fish passes are installed on most of these, allowing migration through most of the catchment (Tweed Foundation, 2005). No known supplementation or stocking programmes have been implemented within the Tweed catchment (Ronald Campbell, *personal communications*).

In contrast to the majority of salmon rivers, the salmon populations of the Tweed are relatively healthy, and are not currently experiencing the common declines experienced in many other rivers (Tweed Foundation, 2005). However, there does appear to be a long-term decline of large fish over 13lbs, with a corresponding increase in the numbers of smaller fish. This may be a natural cyclical phenomenon; however, the tropicalisation of exploited fish stocks, *i.e*. the reduction in the average size of fish and the onset of early maturation induced through the preferential exploitation of larger individuals has been documented previously. Whether this shift to smaller individuals in the Tweed is in any way reflective of this process has not been investigated, but it is a possibility.

There is already some evidence to suggest that there is stock structure within the Tweed catchment (Campbell, 1995; Verspoor *et al.*, 1997). Scale reading analyses was used to identify returning salmon as either early running "spring" fish or later running "autumn" salmon. Results identified the Ettrick to be mostly populated by spring fish, rather than autumn fish, with the Upper Tweed reaches indicating the reverse. These results were then supported in a radio-tracking study. In this study, returning salmon were tagged at the river mouth throughout the year and tracked through their upriver migration to their spawning grounds. Again, these results indicated that the majority of spring salmon, and a few early running grilse, returned to just two tributaries; the Whiteadder and Ettrick. A few were tracked to other tributaries but none returned to the Leader or Upper Tweed. In contrast, autumn fish, mainly grilse, had the reverse pattern of the spring fish, with large numbers entering the Teviot, Upper Tweed, Till and Gala, but few entering the Ettrick and none entering the Whiteadder. Hence, it appears that the Ettrick is largely populated by spring and summer multi sea winter (MSW) salmon, with the other tributaries mainly populated by 1SW, later running grilse. If run-timing has a genetic component then the results of these scale-reading and radio-tracking studies would suggest that populations of the Ettrick and Whiteadder would be genetically distinct from populations in the Teviot, Upper Tweed, Till, Leader and Gala. Indeed, an initial investigation into the genetic structure of Tweed salmon noted that samples from the Ettrick, Gala and Upper tributaries did show some genetic differentiation (Verspoor *et al.*, 1997).

By genotyping Atlantic salmon populations from across the catchment using nuclear microsatellite loci, we test the null hypothesis that Atlantic salmon within the catchment are a single panmictic population, *i.e*. no significant genetic differentiation is observed between samples (H_0). If the null hypothesis is disproven then alternate hypotheses assessing the role of the member-vagrant and/or meta-population evolutionary models can be evaluated; H_1: there are multiple populations of Atlantic salmon within the catchment, *i.e.* significant genetic differentiation between samples is observed; H_{1A}: the samples/populations follow a meta-population evolutionary mechanism, *i.e*. the samples show weak, yet significant genetic differentiation, but there is no isolation-by-distance signal; H_{1B}: the samples/populations follow a member-vagrant evolutionary mechanism, *i.e*. the

samples show highly significant genetic differentiation and there is a strong isolation-by-distance signal. Within both of these alternate hypotheses (H_{1A}, H_{1B}) the differentiation of Ettrick populations from the rest of the Tweed populations will be explored.

METHODS

SAMPLE COLLECTION

Sample collection took place during electric-fishing surveys carried out by the Tweed Foundation from 2004 to 2007. Fin clips from either the adipose or caudal fin were taken from Atlantic salmon and preserved in ethanol. These samples span the whole of the Tweed catchment, except the Whiteadder which was only fully open to salmon in the 1990s (Fig. 2.4; Table 2.2).

DNA ANALYSIS

Genomic DNA was extracted from all fin clips using the Chelex method (Estoup *et al.*). Variation was determined at 13 microsatellite loci: *Ssa*157a (King *et al.*, 2005); *Ssa*197, *Ssa*202, *Ssa*171, *Ssa*85 (O'Reilly *et al.*, 1996); Sssp1605, Sssp 2201 (Paterson *et al.*, 2004); *Ssa* 289, *Ssa*14 (McConnell *et al.*, 1995); SSOSL438, SSOSL85, SSOSL311 (Slettan *et al.*, 1996) and Ssa412 (Cairney *et al.*, 2000). Microsatellite loci were amplified in single 10μL PCR reactions containing 0.25 units *Taq* DNA polymerase (Bioline), 1X reaction buffer, 0.2mM dNTPs, 1.5mM $MgCl_2$, 0.5μM labelled primer and 1μM unlabelled primer. Amplification was undertaken in a thermal cycler block (Thermo) where samples were initially heated to 94˚C for 2 mins for denaturation then exposed to 35 cycles of 94˚C for 2 mins, annealing temperature 58˚C for 30 secs (except Ssa289, Ssa14 at 53˚C; SSOSL85 at 55˚C), 72˚C for 30 secs, with a final extension step of 72˚C for 10 mins. The 13 microsatellite loci were then combined into two multiplexes of seven and six loci and subsequently genotyped on a Beckman-Coulter CEQ 8000 automated sequencing machine.

ASSESSING SPATIAL GENETIC STRUCTURE

The underlying population structure within the river was inferred using STRUCTURE *v2.2* (Pritchard *et al.*, 2000). This Bayesian Markov Chain Monte Carlo (MCMC) based approach uses a clustering based algorithm to partition individuals into groups based on criteria to minimise Hardy Weinberg disequilibrium within populations and linkage disequilibrium between loci within populations. A burn-in period of 30,000 steps was set, followed by 1,000,000 MCMC replicates; allele frequencies were set as ‘weakly correlated’ as recommended by Pritchard *et al.* (2000). The number of groups to be simulated (K) was set as 1 to 15, with 20 replicates for each group. The

optimal number of clusters was estimated using the Δ*K* method of Evanno *et al.* (2005).

Given the number of individuals in some samples (range: 1-31; Table 2.2) it was necessary to group samples in order to undertake further statistical analysis. Results from the spatial autocorrelation analysis (see: Results) suggests that some of the larger tributaries may have the capacity to support more than one population. Hence, hierarchical gene diversity was assessed in a series of AMOVA analyses, as implemented in ARLEQUIN *v3.1* (Schneider *et al.*, 2000); samples were grouped based on geographical criteria and a series of AMOVA analyses were undertaken in order to determine which groupings explained the most spatial variation. This was initially undertaken on a tributary level basis; subsequently, the best estimates at tributary level were incorporated into catchment wide analyses. Three hierarchical levels were defined in order to assess the component of genetic diversity attributable to (i) variance among groups of samples, (ii) variance between samples within groups, and (iii) variance among individuals within samples. Groupings of samples were explored at several spatial scales, including; (i) closest confluence/minor tributary, (ii) major tributary and (iii) whole tributary. In this analysis the samples are partitioned *a priori* based on geographic parameters. Initially this was undertaken at a tributary level; groupings which showed the most spatial variation were thereafter incorporated into catchment wide analysis. The groupings at tributary level were undertaken at several geographical scales from minor tributaries, major tributaries to major regions (*i.e.* upper and lower). If there was some ambiguity into which group a particular sample could belong, several analyses were undertaken for each geographical level (Table 6.1a-2f). The groups that estimated the most genetic variation attributable to spatial variation between groups of samples were then combined into catchment level analyses, with other options explored, such as whole tributaries or combined tributaries (Table 6.1g).

Table 6.1a Groupings of samples defined in the hierarchical gene diversity analyses for the Upper Tweed tributaries and percentages of variance explained within and between groupings (see legend for full details).

ID (Upper Tweed)	Tributary I	Tributary II	Region I	Region II	Region III
1	A	A	A	A	**A**
2	A	A	A	A	**A**
3	A	A	A	A	**A**
4	B	B	B	A	**B**
5	B	B	B	A	**B**
6	B	B	B	A	**B**
7	C	C	C	B	**A**
8	C	C	C	B	**A**
9	C	C	C	B	**A**
10	D	D	C	B	**B**
11	D	D	C	B	**B**
12	E	C	C	B	**A**
among pops between groups	-0.39	-0.33	-0.19	-0.13	**0.12**
among pops within groups	1.12	1.07	0.93	0.87	**0.73**
within pops	99.217	99.26	99.26	99.26	**99.15**

Table 6.1b Groupings of samples defined in the hierarchical gene diversity analyses for the Gala tributaries and percentages of variance explained within and between groupings (see legend for full details).

ID (Gala)	Tributary I	Region I	Region II	Region III
13	A	A	A	**A**
14	A	A	A	**A**
15	A	A	A	**A**
16	B	A	A	**A**
17	C	A	A	**A**
18	D	A	B	**A**
19	E	B	B	**B**
20	E	B	B	**B**
21	F	C	C	**B**
22	F	C	C	**B**
among pops between groups	0.02	1.1	0.32	**0.34**
among pops within groups	1.81	1.16	1.62	**1.65**
within pops	98.17	97.74	98.07	**98.02**

Table 6.1c Groupings of samples defined in the hierarchical gene diversity analyses for the Leader tributaries and percentages of variance explained within and between groupings (see legend for full details).

ID (Leader)	Tributary I	Region I	Region II	Region III	Region IV	Region V	Region VI
28	A	A	**A**	A	A	A	A
29	A	A	**A**	A	A	A	A
30	B	A	**B**	B	A	A	A
31	B	A	**B**	B	A	A	A
32	C	A	**C**	B	C	A	B
33	D	B	**C**	C	B	B	B
34	D	B	**C**	C	B	B	B
35	D	B	**D**	C	B	B	B
36	E	B	**D**	C	C	A	B
among pops between groups	0.64	-0.74	**1.56**	0.31	-0.42	-0.06	-0.4
among pops within groups	3.31	4.36	**2.54**	3.68	4.25	3.94	4.13
within pops	96.05	96.38	**95.9**	96.01	96.17	96.12	96.26

Table 6.1d Groupings of samples defined in the hierarchical gene diversity analyses for the Ettrick tributaries and percentages of variance explained within and between groupings (see legend for full details).

ID (Ettrick)	Tributary I	Region I	Region II
37	**A**	A	A
38	**A**	A	A
39	**A**	A	A
40	**B**	B	B
41	**B**	B	B
42	**B**	B	B
43	**C**	B	B
44	**C**	B	B
45	**D**	C	B
46	**D**	C	B
47	**D**	C	B
among pops between groups	**0.65**	0.38	0.16
among pops within groups	**0.97**	1.2	1.4
within pops	**98.37**	98.42	98.44

Table 6.1e Groupings of samples defined in the hierarchical gene diversity analyses for the Teviot tributaries and percentages of variance explained within and between groupings (see legend for full details).

ID (Teviot)	Tributary I	Tributary II	Region I
48	A	A	**A**
49	A	A	**A**
50	B	A	**A**
51	C	A	**A**
52	C	A	**A**
53	D	A	**A**
54	E	B	**B**
55	E	B	**B**
56	F	C	**B**
57	G	D	**B**
58	G	D	**B**
59	H	E	**B**
60	H	E	**B**
61	H	E	**B**
among pops between groups	-1.2	-0.11	**0.3**
among pops within groups	3.34	2.29	**2.04**
within pops	97.86	97.82	**97.66**

Table 6.1f Groupings of samples defined in the hierarchical gene diversity analyses for the Till tributaries and percentages of variance explained within and between groupings (see legend for full details).

ID (Till)	Tributary I	**Tributary II**	Region I
62	A	**A**	A
63	A	**A**	A
64	A	**A**	A
65	B	**B**	A
66	B	**B**	A
67	C	**C**	B
68	C	**C**	B
69	D	**C**	B
70	D	**C**	B
among pops between groups	0.01	**0.07**	-0.06
among pops within groups	1.07	**1.02**	1.11
within pops	98.93	**98.91**	98.95

Table 6.1g Groupings of samples defined in the hierarchical gene diversity analyses for the Tweed catchment and percentages of variance explained within and between groupings (see legend for full details).

Map ID	Tributary	Whole Tributaries	Best tributary estimates I	Best tributary estimates II	Best tributary estimates III	Inferred Pop
1	Lyn	A	A	**A**	A	Upper
2	Lyn	A	A	**A**	A	Upper
3	Lyn	A	A	**A**	A	Upper
4	Upper	A	B	**A**	A	Upper
5	Upper	A	B	**A**	A	Upper
6	Upper	A	B	**A**	A	Upper
7	Eddlestone	A	A	**A**	A	Upper
8	Eddlestone	A	A	**A**	A	Upper
9	Eddlestone	A	A	**A**	A	Upper
10	Manor	A	B	**A**	A	Upper
11	Manor	A	B	**A**	A	Upper
12	Upper	A	A	**A**	A	Upper
13	Gala	B	C	**B**	B	GalC
14	Gala	B	C	**B**	B	GalC
15	Gala	B	C	**B**	B	GalC
16	Gala	B	C	**B**	B	GalC
17	Gala	B	C	**B**	B	GalC
18	Gala	B	C	**B**	B	GalC
19	Gala	B	D	**C**	C	GalB
20	Gala	B	D	**C**	C	GalB
21	Gala	B	E	**D**	D	GalA
22	Gala	B	E	**D**	D	GalA
23	Caddon	C	F	**E**	E	Cad
24	Caddon	C	F	**E**	E	Cad
25	Caddon	C	F	**E**	E	Cad
26	Caddon	C	F	**E**	E	Cad
27	Middle	D	F	**E**	E	Mid
28	Leader	E	G	**F**	F	LeaA
29	Leader	E	G	**F**	F	LeaA
30	Leader	E	H	**G**	F	LeaB
31	Leader	E	H	**G**	F	LeaB
32	Leader	E	I	**H**	F	LeaD
33	Leader	E	I	**H**	F	LeaC
34	Leader	E	I	**H**	F	LeaC
35	Leader	E	J	**I**	F	LeaC
36	Leader	E	J	**I**	F	LeaD
37	Ettrick	F	K	**J**	G	EttD
38	Ettrick	F	K	**J**	G	EttD
39	Ettrick	F	K	**J**	G	EttD
40	Ettrick	F	L	**K**	H	EttC
41	Ettrick	F	L	**K**	H	EttC
42	Ettrick	F	L	**K**	H	EttC
43	Ettrick	F	M	**L**	I	EttB
44	Ettrick	F	M	**L**	I	EttB
45	Ettrick	F	N	**M**	J	EttA
46	Ettrick	F	N	**M**	J	EttA
47	Ettrick	F	N	**M**	J	EttA
48	Teviot	G	O	**N**	K	TevB
49	Teviot	G	O	**N**	K	TevB
50	Teviot	G	O	**N**	K	TevB
51	Teviot	G	O	**N**	K	TevB
52	Teviot	G	O	**N**	K	TevB
53	Teviot	G	O	**N**	K	TevB
54	Teviot	G	P	**O**	L	TevA
55	Teviot	G	P	**O**	L	TevA
56	Teviot	G	P	**O**	L	TevA
57	Teviot	G	P	**O**	L	TevA
58	Teviot	G	P	**O**	L	TevA
59	Teviot	G	P	**O**	L	TevA
60	Teviot	G	P	**O**	L	TevA
61	Teviot	G	P	**O**	L	TevA
62	Till	H	Q	**P**	M	TilC
63	Till	H	Q	**P**	M	TilC
64	Till	H	Q	**P**	M	TilC
65	Till	H	R	**Q**	N	TilB
66	Till	H	R	**Q**	N	TilB
67	Till	H	S	**R**	O	TilA
68	Till	H	S	**R**	O	TilA
69	Till	H	S	**R**	O	TilA
70	Till	H	S	**R**	O	TilA
among pops between groups		0.12	1.18	**1.2**	1.05	
among pops within groups		0.73	1.26	**1.26**	1.4	
within pops		99.15	97.56	**97.54**	97.55	

Tables 6a-6g: Included are: Map ID and tributary relating to Figure 2.4 and Table 2.2 (Map ID; Tributary); several different groupings of samples based at either the (i) Tributary, *i.e.* minor tributaries within the major tributaries, or (ii) Region, *i.e.* groups of minor tributaries within the major tributaries. Associated with these are the results of the AMOVA analyses, showing the % of genetic variation attributable to (i) variation among populations between groups, (ii) variation among populations within groups, and (iii) variation within populations. Groupings which showed the highest genetic variation attributable to variation among populations between groups are highlighted in bold. In Table 6.1f, an additional column (Inferred Pop) indicates which grouping of populations explains the greatest variation between groups of samples; these groupings were used in subsequent population level analyses.

Analysis of inferred populations

Samples were grouped into inferred populations based on the results of the spatial autocorrelation and AMOVA analyses (see: Results; Table 6.1); population level statistics described below were performed on these inferred populations.

Genetic diversity, Hardy Weinberg equilibrium and genotypic linkage equilibrium

Allele number and allelic richness (allele number corrected for sample size) were calculated for all loci within populations using FSTAT *v2.9.3* (Goudet, 1995). Deviations from Hardy Weinberg expectations across samples (within loci) and across loci (within samples) were estimated using the method of Guo and Thompson (1992) as implemented in GENEPOP *v3.4* (Raymond and Rousset, 1995), with sequential Bonferroni corrections applied for multiple tests across populations according to Rice (1989). The occurrence of linkage disequilibrium between loci was tested using the default parameters in GENEPOP *v3.4* with sequential Bonferroni corrections applied for multiple tests across populations (Rice, 1989).

Genetic differentiation and relationships among populations

Genetic differentiation was measured by calculating F_{ST} values between populations and their significance using FSTAT *v2.9.3* (Goudet, 1995). Genetic distances between populations were estimated according to Neis (1987) DA distance, from which a phylogenetic tree was constructed using the Neighbour Joining method as implemented in POWERMARKER *v3.25* (Liu and Muse, 2005). Strength of support for each node was assessed by undertaking 1000 bootstrap replicates in POWERMARKER, and a consensus tree was obtained using the CONSENSE programme in PHYLIP *v3.67* (Felsenstein, 1993). The extent of genetic differentiation among populations was also measured with assignment tests (Cornuet *et al.*, 1999). Individuals were assigned to their most likely population of origin based on their microsatellite multi locus genotype using GENECLASS2 (Piry *et al.*, 2004). The Bayesian method using the 'leave one out' procedure was implemented (Rannala and Mountain, 1997). Isolation by distance of populations was tested by assessing the association between population genetic differentiation ($F_{ST}/(1- F_{ST})$) and geographic distance between sampling sites (waterway distances in km) by Mantel tests implemented in GENALEX *v6.1* (Peakall and Smouse, 2006).

SPATIAL AUTOCORRELATION ANALYSIS

Spatial genetic structure at the individual level was assessed following Primmer *et al.* (2006) as implemented in GENALEX *v6.1* (Peakall and Smouse, 2006). The multivariate approach as implemented in GENALEX strengthens the spatial signal by simultaneously assessing the spatial signal generated by multiple genetic loci, thereby reducing stochastic noise introduced through more classical allele-to-allele or locus-to-locus comparisons. Genetic distances between individuals were estimated using the squared distance method (Peakall *et al.*, 2003) and compared to waterway distances (km). Individuals from the same pair of sampling sites were assigned the same geographic distance and individuals from the same population were assigned a geographic distance of zero. An autocorrelation coefficient (r) is generated. It provides a measure of the genetic similarity between pairs of individuals where geographic separation falls within the spatial distance class which can be summarised by a correlelogram. In the correlelogram, the autocorrelation coefficient (r) was plotted as a function of ten discrete classes (10, 20, 30, 40, 50, 70, 90, 110, 130, 195 km), with individuals separated by 0-10km included in the first class, individuals separated by 11-20km included in the second class, *etc*. This analysis provides an estimate of the extent of non-random genetic structure across the whole study site. A second genetic autocorrelation analysis performed analysis for multiple distance class sizes in a way that is designed to investigate the interplay between sample size and distance class. The same distance classes were used, however, individuals from more distant groups were added to the previous groups, *i.e.* individuals separated by 0-10km were included in the first class, individuals separated by 0-20km were included in the second class, *etc.* This second analysis enables a more accurate estimate of scale across which spatial genetic structure can be detected.

Statistical significance for these tests was assessed in two ways. First, 1000 permutations were undertaken by shuffling individuals among geographic locations, to calculate 95% confidence intervals (CIs). If r falls outside of these 95% CIs then significant spatial structure can be inferred. Additionally, 1000 bootstrap replicates were undertaken within the set of pairwise comparisons for each distance class, from which 95% CIs were calculated. When the CIs do not straddle r=0, significant spatial structuring can again be inferred.

RESULTS

ASSESSING SPATIAL GENETIC STRUCTURE

The clustering methods implemented in STRUCTURE did not partition individuals into distinct groups. The *ad hoc* methods of Pritchard *et al.* (2000) estimated the optimal number of populations at one, implying that the Tweed has a single panmictic population of Atlantic salmon. The Δ*K* method of Evanno *et al.* (2005) does not have the capacity to predict a situation where only one population is present and it did not produce any robust alternative estimates for the optimum number of clusters; rather, individuals were probabilistically assigned to numerous clusters with more or less equal probabilities, again indicating no overarching structure of populations within the Tweed catchment.

Between 3 and 7 AMOVA analyses were undertaken for samples within each major tributary (Tables 6.1a-f). For each tributary, the grouping that explained the most genetic variation between groups within the tributary were then incorporated into catchment wide analyse, for which four AMOVA analyses were undertaken (Table 6.1g). The overall catchment wide grouping which explained the highest proportion of spatial variation demonstrated in the hierarchical analysis demonstrated that a significant 1.20% ($p<0.001$) of genetic variation was attributable to variance among groups of sampling sites and 1.26% ($p<0.001$) of genetic variance was attributable to variance among samples within groups. The remaining 97.54% ($p<0.001$) was attributable to genetic variance within samples. These groupings were then used in the subsequent population level analyses (Table 6.1g).

ANALYSIS OF INFERRED POPULATIONS

GENETIC DIVERSITY, HARDY WEINBERG EQUILIBRIUM AND GENOTYPIC LINKAGE EQUILIBRIUM

A total of 279 alleles were observed across the 13 loci used, ranging from 5 (Ssa14) to 38 (Ssa157). Within populations, the number of alleles per locus ranged from 2 (Ssa14; Ssa289; Ssa171; Ssa412) to 29 (Sssp2201), the average number of alleles per population ranged from 5.00 (LeaB) to 15.92 (Upp) and population level allelic richness ranged from 4.640 (LeaB) to 5.853 (Upp) (Table 6.2). After sequential Bonferroni corrections were applied, significant deviations from Hardy-Weinberg

expectations for individual loci within populations were observed on 2 occasions, involving the locus Ssa438 in populations LeaC and EttD. At the population level, EttC, GalB, Galc, LeaC and TevA were estimated as out of Hardy-Weinberg equilibrium. Linkage disequilibrium tests revealed 20 (1.3%) of the 1482 tests to be significant after sequential Bonferroni corrections were applied. Of these 20 significant associations, 16 were observed in only one population, while 2 were observed in two populations (Ssa438 *x* Sssp1605; Sssp1605 *x* Ssa85). Hence, linkage disequilibrium was deemed negligible. Collectively, these results are in line with expectations for wild populations, indicating that the grouping of populations undertaken based on the spatial autocorrelation and AMOVA analyses are appropriate for further analysis.

Table 6.2 Diversity indices at each microsatellite loci for inferred populations (see Table 6.1g).

	N		SSSP2201	SSOSL85	SSOSL438	SSA289	SSA171	SSA157	SSA14	SSSP1605	SSOSL311	SSA85	SSA412	SSA202	SSA197	Pop. Ave.
Cad	**39**	**Allele No.**	21	11	13	4	16	17	2	7	20	14	4	12	16	12.1
		A_R	7.804	6.007	6.289	3.527	6.06	7.415	1.921	5.485	7.12	6.172	2.794	6.5	6.875	5.690
		He	0.894	0.852	0.859	0.686	0.844	0.883	0.328	0.854	0.917	0.868	0.569	0.899	0.882	0.795
		Ho	0.923	0.846	0.795	0.641	0.795	0.846	0.308	0.769	0.872	0.821	0.487	0.846	0.949	0.761
		HWE	0.1686	0.8706	0.3048	0.6162	0.0595	0.0114	0.6469	0.0951	0.29	0.0064	0.1422	0.5253	0.7378	0.0098
EttA	**45**	**Allele No.**	24	10	13	4	13	20	4	9	23	11	4	12	16	12.5
		A_R	7.828	5.545	5.343	3.211	5.76	7.763	2.128	5.356	7.899	5.784	2.597	6.075	6.666	5.535
		He	0.917	0.807	0.808	0.638	0.854	0.942	0.359	0.846	0.944	0.821	0.509	0.873	0.894	0.786
		Ho	0.911	0.756	0.622	0.600	0.889	0.911	0.267	0.756	0.978	0.844	0.467	0.911	0.844	0.750
		HWE	0.613	0.2995	0.0015	0.0099	0.9559	0.4023	0.1838	0.345	0.6879	0.3419	0.7037	0.6279	0.4254	0.0465
EttB	**15**	**Allele No.**	16	7	10	4	10	13	3	8	11	9	4	9	16	9.2
		A_R	8.13	4.613	5.634	2.89	5.953	6.71	2.332	5.296	6.155	5.82	3.021	5.872	7.681	5.393
		He	0.954	0.784	0.841	0.582	0.773	0.885	0.536	0.834	0.864	0.848	0.479	0.851	0.936	0.782
		Ho	1.000	0.800	0.867	0.667	0.467	0.867	0.533	0.667	0.867	0.733	0.533	0.800	1.000	0.754
		HWE	0.1403	0.3427	0.4163	0.8931	0.0091	0.75	0.5767	0.0993	0.3333	0.7025	1	0.6563	0.8968	0.3731
EttC	**24**	**Allele No.**	22	10	10	4	10	17	3	10	17	12	3	8	17	11
		A_R	8.195	5.577	6.483	3.132	5.573	7.61	2.194	5.601	7.473	6.609	2.188	5.663	7.51	5.778
		He	0.953	0.848	0.899	0.617	0.725	0.937	0.467	0.843	0.931	0.901	0.451	0.856	0.933	0.797
		Ho	0.958	0.708	0.583	0.792	0.625	0.833	0.417	0.958	0.917	0.917	0.375	0.875	1.000	0.766
		HWE	0.7475	0.0198	**0.000**	0.3891	0.4973	0.0074	0.7601	0.2309	0.212	0.0282	0.5596	0.2254	0.9977	0
EttD	**56**	**Allele No.**	23	14	13	6	17	21	3	8	23	17	5	10	20	13.8
		A_R	7.926	6.053	5.624	3.547	6.481	7.633	2.107	5.321	7.399	6.633	2.614	5.839	7.206	5.722
		He	0.946	0.868	0.836	0.682	0.874	0.935	0.314	0.841	0.926	0.899	0.494	0.851	0.917	0.799
		Ho	0.964	0.839	0.732	0.786	0.804	0.911	0.304	0.839	0.929	0.893	0.464	0.821	0.929	0.786
		HWE	0.589	0.5322	0.0002	0.1687	0.0011	0.6272	0.0513	0.7298	0.5313	0.1309	0.9322	0.5185	0.6987	0.0022
GalA	**32**	**Allele No.**	18	13	9	5	14	22	2	6	20	16	5	10	19	12.2
		A_R	7.817	6.078	4.867	3.652	6.202	7.986	1.807	5.069	7.585	6.444	2.761	5.951	7.466	5.668
		He	0.944	0.856	0.731	0.692	0.835	0.919	0.246	0.811	0.935	0.887	0.463	0.873	0.929	0.778
		Ho	0.906	0.781	0.688	0.625	0.750	0.844	0.281	0.750	0.875	0.938	0.438	0.813	0.906	0.738
		HWE	0.7058	0.0441	0.254	0.6863	0.1007	0.5593	1	0.7497	0.0083	0.7197	0.9119	0.251	0.5147	0.2179
GalB	**23**	**Allele No.**	16	6	10	4	9	14	3	7	14	9	3	9	12	8.9
		A_R	7.121	4.905	5.828	2.916	5.167	6.828	2.21	4.816	6.713	4.921	2.386	5.636	6.085	5.041
		He	0.918	0.823	0.826	0.598	0.771	0.908	0.492	0.798	0.821	0.767	0.529	0.856	0.831	0.765
		Ho	0.913	0.783	0.783	0.435	0.783	0.957	0.609	0.696	0.826	0.826	0.652	0.913	0.870	0.773
		HWE	0.0052	0.6738	0.0103	0.4075	0.3802	0.2788	0.1674	0.0143	0.2817	0.0174	0.3715	0.4004	0.0808	0.0003

	N		SSSP2201	SSOSL85	SSOSL438	SSA289	SSA171	SSA157	SSA14	SSSP1605	SSOSL311	SSA85	SSA412	SSA202	SSA197	Pop. Ave.
GalC	**97**	**Allele No.**	24	14	12	4	18	25	3	9	19	19	4	10	23	14.2
		A_R	7.901	5.674	4.549	3.518	6.077	7.76	2.118	5.514	7.052	6.481	2.62	5.747	7.289	5.562
		He	0.945	0.833	0.744	0.699	0.840	0.911	0.428	0.852	0.883	0.879	0.460	0.829	0.902	0.785
		Ho	0.918	0.814	0.701	0.722	0.763	0.876	0.340	0.866	0.866	0.876	0.443	0.794	0.887	0.759
		HWE	0.1501	0.4471	0.0534	0.6318	0.1751	0.1871	0.092	0.0238	0.0263	0.0872	0.3084	0.0068	0.0459	0.0001
LeaA	**8**	**Allele No.**	8	7	8	5	7	9	3	7	10	8	4	5	9	6.9
		A_R	6.213	5.367	6.295	4.565	5.428	6.831	2.5	5.698	7.589	6.703	3.249	4.374	7.054	5.528
		He	0.875	0.817	0.892	0.825	0.825	0.908	0.342	0.867	0.942	0.798	0.617	0.775	0.925	0.800
		Ho	0.875	0.750	0.875	0.625	1.000	0.875	0.125	1.000	0.875	0.875	0.375	0.500	1.000	0.750
		HWE	0.4939	0.8843	0.0212	0.6518	0.222	0.268	0.0624	0.6596	0.4711	0.6706	0.1507	0.083	0.6301	0.1292
LeaB	**6**	**Allele No.**	7	5	4	2	5	8	2	6	6	6	2	6	6	5
		A_R	6.152	4.652	4	2	4.818	6.985	2	5.636	5.485	5.333	2	5.621	5.636	4.640
		He	0.833	0.788	0.648	0.409	0.848	0.894	0.409	0.879	0.848	0.803	0.545	0.864	0.879	0.742
		Ho	1.000	0.667	0.500	0.500	1.000	1.000	0.500	0.833	0.833	0.833	0.333	0.667	0.833	0.731
		HWE	0.8909	0.2819	0.6954	1	0.8791	0.6882	1	0.3469	0.4367	0.6152	0.4816	0.0986	0.7722	0.9397
LeaC	**12**	**Allele No.**	11	10	7	4	11	11	2	8	12	13	2	9	12	8.6
		A_R	7.1	6.394	5.346	3.365	6.723	7.233	1.67	5.633	7.461	7.021	1.995	6.493	7.449	5.683
		He	0.841	0.880	0.833	0.699	0.810	0.928	0.159	0.830	0.931	0.902	0.417	0.902	0.853	0.768
		Ho	0.917	0.667	0.583	0.500	0.750	0.750	0.167	0.833	0.833	0.833	0.417	0.917	0.750	0.686
		HWE	1	0.0972	**0.000**	0.446	0.1338	0.1754	1	0.6949	0.1324	0.2686	1	0.9463	0.3634	0
LeaD	**13**	**Allele No.**	11	8	5	4	8	11	3	7	14	7	5	7	12	7.8
		A_R	7.157	5.384	4.067	3.589	4.79	7.217	2.015	5.239	8.062	5.137	3.55	5.764	7.196	5.321
		He	0.926	0.797	0.745	0.729	0.769	0.929	0.218	0.825	0.954	0.831	0.677	0.877	0.923	0.785
		Ho	0.846	0.769	0.769	1.000	0.538	0.923	0.077	0.846	0.923	0.692	0.462	0.769	1.000	0.740
		HWE	0.228	0.6701	0.1154	0.125	0.0099	0.2245	0.0362	0.6012	0.2839	0.4628	0.1546	0.5302	0.9045	0.0285
Mid	**23**	**Allele No.**	18	11	13	4	11	18	3	8	16	12	4	9	14	10.8
		A_R	7.68	6.068	6.389	3.597	5.427	7.604	2.195	5.159	7.037	6.768	2.889	6.124	6.485	5.648
		He	0.898	0.877	0.884	0.716	0.805	0.936	0.445	0.814	0.911	0.907	0.540	0.881	0.880	0.807
		Ho	0.957	0.870	0.870	0.565	0.696	0.913	0.435	0.826	1.000	0.870	0.348	0.913	0.826	0.776
		HWE	0.9618	0.3959	0.0129	0.3864	0.1714	0.1765	1	0.4299	0.9507	0.1676	0.084	0.4196	0.4336	0.155
TevA	**54**	**Allele No.**	24	14	11	7	18	24	4	9	20	15	4	10	23	14.1
		A_R	8.136	5.833	4.793	4.066	5.966	7.598	2.179	5.146	7.372	6.251	2.415	6.01	7.735	5.654
		He	0.952	0.840	0.757	0.752	0.837	0.932	0.371	0.810	0.927	0.851	0.388	0.876	0.888	0.783
		Ho	0.926	0.778	0.593	0.722	0.741	0.870	0.315	0.815	0.926	0.833	0.315	0.833	0.870	0.734
		HWE	0.1023	0.5446	0.0008	0.0099	0.2377	0.0754	0.0234	0.7451	0.1623	0.3962	0.0635	0.0637	0.6665	0.0001

	N		SSSP2201	SSOSL85	SSOSL438	SSA289	SSA171	SSA157	SSA14	SSSP1605	SSOSL311	SSA85	SSA412	SSA202	SSA197	Pop. Ave.
TevB	**55**	**Allele No.**	24	16	14	4	15	22	3	7	18	16	4	10	21	13.4
		A_R	7.989	6.067	6.379	3.438	6.515	7.656	2.006	5.301	7.331	6.832	2.13	6.28	6.718	5.742
		He	0.949	0.861	0.888	0.710	0.870	0.918	0.348	0.841	0.923	0.906	0.401	0.889	0.895	0.800
		Ho	0.909	0.818	0.836	0.782	0.855	0.873	0.364	0.855	0.927	0.909	0.418	0.855	0.964	0.797
		HWE	0.9504	0.4097	0.0013	0.6891	0.2432	0.2754	1	0.8362	0.5107	0.2788	1	0.7757	0.4479	0.3734
TilA	**29**	**Allele No.**	17	12	10	4	11	13	4	9	17	13	3	10	16	10.7
		A_R	7.299	6.403	5.93	3.196	5.097	7.066	2.187	5.711	7.587	6.615	2.144	5.731	7.38	5.565
		He	0.924	0.860	0.874	0.655	0.762	0.920	0.324	0.864	0.935	0.899	0.431	0.857	0.895	0.785
		Ho	0.966	0.793	0.690	0.621	0.759	1.000	0.345	0.931	1.000	0.966	0.310	0.828	0.897	0.777
		HWE	0.8363	0.3938	0.0143	0.9723	0.8083	0.9938	0.0306	0.359	0.3636	0.7756	0.2115	0.0404	0.679	0.161
TilB	**9**	**Allele No.**	10	8	5	4	6	13	3	8	11	11	3	7	10	7.6
		A_R	6.922	6.157	4.598	3.51	4.961	8.382	2.549	6.471	7.5	7.647	2.81	5.892	6.99	5.722
		He	0.909	0.882	0.817	0.686	0.830	0.961	0.503	0.909	0.935	0.941	0.569	0.882	0.915	0.826
		Ho	0.667	0.778	0.333	0.556	0.889	1.000	0.444	0.778	1.000	1.000	0.667	0.889	1.000	0.769
		HWE	0.0927	0.5571	0.007	0.53	0.7479	1	1	0.1396	1	1	1	0.9386	1	0.7019
TilC	**24**	**Allele No.**	21	9	9	5	13	18	2	8	16	12	3	10	14	10.8
		A_R	7.783	5.178	5.174	3.902	6.586	7.386	1.903	5.119	6.701	6.376	2.304	5.736	7.116	5.482
		He	0.937	0.771	0.805	0.752	0.892	0.923	0.311	0.803	0.887	0.887	0.401	0.863	0.881	0.778
		Ho	0.958	0.792	0.792	0.750	0.750	0.917	0.208	0.958	0.833	0.833	0.417	0.917	0.792	0.763
		HWE	1	0.5952	0.1534	0.6155	0.0223	0.3461	0.1537	0.8068	0.0698	0.78	1	0.3892	0.2213	0.2521
Upp	**102**	**Allele No.**	29	14	13	7	16	27	3	10	24	19	6	12	27	15.9
		A_R	8.094	6.156	5.545	3.685	6.258	7.98	1.967	5.544	7.513	6.751	2.823	6.249	7.53	5.853
		He	0.951	0.869	0.833	0.693	0.858	0.948	0.314	0.854	0.908	0.886	0.545	0.877	0.921	0.804
		Ho	0.971	0.892	0.725	0.667	0.863	0.951	0.324	0.843	0.892	0.892	0.618	0.863	0.912	0.801
		HWE	0.4114	0.4366	0.0098	0.6976	0.1489	0.4634	1	0.4324	0.124	0.8784	0.3535	0.109	0.6202	0.1821
Total	**666**	**Allele No.**	31	19	22	10	27	38	5	13	31	26	10	13	34	21.5
		A_R	8.001	5.965	6.525	3.569	6.181	7.801	2.084	5.514	7.488	6.608	2.608	6.157	7.336	5.834
		He	0.919	0.838	0.817	0.675	0.822	0.922	0.364	0.841	0.912	0.868	0.499	0.865	0.899	
		Ho	0.920	0.784	0.702	0.661	0.774	0.901	0.335	0.833	0.904	0.862	0.449	0.827	0.907	
		HWE	0.327117	0.324568	0	0.345924	0.000132	0.045233	0.0827	0.150429	0.033063	0.054099	0.606323	0.048525	0.874673	

Included are: Inferred population ID (Pop); Number in sample (N), the number of alleles at each locus (Allele No.), allelic richness (A_R), expected heterozygosity (He); observed heterozygosity (Ho) and the *p* value associated with the loci meeting Hardy-Weinberg Expectations (HWE). Significant deviations from HWE at $p<0.05$ after Bonferroni corrections are indicated in bold.

Table 6.3 Population differentiation estimates; pairwise *F*ST estimates (below diagonal) with associated *p*-value above diagonal.

	Cad	EttA	EttB	EttC	EttD	GalA	GalB	GalC	LeaA	LeaB	LeaC	LeaD	Mid	TevA	TevB	TilA	TilB	TilC	Upp
Cad	-	0.000	0.000	0.021	0.001	0.010	0.000	0.000	0.003	0.015	**0.126**	0.016	**0.405**	0.000	0.000	0.000	**0.054**	0.000	0.000
EttA	0.009	-	0.003	0.000	0.020	0.001	0.000	0.000	0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
EttB	0.023	0.014	-	0.002	0.000	0.000	0.001	0.000	0.000	0.001	0.002	0.000	0.015	0.000	0.000	0.000	0.016	0.000	0.000
EttC	0.007	0.012	0.017	-	0.001	0.000	0.001	0.000	0.000	0.000	0.002	0.000	**0.079**	0.000	0.001	0.000	**0.229**	0.000	0.000
EttD	0.009	0.004	0.017	0.012	-	0.000	0.000	0.000	0.000	0.003	0.007	0.001	0.008	0.000	0.000	0.000	0.018	0.000	0.000
GalA	0.008	0.012	0.026	0.015	0.010	-	0.000	0.041	0.003	0.000	0.008	**0.123**	0.022	0.000	0.000	0.000	0.000	0.000	0.000
GalB	0.018	0.032	0.031	0.027	0.031	0.028	-	0.000	0.004	0.005	0.004	0.001	0.000	0.000	0.000	0.000	0.005	0.000	0.000
GalC	0.009	0.013	0.017	0.013	0.010	0.004	0.029	-	0.000	0.001	0.000	0.000	0.038	0.000	0.000	0.000	0.007	0.000	0.000
LeaA	0.025	0.032	0.044	0.033	0.025	0.024	0.044	0.031	-	0.001	0.008	**0.068**	0.008	0.000	0.000	0.005	0.003	0.000	0.000
LeaB	0.024	0.033	0.046	0.047	0.026	0.039	0.052	0.035	0.060	-	0.018	0.009	0.019	0.000	0.000	0.002	**0.118**	0.000	0.002
LeaC	0.007	0.023	0.029	0.020	0.014	0.015	0.033	0.021	0.032	0.035	-	0.013	**0.106**	0.000	0.000	0.007	0.049	0.001	0.000
LeaD	0.013	0.022	0.029	0.027	0.016	0.007	0.034	0.020	0.018	0.036	0.023	-	0.010	0.000	0.000	0.001	0.012	0.000	0.000
Mid	0.001	0.011	0.015	0.006	0.008	0.008	0.023	0.005	0.023	0.024	0.009	0.016	-	0.000	0.000	0.000	**0.071**	0.000	0.000
TevA	0.013	0.023	0.034	0.012	0.018	0.024	0.036	0.023	0.034	0.046	0.025	0.030	0.018	-	0.000	0.004	**0.127**	**0.230**	0.001
TevB	0.008	0.013	0.029	0.009	0.011	0.015	0.037	0.018	0.029	0.026	0.020	0.025	0.012	0.007	-	0.000	**0.105**	0.000	0.001
TilA	0.016	0.023	0.030	0.017	0.014	0.020	0.038	0.022	0.022	0.032	0.017	0.023	0.016	0.008	0.008	-	**0.593**	0.027	0.000
TilB	0.012	0.024	0.023	0.007	0.014	0.025	0.038	0.018	0.034	0.020	0.019	0.023	0.012	0.009	0.006	0.001	-	**0.150**	**0.188**
TilC	0.019	0.034	0.036	0.020	0.023	0.021	0.036	0.025	0.031	0.047	0.020	0.030	0.019	0.002	0.011	0.006	0.007	-	0.004
Upp	0.008	0.019	0.028	0.008	0.016	0.019	0.032	0.022	0.026	0.034	0.021	0.023	0.015	0.004	0.005	0.011	0.004	0.008	-

Population ID follows Inferred Population ID in Table 6.1g; non-significant *p* values indicated in bold.

Genetic differentiation and relationships among populations

Global F_{ST} was estimated at 0.020879, with pairwise estimates ranging from 0.00083 (Cad-Mid) to 0.06021 (LeaA-LeaB). Although generally low, all but 15 tests were significant at $p<0.05$ after sequential Bonferroni corrections were applied (Table 6.3). The Neighbour Joining population tree generally groups samples from the same tributaries together, however, while this is geographically sensible, bootstrap support overall is low, with the highest support occurring for the grouping of the GalA and GalC populations with 62% (Figure 6.1).

Overall, only 35% of individuals were assigned to the correct population of origin (range: 0% TilB to 53% GalC). When assigned to tributary of origin this rose to an average of 47% correct assignment (range: 9% Mid to 65% GalB) (Table 6.4). Since assignment levels were so poor, tests for associations between genetic divergence were not undertaken.

Spatial autocorrelation analysis

Tests for spatial genetic structure at the individual level revealed that the genetic autocorrelation coefficient, *r*, was significantly positive for the 0-10 to 0-40 distance classes, intercepting the *x* axis at approximately 47km. Significantly negative genetic autocorrelation was observed in the 50km size classes and greater (Figure 6.2). This indicates that more than one population may be present in the larger tributaries of the river Tweed, such as the Teviot (66km in length) and Till (73km in length). The genetic correlation for increasing waterway distance classes indicated that *r* was significantly positive for waterway distance classes up to 90km. Above this positive genetic autocorrelation was no longer detectable (Figure 6.2), indicating that migration and geneflow is unlikely between individuals and populations separated by waterway distances above 90km.

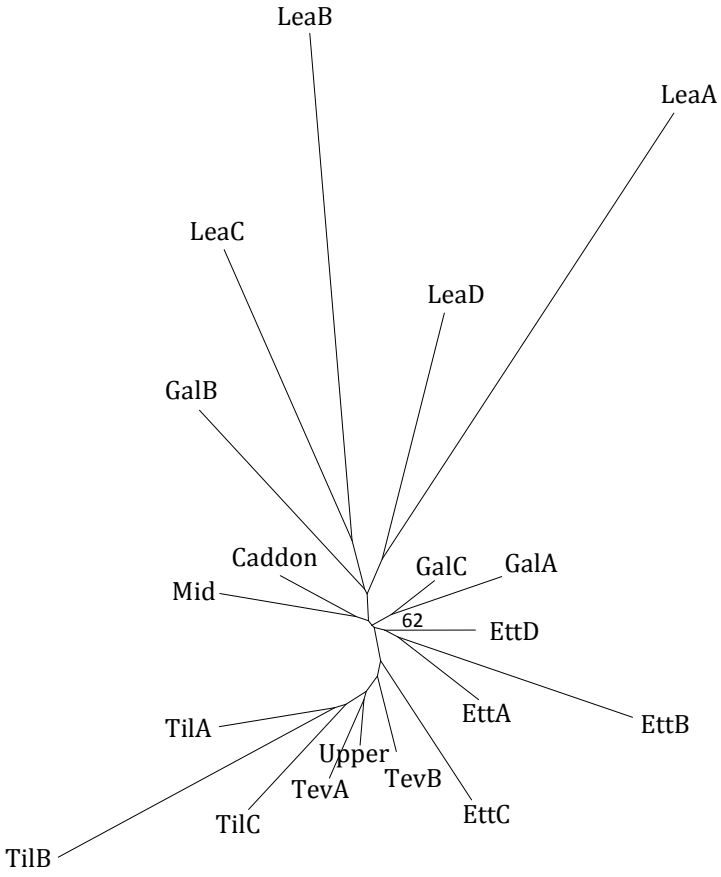


Figure 6.1 A Neighbour-Joining population tree based on Nei's (1987) D_A distance between inferred populations. Population ID follows Inferred Population ID in Table 6.1g; bootstrap values over 50% are indicated.

Table 6.4 Number (percent) of individuals assigned to inferred populations and full tributary of origin (Trib).

	Cad	EttA	EttB	EttC	EttD	GalA	GalB	GalC	LeaA	LeaB	LeaC	LeaD	Mid	TevA	TevB	TilA	TilB	TilC	Upp	Trib
Cad	9 (23)	1 (3)			4 (10)	3 (8)	1 (3)	5 (13)					5 (13)		6 (15)	1 (3)		1 (3)	3 (8)	9 (23)
EttA	5 (11)	19 (42)		1 (2)	7 (16)	2 (4)	1 (2)	4 (9)		1 (2)					1 (2)			1 (2)	3 (7)	27 (60)
EttB	1 (7)	2 (13)	3 (20)	3 (20)	1 (2)		1 (2)	2 (13)	1 (2)										1 (2)	9 (60)
EttC	1 (4)		1 (4)	7 (29)	3 (13)			1 (4)					1 (4)	4 (17)		1 (4)	1 (4)		4 (17)	11 (46)
EttD	5 (9)	5 (9)	2 (4)	3 (5)	21 (38)	2 (4)		4 (7)				1 (2)	1 (2)	2 (4)	6 (11)	1 (2)		1 (2)	2 (4)	31 (55)
GalA	2 (6)	2 (6)	2 (6)	3 (9)	1 (3)	4 (13)		10 (31)			1 (3)	2 (6)	1 (3)		3 (9)				1 (3)	14 (44)
GalB	1 (4)	1 (4)			1 (4)	3 (13)	11 (48)	1 (4)				1 (4)		3 (13)				1 (4)		15 (65)
GalC	5 (5)	7 (7)	3 (3)	1 (1)	6 (6)	8 (8)	1 (1)	51 (53)		1 (1)		1 (1)	6 (6)	1 (1)	3 (3)	1 (1)			2 (2)	60 (62)
LeaA						1 (13)			3 (38)			1 (13)				2 (25)			1 (13)	4 (50)
LeaB										3 (50)					2 (33)				1 (17)	3 (50)
LeaC	2 (17)				1 (8)	1 (8)	1 (8)	1 (8)	1 (8)		3 (25)		2 (17)							4 (33)
LeaD	1 (8)	1 (8)			1 (8)	3 (23)				1 (8)		4 (31)		1 (8)					1 (8)	5 (39)
Mid	3 (13)	2 (9)		3 (13)	3 (13)		1 (4)	3 (13)					2 (9)		3 (13)				3 (13)	2 (9)
TevA	2 (4)	2 (4)	1 (2)	1 (2)		1 (2)				1 (2)	1 (2)			18 (33)	8 (15)	5 (9)	2 (4)	3 (6)	9 (17)	26 (48)
TevB	4 (7)	1 (2)		2 (4)	3 (6)	1 (2)		3 (6)					1 (2)	5 (9)	16 (29)	3 (6)		2 (4)	13 (24)	21 (38)
TilA		1 (3)		1 (3)		1 (3)		1 (3)					2 (7)	1 (3)	2 (7)	10 (35)		5 (17)	5 (17)	15 (51.7)
TilB				1 (11)				1 (11)						2 (22)	1 (11)	1 (11)			3 (33)	1 (11)
TilC					1 (4)									3 (13)	2 (8)	3 (13)	1 (4)	4 (17)	10 (41)	8 (33)
Upp	5 (5)	1 (1)		4 (4)	2 (2)	2 (2)		3 (3)	1 (1)				2 (2)	11 (11)	11 (11)	5 (5)	4 (4)	5 (5)	46 (45)	46 (45)

Population ID follows Inferred Population ID in Table 6.1g. Percentages rounded to nearest whole number.

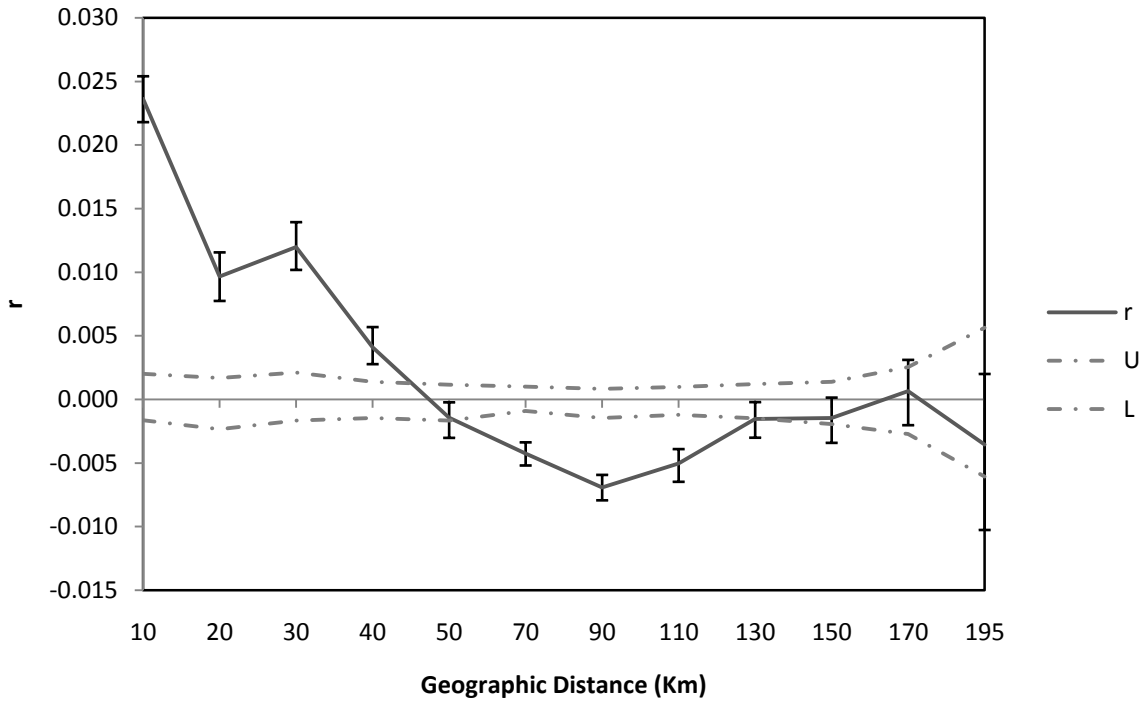


Figure 6.2 Genetic autocorrelation analysis of individual Tweed salmon at different distance classes. Correlelogram showing the genetic correlation, *r*, as a function of distinct distance classes. Dotted grey lines indicate 95% CL about the null hypothesis of no genetic structure and error bars about *r* indicate 95% CI as determined by bootstrapping.

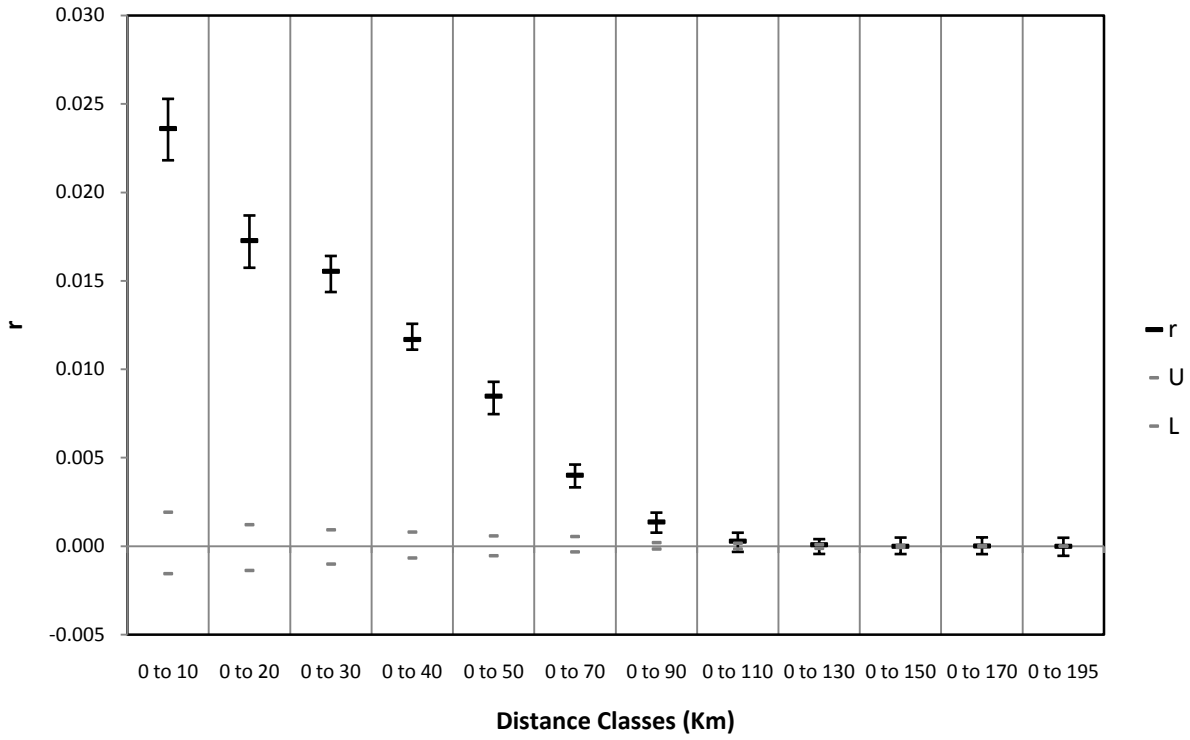


Figure 6.3 Genetic autocorrelation analysis of individual Tweed salmon genetic structure. The genetic correlation *r*, for increasing distance classes. Grey bars indicate 95% CI about the null hypothesis of no genetic structure and error bars about *r* indicate 95% CI as determined by bootstrapping.

DISCUSSION

LIMITATIONS ASSOCIATED WITH THE INFERENCE OF POPULATIONS

This study provided an opportunity to assess whether an opportunistic sampling strategy, whereby samples were taken more or less continuously across the catchment, rather than one which conducted focussed sampling at a particular location, can be successfully utilised for the analysis of population structuring. By the very nature of the sampling strategy, which resulted in samples being separated by 0.4 to 194.4 km and which were comprised of 1 to 31 individuals, it was first necessary to partition individuals into populations. Clustering based analysis failed to reveal any underlying population structuring and results were most consistent with the river Tweed being one single panmictic population (*i.e.* STRUCTURE estimated K=1 using the method of Pritchard *et al.* (2000)). Hence, individuals were partitioned on an *ad hoc* basis, in which hierarchical gene diversity analysis was used to assess the amount of genetic variation attributable to spatial variation between groups of samples. The different groupings of individuals into inferred populations had little bearing on the overall genetic diversity explained by spatial variation (Table 6.1), implying that potentially none of these groupings reflected the true population structuring within the Tweed catchment; a possibility that must be considered when interpreting results. This suggests that opportunistic sampling, such as conducted here, is not necessarily appropriate for population level analysis, since partitioning individuals into inferred populations cannot be achieved satisfactorily, thereby restricting confidence in any further analysis undertaken.

Nevertheless, in order to conduct an assessment of Atlantic salmon populations for the Tweed Foundation, it was necessary to select the most appropriate grouping of samples into inferred populations. The grouping which explained the most variation between the inferred populations was chosen for population level analysis, which involved several within tributary groups of populations (Table 6.1g); here 1.20% of genetic variation was attributable to spatial variation between groups of samples, 1.26% was attributable to genetic variation among samples within groups and 97.54% was attributable to genetic variation within samples. These groups of samples were then used as the inferred populations used in all subsequent analyses, but it is clear that spatial variation explains very

little of the genetic structuring observed and that other groupings would have produced very similar estimates (Tables 6.1a-g).

GENETIC DIVERSITY AND DIVERGENCE OF INFERRED POPULATIONS OF THE RIVER TWEED

The high levels of polymorphism observed with the 13 microsatellite loci used in this study (279 alleles, ranging from 5 to 38 alleles at a single locus within a single sample; Table 6.2) are consistent with those observed in other studies of within-river populations of Atlantic salmon (Garant *et al.*, 2000: 4-33 alleles at a single locus in a single population; Primmer *et al.*, 2006: 2-30 alleles at a single locus within a population; 9-30 alleles in a single population from tributaries throughout the river Dart, see Chapter 4). Genetic diversity, as measured using estimates of allelic richness (allele number corrected for sample size) were lower than reported for the river Dart in southwest England, (river Tweed: Ar=4.640-5.853; river Dart: Ar=7.06-10.60; see Chapter 4) but similar to those reported in previous studies of within river populations undertaken in northeast Europe (Primmer *et al.*, 2006: Ar=6.2-7.2; Vaha *et al.*, 2007: Ar=4.24-7.02), whereas estimates of expected heterozygosity were much more variable in this study compared to previous studies (river Tweed: He=0.16-0.95; river Dart: He=0.68-0.82; see Chapter 4; Primmer *et al.*, 2006: He=0.62-0.68; Vaha *et al.*, 2007: He=0.58-0.70;). Genetic divergence, as measured with pairwise F_{ST} estimates between samples within the river Tweed (global F_{ST} =0.0209; range: 0.00083-0.0602; Table 6.3) lies within the range previously reported for within river studies, being slightly lower than for the river Dart (global F_{ST} =0.033; range: 0.00593-0.078; see Chapter 4), the Sainte Marguerite river in Canada (Global F_{ST} =0.034; range: 0.0077-0.0874; Garant *et al.* 2000) and the river Teno of Norway and Finland (F_{ST}: 0.001-0.138; Vaha *et al*. 2007) and slightly higher than that observed for the Varzuga river on the Kola Peninsula (Global F_{ST} =0.014; range: 0.006-0.07; Primmer *et al.* 2006). Furthermore, Hardy Weinberg Equilibrium within populations and linkage equilibrium between loci was observed. Hence, the structuring inferred by the hierarchical gene diversity analysis of these Tweed samples does not appear to demonstrate departure from panmixia, making population level comparisons between the inferred populations possible.

Despite being weak, significant genetic divergence was observed in 146 of the 162 tests of genetic differentiation between the inferred populations (Table 6.2). Of the 16 non-significant results, four involved the sample from the main branch of the river (Mid) which may be comprised of individuals straying from other locations into the main channel to feed. Nine involved the inferred populations of the Till tributary, which is the tributary closest to the river mouth. This may indicate that returning adults may enter the Till and not progress onto their natal tributary; indeed, Vaha *et al.* (2007) noted that the headwater tributaries tended to be the least diverse and most diverged for this reason.

In the population tree, inferred populations from the same tributaries tended to group together, with geographically proximate tributaries also occurring close together. However, despite this being intuitively correct from a geographic perspective, statistical support was generally low (Figure 6.1). No correlation between population level genetic and geographic distance was observed in the Mantel isolation-by-distance tests at a catchment or a tributary level (results not shown). Self assignment to the correct inferred population was low (average: 35%; range: 0-53%) but rose slightly when assigned to tributary of origin (average: 47%; range: 9-65%), indicating that the geographic proximity of inferred populations has a small bearing on the genetic relationships between populations (Table 6.4).

Spatial autocorrelation analysis was undertaken on individual samples rather than the inferred populations, hence, confidence in these results is high since it is not subject to the same limitations as analysis undertaken on the inferred populations. Here, a significant positive association between individuals separated by waterway distances of up to 40km was observed, indicating that migration and geneflow between individuals is likely up to this distance (Figure 6.3). For individuals separated by waterway distances of over 50km there is a negative association between genetic similarity and geographic separation, indicating that migration and geneflow is limited over distances of 50km or more. The genetic correlation for increasing waterway distance classes indicated that there was significant genetic correlation for waterway distance classes up to 90km, above this positive genetic autocorrelation was no longer detectable (Figure 6.3), indicating that migration and

geneflow is unlikely between individuals and populations separated by waterway distances above 90km. Spatial autocorrelation analysis has been advocated as a useful tool in defining Operational Conservation Units (OCUs). Specifically, the intercept of the *x* axis on the correlelogram has been used as an indication of the minimum distance between samples that genetic diversity can be assessed and conserved effectively at low costs (Diniz-Filho and Telles, 2002); for the river Tweed this would equate to 47km (Figure 6.3). If the upper Tweed tributaries (Eddleston, Lyn, Manor and Upper) are grouped together as done for the inferred populations here, then the tributaries in the river Tweed are on average (S.D.) 55(17) km in length (range: 33-73km), hence, based on the results from the spatial autocorrelation analysis, the designation of each main tributary of the river Tweed as an OCU would be a reasonable approach to maintain genetic diversity within the catchment.

Role of evolutionary mechanisms acting to maintain genetic structuring within the Tweed

The grouping of samples into inferred populations was arbitrary and it is unlikely that it reflects the true structuring of populations within the Tweed catchment. However, given the results presented above, it is unlikely that the salmon of the river Tweed represent one single panmictic population (*i.e.* H_0 can be rejected). Rather, these findings are most consistent with Tweed salmon exhibiting a meta-population structuring system (*i.e.* H_{1A} can be accepted), whereby weak but significant genetic differentiation of inferred populations is observed. There is no evidence of any correlation between population level genetic divergence and geographic distance, as would have been observed if the Tweed salmon had progressed onto a member-vargrant system (*i.e.* H_{1B} can be rejected). This therefore suggests that, while there are distinct salmon populations within the Tweed catchment, there is considerable straying and migration between them. Indeed, this is also evidenced in the individual level spatial autocorrelation analysis, which suggests that straying is common between individuals separated by up to 40km and only ceases when distances exceed 90km (Figure 6.2; Figure 6.3).

These findings are consistent with those observed in other catchment level studies, whereby many rivers appear to support a meta-population hierarchy of Atlantic salmon populations (*i.e.* the river Dart in southwest England, see Chapter 4;

the Sainte Marguerite river in Canada, (Garant *et al.*, 2000)) and it is only the very large river systems, such as the river Teno in Norway/Finland (catchment area: 16,386km^2; Vaha *et al.*, 2007) and the Varzuga river in Russia (Catchment area: 5000km^2; Primmer *et al.*, 2006), which have the capacity to support a member-vagrant hierarchy of salmon populations (Vaha *et al.*, 2007).

Previous scale-reading and radio-tracking studies undertaken with Atlantic salmon on the river Tweed have implied that salmon returning to the Ettrick and Whiteadder tributaries have different run-times to salmon returning to the other tributaries of the Tweed; spring running and multi-sea-winter salmon seem to preferentially return to the Ettrick and Whiteadder, with autumn running fish and grilse returning to the other tributaries. If this has a genetic component detectable by the genotyping techniques used in this study, it would be expected that the Ettrick samples would be more differentiated from the other Tweed samples than they are from each other (the Whiteadder was not genotyped in this study). No evidence of this was found, hence it is impossible to determine if run-timing of Tweed salmon has a genetic component from this data.

Implications for management and conservation of Atlantic salmon populations in the Tweed

These results are indicative that the river Tweed supports multiple populations of Atlantic salmon within its tributaries. Populations seem to be supported by substantial straying of individuals within 40km of each other, but once distances exceed 90km, little effective straying occurs. Based on the results of the spatial autocorrelation analysis and in line with recommendations for the river Dart in southwest England (see Chapter 4) and the river Teno in Norway/Finland (Vaha *et al.*, 2007; 2008), designating each major tributary as an Operational Conservation Unit (OCU) would appear to be a reasonable approach to management and conservation of Atlantic salmon in the river Tweed. Given the range of catchment sizes between the three rivers (Dart: 475km^2; Tweed: 5000km^2; Teno: 16,386km^2) this may be a management and conservation unit applicable to other salmon rivers. Under these circumstances it would be recommended that movement of fish between tributaries should be avoided and that where supplementation programmes are used, adults used for broodstock should be taken from the same tributary and the progeny should be released in that same tributary in order to maximise the chance of survival.

From a conservation perspective, multiple tributaries from across the catchment would need to be preserved in order to maintain the genetic diversity of salmon populations within the catchment.

CONCLUSIONS

Some of these findings and the associated recommendations are based on results from inferred populations, where samples were grouped together in order to be able to undertake population level analyses. Analysis of the spatial distribution of genetic variance indicated that several different groupings of samples explained a similar amount of spatial variation between groups. Accordingly, these groupings should not be considered as precise boundaries for populations. However, the grouping of samples into inferred populations has produced results which are consistent with findings for other rivers and therefore the general results, such as the presence of meta-population structuring, may reflect the true situation for the salmon populations of the river Tweed. Furthermore, the spatial autocorrelation analysis does not require individuals to be grouped, hence this analysis, and the designation of each tributary as a distinct OCU, is not affected by the limitations introduced through grouping samples into inferred populations. Nevertheless, significant limitations were introduced through the sampling strategy employed here, indicating that it is not an appropriate sampling strategy for population genetic studies. The sampling of spawning grounds has proved very successful in other studies (Garant *et al.*, 2000; Dillane *et al.*, 2008) and would therefore be recommended for the Tweed in any future studies.

CHAPTER 7: RE-COLONISATION OF TWO TRIBUTARIES OF THE TWEED BY ATLANTIC SALMON AFTER MORE THAN 120 YEARS OF OBSTRUCTION.

The genetic analysis undertaken in this chapter was funded by the Atlantic Salmon Trust and the Tweed Foundation.

Chapters 6 and 7 have been combined with a non-technical summary and submitted to the Tweed Foundation and Atlantic Salmon Trust.

INTRODUCTION

The existence of multiple populations of Atlantic salmon within a river catchment is now well documented (Garant *et al.*, 2000; Primmer *et al.*, 2006; Vaha *et al.*, 2007; Dillane *et al.*, 2008; Chapters 4 and 6). This is believed to be due to the precise homing of Atlantic salmon to their tributary of origin after long marine migrations, and can be facilitated by habitat features within the river catchment (Stabell, 1984; Vaha *et al.*, 2007; Dillane *et al.*, 2008). However, whilst homing to their natal tributary is key to the development of within-river population structuring, the migration of individuals between populations is often an intrinsic component in maintaining populations (Garant *et al.*, 2000; Primmer *et al.*, 2006). Indeed, populations within catchments often appear to be maintained by a degree of connectivity and migration of individuals between populations, as described in the meta-population evolutionary mechanism (see Chapter 1; Figure 1.9; Chapters 4 and 6). In this situation, young fish are retained and adult fish return to the natal tributary, but the meta-population theory assumes that there is a small amount of natural connectivity between populations, but also that these populations suffer intermittent local extinction events and subsequent re-colonisation from other populations (McQuinn, 1997; Rieman and Dunham, 2000). Intrinsic to this situation are localised extinction events, which can be relatively common for Atlantic salmon populations within rivers due to factors such as pollution, flooding, drought and freeze/thaw events, with re-colonisation of the affected area occurring by individuals straying from other populations once the problem has been resolved. Hence, in catchments where meta-population models of population structuring prevail, vagrants, or migrants, are an intrinsic part in maintaining each population. The rivers Dart in southwest England (Chapter 4), the Tweed in Eastern Scotland (Chapter 6), and the Sainte Marguerite in Canada (Garant *et al.*, 2000) are all examples of systems in which connectivity between populations in the catchment is high, which can be explained by the meta-population theory of evolution (see Chapter 1; Figure 1.9 and associated text).

The concept that migration of individuals between populations is necessary to maintain populations might appear counter intuitive to the theory that salmon exhibit precise homing to their natal tributaries. However, it has been proposed that migrants may have a lower fitness than residents in a particular environment, hence

migrants will be less successful if they migrate into a stable population, but can be extremely successful if they migrate into virgin habitat after an extinction event; a concept referred to as 'selection against immigrants' (Hendry, 2004; Nosil, 2005; Bolnick and Nosil, 2007). In this way, local adaptation to natal tributaries and the associated within-catchment population structuring occurs through the homing of returning adults, but vulnerable and/or unstable populations can be maintained or re-colonised by straying and migration of salmon from other populations.

Historically, the construction of dams and weirs has been common in many river catchments to manage flow rates in order to power mills, factories and, more recently, hydro-power developments. If adequate fish passes are not installed, the presence of such obstacles interrupts the connectivity of riverine systems and, for migratory, fish, such as the Atlantic salmon, this can dramatically reduce the extent of habitat available to them within rivers. As such, the presence of dams and weirs has been cited as one of the major contributing factors to the decline of Atlantic salmon populations within rivers (MacCrimmon and Gots, 1979; Parrish *et al.*, 1998). However, when fish passes are installed, this can mitigate against some of the negative impacts of the dams and weirs by allowing passage of fish upstream and downstream of the obstruction during their migrations (Scruton *et al.*, 2006). Often, particularly for older dams and weirs constructed before environmental and ecological issues were considered in industrial design, fish passes were not included, and thus have been fitted retrospectively (Scruton *et al.*, 2006). Hence, it is common for populations of Atlantic salmon upstream of an obstruction to become locally extinct before a fish pass is installed (MacCrimmon and Gots, 1979; Parrish *et al.*, 1998). Then, once a fish pass is installed, the upstream habitat becomes available to salmon again, and individuals straying from neighbouring tributaries can instigate the formation of populations in the previously inaccessible habitat; such extinction and re-colonisation, albeit man-made, is an intrinsic component in the meta-population theory.

On the river Tweed, two tributaries were inaccessible to salmon for extended periods of time before fish passes were installed. In 1821 one of the major northern tributaries, the Gala Water, was closed to migrating fish for 130 years as a weir was installed for water abstraction. Following flooding in 1948, the weir was re-built with a fish pass and the 40km Gala Water was re-opened to migrating salmon

(Figure 7.1). Similarly, a second northern tributary of the Tweed, the Leader Water, was inaccessible to salmon for 120 years, from 1840 until 1959 (Figure 7.2).

Figure 7.1. The Gala Cauld (Weir) after being rebuilt in the 1940s with fishpass.

Figure 7.2 One of three obstructions to fish passage on the Leader.

No known supplementation programmes have been employed on either tributary and salmon have been allowed to re-colonise these previously inaccessible stretches of river naturally (R. Campbell, *personal communication*). Routine electric-fishing surveys undertaken by the Tweed Foundation find low juvenile densities in both the Gala and Leader tributaries, hence suggesting that there is some successful local recruitment in both tributaries, but not necessarily enough to support a full

population (R. Campbell, *personal communication*). This paper aims to assess which populations within the Tweed catchment the current Gala and Leader populations may have been re-colonised from, and whether the new Gala population (for which temporal samples are available) has reached a stable, self-sustaining state in the 50 years since it has been accessible.

Previous studies have identified adult salmon straying from the most proximate rivers to be the initial colonisers of previously inaccessible rivers (Jonsson *et al.*, 2003; Saura *et al.*, 2008). Whilst this phenomenon has not, to our knowledge, been assessed within a catchment, it would appear logical to hypothesise that individuals straying from the closest tributaries to the Gala and Leader tributaries, *i.e.* the Ettrick and Caddon tributaries, would be the first to colonise the Gala and Leader once they were made accessible to salmon. Hence, the null hypothesis tested here (H_0) is that all populations across the Tweed catchment contributed to the re-colonisation of the Gala and Leader populations, whilst the alternate hypotheses are, H_1: individuals straying from the geographically closest tributaries, *i.e.* the Ettrick and Caddon, are the principle colonisers of the Gala and Leader tributaries, or H_2: no signal of the original colonisers can be detected as the Gala and Leader populations have become self-recruiting, hence contemporary populations have become genetically differentiated from their founding populations. To this end, results from the previous chapter (Chapter 6) can be compared to analysis undertaken in this chapter. Furthermore, some temporal samples are available from the Gala tributary; hence, temporal stability of the current Gala population can be investigated. If the Gala populations are found to be temporally stable, this may be indicative that the population has reached a stable, self-sustaining state in the 60 years since it has been accessible to salmon, indicating H_2 to be more likely; unfortunately, similar samples are not available for the Leader.

METHODS

Details of sample collection, DNA extraction and the genotyping undertaken are given in Chapter 6.

DATA ANALYSIS

ASSESSMENT OF BASELINE DATASET AND THE ASSIGNMENT OF GALA AND LEADER SAMPLES TO THE BASELINE DATASET

The baseline dataset is comprised of potential source populations within the river Tweed, *i.e.* all samples except those from the Gala and Leader tributaries. Samples were grouped into inferred populations based on the results of the previous section (Chapter 6; Table 6.1.). Before undertaking the assignment analysis it is necessary to validate the robustness of the baseline dataset and therefore assess the power of assignment achievable with the dataset. This is achieved through a process of 'self-assignment' whereby each individual fish is removed from the baseline dataset and assigned to their most likely source population based on their individual genotype and that of the populations in the baseline. If there is strong genetic differentiation between the populations in the baseline dataset, then each individual should assign back to their known population with a high degree of certainty. This 'self assignment' procedure therefore gives an estimate of the accuracy to expect in the actual assignment test if an individual is truly from a certain population. Once these estimates are in place, it is then a case of assigning the unknown Gala and Leader samples to the baseline dataset. All assignment methods were undertaken using the Bayesian method of Rannala and Mountain (1997) as implemented in GENECLASS *v2.0* (Cornuet *et al.*, 1999; Piry *et al.*, 2004), with the self assignment analysis undertataken using the 'leave one out' option.

Potential differences in assignment of individuals from populations within the Gala and Leader tributaries, to the baseline populations (Chapter 6; Table 6.4) was assessed using Mann Whitney U tests implemented in MINITAB.

ASSESSING THE TEMPORAL STABILITY OF THE CURRENT GALA POPULATION

Genetic differentiation between samples collected from the same locations over a number of years was assessed by pairwise *F*ST analysis and Fishers exact test as implemented in ARLEQUIN *v3.1* (Schneider *et al.*, 2000). Potential differences in

assignment of individuals from temporal replicates within a sampling location on the Gala were assessed using Mann Whitney U tests implemented in MINITAB. To quantify the relative contribution of spatial variation in gene diversity among sites across the Gala tributary versus the temporal variance in gene diversity within sites of the Gala tributary, hierarchical gene diversity was assessed in an AMOVA, as implemented in ARLEQUIN *v3.1* (Schneider *et al.*, 2000). Three hierarchical levels were defined in order to assess the component of genetic diversity attributable to (i) variance among sample sites within the Gala tributary (geographic component), (ii) variance between temporal samples within sites of the Gala tributary (temporal component), and (iii) variance among individuals within sites.

ASSESSING PAST POPULATION BOTTLENECKS

Populations that have experienced a recent reduction in their effective population size exhibit a reduction in the number of alleles and gene diversity at polymorphic microsatellite loci. However, the number of alleles is reduced faster than the gene diversity; thus, in a recently bottlenecked population, the observed gene diversity is higher than the expected equilibrium gene diversity (Luikart et al. 1998). BOTTLENECK (Cornuet and Luikart, 1996; Piry *et al.*, 1999) was used to compute the observed number of alleles, under the assumption of a population in equilibrium using the Two Phased Model (TPM) of mutation, as recommended by Luikart *et al.*, (1998). The TPM consists of mostly one-step mutations with a small percentage of multi-step changes, thus it is more suited to microsatellite data than either the Infinite Allele Model or Stepwise Mutation Model of mutation (Luikart *et al.*, 1998). The Wilcoxon sign-rank test was used to assess the significance of heterozygosity excess over all loci. A mode-shift away from a normal L-shaped distribution of allele frequencies was qualitatively assessed also using BOTTLENECK (Cornuet and Luikart, 1996; Piry *et al.*, 1999), which discriminates bottlenecked populations from stable populations (Luikart and Cornuet, 1998). This analysis was undertaken on the inferred populations as described in Chapter 6 (Table 6.1g).

RESULTS

ASSESSMENT OF BASELINE DATASET AND THE ASSIGNMENT OF GALA AND LEADER SAMPLES TO THE BASELINE DATASET

Correct self-assignment of the baseline samples to their population of origin was low, with an average of 26% correct assignment (range: 0% (TilB) to 47% (EttA); Table 7.1). When assigned to tributary of origin this rose to an average of 50% correct assignment (range: 11% (TilB) to 80% EttB; Table 7.1). These results would certainly compromise our confidence in the assignment power of this technique with this dataset. For example, correct self-assignment of TilB samples was 0%, rising to only 11% correct assignment to tributary of origin. Hence, confidence in the assignment of Gala or Leader samples to the TilB sample would be low. Conversely, assignment power of Ettrick samples to their sample and tributary of origin was considerably higher (Table 7.1) therefore confidence in the assignment of Gala and Leader samples to these populations would be correspondingly higher.

Samples from the Gala tributary assigned to 9 of the 12 baseline populations (Table 7.2), ranging from 1% to the Cad and EttC samples to 42% to the TevA sample. The Gala samples assigned preferentially to TevA (42%) and the Ettrick samples (total 29%). However, 18% assigned to the TilB sample, for which confidence is low. When these results were assessed by individual populations defined in the Chapter 6 (Table 6.4), there was no significant difference in where individuals from different Gala populations assigned to in the baseline (*i.e.* no difference in where the GalA, GalB and GalC samples assigned to) (Table7.2).

Samples from the Leader tributary also assigned to 9 of the 12 baseline populations, but with a much more even spread than the Gala samples, ranging from 5% assigning to the EttC and TevA samples to 21% to the Upper sample and a total of 28% to the Ettrick tributary (Table 7.3). There was no significant difference in where individuals from different Leader populations assigned to in the baseline (*i.e.* no difference in where the LeaA, LeaB, LeaC and LeaD samples assigned to).

Table 7.1 Self assignment of baseline Tweed populations; number (percentage). Populations defined in Chapter 6, see Table 6.1.

	Cad	EttA	EttB	EttC	EttD	Mid	TevA	TevB	TilA	TilB	TilC	Upp	Trib
Cad	11 **(28)**	1 **(3)**		2 **(5)**	6 **(15)**	6 **(15)**	1 **(3)**	7 **(18)**	1 **(3)**		1 **(3)**	3 **(8)**	
EttA	6 **(13)**	21 **(47)**		1 **(2)**	10 **(22)**		1 **(2)**	1 **(2)**			1 **(2)**	4 **(9)**	32 **(71)**
EttB	1 **(7)**	3 **(20)**	5 **(33)**	3 **(20)**	1 **(7)**	1 **(7)**						1 **(7)**	12 **(80)**
EttC	1 **(4)**		1 **(4)**	7 **(29)**	3 **(12)**	1 **(4)**	4 **(17)**	1 **(4)**	1 **(4)**	1 **(4)**		4 **(17)**	11 **(46)**
EttD	5 **(9)**	6 **(11)**	3 **(5)**	3 **(5)**	22 **(39)**	4 **(7)**	3 **(5)**	6 **(11)**	1 **(2)**		1 **(2)**	2 **(4)**	34 **(61)**
Mid	6 **(26)**	2 **(9)**		2 **(9)**	4 **(17)**	3 **(13)**		3 **(13)**				3 **(13)**	
TevA	3 **(6)**	2 **(4)**	1 **(2)**	1 **(2)**			19 **(35)**	8 **(15)**	5 **(9)**	2 **(4)**	3 **(6)**	10 **(17)**	27 **(50)**
TevB	4 **(7)**	1 **(2)**		2 **(4)**	3 **(5)**	3 **(5)**	5 **(9)**	18 **(33)**	3 **(5)**		2 **(4)**	14 **(25)**	23 **(42)**
TilA		1 **(3)**		1 **(3)**		2 **(7)**	1 **(3)**	3 **(10)**	10 **(35)**		6 **(21)**	5 **(17)**	16 **(55)**
TilB				1 **(11)**			2 **(22)**	1 **(11)**	1 **(11)**			4 **(44)**	1 **(11)**
TilC					1 **(4)**		3 **(13)**	2 **(8)**	3 **(13)**	1 **(4)**	4 **(17)**	10 **(42)**	8 **(33)**
Upp	6 **(6)**	1 **(1)**	1 **(1)**	4 **(4)**	2 **(2)**	2 **(2)**	11 **(11)**	15 **(15)**	5 **(5)**	4 **(4)**	5 **(5)**	46 **(45)**	

Table 7.2 Assignment of Gala samples to baseline Tweed populations; number (percentage). Shown as original samples in different years (Figs x table y), as inferred populations as defined in Chapter 4 (Table 6.1) and as a total.

Inferred Pop	Sample ID (year)	Cad	EttA	EttB	EttC	EttD	Mid	TevA	TevB	TilA	TilB	TilC	Upp
GalA	**22 (1997)**							1					
	22 (2001)			1				1					
	22 (2003)	1		4				10			2	1	
	21 (1997)			1				1				1	
	21 (2001)			1				2					
	21 (2003)							3				2	
GalB	**19 (1997)**			1				2			1		
	19 (2001)					1							
	20 (1997)			1				1			1		
	20 (2001)							2					
	20 (2003)	1		2		4	2	4					
GalC	**13 (1997)**							2			1		
	13 (2003)			1				1			1		
	14 (2001)					1		1					
	14 (2003)			6		2		7			3	2	1
	15 (1997)			3				3			1		
	15 (2001)										1		1
	15 (2003)			2		1		5			2	2	
	16 (1997)			1				1					
	16 (2001)							3					1
	16 (2003)			2							2		
	17 (2001)							1			2		
	17 (2003)			5	1	1		8			8		1
	18 (1997)			1				2					
	18 (2003)							3			2		1
Total		2 (1)		32 (21)	1 (1)	10 (7)	2 (1)	64 (42)			27 (18)	8 (5)	5 (3)

Table 7.3 Assignment of Leader samples to baseline Tweed samples; number (percentage). Shown as inferred populations as defined in Chapter 4 (Table 6.1) and as a total.

	Cad	EttA	EttB	EttC	EttD	Mid	TevA	TevB	TilA	TilB	TilC	Upp
LeaA		1		1					3			3
LeaB					1			3				2
LeaC	3			1	2	2	1	1				2
LeaD	2	2			3	3	1	1				1
Total	5 **(13)**	3 **(8)**		2 **(5)**	6 **(15)**	5 **(13)**	2 **(5)**	5 **(13)**	3 **(8)**			8 **(21)**

Assessing the temporal stability of the current Gala population

One pairwise *F*ST test was found to be significant between the 1997 and 2003 samples taken from the GalA sample (F_{ST}=0.13113, p=0.00391), however, this was not found to be significant in the Fishers exact test and the 1997 sample consists of only one fish, hence this is an unreliable result. All other pairwise F_{ST} estimates and Fishers exact tests between temporal replicates of samples taken from the same location were found to be non-significant, hence genetic differentiation between temporal replicates is deemed, overall, to be non-significant. Furthermore, there was no significant difference in where individuals from samples taken in different years assigned to the baseline populations (Table 7.2). Hierarchical analysis demonstrated that a significant component (2.34%, p<0.001) of genetic variation was attributable to variance among sampling sites (geographic component), whereas only a small, non-significant component (0.06%, p=0.53666) of genetic variance was attributable to variance among temporal replicates within sampling sites. The remaining 97.6% was attributable to genetic variance within samples.

Assessing past population bottlenecks

There was no evidence for contemporary population size reductions, *i.e.* no recent population bottlenecks were detected, in any of the Gala or Leader populations. However, heterozygosity excess, indicative of recent declines, were observed in the EttC (p=0.01314) and TilB (0.04468) inferred populations. The TilB sample is comprised of only 9 individuals, which again may decrease confidence in this result. Mode shifts in allele frequency distributions were not evident in any of the samples.

DISCUSSION

LIMITATIONS OF THE BASELINE DATASET

The results discussed in this chapter must be viewed in full appreciation of the limitations inherent in the dataset, which were discussed at length in Chapter 6. Despite these limitations, an attempt has been made to determine the origins of the Gala and Leader populations from the baseline dataset.

The results of the self-assignment of the baseline samples (which acts as a validation exercise) are low, with an average of 26% correct assignment (range: 0% (TilB) to 47% (EttA); Table 7.1); when assigned to tributary of origin this rose to an average of 50% correct assignment (range: 11% (TilB) to 80% EttB; Table 7.1). These low self-assignment scores compromise the confidence in the power of assignment achievable, particularly to the Till tributary. However, it is not unexpected to get low accuracy in self-assignment tests in a catchment with genetically homogenous populations, such as the Tweed (see Chapter 6; Figure 6.1). Since the populations are only weakly genetically differentiated (see Chapter 6; Table 6.2), it decreases the likelihood that individuals would assign correctly, thereby limiting confidence in the findings.

ORIGIN OF GALA AND LEADER POPULATIONS

Based on the results of the assignment analysis, the most likely source of salmon colonising the Gala are individuals straying from the Teviot (to which 42% of Gala individuals assigned) the Ettrick (29% assignment) and the Till (18% assignment), with few individuals assigning to the Upper or Caddon tributaries. In contrast, colonisers of the Leader appear to have strayed from all tributaries across the Tweed catchment, with individuals straying from the Ettrick (28% assignment), Upper (21% assignment) and Teviot (19% assignment) providing the majority of colonisers. Despite relatively high proportions of Gala and Leader salmon assigning to the baseline Ettrick samples, very few assigned to the Caddon baseline sample and large proportions assigned elsewhere. Hence, these results are inconsistent with the hypothesis that the Gala and Leader tributaries would have predominantly been colonised by individuals straying from the most proximate tributaries (*i.e.* the Caddon and Ettrick tributaries). However, the lack of statistical power due to problems associated with the grouping of samples into inferred populations (see

previous chapter for further discussion) and the small sample sizes reduces confidence of all results in this chapter. However, from the available data, the hypothesis that salmon colonising the Gala and Ettrick tributaries did so from the most proximate populations (H_1) would be rejected. This means that the null hypothesis (H_0) that salmon colonising the Gala and Leader originated from populations across the catchment could be true. Alternatively, it could be argued that the genetic signal of the original colonisers may have been lost as the Gala and Leader populations may have become genetically differentiated from the founding populations since they have been re-colonised (H_2), but this would be difficult to prove with the current dataset.

Genetic differentiation of Gala and Leader populations

The assignment of individuals from the Leader tributary to many baseline tributaries may be an indication that the true origin of the Leader population is not present in the baseline dataset. Alternatively, the low self assignment scores of both the Gala and Leader individuals could be evidence that these populations have become genetically differentiated from their founding populations and the other populations across the rest of the Tweed catchment over the 50-60 year period that the tributaries have been accessible to salmon. Indeed, when included in the baseline populations, self-assignment of Gala and Leader populations to the correct population of origin (*i.e.* to the Gala and Leader populations, respectively) is comparable to the other baseline populations (see Chapter 6; Table 6.4). In fact, the Gala populations show the highest self-assignment scores for both individual population (GalC: 53%) and tributary level (GalB 65%) analysis. Such findings are consistent with the theory that, particularly the Gala populations, but also potentially the Leader populations, have become stable self-recruiting populations in their own right in the 50-60 years since they have been accessible to salmon. Furthermore, in the catchment-wide analysis conducted in Chapter 6, hierarchical gene diversity analysis determined that there were potentially multiple, genetically differentiated populations within the Gala and Leader tributaries (see Chapter 6; Table 6.1). Significant genetic differentiation between these populations and other inferred populations across the catchment was observed in the pairwise F_{ST} estimates (see Chapter 6; Table 6.2). None of the inferred Gala or Leader populations showed any signs of recent genetic bottlenecks; hence, there is no evidence of contemporary population size reductions, as might be

expected in a recently founded population. This may be indicative that the Gala and Leader populations have reached a stable, equilibrium state, thus corroborating the theory that the Gala and Leader populations have become genetically distinct and stable populations.

Temporal stability of the Gala populations

There are numerous lines of evidence which support the theory that populations within the Gala tributary are temporally stable. Firstly, there were no significant differences in where temporal replicates from the same sampling location assigned to in the baseline populations. Secondly, there was no significant genetic differentiation between temporal samples within sites as estimated in pairwise *F*ST estimates, except for sample ID 22 taken in 1997 and 2003 (Table 2.4). However, sample ID 22 taken in 1997 had only one individual and therefore is not a representative sample. Finally, the hierarchical gene diversity analysis determined that a significant 2.34% of genetic variation within the Gala tributary was due to spatial variation between sampling sites, while only a non-significant 0.06% of genetic variation was attributable to temporal variation within sampling locations. The remaining 97.6% was due to genetic variation within samples. Collectively these results indicate that the Gala populations may be stemporally stable, however, some caution must be exercised in any interpretation of this data since many sample sizes are extremely small (Table 2.2).

Implications for management and conservation of Tweed salmon

Studying the within river genetic variation of Atlantic salmon populations is a relatively new avenue of research (Garant *et al.*, 2000; Primmer *et al.*, 2006; Vaha *et al.*, 2007; Dillane *et al.*, 2008). Hence, determining the origin of colonisers within catchments, which have founded populations in tributaries that have only recently been made accessible to Atlantic salmon, has not, to our knowledge, been attempted previously. Instrumental to the success of such a project is the level of genetic diversity observed within a catchment. In the over-arching investigation into genetic diversity of salmon populations across the Tweed catchment, only weak genetic differentiation was observed (see Chapter 6), which certainly hampered attempts to determine the origins of the current Gala and Leader populations. Furthermore, the low statistical power when samples were grouped into inferred populations significantly reduces confidence in the validity of results in both investigations.

Previous research has attempted to assess the source of Atlantic salmon colonising whole river catchments. In these earlier studies, baseline populations included neighbouring rivers and results suggested that it is salmon straying from the closest neighbouring tributaries that initially colonised newly accessible rivers (Jonsson *et al.*, 2003; Saura *et al.*, 2008). Hence, the first alternative hypothesis tested here was that salmon from the closest neighbouring tributaries to the Gala and Leader (*i.e.* the Ettrick and Caddon tributaries; H_1) would have been the principle colonisers of the Gala and Leader populations. However, this was not evidenced in the assignment analysis, neither are associations between these tributaries observed in the population tree (see Chapter 6; Figure 6.1). This may be interpreted in several ways: firstly, colonisation of the Gala and Leader tributaries may have been from salmon straying from all populations across the Tweed catchment; alternatively, the weak differentiation of baseline populations may be obscuring the true signal; and finally, there may be no genetic signal of the original colonisers left in the Gala and Leader populations as they have progressed onto stable, self-recruiting populations which no longer rely on migrants from other populations to maintain them (H_2). The apparent temporal stability of the Gala samples may indicate that H_2 is correct, but caution must be taken in interpretation here since sample sizes were very small. Data from Chapter 6 were most consistent with the theory that the Tweed catchment supports a meta-population structure of Atlantic salmon populations. By definition, straying and migration of salmon between populations across the catchment helps to maintain the overall population structuring. Hence, it is unlikely that the Gala and Leader populations receive no migrants from other populations, but it is impossible to determine from this data which the true situation is, and indeed there might be interplay between these possibilities. In reality, it is impossible to be confident in these findings given the limitations of the whole dataset, as discussed in depth in Chapter 6.

While there is no evidence to suggest that the current Gala population has not reached a stable state, due to the original sample sizes of the temporal replicates, it is impossible to declare the population as stable with absolute confidence. However, these findings certainly warrant further investigation through the genotyping of much larger samples of temporal replicates of Gala and Leader samples, as also advised for the catchment wide analysis in Chapter 6. Nevertheless, despite the low confidence

in these results, if they can be verified in further analysis of larger samples, then this would indicate that salmon can re-establish previously extirpated tributaries, and form stable, self-recruiting populations in relatively short periods of time, *i.e.* in 50-60 years. This adds weight to the theory that stocking of previously extirpated rivers with captive reared salmon is an un-necessary risk to native stocks and a waste of financial resources for fisheries managers.

To address the initial aim of this study, which was to determine the origins of the recently founded Gala and Leader populations, has proved extremely difficult. Managerial and conservation recommendations would have to incorporate the Precautionary Principle and take into account the results of the catchment wide analysis (Chapter 6). Thus, treating the Gala and Leader tributaries as discrete Operational Conservation Units, as recommended in Chapter 6, would be a logical and practical approach to catchment wide management and conservation of Atlantic salmon within the Tweed, whilst recognising the requirement for further analysis of larger sample sizes to improve the robustness of data underpinning such recommendations for the Tweed.

CHAPTER 8: GENERAL DISCUSSION

The overall objective of this research project was to advance understanding of the genetic population structuring of Atlantic salmon in northwest Europe. To this end, a series of investigations were undertaken at both regional and local scales, with varying degrees of success. These studies, and their application in management and conservation terms, are discussed herewith.

UNRAVELLING THE PHYLOGEOGRAPHIC HISTORY OF ATLANTIC SALMON IN NORTHWEST EUROPE

The utilisation of both nuclear microsatellite and mtDNA PCR-RFLP markers proved very successful in elucidating a robust glacial history and post-glacial colonisation routes of Atlantic salmon in northwest Europe since the last glacial maximum (Chapter 3). For the first time, evidence for a glacial refuge north of the Iberian Peninsula, located in northwest France, was presented, suggesting that colonisation of Britain and Ireland originated from salmon expanding out of both refugial locations. This is the first phylogeographic study to focus on this region since the 1970s and the concordance of both the microsatellite and mtDNA data increases the confidence in these findings.

Since no Baltic populations had been genotyped during the ASAP project in which the microsatellite data was generated, it was not possible to determine whether French and Baltic salmon diverged before the last glacial maximum, or whether one derived from the other. In an attempt to address this issue, two individuals from six populations across the study area were chosen for sequencing at the ITS1 region; this included two individuals from the river Tornio in Finland, which displayed the traditional Baltic haplotype. An internal transcribed spacer (ITS) is a piece of non-functional RNA situated between ribosomal RNAs (rRNA) on a common precursor transcript. Genes encoding rRNA and spacers are recombining, bi-parentally inherited markers which occur in tandem repeats. Due to their high evolutionary rates, sequence comparison of the ITS region has been used extensively in inter-specific phylogenetic studies between closely related species. More recently, some studies have shown variation within species on large geographic scales, highlighting their potential for use in phylogeographic studies. Presa *et al*. (2002) sequenced the ITS1 region from brown trout taken from populations spanning their native range

and found there to be a comparable amount of genetic variation within the mtDNA control region. Genetic variation was observed on large geographic scales, which was partially congruent with mtDNA data, and sequencing of the ITS region proved useful for phylogeographic interpretation in this closely related species. However, sequencing of the ITS1 gene in these samples did not produce any variants and so the technique was omitted from the main results chapter. However, if a gene can be located which exhibits regional variation, this may help to determine if the French populations have descended from Baltic populations, or whether they became differentiated from the Spanish refugial populations during the last glacial period.

Evidence for small refugia located north of the traditionally cited southern refugia of the Iberian, Italian or Balkan peninsulas has been presented for many cold tolerant species with a current northern distribution. Indeed, in an analogous situation, Atlantic salmon are believed to have persisted in small refugial locations throughout the Baltic region during the last glacial period. Being the first study since the 1970s to extensively sample from northwest Europe, it is perhaps not surprising that a similar situation has been observed in this region. Extending the sample set to include additional samples from France and southwest England may help to delineate geographical boundaries of the refuge, and including samples from the Baltic may help to determine the true origin of the French populations. As for Atlantic salmon in the Baltic, this information can be useful for informing regional management and conservation strategies in order to maintain the genetic diversity of the species across the region in the long-term.

WITHIN RIVER STRUCTURING OF ATLANTIC SALMON POPULATIONS

As population geneticists, we are continuously advocating the value of using genetics techniques to inform the management of Atlantic salmon populations across all geographic scales. However, conducting a detailed catchment level analysis is costly and time consuming. It is therefore unrealistic to assume that all rivers managers can afford (or would like) this level of detailed investigation. Hence, there will come a time when if we want to help inform management issues on a broad scale, we must be able and willing to make generalisations. By undertaking the same analyses on two catchments of very different characters, I (and the co-authors on the resulting papers) have attempted to move some way towards this end. Two

catchment level investigations were conducted, one on the river Dart in southwest England (Chapter 4) and one on the river Tweed in eastern Scotland (Chapter 6). The Dart is a small catchment (catchment area: 475km^2) draining the granite massif of Dartmoor, whilst the Tweed is the second largest salmon river in Scotland (catchment area: 5000km^2) and is highly dendritic, draining the Lammermoor and Moorfoot Hills to the north and the Cheviot Hills to the south. By genotyping multiple samples of Atlantic salmon across these two very different catchments using the same microsatellite markers, the aim was to assess the extent of population structuring within these catchments of differing nature and size. In doing so, it was intended that inferences could be drawn on whether the geological nature of a catchment would have any bearing on the extent of population structuring of Atlantic salmon populations within catchments, *i.e.* would the larger, more dendritic Tweed catchment have several highly genetically differentiated populations of Atlantic salmon, analogous to the member-vagrant theory of evolution, while the smaller, less dendritic Dart catchment might have several weakly genetically differentiated populations, more analogous to the meta-population theory of evolution?

It was determined that within both catchments, several weakly differentiated populations of Atlantic salmon were present, analogous to the meta-population evolutionary theory. Furthermore, individual spatial autocorrelation analysis on both rivers determined that the ideal Operational Conservation Unit (OCU) would be at the level of major tributary. Encouragingly, this has also been recommended for one of the largest salmon rivers in Europe, the river Teno (catchment area: 16,386km^2) (Vaha *et al.*, 2007). This may therefore be an indication that population geneticists could advise catchment level managers to consider each major tributary as a distinct population or unit in order to maintain genetic diversity across a catchment.

LIMITATIONS OF THE DATASET

Limitations in this study are introduced through small sample sizes and the opportunistic, rather than targeted, nature of sampling. The comparisons drawn between the two rivers must be viewed with full knowledge of the differing sampling strategies employed.

For the river Dart, sampling locations were chosen based on pre-existing data held by the Environment Agency (EA). The EA have a number of sites they visit

annually or every five years, and further sites which historically were part of the sampling strategy but are no longer included in the programme. From available data provided by the EA, sites were selected for sampling based on the likelihood of salmon being present. These sites were originally chosen by the EA probably due to ease of access, rather than based on any relevant biological parameters, such as proximity to spawning grounds. Most of the sampling undertaken on the river Dart was done with the Devon EA electric-fishing team during annual surveys, while the remainder were EA sites, with sampling undertaken by the Universities electric-fishing team.

Almost all sampling locations in the river Dart were in the upper reaches and many of the tributaries sampled are very narrow (*i.e.* <1m in width), shallow streams. Hence, obtaining sufficient numbers of salmon to sample was problematic. If time allowed, the EA electric-fishing team would fish additional runs to increase the sample size, but as they have a tight schedule to keep to this was not always possible. At some of the more inaccessible sites fished by the University electric-fishing team, the length of river which could feasibly be fished was not always long enough to catch high numbers of fish, and often the team would spend a number of hours fishing a stretch with poor return.

Hence, it is acknowledged that sample sizes for this type of analysis would be greater than achieved for the river Dart, but this was largely due to the nature of the river and the use of the EA electric-fishing team, without whom it would have been very difficult to visit many of the sites. Since the sampling of spawning grounds has proven successful in other recent studies (Garant *et al*., 2000; Dillane *et al*., 2008), it would be recommended that this be employed in future work, rather than the opportunistic sampling as was undertaken in this study. Since large areas of the upper tributaries would need to be sampled in order to dramatically increase sample sizes, it would also be prudent to map the precise extent of the sampling site in order to take account of this in any analyses, *i.e.* in a Geographic Information System (GIS).

Samples from the river Tweed were collected by the Tweed Foundation prior to conception of this project. Small samples, ranging between 1 to 31 individuals, were taken more or less continuously across the Tweed catchment. This provided an

interesting opportunity to assess the efficacy of this sampling method for population genetics studies. It is an appropriate sampling strategy with which to conduct clustering (*i.e.* STRUCTURE or BAPS) or spatial autocorrelation analysis, since in these types of analyses the relationships between individual fish are explored. However, it is less appropriate for the estimation of standard population genetics parameters, such as Hardy Weinberg expectations, F_{ST}, allelic richness, isolation by distance, and so on, since these assume that each sample is representative of a true population and require an adequate number of representatives of the population. Had the clustering analyses been able to partition the raw data into discrete groups (inferred populations), this sampling strategy might have been successful. However, it was not possible to isolate geographic boundaries using clustering analyses because significant genetic differentiation was not detectable with the molecular markers used. Hence, samples were grouped almost arbitrarily, based loosely on a series of AMOVA analyses. It is therefore unlikely that this reflects the true population structuring of Atlantic salmon in the Tweed catchment. Nevertheless, significant genetic differentiation was observed between the inferred populations, and results were analogous to those generated for the river Dart. This therefore raises the issue of whether opportunistic sampling, as conducted on the Dart, is a true representation of the population being sampled. Similarly for recommendations made for future studies on the river Dart, it would also be recommended that sampling was focussed on spawning habitat for any future work undertaken in the Tweed.

ASSESSING THE IMPACT OF ANTHROPOGENIC ACTIVITIES ON WITHIN RIVER ATLANTIC SALMON POPULATIONS

For the last 100 years or more, it has been common for wild populations of Atlantic salmon to be supplemented with fish from different rivers and/or with artificially reared fish. The reasons have largely been to bring more and/or larger fish to the recipient river; the justifications have been many and varied, including improving the number of returns, in the hope that the larger donor fish will replace the smaller native fish, to improve 'hybrid vigour', and so on. There is very little evidence to corroborate the success of such practices and mounting evidence, particularly from the field of genetics, that individual stocking/supplementation programmes are ineffective and, more importantly, that these practices can have detrimental impacts on native populations. Regardless of this evidence, these

practices continue across the species range. In Chapter 5, the outcome of historical supplementation practices undertaken on the river Dart on contemporary population structure was assessed. During the 1960s, and probably for long periods before and after this, native Dart stocks were supplemented by eyed ova imported from Iceland and Scotland. Admixture analyses was undertaken on historical populations from the donor stocks (*i.e.* from the 1960s) and compared to contemporary Dart populations.

Overall, admixture between the historical donor stocks and the contemporary recipient stocks was deemed to be very low, which possibly reflected the underlying natural genetic relationships of populations across the region. Genetic differentiation of Atlantic salmon populations, though often described by neutral molecular markers such as microsatellites, is believed to reflect underlying adaptive differentiation to the natal habitat. Accordingly, what constitutes suitable habitat for one population may not be suitable for another. Obvious differences in habitat type between the donor (Icelandic and Scottish) and recipient (Dart) rivers are evident. For example, the river Dart is a much smaller, less dendritic system compared to the Tweed or Ellidaar. It is also highly acidic, has shallow gradients (which would influence flow regimes) and has a considerably warmer climate. Each population will have specific adaptations for their native habitat and these local adaptations are likely to be a contributing factor in the general failure of the transplanted Icelandic and Scottish salmon into non-native river Dart (Verspoor *et al*, 2007).

However, admixture of the donor stocks with samples from one tributary was consistently higher, potentially indicating that the donor stocks did persist in some tributaries at a very local scale. Why this would be is difficult to determine, but speculation on whether there was a greater carrying capacity available in this tributary, or whether conditions here were more suitable for the transplanted fish, are two possibilities of many which are impossible to discern after such a long time period. Hence, over long time-scales, exogenous fish do not appear to persist in their recipient rivers over large spatial scales, but there may be some localised success.

Since historical samples for the river Dart from before the time of the supplementation programmes were not available, this study effectively lacked a complete baseline (*i.e.* samples from the Dart from before the stocking took place).

Nevertheless, the statistical technique of admixture analysis did prove useful in this situation where the baseline samples were lacking.

One of the major factors associated with habitat loss for Atlantic salmon within river catchments are man-made barriers to migration, constructed for such things as water abstraction and hydropower development. Two of the major northern tributaries of the river Tweed, the Gala and Leader tributaries, were inaccessible to salmon for long periods during the 1800-1900s due to hydro-power developments. With fish passes installed in the 1940s, salmon have been able to migrate into the Gala and Leader tributaries for the past 60 years. In Chapter 7 the aim was to determine from which populations across the Tweed catchment the Gala and Leader were colonised, and if the Gala had become temporally stable. Unfortunately, since the inferred populations across the Tweed were only found to be weakly genetically differentiated validation of the baseline samples in the self-assignment analysis was low, and hence, this lowered the confidence in the assignment of the Gala and Leader samples to the baseline populations. Contrary to what has been observed in between-catchment studies (Jonsson *et al.*, 2003; Saura *et al.*, 2008), there was no evidence that salmon straying from the most proximate tributaries were the initial colonisers of the re-opened tributaries. Whether this is a true reflection of the situation or is a product of sample collection and the genotyping methods used is impossible to determine in this investigation. Nevertheless, there was no evidence to suggest that the Gala populations have not reached a temporally stable state; therefore, this may indicate that this recently re-colonised tributary has established a self-recruiting population in the 60 years since it has been re-opened. Similarly to the catchment-wide Tweed analysis there are limitations in the data which need to be acknowledged and the collection of larger samples of individuals is required in order to verify these results.

With results from Chapter 5 indicating that non-native salmon stocked into rivers do not contribute in the long-term to the recipient population, alongside the results from Chapter 7, which indicate that re-colonised populations might have the capacity to become temporally-stable, self-recruiting populations in less than 60 years, the case

against supplementation of wild stocks with non-native or hatchery reared fish is strongly supported. These additions to the debate will not buck the trend of the last 100 years or more, but by adding to the body of evidence that argues against the use of stocking and supplementation programmes as a management practice, hopefully there will come a time in the not too distant future when fisheries managers can no longer dismiss the evidence so easily.

POTENTIAL IMPROVEMENTS AND ADDITIONAL AVENUES OF RELATED RESEARCH

There are always aspects of a project which could benefit from modifications, enhancements or additional research. For this project, modifications to the sampling strategies would have been beneficial and would be advised for future studies, as has been discussed at length earlier in this chapter. In hindsight, alterations to the genotyping methods used and modifications to laboratory protocols would also be made. There is a will within this field of research to achieve the best possible outcomes for the species based on the best possible science (Nielsen *et al.*, 2007). In line with this principal, there is a desire within the field to adopt a standardised panel of microsatellites, named the Virginia panel, to be used across the species range, where standardisation procedures can easily be undertaken, thus enabling cross-laboratory comparisons to be made which is usually difficult to achieve with standard microsatellite data. Using such a multiplex would not only have increased laboratory efficiency, therefore resulting in more effective use of the limited time available, but would also have made a contribution to a range-wide database for the use by other researches for years to come.

Had extra time been available, additional analyses could have been undertaken. It was initially envisaged that the genetic data would be incorporated into a Geographic Information System (GIS), and as such would open up possibilities for landscape genetic analyses; an area of research which has receieved considerable attention in recent publications (Dionne *et al.*, 2008; Dillane *et al.*, 2008). This new avenue for research enables the analysis of disperate datasets concurrently, for example, genetic data, geographic distances and any other number of biotic or abiotic factors, such as temperature, flow, stocking history. A range of meaningful geographic scales can be explored, irrespective of geographical and/or

political boundaries, and can provide insights into regional eveolutionary histories for the species in relation to historic patterns of glaciation, volcanism or other geological activity leading to vicariant or discontinuous biogeographic boundaries. These capabilities offer managers and conservationists a diverse array of tools to aid decisions on how to safeguard biodiversity within the species. Whilst a GIS was initially set up for this study, there was insufficient time to develop the system into one sophisticated enough to address such issues. However, this is certainly an exciting area of developing research.

CONCLUSIONS

Whilst individual studies such as the ones undertaken in this thesis might only have direct relevance for the individual study site, it is always important to strive to use the information in a wider context. It is hoped that the knowledge of the genetic structuring of Atlantic salmon populations gained in this research project will not only help inform interested parties, managers and conservationists directly involved with these study locations, but that these studies will help to further the rapidly increasing body of knowledge in the genetic structuring of Atlantic salmon populations across their native range, and in this way contribute to informing those who strive to manage and conserve this enigmatic species.

REFERENCES

Allendorf, F. W. & Thorgaard, G. H. (1984). Tetraploidy and the evolution of salmonid fishes. In *Evolutionary Genetics of Fishes.* (Turner, B. J., ed.), pp. 1-53. New York: Plenum.

Andersen, B. G. & Borns, W. W. (1994). *The Ice Age World.*: Scandinavian University Press, Oslo, Norway.

Aprahamian, M. W., Smith, K. M., McGinnity, P., McKelvey, S. & Taylor, J. (2003). Restocking of salmonids - opportunities and limitations. *Fisheries Research* **62**, 211-227.

Araki, H., Cooper, B. & Blouin, M. S. (2007). Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. *Science* **318**, 100-103.

Arnesen, A. M., Toften, H., Agustsson, T., Stefansson, S. O., Handeland, S. O. & Bjornsson, B. T. (2003). Osmoregulation, feed intake, growth and growth hormone levels in 0+ Atlantic salmon (*Salmo salar* L.) transferred to seawater at different stages of smolt development. *Aquaculture* **222**, 167-187.

Asplund, T., Veselov, A., Primmer, C. R., Bakhmet, I., Potutkin, A., Titov, S., Zubchenko, A., Studenov, I., Kaluzchin, S. & Lumme, J. (2004). Geographical structure and postglacial history of mtDNA haplotype variation in Atlantic salmon (*Salmo salar* L.) among rivers of the White and Barents Sea basins. *Annales Zoologici Fennici* **41**, 465-475.

Avise, J. C. (1994). *Molecular Markers, Natural History and Evoloution.*: Chapman & Hall.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Bakke, T. A., Jansen, P. A. & Hansen, L. P. (1990). Differences in the Host-Resistance of Atlantic Salmon, *Salmo salar* L, Stocks to the Monogenean *Gyrodactylus salaris* Malmberg, 1957. *Journal of Fish Biology* **37**, 577-587.

Ballard, J. W. O. & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.

Beacham, T. D. & Dempson, J. B. (1998). Population structure of Atlantic salmon from the Conne river, Newfoundland as determined from microsatellite DNA. *Journal of Fish Biology* **52**, 665-676.

Beaugrand, G. E. G. & Reid, P. C. (2003). Long-term changes in phytoplankton, zooplankton and salmon related to climate. *Global Change Biology* **9**, 801-817.

Beebee, T. J. C. & Rowe, G. (2004). *An Introduction to Molecular Ecology*: Oxford University Press.

Berglund, I., Hansen, L. P., Lundqvist, H., Jonsson, B., Eriksson, T., Thorpe, J. E. & Eriksson, L. O. (1991). Effects of Elevated Winter Temperature on Seawater Adaptability, Sexual Rematuration, and Downstream Migratory Behavior in Mature Male Atlantic Salmon Parr (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 1041-1047.

Bernatchez, L. & Wilson, L. L. (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology* **7**, 431-452.

Bhagwat, S. A. & Willis, K. J. (2008). Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *Journal of Biogeography* **35**, 464-482.

Blanco, G., Sanchez, J. A., Vazquez, E., Rubio, J. & Utter, F. M. (1992). Genetic Differentiation among Natural European Populations of Atlantic Salmon, *Salmo salar* L, from Drainages of the Atlantic- Ocean. *Animal Genetics* **23**, 11-18.
Boeuf, G. (1994). Salmonid smolting: a pre-adaptation to the oceanic environment. In *Fish Ecophysiology* (Rankin, J. C. & Jensen, F. B., eds.), pp. 105-135. London: Chapman & Hall.
Bolnick, D.I. & Nosil, P. (2007). Natural selection in populations subject to a migration load. *Evolution*. **61**, 2229-2243.
Bourke, E. A., Coughlan, J., Jansson, H., Galvin, P. & Cross, T. F. (1997). Allozyme variation in populations of Atlantic salmon located throughout Europe: diversity that could be compromised by introductions of reared fish. *Ices Journal of Marine Science* **54**, 974-985.
Bowen, D. Q., Phillips, F. M., McCabe, A. M., Knutz, P. C. & Sykes, G. A. (2002). New data for the Last Glacial Maximum in Great Britain and Ireland. *Quaternary Science Review* **21**, 89-101.
Boyce, M. S. (1992). Populations viability analysis. *Annual Review of Ecology and Systematics* **23**, 481-506.
Bridgeland, D. R. & D'Olier, B. (1995). The Pleistocene evolution of the Thames and Rhine drainage systems in the southern North Sea Basin. In *Island Britain: a Quternary perspective* (Preece, R. C., ed.), pp. 27-45: Geological Society Special Publications.
Brumfield, R. T., Beerli, P., Nickerson, D. A. & Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology & Evolution* **18**, 249-256.
Cairney, M., Taggart, J. B. & Hoyheim, B. (2000). Characterization of microsatellite and minisatellite loci in Atlantic salmon (*Salmo salar* L.) and cross-species amplification in other salmonids. *Molecular Ecology* **9**, 2175-2178.
Child, A. R., Burnell, A. M. & Wilkins, N. P. (1976). The existence of two races of Atlantic salmon (*Salmo salar* L.) in the British Isles. *Journal of Fish Biology* **8**, 35-43.
Chistiakov, D. A., Hellemans, B. & 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.
Clair, T. A., Dennis, I. F., Amiro, P. G. & Cosby, B. J. (2004). Past and future chemistry changes in acidified Nova Scotian Atlantic salmon (Salmo salar) rivers: a dynamic modeling approach. *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 1965-1975.
Consuegra, S., De Leaniz, C. G., Serdio, A., Morales, M. G., Straus, L. G., Knox, D. & Verspoor, E. (2002). Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular Ecology* **11**, 2037-2048.
Cornuet, J. M. & Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**, 2001-2014.
Cornuet, J. M., Piry, S., Luikart, G., Estoup, A. & Solignac, M. (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**, 1989-2000.
Crespi, B. J. & Fulton, M. J. (2004). Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. *Molecular Phylogenetics and Evolution* **31**, 658-679.

Crozier, W. W., Schon, P. J., Chaput, G., Potter, E. C. E., Maoileidigh, N. O. & MacLean, J. C. (2004). Managing Atlantic salmon (*Salmo salar* L.) in the mixed stock environment: challenges and considerations. *ICES Journal of Marine Science* **61**, 1344-1358.
Cutler, M. G., Bartlett, S. E., Hartley, S. E. & Davidson, W. S. (1991). A Polymorphism in the Ribosomal-Rna Genes Distinguishes Atlantic Salmon (*Salmo Salar*) from North-America and Europe. *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 1655-1661.
Dalgaard, M. B., Nielsen, C. V. & Buchmann, K. (2003). Comparative susceptibility of two races of *Salmo salar* (Baltic Lule river and Atlantic Conon river strains) to infection with *Gyrodactylus salaris*. *Diseases of Aquatic Organisms* **53**, 173-176.
Dillane, E., McGinnity, P., Coughlan, J. P., Cross, M. C., de Eyto, E., Kenchington, E., Prodohl, P. & Cross, T. F. (2008). Demographics and landscape features determine intrariver population structure in Atlantic salmon (*Salmo salar* L.): the case of the River Moy in Ireland. *Molecular Ecology* **17**, 4786-4800.
Diniz-Filho, J.A.F. & M, Pires De Campos Telles (2001). Spatial autocorrelation analysis and the identification of operational units for conservation in continuous populations. *Conservation Biology* **16**, 924-935.
Dodson, J. J., Gibson, R. J., Cunjak, R. A., Friedland, K. D., de Leaniz, C. G., Gross, M. R., Newbury, R., Nielsen, J. L., Power, M. E. & Roy, S. (1998). Elements in the development of conservation plans for Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 312-323.
Ducharme, L. J. A. (1969). Atlantic salmon returning for their fifth and sixth consecutive spawning trips. *Journal of the Fisheries Research Board of Canada* **26**, 1661-1664.
Environment Agency (2003). Salmon Action Plan: River Dart. pp. 18-20: Environment Agency.
Einum, S. & Fleming, I. A. (2001). Implications of stocking: ecological interactions between wild and released salmonids. *Nordic Journal of Freshwater Research* **75**, 56-70.
Einum, S., Fleming, I. A., Cote, I. M. & Reynolds, J. D. (2003). Population stability in salmon species: effects of population size and female reproductive allocation. *Journal of Animal Ecology* **72**, 811-821.
Ellegren, H. (2000). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics* **16**, 551-558.
Elliott, J. M. (1994). *Quantitative Ecology and the Brown Trout*: Oxford University Press.
Elo, K., Vuorinen, J. A. & Niemela, E. (1994). Genetic resources of Atantic salmon (*Salmo salar* L.) in Teno and Naatamo Rivers, northernmost Europe. *Hereditas* **120**, 19-28.
Estoup, A., Largiader, C. R., Perrot, E. & Chourrout, D. (1996). Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. *Molecular Marine Biology and Biotechnology* **5**, 295-298.
Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
Felsenstein, J. (1993). PHYLIP: Phylogeny inference package. Version 4.0.
Fleming, I. A. (1996). Reproductive strategies of Atlantic salmon: Ecology and evolution. *Reviews in Fish Biology and Fisheries* **6**, 379-416.

Fleming, I. A. (1998). Pattern and variability in the breeding system of Atlantic salmon (*Salmo salar*), with comparisons to other salmonids. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 59-76.
Fleming, I. A., Lamberg, A. & Jonsson, B. (1997). Effects of early experience on the reproductive performance of Atlantic salmon. *Behavioral Ecology* **8**, 470-480.
Fleming, I. A. & Petersson, E. (2001). The ability of released, hatchery salmonids to breed and contribute to the natural productivity of wild populations. *Nordic Journal of Freshwater Research* **75**, 71-98.
Friedland, K. D., Chaput, G. & MacLean, J. C. (2005). The emerging role of climate in post-smolt growth of Atlantic salmon. *ICES Journal of Marine Science* **62**, 1338-1349.
Friedland, K. D., Hansen, L. P., Dunkley, D. A. & MacLean, J. C. (2000). Linkage between ocean climate, post-smolt growth, and survival of Atlantic salmon (*Salmo salar* L.) in the North Sea area. *ICES Journal of Marine Science* **57**, 419-429.
Garant, D., Dodson, J. J. & Bernatchez, L. (2000). Ecological determinants and temporal stability of the within-river population structure in Atlantic salmon (*Salmo salar* L.). *Molecular Ecology* **9**, 615-628.
Garcia de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Consuegra, S., Jordan, W. C., Aubin-Horth, N., Lajus, D. L., Villanueva, B., Ferguson, A., Youngson, A. F. & Quinn, T. P. (2007). Local Adaptation. In *The Atlantic Salmon: Genetics, Conservation and Management* (Verspoor, E., Stradmeyer, S. & Nielsen, J. L., eds.), pp. 196-221. Oxford: Blackwell Publishing Ltd.
Gemmell, N. J., Metcalf, V. J. & Allendorf, F. W. (2004). Mother's curse: the effect of mtDNA on individual fitness and population viability. *Trends in Ecology & Evolution* **19**, 238-244.
Gibbard, P. L. (1988). The history of the great northwest European rivers during the last three million years. *Philosophical Transactions of the Royal Society of London. B.* **318**, 559-602.
Gibbard, P. L. & Lewin, J. (2003). The history of the major rivers of southern Britain during the Tertiary. *Jounal of the Geological Society of London.* **160**, 829-845.
Gilbey, J., Verspoor, E., McLay, A. & Houlihan, D. (2004). A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Animal Genetics* **35**, 98-105.
Goldstein, D. B. & Pollock, D. D. (1997). Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* **88**, 335-342.
Goldstein, D. B. & Schlotterer, C., eds. (2001). *Microsatellites: Evolution and Applications*: Oxford University Press.
Goudet, J. (1995). FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.
Gowans, A. R. D., Armstrong, J. D., Priede, I. G. & McKelvey, S. (2003). Movements of Atlantic salmon migrating upstream through a fish- pass complex in Scotland. *Ecology of Freshwater Fish* **12**, 177-189.
Griffiths, A. M., Koizumi, I., Bright, D. & Stevens, J. R. (In press). A case of within catchment isolation by distance and short-term temporal stability of population structure in brown trout (*Salmo trutta*). *Ecological Applications*.
Guo, S. W. & Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* **48**, 361-372.
Hall, H. J. & Nawrocki, L. W. (1995). A rapid method for detecting mitochondrial DNA variation in the brown trout, *Salmo trutta*. *Journal of Fish Biology* **46**, 360-364.

Hanfling, B., Hellemans, B., Volckaert, F. A. M. & Carvalho, G. R. (2002). Late glacial history of the cold-adapted freshwater fish *Cottus gobio*, revealed by microsatellites. *Molecular Ecology* **11**, 1717-1729.
Hansen, L. P., Jonsson, N. & Jonsson, B. (1993). Oceanic Migration in Homing Atlantic Salmon. *Animal Behaviour* **45**, 927-941.
Hansen, L. P. & Pethon, P. (1985). The food of Atlantic salmon, *Salmo salar* L. caught by long-line in Northern Norwegian waters. *Journal of Fish Biology* **26**.
Hansen, L. R. & Quinn, T. R. (1998). The marine phase of the Atlantic salmon (*Salmo salar*) life cycle, with comparisons to Pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 104-118.
Hansen, M. M., Nielsen, E. E., Bekkevold, D. & Mensberg, K. L. D. (2001). Admixture analysis and stocking impact assessment in brown trout (*Salmo trutta*), estimated with incomplete baseline data. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 1853-1860.
Hansen, M. M., Ruzzante, D. E., Nielsen, E. E., Bekkevold, D. & Mensberg, K. L. D. (2002). Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. *Molecular Ecology* **11**, 2523-2535.
Hardy, O. J., Charbonnel, N., Freville, H. & Heuertz, M. (2003). Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation. *Genetics* **163**, 1467-1482.
Hardy, O. J. & Vekemans, X. (2002). SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618-620.
Hendry, A.P. (2004). How much of the variation in adaptive divergence can be explained by gene flow? - An evaluation using lake-stream stickleback pairs. *Evolution* **58**, 2319-2331.
Hendry, A. P., Letcher, B. H. & Gries, G. (2003). Estimating natural selection acting on stream-dwelling Atlantic salmon: Implications for the restoration of extirpated populations. *Conservation Biology* **17**, 795-805.
Hewitt, G. M. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.
Hindar, K., Ryman, N. & Utter, F. M. (1991). Genetic effects of cultured fish on natural fish populations. *Canadian Journal of Fisheries and Aquatic Sciences* **48**, 945-957.
Horreo, J. L., Machado-Schiaffion, G., Griffiths, A. M., Bright, D., Stevens, J. R. & Garcia-Vazquez, E. (2008). Identification of differential broodstock contribution affecting genetic variability in hatchery stocks of Atlantic salmon (*Salmo salar*). *Aquaculture* **280**, 89-93.
Hutchings, J. A. & Jones, M. E. B. (1998). Life history variation and growth rate thresholds for maturity in Atlantic salmon, *Salmo salar*. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 22-47.
ICES (2006). Extract of the Report of the Working Group on North Atlantic Salmon Stocks (WGNAS), May 2006, ICES Heardquaters..69pp.
ICES (2008). Report of the Working Group on the North Atlantic Salmon (WAGNAS). ICES.
Iles, T. D. & Sinclair, M. (1982). Atlantic herring: stock discreteness and abundance. *Science* **215**, 627-633.
Ingerslev, H. C., Pettersen, E. F., Jakobsen, R. A., Petersen, C. B. & Wergeland, H. I. (2006). Expression profiling and validation of reference gene candidates in

immune relevant tissues and cells from Atlantic salmon (*Salmo salar* L.). *Molecular Immunology* **43**, 1194-1201.
Jonsson, B. & Jonsson, N. (2004). Factors affecting marine production of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 2369-2383.
Jonsson, B., Jonsson, N. & Hansen, L. P. (2003). Atlantic salmon straying from the River Imsa. *Journal of Fish Biology* **62**, 641-657.
Jordan, W. C., Verspoor, E. & Youngson, A. F. (1997). The effect of natural selection on estimates of genetic divergence among populations of the Atlantic salmon. *Journal of Fish Biology* **51**, 546-560.
Kadri, S., Mitchell, D. F., Metcalfe, N. B., Huntingford, F. A. & Thorpe, J. E. (1996). Differential patterns of feeding and resource accumulation in maturing and immature Atlantic salmon, *Salmo salar*. *Aquaculture* **142**, 245-257.
Kazakov, R. V. & Titov, S. F. (1991). Geographical Patterns in the Population-Genetics of Atlantic Salmon, *Salmo salar* L, on USSR Territory, as Evidence for Colonization Routes. *Journal of Fish Biology* **39**, 1-6.
Kazakov, R. V. & Titov, S. F. (1993). Population genetics of salmon, *Salmo salar* L., in northern Russia. *Aquaculture and Fisheries Management* **24**, 495-506.
Kenchington, E., Heino, M. & Nielsen, E. E. (2003). Managing marine genetic diversity: time for action? *ICES Journal of Marine Science* **60**, 1172-1176.
King, T. L., Eackles, M. S. & Letcher, B. H. (2005). Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. *Molecular Ecology Notes* **5**, 130-132.
King, T. L., Kalinowski, S. T., Schill, W. B., Spidle, A. P. & Lubinski, B. A. (2001). Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology* **10**, 807-821.
King, T. L., Spidle, A. P., Eackles, M. S., Lubinski, B. A. & Schill, W. B. (2000). Mitochondrial DNA diversity in North American and European Atlantic salmon with emphasis on the Downeast rivers of Maine. *Journal of Fish Biology* **57**, 614-630.
Klemetsen, A., Amundsen, P. A., Dempson, J. B., Jonsson, B., Jonsson, N., O'Connell, M. F. & Mortensen, E. (2003). Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. *Ecology of Freshwater Fish* **12**, 1-59.
Koljonen, M. L., Jansson, H., Paaver, T., Vasin, O. & Koskiniemi, J. (1999). Phylogeographic lineages and differentiation pattern of Atlantic salmon (*Salmo salar*) in the Baltic Sea with management implications. *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 1766-1780.
Koljonen, M. L., Tahtinen, J., Saisa, M. & Koskiniemi, J. (2002). Maintenance of genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding programmes and the geographic distribution of microsatellite variation. *Aquaculture* **212**, 69-92.
Koskinen, M. T., Knizhin, I., Primmer, C. R., Schlotterer, C. & Weiss, S. (2002). Mitochondrial and nuclear DNA phylogeography of *Thymallus* spp. (grayling) provides evidence of ice-age mediated environmental perturbations in the world's oldest body of fresh water, Lake Baikal. *Molecular Ecology* **11**, 2599-2611.
Liu, K. J. & Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128-2129.
Liu, Z. J. & Cordes, J. F. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1-37.

Luikart, G., Allendorf, F. W., Cornuet, J. M. & Sherwin, W. B. (1998). Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity* **89**, 238-247.
Luikart, G. & Cornuet, J. M. (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12**, 228-237.
MacCrimmon, H. R. & Gots, B. L. (1979). World distribution of Atlantic salmon, *Salmo salar*. *Journal of the Fisheries Research Board of Canada* **36**, 422-457.
MAFF (2001). Review of Salmon and Freshwater Fisheries (Government Response). p. 39: Ministry of Agriculture, Fisheries and Food.
Makhrov, A. A., Verspoor, E., Artamonova, V. S. & O'Sullivan, M. (2005). Atlantic salmon colonization of the Russian Arctic coast: pioneers from North America. *Journal of Fish Biology* **67**, 68-79.
Martinez, J. L., Dumas, J., Beall, E. & Garcia-Vazquez, E. (2001a). Assessing introgression of foreign strains in wild Atlantic salmon populations: variation in microsatellites assessed in historic scale collections. *Freshwater Biology* **46**, 835-844.
Martinez, J. L., Gephard, S., Juanes, F., Perez, J. & Garcia-Vazquez, E. (2001b). Genetic and life history differentiation between donor and derivative populations of Atlantic salmon. *Transactions of the American Fisheries Society* **130**, 508-515.
McConnell, S. K., Oreilly, P., Hamilton, L., Wright, J. N. & Bentzen, P. (1995). Polymorphic Microsatellite Loci from Atlantic Salmon (*Salmo salar*) - Genetic Differentiation of North-American and European Populations. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 1863-1872.
McCormick, S. D., Hansen, L. P., Quinn, T. P. & Saunders, R. L. (1998). Movement, migration and smolting of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 77-92.
McElroy, D., Moran, P., Bermingham, E. & Kornfield, I. (1992). REAP - An integrated environment for the manipulation and phylogenetic analysis of restriction data. *Journal of Heredity* **83**, 157-158.
McGinnity, P., Prodohl, P., Ferguson, K., Hynes, R., O'Maoileidigh, N., Baker, N., Cotter, D., O'Hea, B., Cooke, D., Rogan, G., Taggart, J. & 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 Series B-Biological Sciences* **270**, 2443-2450.
McGinnity, P., Stone, C., Taggart, J. B., Cooke, D., Cotter, D., Hynes, R., McCamley, C., Cross, T. & Ferguson, A. (1997). Genetic impact of escaped farmed Atlantic salmon (*Salmo salar* L.) on native populations: use of DNA profiling to assess freshwater performance of wild, farmed, and hybrid progeny in a natural river environment. *ICES Journal of Marine Science* **54**, 998-1008.
McQuinn, I. H. (1997). Metapopulations and the Atlantic herring. *Reviews in Fish Biology and Fisheries* **7**, 297-329.
Metcalfe, N. A. (1998). The interaction between behavior and physiology in determining life history patterns in Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 93-103.
Metcalfe, N. B. & Thorpe, J. E. (1990). Determinants of Geographical Variation in the Age of Seaward Migrating Salmon, *Salmo salar*. *Journal of Animal Ecology* **59**, 135-145.
Mills, D. (1971). *Salmon and trout: a resource, its ecology, conservation and management.* Edinburgh: Oliver and Boyd.

Moore, A. & Waring, C. P. (1996). Electrophysiological and endocrinological evidence that F-series prostaglandins function as priming pheromones in mature male Atlantic salmon (*Salmo salar*) Parr. *Journal of Experimental Biology* **199**, 2307-2316.
Moore, A. & Waring, C. P. (2001). The effects of a synthetic pyrethroid pesticide on some aspects of reproduction in Atlantic salmon (*Salmo salar* L.). *Aquatic Toxicology* **52**, 1-12.
Morin, P. P., Dodson, J. J. & Dore, F. Y. (1989). Cardiac Responses to a Natural Odorant as Evidence of a Sensitive Period for Olfactory Imprinting in Young Atlantic Salmon, *Salmo salar*. *Canadian Journal of Fisheries and Aquatic Sciences* **46**, 122-130.
Morin, P. P. & Doving, K. B. (1992). Changes in the Olfactory Function of Atlantic Salmon, *Salmo salar*, in the Course of Smoltification. *Canadian Journal of Fisheries and Aquatic Sciences* **49**, 1704-1713.
Murton, J. B. & Lautridou, J. P. (2003). Recent advances in the understanding of Quaternary periglacial features of the English Channel coastlands. *Journal of Quaternary Science* **18**, 301-307.
Nei, M. (1983). Genetic Diversity and the Neutral Mutation Theory. *Heredity* **51**, 531-531.
Nei, M. (1987). *Molecular Evolutionary Genetics*. New York: Columbia University Press.
Nielsen, E. E., Hansen, M. M. & Loeschcke, V. (1996). Genetic structure of European populations of *Salmo salar* L (Atlantic salmon) inferred from mitochondrial DNA. *Heredity* **77**, 351-358.
Nielsen, E. E., Hansen, M. M. & Loeschcke, V. (1997). Analysis of microsatellite DNA from old scale samples of Atlantic salmon, *Salmo salar*: A comparison of genetic composition over 60 years. *Molecular Ecology* **6**, 487-492.
Nielsen, E. E., Hansen, M. M. & Loeschcke, V. (1999). Genetic variation in time and space: Microsatellite analysis of extinct and extant populations of Atlantic salmon. *Evolution* **53**, 261-268.
Nielsen, J. L. (2007). Atlantic salmon genetics: past, present and what's in the future? In *The Atlantic Salmon: Genetics, Conservation and Management* (Verspoor, E., Stradmeyer, S. & Nielsen, J. L., eds.), pp. 196-221. Oxford: Blackwell Publishing Ltd.
Nilsson, J., Gross, R., Asplund, T., Dove, O., Jansson, H., Kelloniemi, J., Kohlmann, K., Loytynoja, A., Nielsen, E. E., Paaver, T., Primmer, C. R., Titov, S., Vasemagi, A., Veselov, A., Ost, T. & Lumme, J. (2001). Matrilinear phylogeography of Atlantic salmon (*Salmo salar* L.) in Europe and postglacial colonization of the Baltic Sea area. *Molecular Ecology* **10**, 89-102.
Nosil, P. (2005). Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**, 705-719.
O'Brien, S. J. (1994). A role for molecular genetics in biological conservation. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5748-5755.
O'Reilly, P., Hamilton, L., McConnell, S. K. & Wright, J. N. (1996). Rapid analysis of genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 2292-2298.

Oakley, T. H. & Phillips, R. B. (1999). Phylogeny of salmonine fishes based on growth hormone introns: Atlantic (*Salmo*) and Pacific (*Oncorhynchus*) salmon are not sister taxa. *Molecular Phylogenetics and Evolution* **11**, 381-393.
Ostergaard, S., Hansen, M. M., Loeschcke, V. & Nielsen, E. E. (2003). Long-term temporal changes of genetic composition in brown trout (*Salmo trutta* L.) populations inhabiting an unstable environment. *Molecular Ecology* **12**, 3123-3135.
Parrish, D. L., Behnke, R. J., Gephard, S. R., McCormick, S. D. & Reeves, G. H. (1998). Why aren't there more Atlantic salmon (*Salmo salar*)? *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 281-287.
Paterson, S., Piertney, S. B., Knox, D., Gilbey, J. & Verspoor, E. (2004). Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. *Molecular Ecology Notes* **4**, 160-162.
Payne, R. H., Child, A. R. & Forrest, A. (1971). Geographical variation in the Atlantic salmon. *Nature* **231**, 250-252.
Peakall, R., Ruibal, M. & Lindenmayer, D. B. (2003). Spatial autocorrelation analysis offers new insights into gene flow in the Australian bush rat, *Rattus fuscipes*. *Evolution* **57**, 1182-1195.
Peakall, R. & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.
Pendas, A. M., Moran, P., Martinez, J. L. & Garcia-Vazquez, E. (1995). Applications of 5s-rDNAin Atlantic Salmon, Brown Trout, and in Atlantic Salmon X Brown Trout Hybrid Identification. *Molecular Ecology* **4**, 275-276.
Peress, J. (1999). Stock assessment review on the River Dart. (Environment Agency).
Peress, J. (2002). Amendments to River Dart conservation measures to protect salmon stock. (Environment Agency).
Phillips, R. B. & Hartley, S. E. (1988). Flouescent banding patterns in the chromosomes of the genus *Salmo*. *Genome Research* **30**, 193-197.
Piry, S., Alapetite, A., Cornuet, J. M., Paetkau, D., Baudouin, L. & Estoup, A. (2004). GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity* **95**, 536-539.
Piry, S., Luikart, G. & Cornuet, J. M. (1999). BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90**, 502-503.
Presa, P., Pardo, B. G., Martinez, P. & Bernatchez, L. (2002). Phylogeographic congruence between mtDNA and rDNA ITS markers in brown trout. *Molecular Biology and Evolution* **19**, 2161-2175.
Primmer, C. R., Veselov, A. J., Zubchenko, A., Poututkin, A., Bakhmet, I. & Koskinen, M. T. (2006). Isolation by distance within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River in northwest Russia. *Molecular Ecology* **15**, 653-666.
Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.
Rand, D. M. (2001). The units of selection on mitochondrial DNA. *Annual Review of Ecology and Systematics* **32**, 415-448.
Rannala, B. & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 9197-9201.

Raymond, M. & Rousset, F. (1995). Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86**, 248-249.
Reddin, D. G. & Friedland, K. D. (1999). A history of identification to continent of origin of Atlantic salmon (*Salmo salar* L.) at west Greenland, 1969-1997. *Fisheries Research* **43**, 221-235.
Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* **43**, 223-225.
Rieman, B. E. & Dunham, J. B. (2000). Metapopulations and salmonids: a synthesis of life history patterns and empirical observations. *Ecology of Freshwater Fish* **9**, 51-64.
Saisa, M., Koljonen, M. L., Gross, R., Nilsson, J., Tahtinen, J., Koskiniemi, J. & Vasemagi, A. (2005). Population genetic structure and postglacial colonization of Atlantic salmon (*Salmo salar*) in the Baltic Sea area based on microsatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 1887-1904.
Saisa, M., Koljonen, M. L. & Tahtinen, J. (2003). Genetic changes in Atlantic salmon stocks since historical times and the effective population size of a long-term captive breeding programme. *Conservation Genetics* **4**, 613-627.
Saltveit, S. J. (2006). The effects of stocking Atlantic salmon, *Salmo salar*, in a Norwegian regulated river. *Fisheries Management and Ecology* **13**, 197-205.
Saunders, R. L. & Gee, J. H. (1964). Movements of young Atlantic salmon in a small stream. *Journal of the Fisheries Research Board of Canada* **21**, 27-36.
Saura, M., Caballero, P. & Moran, P. (2008). Are there Atlantic salmon in the river Tambre? *Journal of Fish Biology* **72**, 1223-1229.
Schlotterer, C. (2004). The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics* **5**, 63-69.
Schneider, S., Roessli, D. & Excoffier, L. (2000). Arlequin ver. 2.000: A software for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.
Scruton, D. A., Pennell, C. J., Bourgeois, C. E., Goosney, R. F., King, L., Booth, R. K., Eddy, W., Porter, T. R., Ollerhead, L. M. N. & Clarke, K. D. (2006). Hydroelectricity and fish: a synopsis of comprehensive studies of upstream and downstream passage of anadromous wild Atlantic salmon, *Salmo salar*, on the Exploits River, Canada. In *Symposium on Hydropower, Flood Control and Water Abstraction*, pp. 225-239. Mondsee, AUSTRIA.
Slettan, A., Olsaker, I. & Lie, O. (1996). Atlantic salmon, *Salmo salar*, microsatellites at the SSOSL25, SSOSL85, SSOSL311, SSOSL412 loci. *Animal Genetics* **26**, 281-282.
Smith, I. P., Johnstone, A. D. F. & Smith, G. W. (1997). Upstream migration of adult Atlantic salmon past a fish counter weir in the Aberdeenshire Dee, Scotland. *Journal of Fish Biology* **51**, 266-274.
Soulsby, C., Youngson, A. F., Moir, H. J. & Malcolm, I. A. (2001). Fine sediment influence on salmonid spawning habitat in a lowland agricultural stream: a preliminary assessment. *Science of the Total Environment* **265**, 295-307.
Stabell, O. B. (1984). Homing and Olfaction in Salmonids - a Critical-Review with Special Reference to the Atlantic Salmon. *Biological Reviews of the Cambridge Philosophical Society* **59**, 333-388.
Ståhl, G. (1987). Genetic Population Structure of Atlantic Salmon. In *Population genetics and fisheries management.* (Ryman, N. & Utter, F. M., eds.), pp. 121-140. Seattle: University of Washington Press.
Stergiou, K. I. (2002). Overfishing, tropicalisation of fish stocks, uncertainty and ecosystem management: resharpening Ockham's razor. *Fisheries Research* **55**, 1-9.

Stewart, J. R. & Lister, A. M. (2001). Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* **16**, 608-613.
Taberlet, P., Fumagalli, L., Wust-Saucy, A. G. & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**, 453-464.
Taggart, J. B., Hynes, R. A., Prodohl, P. A. & Ferguson, A. (1992). A Simplified Protocol for Routine Total DNA Isolation from Salmonid Fishes. *Journal of Fish Biology* **40**, 963-965.
Taylor, E. B. (1991). A Review of Local Adaptation in Salmonidae, with particular reference to Pacific and Atlantic Salmon. *Aquaculture* **98**, 185-207.
Tonteri, A., Titov, S., Veselov, A., Zubchenko, A., Koskinen, M. T., Lesbarreres, D., Kaluzhin, S., Bakhmet, I., Lumme, J. & Primmer, C. R. (2005). Phylogeography of anadromous and non-anadromous Atlantic salmon (*Salmo salar*) from northern Europe. *Annales Zoologici Fennici* **42**, 1-22.
Tweed Foundation (1995). Evidences for stock differences within the salmon of the Tweed catchment. Tweed Foundation.
Tweed Foundation (2005). The Fisheries of the Tweed Report. The Tweed Foundation.
Vaha, J. P., Erkinaro, J., Niemela, E. & Primmer, C. R. (2007). Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology* **16**, 2638-2654.
Vaha, J. P., Erkinaro, J., Niemela, E. & Primmer, C. R. (2008). Temporally stable genetic structure and low migration in an Atlantic salmon population complex: implications for conservation and management. *Evolutionary Applications* **1**, 137-154.
Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. (2004). Micro-checker: software for identifying and correcting genotyping errors in microsatelite data. *Molecular Ecology Notes* **4**, 535-538.
Vasemagi, A., Gross, R., Paaver, T., Koljonen, M. L. & Nilsson, J. (2005a). Extensive immigration from compensatory hatchery releases into wild Atlantic salmon population in the Baltic sea: spatio-temporal analysis over 18 years. *Heredity* **95**, 76-83.
Vasemagi, A., Nilsson, J. & Primmer, C. R. (2005b). Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution* **22**, 1067-1076.
Verspoor, E. (1986a). Identification of stocks in the Atlantic salmon. In *Proceedings of the symposium on future Atlantic salmon management.* (Stroud, R. H., ed.), pp. 37-46. Savannah, Georgia.: Marine Recreational Fisheries Series, National Coalition for Marine Conservation.
Verspoor, E. (1986b). Spatial correlations of transferrin allele frequencies in Atlantic salmon (*Salmo salar*) populations from North America. *Canadian Journal of Fisheries and Aquatic Sciences* **43**, 1074-1078.
Verspoor, E. (1988). Reduced genetic variability in 1st generation hatchery populations of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* **45**, 1686-1690.
Verspoor, E. (1997). Genetic diversity among Atlantic salmon (*Salmo salar* L.) populations. In *ICES/NASCO Symposium on Interactions Between Salmon Culture and Wild Stocks of Atlantic Salmon - THe Scientific and Management Issues*, pp. 965-973. Bath, England.

Verspoor, E., Beardmore, J. A., Consuegra, S., De Leaniz, C. G., Hindar, K., Jordan, W. C., Koljonen, M. L., Mahkrov, A. A., Paaver, T., Sanchez, J. A., Skaala, O., Titov, S. & Cross, T. F. (2005). Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. *Journal of Fish Biology* **67**, 3-54.
Verspoor, E., Corse, I., McCarthy, E. M. & Youngson, A. F. (1997). Atlantic salmon population structure in the River Tweed: allozyme, mitochondrial DNA and minisatellite locus differentiation among fish from three areas of the catchment. Tweed Foundation.
Verspoor, E. & Jordan, W. C. (1989). Genetic-Variation at the Me-2 Locus in the Atlantic Salmon within and between Rivers - Evidence for Its Selective Maintenance. *Journal of Fish Biology* **35**, 205-213.
Verspoor, E., McCarthy, E. M. & Knox, D. (1999). The phylogeography of European Atlantic salmon (*Salmo salar* L.) based on RFLP analysis of the ND1/16sRNA region of the mtDNA. *Biological Journal of the Linnean Society* **68**, 129-146.
Wang, J. L. & Ryman, N. (2001). Genetic effects of multiple generations of supportive breeding. *Conservation Biology* **15**, 1619-1631.
Wang, S. Z., Hard, J. & Utter, F. (2001). Salmonid inbreeding: a review. *Reviews in Fish Biology and Fisheries* **11**, 301-319.
Wang, S. Z., Hard, J. J. & Utter, F. (2002). Genetic variation and fitness in salmonids. *Conservation Genetics* **3**, 321-333.
Wennevik, V., Skaala, O., Titov, S. F., Studyonov, I. & Naevdal, G. (2004). Microsatellite variation in populations of Atlantic salmon from North Europe. *Environmental Biology of Fishes* **69**, 143-152.
Wingfield, R. T. R. (1995). A model of sea-levels in the Irish and Celtic seas during the end-Pleistocene to Holocene transition. In *Island Britain: a Quaternary perspective* (Preece, R. C., ed.), pp. 209-242: Geological Society Special Publications.
Youngson, A. F., Jordan, W. C., Verspoor, E., McGinnity, P., Cross, T. & Ferguson, A. (2003). Management of salmonid fisheries in the British Isles: towards a practical approach based on population genetics. *Fisheries Research* **62**, 193-209.
Youngson, A. F., Martin, S. A. M., Jordan, W. C. & Verspoor, E. (1991). Genetic Protein Variation in Atlantic Salmon in Scotland - Comparison of Wild and Farmed Fish. *Aquaculture* **98**, 231-242.
Youngson, A. F. & Verspoor, E. (1997). Interactions between wild and introduced Atlantic salmon (*Salmo salar*). In *International Workshop on Integrating Across Scales - Predicting Patterns of Change in Atlantic Salmon*, pp. 153-160. Braemar, Scotland.
Youngson, A. F. & Verspoor, E. (1998). Interactions between wild and introduced Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 153-160.