REVIEW PAPER

Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation

E. Verspoor*, J. A. Beardmore‡, S. Consuegra§, C. García de Leániz‡, K. Hindar¶, W. C. Jordan§, M.-L. Koljonen**, A. A. Mahkrov††, T. Paaaver‡‡, J. A. Sánchez§§, Ø. Skaala¶¶, S. Titov*** and T. F. Cross†††

*FRS Freshwater Laboratory, Pitlochry, Perthshire, Scotland PH16 5LB; ‡University of Wales, Swansea, School of Biological Sciences, Singleton Park, Swansea SA2 8PP; §Institute of Zoology, Zoological Society of London, London, UK; ¶Norwegian Institute for Nature Research (NINA), Tungasletta 2, N-7485 Trondheim Norway; **Finnish Game and Fisheries Research Institute, P. O. Box 2, FIN-00791 Helsinki, Finland; ††Vavilov Institute of General Genetics, Russian Academy of Science, Gubkin St. 3, GSP-1, Moscow, 119991, Russia; ‡‡Aquaculture Research Unit, Institute of Veterinary Medicine and Animal Sciences, Estonian Agricultural University, 48 Kreutzwaldi Street, 51014, Tallinn, Estonia; §§Universidad de Oviedo, Departamento de Biología Funcional, Area de Genética, Laboratorio de Genética Acuícola, 33071, Oviedo, Spain; ¶¶Institute of Marine Research, Bergen, Norway; ***GosNIOHKH, Makarova Embankment 26, 199053, St. Petersburg, Russia and †††Department of Zoology & Animal Ecology, National University of Ireland, Cork, Republic of Ireland

Electrophoretic studies of proteins remain a primary source of insight into genetic diversity in many species including the Atlantic salmon *Salmo salar*, one of the most culturally and economically important fish species of the North Atlantic region. Since 1966, >350 scientific papers on protein variation have been published encompassing 25 000+ salmon from over 400 locations in >200 river systems across the species’ distribution. Variation has been detected at >40% of the 110 protein loci screened, though most studies examine <40. The method has been applied largely to the investigation of population structure and differentiation, but work has also led to the systematic revision of the genus *Salmo* and remains the primary source of insight into hybridization in the wild with brown trout *Salmo trutta*. Spatial patterns of differentiation show temporal stability, both within and among river systems, and strongly support structuring of the species into river and tributary specific populations and the designation of European and North American populations as distinct sub-species. They also show widespread regional differentiation within both continents, beyond the marked subcontinental differences between Baltic Sea and Atlantic Ocean populations in Europe. Most of the differentiation probably reflects gene flow and founder events associated with colonization following the retreat of the glaciers from much of the species’ modern range. However, variation at MEP-2* shows strong correlations with environmental temperature, both within and among rivers, and associations with phenotypic performance. This suggests selection is acting on the locus and provides

†Author to whom correspondence should be addressed. Tel.: +44 (0) 796 472060; fax: +44 (0) 1796 473523; email: verspoor@marlab.ac.uk
compelling evidence for the local adaptation of populations. Protein studies have led to more population centred management of the species and have been exploited in the discrimination of regional stocks in mixed stock analysis in high seas fisheries, particularly in the Baltic Sea, and as markers for the assessment of stocking success. They have also advanced insight into how the genetic character of populations can be changed in cultivation and the potential impact of salmon aquaculture and stocking on wild populations. The method has been largely superseded by DNA based analyses, but the results remain highly relevant to Atlantic salmon management and conservation and are an irreplaceable data set for studying genetic stability of populations over time.

Key words: allozymes; management; population differentiation; Salmo salar.

INTRODUCTION

Proteins are among the most diverse of biological molecules and are central to the expression of hereditary information. Tens of thousands of genes code for unique polypeptides which, either singly or in combination, generate a myriad of proteins, a variety further embellished by allelic diversity. Until the last decade, studies of allelic diversity revealed by protein electrophoresis provided the main source of insight into the genetic character of species in the wild. This work began with the seminal papers by Lewontin & Hubby (1966) and Harris (1966) on protein variation in fruit flies (Drosophila spp.) and humans, respectively, and provoked the great neutralist-selectionist debate of the 1970s regarding how such variation was maintained. This debate continues today (Wagner, 2000; Van Oosterhout et al., 2004).

Historically, protein work has focused on allozymes (i.e. allelic variants of enzymes) which can be specifically resolved using histochemical staining methods (Utter et al., 1987; Buth, 1990). This research, now spanning over 40 years, has confirmed intraspecific genetic polymorphism to be widespread and has dramatically advanced understanding of genetic and evolutionary processes. At a practical level, for many threatened and exploited species, it has clarified issues of taxonomy, hybridization, population structure and population adaptation, which has helped to improve species management and conservation (Ryman & Utter, 1987). The latter motivation underlies much of the work carried out on the Atlantic salmon Salmo salar L., a North Atlantic salmonid and one of the most studied of all fish species (Saunders, 1988). This work has focused on the more culturally and economically important anadromous stocks but has also included freshwater resident forms which occur in many parts of the species range (MacCrimmon & Gots, 1979).

More than 350 scientific studies on electrophoretic protein variation in the Atlantic salmon have been published since 1966 (Wilson & Verspoor, 1996). This effort has been driven by the species’ cultural and economic importance in commercial and recreational fisheries and, more recently, in salmon farming, and by dramatic declines in abundance in rivers in many parts of the native range (Parish et al., 1998). Early studies, focused on blood proteins, sought genetic markers which could identify the river or region of origin of salmon (Wilkins, 1972a, b). A major objective was to assist management of the then new high seas fisheries developed in the 1960s off West Greenland which, based on tagging studies, exploited fish from across much of the species’ range. The
concern was that overexploitation of some regional stocks might compound existing declines in many rivers caused by habitat destruction from dams, pollution, and acid rain. Studies of natal homing (Stabell, 1984) suggested that Atlantic salmon were structured into river specific populations, though whether this view was valid was the subject of considerable debate. Genetic studies of the distribution of protein variation were seen as a potential source of insight into this aspect of species biology (Saunders, 1981).

The main insight from early work came from serum transferrin studies (Møller, 1970; Payne et al., 1971). Later, research shifted to enzymatic proteins in body tissues with the first paper describing such work published by Cross & Ward (1980). They reported on variation in a single population, in relation to 23 enzymes and two serum proteins, representing 59 putative loci of which nine showed allelic variation. Soon after, Ståhl (1981, 1987) published a study of 38 loci in c. 4000 fish from 30 samples. Since this early work >25 000+ salmon from over 200 river systems across the entire species’ range have been screened. The last detailed review of published work is that of Ståhl (1987) with more recent reviews providing only potted overviews (Wilkins, 1985; Cross, 1989; Davidson et al., 1989; Verspoor, 1997). Thus, regarding protein variation in the species, the full implications of this body of information remains to be assessed.

This paper presents a detailed review of studies of genetic protein variation in Atlantic salmon carried out over the last four decades. To put the work in context, it starts with an overview of the limitations of protein electrophoresis as a tool for understanding the population genetics of species in the wild. The review then highlights what has been learned about salmon systematics before addressing what has been learned from these studies about genetic differentiation in Atlantic salmon, both within and among rivers in the wild. This includes new analyses of data across studies to provide insight into overall regional differentiation. Then, the findings of studies on the differences between wild and cultivated salmon are reviewed and integrated. The paper ends by considering what has been learned from the work regarding the origins of protein differentiation, population structuring and population adaptation, and how it has been applied as a tool in management.

**LIMITATIONS OF PROTEIN ELECTROPHORESIS**

The genetic variation detected by protein electrophoresis arises from amino acid substitutions generated by base sequence variation (Utter et al., 1987). These substitutions alter the charge state or conformational character of the protein and change its mobility when placed in a gel matrix subject to an electrical field. However, not all DNA sequence variation leads to amino acid changes and not all amino acid changes are detectable by electrophoresis. Thus electrophoretic screening can detect only part of the amino acid sequence variation which might be present. Furthermore, amino acid variation represents only part of the variation in the DNA, and protein coding regions represent only a small part of the overall genome. At best, only a third of amino acid substitutions and a tenth of DNA base substitutions will be detected by electrophoresis.
(Lewontin, 1974). This means protein electrophoresis gives a limited and potentially biased picture of overall genome variation.

Whether a change in charge state or conformation is actually detected, depends on factors such as the gel type, gel concentration and electrophoresis buffer chemistry (e.g. pH, ion concentration) as well as electrical conditions. Most studies have screened for variation using a limited set of conditions and much potentially detectable variation is likely to remain unresolved. Additionally, chemical methods for selectively resolving variation once proteins are separated have been developed for only a few classes of enzymes. Even then if variation is detected, it may not always be reliably resolved to allow typing, particularly where protein expression varies with developmental stage or physiological state, or proteins undergo post-translational modification due to tissue ageing and storage conditions (Utter et al., 1987). Typing can also be difficult, even if phenotypes are resolved, when electrophoretic banding patterns are complex, for example, or when there is a simultaneous resolution of several variable loci coding for interacting multimeric proteins (Utter et al., 1987; Waples, 1988; Shaklee & Phelps, 1992).

Work in practice has often also been constrained by the need for fresh, or freshly frozen, tissue for analysis, particularly in relation to species in remote regions. In general, protein work requires killing fish to get the required tissues. This can pose problems when trying to study threatened populations, limiting sample sizes or the loci screened. Only a few loci can be screened non-destructively (Wilson et al., 1995). The need for fresh tissue has also precluded the analysis of archived material such as scales, or prehistoric bones, both possible using DNA analyses, (Nielsen et al., 1997; Consuegra et al., 2002; Knox et al., 2002).

Numbers of polymorphic loci identified using protein electrophoresis are for most species small (e.g. <20), and most polymorphisms tend to be diallelic with one common, high frequency allele. This limits genetic analyses to comparisons of allelic diversity and heterozygosity, and genotype and allele frequencies, among samples (i.e. to ‘bean bag genetics’; Crow, 2000). More powerful statistical methods (Beaumont, 2001) based on estimates of the relatedness of individuals require higher levels of allelic diversity, as displayed by microsatellite loci, or larger numbers of variable loci. Protein analyses thus focus on rejecting the null hypothesis that samples fit expectations of random genetic assortment (e.g. Hardy-Weinberg equilibrium) or derive from different populations. This requires careful consideration of sampling theory and hypothesis testing (Weir, 1996), particularly when drawing inferences regarding the species or genome as a whole, and the rejection of null hypotheses is based on the balance of probability. Despite this, electrophoretic analyses still have the potential to provide valuable insight into the relative amount and distribution of variation which, in the context of population genetic theory, in turn can give significant insight into the genetic character of a species (Lewontin, 1974; Nei, 1987).

**SYSTEMATICS**

Electrophoretic studies of protein variation have been able to resolve evolutionary relationships at the interspecific level as a high proportion of loci tend to be informative and diagnostic at the species level. Work on salmonid species (Utter et al., 1973; Ferguson & Fleming, 1983; Johnson, 1984; Osinov &
Lebedev, 2000) supports the Atlantic salmon being a species distinct from brown *S. trutta* Richardson, rainbow *S. gairdneri* Richardson and cutthroat *S. clarkii* Richardson trout, the other species originally placed in the genus *Salmo*. However, the studies showed Atlantic salmon forming a distinct cluster with brown trout (Ferguson & Fleming, 1983; Fig. 1), with rainbow and cutthroat trout clustering with the Pacific salmon species (genus *Oncorhynchus*). With other lines of evidence, this led to the reclassification of the latter two species, as *O. mykiss* (Walbaum) and *O. clarki* Richardson (Smith & Stearley, 1989).

Until 1947, *Salmo salar* was viewed as composed of two distinct sub-species *S. salar salar* and *S. s. sebago*, encompassing anadromous and freshwater (i.e. non-anadromous) forms, respectively (Scott & Crossman, 1973). This taxonomic division was rejected when a detailed study found no consistent morphological or meristic differences between populations of the two types (Wilder, 1947). The lack of consistent differentiation between the two forms is also seen at allozyme loci (Stålhl, 1987; Verspoor, 1994), supporting this view, though as discussed later, the two forms do form genetically distinct populations.

**HYBRIDIZATION WITH BROWN TROUT**

The validity of the Atlantic salmon as a genetically distinct species has been demonstrated by studies of its hybridization in the wild with *S. trutta*. Like many other congeneric salmonids (Chevassus, 1979), and other fish species (Verspoor & Hammar, 1991), the two species can produce viable hybrids. This raised questions as to whether salmon and brown trout hybridize naturally and, if so, whether there was gene introgression between them. These issues could not be addressed prior to protein based methods, as morphological and meristic characters are not diagnostic. Certainly, early protein work showed that congeneric rainbow and cutthroat trout hybridize extensively where the former are introduced into the range of the latter, something which can lead to the loss of the native cutthroat populations (Campton & Johnston, 1985). However, these questions could be addressed once protein variation able to identify F₁ hybrids was developed.

Initial protein studies were able to identify potentially diagnostic differences between these two species for esterases, transferrin and a few other proteins (Nyman, 1967a; Payne et al., 1972). Later, Vuorinen & Piironen (1984) reported five of 14 enzyme systems surveyed in European salmon to provide diagnostic patterns; in a parallel study in North America, Johnson (1984) found 13 of 33 enzymes to be diagnostic. However, ease of typing, and ability to use adipose fin clips led to a focus on glucosephosphate isomerase (GPI) (Fig. 2), which can also be used for screening eggs (Mork & Heggeberget, 1984), and to a lesser extent on phosphoglucomutase (PGM), which is less strongly expressed in fin tissue and eggs.

Protein surveys have shown hybrids occurring throughout much of the zone of overlap of Atlantic salmon and brown trout in Europe, as well as where trout have been introduced and colonized the salmon’s range in North America (Scott & Crossman, 1973). In Europe, their occurrence is sporadic and they make up <2% of salmonids in most samples analysed (Verspoor & Hammar, 1991; Elo et al., 1995; Makhrov et al., 1998a; Matthews et al., 2000). This sporadic distribution may reflect a low frequency of occurrence and small sample sizes (Verspoor,
Fig. 1. Genetic relatedness of species assigned to the genera *Oncorhynchus* and *Salmo* based on differentiation at 18 allozyme loci; adapted from Ferguson & Fleming (1983).
In only a few locations have hybrids been found to occur regularly at high frequencies. In the Leader tributary of the River Tweed, Scotland (Jordan & Verspoor, 1993) and the River Grönan, Sweden (Jansson et al., 1991), allozyme studies encompassing consecutive year classes showed 8·5 and 13·2% of F1 hybrids, respectively. The latter was attributed to the lack of temporal and spatial separation of spawning of the two species, along with a high incidence of early maturing salmon parr.

More focused studies have shown that the frequency of hybrids in Norway from the early 1980s to the early 1990s has increased for reasons that are only partly understood (Youngson et al., 1993; Hindar & Balstad, 1994). Factors such as the introduction of non-native fish (Verspoor, 1988a; García de Leániz & Verspoor, 1989), habitat disturbance (Jansson & Oest, 1997), lack of spawners of one species (Makhrov et al., 1998a; Paaver et al., 2001), and farm escapes (Hindar & Balstad, 1994) have been found to be associated with an increased level of hybridization. There is direct evidence from releases of genetically marked salmon into a natural stream, that farm salmon are more likely to breed with trout than are wild salmon (Hindar & Fleming, in press). Artificial hybrids have been found in some rivers draining into the Baltic Sea (Semenova & Slyn’ko, 1988; Makhrov et al., 2004).

Little insight has been gained from protein electrophoresis into introgressive hybridization. The independent segregation of variation at most loci in post F1 hybrid generations means large numbers of diagnostic loci are needed for their identification. Protein variants common to one or other species have in some cases been detected in the other species where there is also a high incidence of hybrids (Verspoor & Hammar, 1991). However, more detailed studies of allozyme variation in populations with high levels of hybrids (Jansson & Dannewitz, 1995), and studies of artificial hybridization using allozymes (Galbreath & Thorgaard, 1994), suggest that post F1 hybrids may not survive. If so, the shared
variants may reflect parallel mutations rather than introgression. Alternatively, such fish may be first or second generation hybrids, or backcrosses, with disrupted patterns of allozyme expression caused by the mixing of two divergent genomes (Johnson & Wright, 1986; Wilkins et al., 1993; Dannewitz & Jansson, 1996; Jansson & Dannewitz, 1995).

**INTRASPECIFIC POLYMORPHISM AND ALLELIC VARIATION**

Overall, studies have screened 110 putative loci for allelic variation in Atlantic salmon of which 32 (30%) show probable genetic variation. Wilson et al. (1995) identified 91 loci in a survey of 40 enzymes using six different buffer systems. Of these, 13 showed variation consistent with genetic models based on enzyme quaternary structure (Utter et al., 1987; Buth, 1990), with ten diallelic and three triallelic. A further 19 variable loci have been reported by others. Identification of allelic variation is based on the conformation of banding patterns to expectations for enzyme quaternary structure and of phenotypic proportions to genetic models (e.g. Hardy–Weinberg, H–W). For most loci this is supported by breeding studies (Møller, 1970; Johnson, 1984; Crozier & Moffett, 1990; Skaala & Jorstad, 1994; Verspoor & Jordan, 1994; Verspoor & Moyes, this volume). However, support from breeding studies is lacking for most rare variants and for recently identified polymorphisms such as *FBALD-3*<sup>*</sup>, *TPI-3*<sup>*</sup> and *ESTD-2*<sup>*</sup>. Only a small proportion of loci can be screened non-destructively (Moffett & Crozier, 1991; Wilson et al., 1995).

There appears to be no physical linkage of the polymorphic loci identified where studied (Johnson, 1984). Mean levels of polymorphism (P) and heterozygosity (H) among four populations screened by Wilson et al. (1995) ranged from 9.7–16.7 to 2.7–3.4%, respectively, with most polymorphic loci diallelic in most populations. However, up to four alleles have been reported at some loci in some cases (Jordan et al., 1992; Verspoor, this volume). Despite the large number of loci examined overall, for practical reasons most population studies have looked at only 20–40 loci; of which usually five or six are polymorphic. The loci most commonly found to be polymorphic are listed in Table I.

Reference to EC number (IUB, 1984) in most studies makes it clear where the same enzymes have been screened in different studies. However, the equivalence of loci and alleles is less clear, given that the naming and numbering of loci and alleles is based on the relative mobilities of bands which can vary depending on the specific electrophoretic conditions used. This does not present a problem for enzymes with a single variable locus with two allelic variants and, in later work, the equivalence of loci is clearer due to the widespread adoption of a standard system of nomenclature (Shaklee et al., 1990a). Particular confusion exists in the literature regarding IDDH (iditol dehydrogenase) and PGM (phosphoglucomutase). Problems at other loci are confined to the equivalence of rare variants found in different studies. IDDH, referred to as SDH (sorbitol dehydrogenase) prior to 1991, shows two variable loci and the common polymorphism has been variously assigned to *IDDH-1*<sup>*</sup> or *IDDH-2*<sup>*</sup>. There are also inconsistencies in the naming of allelic variants. For PGM, the assignment of *PGM-1*<sup>*</sup> and *PGM-2*<sup>*</sup> was initially reversed from what is now widely accepted, though this issue still lacks clarity across studies.
<table>
<thead>
<tr>
<th>Enzyme (abbreviation)</th>
<th>E.C. NO.</th>
<th>Quaternary structure</th>
<th>Locus</th>
<th>Tissue</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartate aminotransferase (AAT)</td>
<td>2.6.1.1</td>
<td>Dimer</td>
<td>sAAT-1,2*</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>sAAT-4*</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>Creatinekinase (CK)</td>
<td>2.7.3.2</td>
<td>Dimer</td>
<td>CK-1*</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>Esterase D (ESTD)</td>
<td>3.1.-.-</td>
<td>Dimer</td>
<td>ESTD*</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>Fructose-bisphosphate aldolase (FBALD)</td>
<td>4.1.2.13</td>
<td>Tetramer</td>
<td>FBALD-3*</td>
<td>E</td>
<td>FBALD-2* in Cordes et al. (this volume)</td>
</tr>
<tr>
<td>Glucose-6-phosphate isomerase (GPI)</td>
<td>5.3.1.9</td>
<td>Dimer</td>
<td>GPI-1,2*</td>
<td>M</td>
<td>Same as PGI (phosphogluucose isomerase)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GPI-3*</td>
<td>L, M</td>
<td></td>
</tr>
<tr>
<td>L-Iditol dehydrogenase (IDDH)</td>
<td>1.1.1.14</td>
<td>Tetramer</td>
<td>IDDH-1*</td>
<td>L</td>
<td>Previously referred to as SDH-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IDDH-2*</td>
<td>L</td>
<td>Previously referred to as SDH-1</td>
</tr>
<tr>
<td>Isocitrate dehydrogenase (IDHP)</td>
<td>1.1.1.42</td>
<td>Dimer</td>
<td>IDHP-2*</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IDHP-3*</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>L-Lactate dehydrogenase (LDH)</td>
<td>1.1.1.27</td>
<td>Tetramer</td>
<td>LDH-4*</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>Malate dehydrogenase (MDH)</td>
<td>1.1.1.37</td>
<td>Dimer</td>
<td>MDH-1*</td>
<td>L, E</td>
<td>sMDH-A1* in Cordes et al. (this volume)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MDH-3,4*</td>
<td>M</td>
<td>sMDH-B1* in Cordes et al. (this volume)</td>
</tr>
<tr>
<td>Malic enzyme (NADP+) (MEP)</td>
<td>1.1.1.40</td>
<td>Tetramer</td>
<td>mMEP-1*</td>
<td>M</td>
<td>Previously rebered to as ME</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>mMEP-2*</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>Malic enzyme (NAD+) (ME)</td>
<td>1.1.1.38</td>
<td>Tetramer</td>
<td>ME*</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>Phosphoglucomutase (PGM)</td>
<td>5.4.2.2</td>
<td>Monomer</td>
<td>PGM-1*</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PGM-2*</td>
<td>M, L</td>
<td>Referred to as PGM-1 in some early studies</td>
</tr>
<tr>
<td>Triose-phosphate isomerase (TPI)</td>
<td>5.3.1.1</td>
<td>Dimer</td>
<td>TPI-3*</td>
<td>L, E</td>
<td></td>
</tr>
<tr>
<td>Transferrin (TF)</td>
<td></td>
<td>Monomer</td>
<td>TF*</td>
<td>S</td>
<td></td>
</tr>
</tbody>
</table>

E, eye; L, liver; M, muscle; S, blood serum.
The use of IDDH, one of the most polymorphic enzymes, in population studies has proved generally problematic. Three easily typed variants [Fig. 3(a) – A, B, C] occur in most rivers and have been shown by breeding studies to arise from two alleles at one of the two loci (Johnson, 1984). However, the presence of additional banding variants [Fig. 3(a) – D] in some areas suggest further alleles occur (Cross & Ward, 1980) which have been attributed to different loci by different workers (Verspoor et al., 1991; Elo et al., 1994). Other as yet unnamed variants have also been found [Fig. 3(b) – ?]. Complex electrophoretic patterns are expected if allelic variation occurs at both loci due to the enzyme’s tetrameric structure and being coded for by two loci whose polypeptide products form interlocus heterotetramers. Problems in distinguishing allelic variants and assigning variation to loci have led to a patchy and geographically biased coverage, and often only the frequency of the common variant allele for the main polymorphism has been estimated. In the south of France and Spain, complex patterns are particularly common and data from these regions is correspondingly limited.

**VARIATION AMONG RIVER SYSTEMS**

Genotype proportions within most samples collected from a single locality have been consistent with the random allelic assortment (i.e. H–W expectations) and derivation from a single population, though it is widely recognized that such tests lack statistical power (Chakraborty & Zhong, 1994). As such, derived sample allele frequencies can be used for analysing variation among samples. In general these show heterogeneity among samples from different river systems in North America (Møller, 1970; Payne, 1974; Verspoor, 1988b, this volume; Cordes et al., this volume) as well as Europe (Ståhl, 1987; Kazakov & Titov, 1991, 1993; McElligott & Cross, 1991; Sanchez et al., 1991; Jordan et al., 1992, this volume; Morán et al., 1994a; O’Connell et al., 1995, 1996; Bourke et al., 1997; Danielsdottir et al., 1997; Koljonen et al., 1999).

Most allozyme diversity is found within samples. The relative amount depends on the geographical scale of the study. Ståhl (1987) found 59% to reside within samples across a set of samples spanning the species’ range with 3-6% residing within drainages. In a study of Scottish populations, Jordan et al. (1992) found 96.4% of variation within samples, 1-2% among year classes within tributaries, 0.7% among tributaries within rivers, and 1.6% among rivers, consistent with
the findings of other studies (Elo *et al*., 1994; Koljonen *et al*., 1999; Verspoor, this volume) if geographical scale is taken into account. This pattern is broadly similar to other salmonid species (Gyllensten, 1985) though the congeneric brown trout appears to show greater differentiation among localities (Ryman, 1983). This may reflect the general observation that resident freshwater fish species generally show more variation among localities than is seen in anadromous species (Gyllensten, 1985).

**BETWEEN CONTINENTS**

Differentiation among anadromous European and North American Atlantic salmon was the main focus of early studies of genetic protein variation, with one of the first studies by Nyman (1967b) reporting differences in the quantitative and qualitative expression of a number of serum proteins between a Canadian and Swedish population. However, most informative was the early work on serum transferrin, \( TF^* \). Payne *et al*. (1971) found that rivers in the Gulf of St Lawrence in eastern Canada had two variants absent in salmon from the British Isles and the latter area to have one variant not found in the former, though both shared the same common variant. This differentiation was confirmed by later work (Møller, 1970; Payne, 1974; Child *et al*., 1976). A decade later Cross (1981), comparing three European and two Canadian rivers showed much greater variation between than within continents at four of five enzyme loci. Ståhl (1987), comparing a large number of European samples with samples from the Miramichi river in eastern Canada found genetic distances between continents to be three times greater than between Baltic Sea and Atlantic Ocean salmon rivers in Europe; Ståhl found the latter divergence to be the largest observed among European rivers. In a more extensive survey of loci, Bourke *et al*. (1997) found genetic distances \( D_A \) between one typical Canadian river (the St John in New Brunswick) and rivers in Europe to be twice that observed between any European rivers. The St John, with one of the largest runs of salmon in the region, had a lower level of heterozygosity than Atlantic Ocean European rivers, though higher than found for Baltic Sea rivers. It also had the highest number of polymorphic loci and the highest mean number of alleles per locus, a difference noted by Ståhl (1987). In contrast, a broader survey of Canadian populations, based on 23 loci, (Verspoor, this volume) found mean heterozygosity in Canadian rivers to be similar to the European populations based on data for the same loci from the study by Bourke *et al*. (1997). This survey found among river variation to be more than twice as great in Europe \( F_{ST} = 0.076 \) in North America and 0.176 in Europe), as expected from the species’ much larger European range (MacCrimmon & Gots, 1979).

No fixed allelic differences have been found between Europe and North America but allele frequencies are highly divergent at many loci. At some loci, such as \( TF^* \), \( ME^* \) and \( MDH-3,4^* \), North American populations share common allelic variants which are rare or absent in Europe (Payne *et al*., 1971; Verspoor & McCarthy, 1997; Verspoor, this volume). For \( ESTD-2^* \), though the number of rivers screened is limited, European salmon appear to be largely fixed for the \( ^*100 \) variant and in North American salmon the \( ^*80 \) variant dominates. The only polymorphic European rivers occur on the Kola Peninsula and adjacent
northern Norway and the White Sea (Semenova & Slyn’ko, 1988; Kazakov & Titov, 1991, 1993; Skaala et al., 1998; Makhrov et al., 1998a, this volume), plus one population in northern Scandinavia (Bourke et al., 1997) where Canadian fish have been introduced.

European rivers from northern Russia and northern Norway are the most closely related to Canadian rivers (Bourke et al., 1997), something which probably reflects historical gene flow between the two regions. This possibility is suggested by the distribution of mtDNA variants (Verspoor et al., 1999; Makhrov et al., this volume) as well as ESTD-2* alleles. In North America, ESTD-2*100 appears to occur sporadically at low frequencies. However, it is most common in Newfoundland and Labrador where variants typical of Europe are also most common both for allozymes (Verspoor, this volume) and for mtDNA (Birt et al., 1991; King et al., 2000; Nilsson et al., 2001).

WITHIN NORTH AMERICA

Regional differentiation within continents was first identified by Møller (1970; this volume) in relation to TF* in North American populations. Later Payne (1974) reported variation among rivers at this locus to show a latitudinal cline. Combining data from the two studies this appeared to be an artefact of a stepwise north-south regional divergence (Verspoor, 1986). TF*-I, the common variant in European salmon was higher in Newfoundland and southern Labrador compared to regions to the south and west; no data was available for rivers from the Quebec North Shore, Ungava Bay and northern Labrador.

Regional differentiation in North America is also revealed by allozyme studies. Verspoor (this volume), in a study of 23 loci in 53 populations from across the species’ Canadian range, of which 15 loci were polymorphic, found interriver heterogeneity to be regionally distributed. Though some of the largest genetic differences occurred between nearby rivers, there was a weak though significant correlation of genetic and geographic distance. This was in part linked to a significant increase in MEP-2*125 frequency with latitude (Verspoor & Jordan, 1989). Furthermore, the rivers clustered into six, largely distinct regional groups, both a result of regional allele frequency divergence at shared polymorphisms and the occurrence of regionally restricted variants. Cordes et al. (this volume) studied variation at 54 loci, of which eight loci were variable, in a more geographically restricted study of seven Maine rivers plus the nearby St John River in New Brunswick. While genetic differences were detected among all rivers, no association between genetic and geographic distance, or regional differences, were found.

An overview of the pattern of North American variation can be gained by combining data from Bourke et al. (1997), Cordes et al. (this volume) and Verspoor (this volume) for a geographically representative set of four Maine and 44 Canadian rivers (Fig. 4); this includes four replicate samples from the St John river stock to show the impact of sampling variation on the patterns revealed. All the protein polymorphisms found in this part of the species range are included except ME*, (which was not screened by Cordes et al. and only typed in some populations surveyed by Verspoor, this volume), ESTD-2* and TF*, which were not included in either study. The NJ (nearest-neighbour
joining) tree based on genetic distance \( (D_A; \text{Nei, 1978}) \) for the combined data set (Fig. 5) shows that the Maine samples cluster with those from Outer Bay of Fundy (oBoF) and Gulf of St Lawrence (GoSL). In the MDS (multi-dimensional

Fig. 4. Rivers from which samples used in the regional analysis of variation in North America were derived: A, Verspoor (this volume); B, Bourke et al. (1997); C, Cordes et al. (this volume).
scaling) plot of $D_A$ [Fig. 6(a)], the Maine/oBoF samples are revealed to be geographically disjunct and separated by the distinctive Atlantic shore/Southern Uplands of Nova Scotia rivers.

No regional pattern is seen if $F_{ST}$ is used as the measure of genetic distance. As $D_A$ is more sensitive to allele presence or absence, this probably underlies much of the observed regional differentiation (Verspoor, this volume). However,
Fig. 6. (a) Multi dimensional scaling (MDS) plot for North American river samples based on genetic distance, $D_A$ and (b) plot of PCA scores for components 3 and 4 for the same river samples based on arcsine transformed (Sokal & Rohlf, 1981) allele frequencies constructed with Systat Version 9.0 (SPSS, 1999). Lines enclose regional geographic groups. Numbers in both plots refer to Fig. 4 (○, Labrador; ●, Newfoundland; ●, Gulf of St Lawrence; ●, Nova Scotia; ●, Inner Bay of Fundy; ●, Outer Bay of Fundy; ●, Maine).

separation of the Maine samples from the oBoF is seen in the principal compo-
nent analysis (PCA) of allele frequencies [Fig. 6(b)] based on components 3 and
5, along with the same separation of the Canadian rivers into more or less the
same regional groups as seen in the MDS analyses. The main loading for

\[ \text{AAT-1}^{*}70, \text{IDDH-2}^{*}-72, \text{MEP-2}^{*}125 \]

and for component 5 is on \[ \text{AAT-4}^{*}50 \text{ IDDH-2}^{*}72, \text{MDH-3,4}^{*}133, \text{MDH-3,4}^{*}120 \]
and \[ \text{PGM-1r}^{*}a \] (Table II). The River Hammond, a lower tributary of the St John system,
clusters separately from the other oBoF samples in the MDS analysis, but with
the oBoF samples in the NJ and PCA analysis. This reflects the different way in
which the statistical methods deal with the anomalous high frequency private
\[ \text{GPI-3}^{*} \] allele found in this sample (Verspoor, this volume), something which
needs to be taken into account when interpreting the results.

**WITHIN EUROPE**

An up-to-date overview of geographical patterns of protein variation among
salmon in European rivers is lacking. Most studies of European salmon have
been regionally restricted (e.g. Ireland: McElligott & Cross, 1991; Wales:
O’Connell *et al.*, 1995, 1996; Scotland/British Isles: Jordan *et al.*, 1992, this
volume; Iceland: Danielsdottir *et al.*, 1997; Russia: Kazakov & Titov, 1991,

---

**Table II. Coefficients for standardised factor scores from the principle component ana-
lysis (PCA) of North American samples**

<table>
<thead>
<tr>
<th>Allele</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
<th>Factor 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT-1*70</td>
<td>0.021</td>
<td>-0.005</td>
<td>-0.335</td>
<td>0.090</td>
<td>-0.036</td>
</tr>
<tr>
<td>AAT-4*25</td>
<td>0.028</td>
<td>0.440</td>
<td>-0.074</td>
<td>-0.013</td>
<td>-0.044</td>
</tr>
<tr>
<td>AAT-4*50</td>
<td>0.203</td>
<td>0.009</td>
<td>-0.077</td>
<td>0.045</td>
<td>0.251</td>
</tr>
<tr>
<td>CK-1*65</td>
<td>-0.393</td>
<td>0.006</td>
<td>0.006</td>
<td>0.034</td>
<td>-0.026</td>
</tr>
<tr>
<td>CK-1*87</td>
<td>0.045</td>
<td>0.008</td>
<td>0.005</td>
<td>0.025</td>
<td>-0.007</td>
</tr>
<tr>
<td>GPI-1,2*140</td>
<td>-0.048</td>
<td>-0.018</td>
<td>0.146</td>
<td>0.050</td>
<td>-0.074</td>
</tr>
<tr>
<td>GPI-1,2*180</td>
<td>-0.008</td>
<td>0.111</td>
<td>0.155</td>
<td>0.066</td>
<td>0.188</td>
</tr>
<tr>
<td>GPI-3*90</td>
<td>0.022</td>
<td>-0.034</td>
<td>0.057</td>
<td>0.062</td>
<td>-0.107</td>
</tr>
<tr>
<td>GPI-3*110</td>
<td>-0.393</td>
<td>0.006</td>
<td>0.006</td>
<td>0.034</td>
<td>-0.026</td>
</tr>
<tr>
<td>IDDH-1*28</td>
<td>-0.004</td>
<td>0.067</td>
<td>0.121</td>
<td>-0.275</td>
<td>-0.139</td>
</tr>
<tr>
<td>IDDH-2*72</td>
<td>-0.147</td>
<td>0.042</td>
<td>0.217</td>
<td>0.087</td>
<td>0.272</td>
</tr>
<tr>
<td>IDHP-2*</td>
<td>0.057</td>
<td>-0.095</td>
<td>0.139</td>
<td>0.007</td>
<td>0.320</td>
</tr>
<tr>
<td>IDHP-3*116</td>
<td>-0.004</td>
<td>0.063</td>
<td>-0.351</td>
<td>0.046</td>
<td>-0.012</td>
</tr>
<tr>
<td>LDH-4*60</td>
<td>-0.008</td>
<td>-0.001</td>
<td>0.028</td>
<td>-0.310</td>
<td>-0.010</td>
</tr>
<tr>
<td>MDH-1*50</td>
<td>-0.042</td>
<td>-0.046</td>
<td>-0.042</td>
<td>-0.333</td>
<td>0.178</td>
</tr>
<tr>
<td>MDH-3,4*75</td>
<td>0.016</td>
<td>0.003</td>
<td>0.006</td>
<td>-0.368</td>
<td>0.041</td>
</tr>
<tr>
<td>MDH-3,4*80</td>
<td>0.026</td>
<td>0.452</td>
<td>-0.019</td>
<td>0.015</td>
<td>0.029</td>
</tr>
<tr>
<td>MDH-3,4*120</td>
<td>0.048</td>
<td>-0.066</td>
<td>0.084</td>
<td>0.046</td>
<td>-0.293</td>
</tr>
<tr>
<td>MEP-2*125</td>
<td>-0.153</td>
<td>-0.025</td>
<td>-0.291</td>
<td>-0.014</td>
<td>0.036</td>
</tr>
<tr>
<td>PGM-1r*a</td>
<td>-0.071</td>
<td>0.185</td>
<td>0.071</td>
<td>-0.093</td>
<td>-0.205</td>
</tr>
<tr>
<td>PGM-2*45</td>
<td>-0.001</td>
<td>0.016</td>
<td>0.020</td>
<td>0.029</td>
<td>-0.153</td>
</tr>
</tbody>
</table>
(1987), in an early overview, identified the major regional differentiation between salmon from Baltic Sea and Atlantic Ocean drainages, while Verspoor & Jordan (1989) reported a strong continent wide latitudinal cline at MEP-2* associated with temperature. Analyses by Blanco et al. (1992) and Elo (1993) showed there was a general correlation between genetic and geographic distance. These observations were later supported by independent work by Bourke et al. (1997). The distribution of AAT-4* in Europe appears to be confined to more northerly populations (Table III) and, in the British Isles, it increases in frequency from south to north (Jordan et al., this volume).

Within the Baltic Sea, Koljonen et al. (1999) observed a divergence of salmon in southeastern and northwestern rivers as well as a local correlation within the sea between genetic and geographic distance. Among 19 Atlantic Ocean rivers, Blanco et al. (1992), based on data for four polymorphic loci, but excluding MEP-2*, showed a clustering of Atlantic rivers into northern and southern groups. Variation at MEP-2* is greater than that seen for the other three main polymorphic loci, AAT-4*, IDDH-2*, and IDHP-3*, though within rivers and on smaller geographical scales it is generally lower (Jordan et al., 1997; but see Gilbey et al., 1999).

An overall picture of regional differentiation in Europe is lacking but can be gained by synthesizing data from published and unpublished studies (Table III and Fig. 7) with overlapping coverage of rivers, where the equivalence of allele and locus assignments can be established. The data are for 12 polymorphic loci from 56 rivers broadly representative of the species European range; for a few of these rivers, samples from consecutive years have been pooled to ensure adequate sample sizes and, for three other rivers, samples of nonanadromous salmon are included to assess differentiation associated with life history as opposed to geographical factors. For all rivers, data are available for six polymorphisms. MDH-3,4*75 and *80 have been treated as one allele, though the latter are distinct variants, as it is not certain that these were unambiguously typed in all studies. For IDDH-1* and for IDDH-2*, variant alleles have been treated as a single allele class given that for some rivers only the frequency of the common variant is provided. Unfortunately, this was the only option for treatment of the IDDH data and valuable information about geographic diversity in these highly variable loci was lost due to these interpretation difficulties.

The NJ analysis of genetic distances shows the basic division of the selected rivers into Baltic Sea and Atlantic Ocean groups (Fig. 8) as revealed by others (Ståhl, 1987), though two Atlantic rivers, one from Iceland and one from northern Russia, occur in the Baltic Sea cluster. Among Atlantic Ocean rivers, those of southern France and Spain form a distinct group most closely associated with the rivers of northern France and the southern British Isles. The rivers of Iceland form a largely distinct cluster with the Kapisidlit River in West Greenland which is most closely related to rivers of the northern British Isles, Atlantic Scandinavia, and northern Russia. Differentiation among rivers in the British Isles and adjacent Scandinavia is relatively small while within the Baltic Sea there remains the regional dichotomy noted by Koljonen et al. (1999). Between anadromous and non-anadromous populations from the same river catchments, differentiation is greater than seen among many regions. The nonanadromous samples from Lake Ladoga in the Neva basin and the River Namsen cluster with


<table>
<thead>
<tr>
<th>Map number</th>
<th>River</th>
<th>Lat. number</th>
<th>AAT-4</th>
<th>ESTD</th>
<th>FBALD-3</th>
<th>GPI-3</th>
<th>IDDH-1</th>
<th>IDDH-2</th>
<th>IDHP-3</th>
<th>MDH-1</th>
<th>MDH-3,4</th>
<th>MEP-2</th>
<th>PGM-2</th>
<th>TPI-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Greenland and Iceland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Kapisidlit</td>
<td>62-5</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
</tr>
<tr>
<td>2</td>
<td>Holmsa</td>
<td>62-5</td>
<td>0-243</td>
<td>0-021</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
</tr>
<tr>
<td>3</td>
<td>LaxaaDolum</td>
<td>62-5</td>
<td>0-330</td>
<td>0-102</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
</tr>
<tr>
<td>4</td>
<td>Laugardalsa</td>
<td>67-5</td>
<td>0-100</td>
<td>0-450</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-025</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
</tr>
<tr>
<td>5</td>
<td>LaxaaAsum</td>
<td>67-5</td>
<td>0-375</td>
<td>0-222</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-012</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
</tr>
<tr>
<td>6</td>
<td>Fjoska</td>
<td>67-5</td>
<td>0-263</td>
<td>0-150</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-012</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
</tr>
<tr>
<td>7</td>
<td>Svalbardsa</td>
<td>67-5</td>
<td>0-236</td>
<td>0-319</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-014</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
</tr>
<tr>
<td>8</td>
<td>Hofsø</td>
<td>67-5</td>
<td>0-107</td>
<td>0-311</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-033</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>Northern Russia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Mezen</td>
<td>62-5</td>
<td>0-070</td>
<td>0-170</td>
<td>0-000</td>
<td>ND</td>
<td>0-000</td>
<td>0-300</td>
<td>0-230</td>
<td>0-000</td>
<td>0-030</td>
<td>0-000</td>
<td>0-740</td>
<td>ND</td>
</tr>
<tr>
<td>10</td>
<td>Northern Dvina</td>
<td>57-5</td>
<td>0-010</td>
<td>0-050</td>
<td>0-000</td>
<td>ND</td>
<td>0-000</td>
<td>0-220</td>
<td>0-080</td>
<td>0-010</td>
<td>0-030</td>
<td>0-000</td>
<td>0-660</td>
<td>ND</td>
</tr>
<tr>
<td>11</td>
<td>Onega</td>
<td>57-5</td>
<td>0-010</td>
<td>0-029</td>
<td>0-000</td>
<td>ND</td>
<td>0-000</td>
<td>0-250</td>
<td>0-250</td>
<td>0-010</td>
<td>0-000</td>
<td>0-000</td>
<td>0-856</td>
<td>ND</td>
</tr>
<tr>
<td>12</td>
<td>Varzuga</td>
<td>62-5</td>
<td>0-025</td>
<td>0-175</td>
<td>0-313</td>
<td>ND</td>
<td>0-000</td>
<td>0-024</td>
<td>0-371</td>
<td>0-017</td>
<td>0-000</td>
<td>0-000</td>
<td>0-432</td>
<td>ND</td>
</tr>
<tr>
<td>13</td>
<td>Ponoy</td>
<td>62-5</td>
<td>0-000</td>
<td>0-127</td>
<td>0-539</td>
<td>ND</td>
<td>0-000</td>
<td>0-069</td>
<td>0-461</td>
<td>0-108</td>
<td>0-000</td>
<td>0-000</td>
<td>0-549</td>
<td>ND</td>
</tr>
<tr>
<td>14</td>
<td>Pecha</td>
<td>62-5</td>
<td>0-000</td>
<td>0-185</td>
<td>0-130</td>
<td>0-430</td>
<td>0-000</td>
<td>0-000</td>
<td>0-478</td>
<td>0-120</td>
<td>0-000</td>
<td>0-000</td>
<td>0-674</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>Norway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Tana</td>
<td>62-5</td>
<td>0-000</td>
<td>0-085</td>
<td>0-190</td>
<td>0-270</td>
<td>0-000</td>
<td>0-094</td>
<td>0-623</td>
<td>0-019</td>
<td>0-000</td>
<td>0-000</td>
<td>0-651</td>
<td>0-110</td>
</tr>
<tr>
<td>10</td>
<td>Nansen</td>
<td>57-5</td>
<td>0-021</td>
<td>0-049</td>
<td>ND</td>
<td>0-071</td>
<td>0-000</td>
<td>0-000</td>
<td>0-472</td>
<td>0-000</td>
<td>0-021</td>
<td>0-000</td>
<td>0-500</td>
<td>0-172</td>
</tr>
<tr>
<td>16b</td>
<td>Nansen NA</td>
<td>62-5</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-002</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-695</td>
<td>ND</td>
</tr>
<tr>
<td>17</td>
<td>Vossø</td>
<td>57-5</td>
<td>0-108</td>
<td>0-011</td>
<td>0-000</td>
<td>0-088</td>
<td>0-000</td>
<td>0-026</td>
<td>0-490</td>
<td>0-004</td>
<td>0-000</td>
<td>0-000</td>
<td>0-558</td>
<td>0-432</td>
</tr>
<tr>
<td>18</td>
<td>Numendal</td>
<td>52-5</td>
<td>0-060</td>
<td>0-070</td>
<td>0-000</td>
<td>0-140</td>
<td>0-000</td>
<td>0-020</td>
<td>0-640</td>
<td>0-000</td>
<td>0-030</td>
<td>0-000</td>
<td>0-490</td>
<td>0-130</td>
</tr>
<tr>
<td></td>
<td>Sweden – Atlantic Ocean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19a</td>
<td>Göt (Grönån)</td>
<td>52-5</td>
<td>0-092</td>
<td>0-061</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-420</td>
<td>0-010</td>
<td>0-010</td>
<td>0-000</td>
<td>0-657</td>
<td>0-088</td>
</tr>
<tr>
<td>19b</td>
<td>Gullspång/L. Varner NA</td>
<td>52-5</td>
<td>0-000</td>
<td>0-000</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-200</td>
<td>0-020</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-500</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>Sweden – Baltic Sea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Mörumsån</td>
<td>52-5</td>
<td>0-254</td>
<td>0-000</td>
<td>ND</td>
<td>0-040</td>
<td>0-000</td>
<td>0-009</td>
<td>0-769</td>
<td>0-485</td>
<td>0-000</td>
<td>0-000</td>
<td>0-119</td>
<td>0-000</td>
</tr>
<tr>
<td>22</td>
<td>Emån</td>
<td>52-5</td>
<td>0-362</td>
<td>0-000</td>
<td>ND</td>
<td>ND</td>
<td>0-000</td>
<td>0-000</td>
<td>0-731</td>
<td>0-435</td>
<td>0-000</td>
<td>0-000</td>
<td>0-037</td>
<td>0-000</td>
</tr>
<tr>
<td>23</td>
<td>Dal</td>
<td>57-5</td>
<td>0-062</td>
<td>0-000</td>
<td>ND</td>
<td>0-130</td>
<td>0-000</td>
<td>0-042</td>
<td>0-551</td>
<td>0-000</td>
<td>0-000</td>
<td>0-000</td>
<td>0-170</td>
<td>0-050</td>
</tr>
</tbody>
</table>

Table III. Frequencies of uncommon alleles of polymorphic enzyme loci in the salmon populations of Northern Europe, from literature and unpublished data, used for analysis of regional variation across the species European range. Lat. North refers to mean latitude north to 5°.
<table>
<thead>
<tr>
<th>Country</th>
<th>Region</th>
<th>Sample Code</th>
<th>Protein Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>Ljungan 0.196</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>Angerman 0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>26</td>
<td>Ume 0.133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27</td>
<td>Skellefte 0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>28</td>
<td>Lule 0.133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>29</td>
<td>Kalix 0.133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>Torne 0.133</td>
</tr>
<tr>
<td>Russia</td>
<td>Baltic Sea</td>
<td>31</td>
<td>Simo 0.250</td>
</tr>
<tr>
<td>Estonia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>32a</td>
<td>Neva 0.229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32b</td>
<td>Lizhma NA 0.118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32c</td>
<td>Burnaya NA 0.359</td>
</tr>
<tr>
<td>Latvia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33</td>
<td>Kunda 0.413</td>
</tr>
<tr>
<td></td>
<td></td>
<td>34</td>
<td>Keila 0.325</td>
</tr>
<tr>
<td>Scotland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>35</td>
<td>Daugava 0.196</td>
</tr>
<tr>
<td>Ireland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>36</td>
<td>Tweed 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>37</td>
<td>Tay 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38</td>
<td>Nesk 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39</td>
<td>Spey 0.014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
<td>Shin 0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41</td>
<td>Dionard 0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42</td>
<td>Grimmersta 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>43</td>
<td>Crean 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44</td>
<td>Nith 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45</td>
<td>Bush 0.041</td>
</tr>
<tr>
<td></td>
<td></td>
<td>46</td>
<td>Spaddagh 0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47</td>
<td>Blackwater 0.010</td>
</tr>
<tr>
<td>England</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>48</td>
<td>Dee-w 0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>49</td>
<td>Usk 0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>Teign 0.000</td>
</tr>
<tr>
<td>Map number</td>
<td>River</td>
<td>Lat.</td>
<td>AAT-4</td>
</tr>
<tr>
<td>------------</td>
<td>----------------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*25</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>Douron</td>
<td>47.5</td>
<td>0.000</td>
</tr>
<tr>
<td>56</td>
<td>Elorn</td>
<td>47.5</td>
<td>0.000</td>
</tr>
<tr>
<td>57</td>
<td>Adour (Nive)</td>
<td>42.5</td>
<td>0.000</td>
</tr>
<tr>
<td>Spain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>Bidasoa</td>
<td>42.5</td>
<td>0.000</td>
</tr>
<tr>
<td>59</td>
<td>Pas</td>
<td>42.5</td>
<td>0.000</td>
</tr>
<tr>
<td>60</td>
<td>Narcea</td>
<td>42.5</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Ref.: 1, Verspoor (unpubl.); 2, Bourke et al. (1997); 3, Danielsdottir et al. (1997); 4, Kazakov & Titov (1991); Titov (unpubl.); 5, Vuorinen & Berg (1989); 6, Hindar (unpubl.); 7, Jansson (unpubl.); 8, Koljonen et al. (1999); 9, Jordan et al. (1992); 10, Wilson et al. (1995). *X and *Y refer to the collective frequencies in each case of all the variant alleles i.e. 1 minus the frequency of the *100 allele. ND: no data.
However, the sample from Lake Vänern (Gullspång) clusters with samples from northern Russia. The MDS plot [Fig. 9(a)] shows the same regional differentiation. France/Spain, and Iceland/West Greenland occur in non-overlapping clusters radiating out from a central group of rivers from northern France/southern British Isles, northern British Isles, Norway and northern Russia, which overlap. The nonanadromous samples from the River Namsen and Lake Ladoga in the Neva basin occur on the periphery of the Baltic cluster, while the Gullspång/Lake Vänern sample now clusters with the southern European rivers. PCA analysis [Fig. 9(b)] is less informative. The most marked regional differentiation, revealed by factor 3 (dominated by MEP-2*125 and PGM-2*120) and factor 4 (AAT-4*25 and *50 variation; Table IV), shows only the distinct Icelandic and Baltic Sea clusters and the nonanadromous samples to be regional outliers.

Fig. 7. Map of rivers from which samples were used in the regional analysis of allozyme variation in Europe. Numbers refer to Table III.
The river samples, excluding nonanadromous populations, show heterogeneity across latitude classes (Table V) for \( AAT-4^{*50} \), \( MDH-3,4^{*75} \) and \( MEP-2^{*} \), both including and excluding the Baltic Sea samples, though the analysis is biased given that the number of rivers analysed for \( ESTD-2^{*} \), \( FBALD-3^{*} \) and \( TPI-3^{*} \) is limited. However, only \( MEP-2^{*} \) shows a clinal change (Fig. 10), with the mean frequency of \( *125 \) increasing from 0·1 in the south of the species’ range to 0·9 in the north, something found previously (Verspoor & Jordan, 1989; Bourke et al., 1997). At the same time, no polymorphisms are regionally restricted, though
Fig. 9. (a) MDS plot for European river samples based on genetic distance, $D_A$ and (b) plot of PCA scores for components 3 and 4 for the same river samples based on variant allele frequencies. Lines enclose regional geographic groups; numbers refer to Fig. 7 and Table III (•, Iceland/Greenland; ○, Northern Russia/Norway; ●, Southern Norway/Western Sweden; □, Baltic Sea; ●, Northern British Isles; □, Southern British Isles/Northern France; ●, Southern France/Spain).

AAT-4*25 is rare in southern France and Spain. It is also the main AAT-4* variant in the Baltic Sea drainage area with a mean frequency of 0.213 in the anadromous populations and occurs at high frequency in Iceland (mean 0.236), similar to the *50 variant (0.225); otherwise AAT-4*25 averages from 0.009 in northern France/southern British Isles to 0.052 in southern Scandinavia. MDH-3,4*75/80 is relatively common in southern France and Spain but rare elsewhere.

The non-anadromous samples encompassed by the analysis showed significantly lower heterozygosity (H; Table VI) than anadromous samples (K–W test, d.f. = 1, P = 0.022), consistent with other studies (Vuorinen, 1982; Ståhl, 1987; Koljonen, 1989; Vuorinen & Berg, 1989; Verspoor, 1994); these studies also

<table>
<thead>
<tr>
<th>Variable</th>
<th>With Baltic Sea</th>
<th>Without Baltic Sea</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>H</td>
<td>P</td>
</tr>
<tr>
<td>AAT-4*25</td>
<td>56</td>
<td>14.8</td>
</tr>
<tr>
<td>AAT-4*50</td>
<td>56</td>
<td>21.8</td>
</tr>
<tr>
<td>ESTD*80</td>
<td>17</td>
<td>6.9</td>
</tr>
<tr>
<td>FBALD-3*50</td>
<td>18</td>
<td>7.5</td>
</tr>
<tr>
<td>GPI-3*90</td>
<td>56</td>
<td>13.0</td>
</tr>
<tr>
<td>IDDH-1*X</td>
<td>56</td>
<td>14.6</td>
</tr>
<tr>
<td>IDDH-2*Y</td>
<td>56</td>
<td>7.5</td>
</tr>
<tr>
<td>IDHP-3*116</td>
<td>56</td>
<td>10.9</td>
</tr>
<tr>
<td>MDH-1*-200</td>
<td>56</td>
<td>6.5</td>
</tr>
<tr>
<td>MDH-3,4*75/80</td>
<td>56</td>
<td>20.4</td>
</tr>
<tr>
<td>MDH-3,4*120</td>
<td>56</td>
<td>3.7</td>
</tr>
<tr>
<td>MEP-2*125</td>
<td>56</td>
<td>15.7</td>
</tr>
<tr>
<td>PGM-2*120</td>
<td>56</td>
<td>5.4</td>
</tr>
<tr>
<td>TPI-3*103</td>
<td>21</td>
<td>0.5</td>
</tr>
</tbody>
</table>

show allelic diversity to be generally lower in non-anadromous populations. Among the defined regions, there was also significant heterogeneity (K–W test, d.f. = 6, \( P < 0.001 \)) for anadromous samples, with \( H \) lowest for Baltic Sea rivers as found previously (Ståhl, 1987; Bourke et al., 1997). However, excluding Baltic Sea rivers, \( H \) still shows heterogeneity among regions (K-W test, d.f. = 5, \( P = 0.004 \)), with the highest levels in northern France/southern British Isles and northern Russia, and lowest in Iceland and southern Scandinavia.

Most studies report the occurrence of rare alleles (Ståhl, 1987; Jordan et al., 1992; this volume; Wilson et al., 1995; Bourke et al., 1997; Mjølnerød et al., 1997), but little is known in most cases about their overall distributions in Europe; even where the same loci have been screened across studies, the equivalence of rare alleles remains uncertain. At the \( ME^* \) locus, \(*110\), the only European variant appears restricted to the north of France (Verspoor & McCarthy, 1997), while within the British Isles Child et al. (1976) detected

![Figure 10](https://example.com/figure10.png)

**Fig. 10.** Mean frequency of the \( AAT-4^*50 \) (—), \( MDH-3,4^*75/80 \) (---) and \( MEP-2^*125 \) (--) for European populations in relation to latitude based on their classification into 5° categories.

<table>
<thead>
<tr>
<th>Grouping</th>
<th>Mean ( H )</th>
<th>S.E. of mean</th>
<th>Number of rivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iceland</td>
<td>0.156</td>
<td>0.014</td>
<td>8</td>
</tr>
<tr>
<td>Northern Russia/Northern Norway</td>
<td>0.195</td>
<td>0.009</td>
<td>7</td>
</tr>
<tr>
<td>Southern Norway/Western Sweden</td>
<td>0.150</td>
<td>0.011</td>
<td>5</td>
</tr>
<tr>
<td>Northern British Isles</td>
<td>0.180</td>
<td>0.006</td>
<td>10</td>
</tr>
<tr>
<td>Northern France/Southern British Isles</td>
<td>0.215</td>
<td>0.011</td>
<td>7</td>
</tr>
<tr>
<td>Southern France/Spain</td>
<td>0.191</td>
<td>0.024</td>
<td>4</td>
</tr>
<tr>
<td>Baltic Sea</td>
<td>0.139</td>
<td>0.006</td>
<td>15</td>
</tr>
<tr>
<td>Non-anadromous populations</td>
<td>0.115</td>
<td>0.025</td>
<td>4</td>
</tr>
</tbody>
</table>

regional differences in the distribution of the rare $TF^*2$ variant in the British Isles. The $TF^*2$ defined regions also show an overall difference across allozyme loci (Jordan et al., this volume). Danielsdottir et al. (1997) report a frequent low level $GLUDH-2^*115$ variant in some Icelandic rivers which, though not widely screened, Ståhl (1987) did not find in Sweden. In the Baltic Sea area and northern Russia, Koljonen (1989) found five of 25 loci polymorphic. For the same set of loci, in Scottish populations Jordan et al. (1992) found 12 polymorphisms while eight were found by McElligott & Cross (1991) in Irish rivers. Such general differences suggest, as observed in North America, that rare alleles may often be regionally distributed. Thus the regional analysis presented above, which excludes rare alleles, is likely to underestimate the actual degree of regional differentiation.

VARIATION WITHIN RIVER SYSTEMS

ANADROMOUS LIFE-HISTORY FORMS

Møller (1970; current volume) was the first to investigate whether individual river systems contain multiple populations of anadromous salmon using protein electrophoresis. He found significant heterozygote deficiencies at the $TF^*$ locus in pooled as well as in some individual samples from two large rivers, the Miramichi and St John, in New Brunswick in eastern Canada, but not in smaller and less extensively sampled rivers. Deficiencies may have a number of causes but are expected where samples are composed of mixtures of genetically different populations (i.e. Wahlund Effect; Hartl & Clark, 1989). In the Miramichi, population structure was supported by allele frequency heterogeneity among tributary samples as well as among smolt samples taken during different stages of the smolt run in the river. The latter suggested that different tributary populations migrated at different times. Later multi-locus studies by Ståhl (1987) in the Miramichi, and by Verspoor (this volume) in the St John, also showed significant variation among tributaries. Indeed, in the latter river, tributary differences were greater than between most river systems in North America.

Allozyme studies of medium to large river systems elsewhere in the species range support the view that population structuring among anadromous stocks within river systems is common. Ståhl (1981, 1987; see also Heggberget et al., 1986; Ståhl & Hindar, 1988) analysed variation within a number of Swedish and Norwegian drainages and in 9 of 11 river systems found significant heterogeneity across loci. The same was reported by Elo et al. (1994) among geographical samples in both the Teno and Näättämö rivers in northern Finland. Significant heterogeneity has also been reported among tributaries in rivers in Scotland (Verspoor et al., 1991), Ireland (McElligott & Cross, 1991; Galvin et al., 1994) and Wales (O’Connell et al., 1995, 1996). The only exception in the literature is the River Tweed in Scotland (Jordan et al., 1992), possibly because many of the tributaries have been newly recolonized following the removal of dams and weirs. However, the distribution of an $IDHP-2^*$ variant was restricted to only a few tributaries of the Tweed where it occurred at frequencies of up to 0·11, suggesting gene flow among tributaries is still constrained to a large degree.

NON-ANADROMOUS (RESIDENT) LIFE-HISTORY FORMS

Allozyme studies of parts of the species’ range with both anadromous and resident (i.e. non-anadromous) forms suggest that, in most cases at least, within river systems these two life-history forms belong to distinct populations. Early on, Payne (1974) showed significant divergence between the two forms within watersheds in Newfoundland at the TF* locus. Later Verspoor & Cole (1989) showed that levels of divergence between sympatric populations of the two forms in Little Gull Lake in the Gander River system in Newfoundland approached those found between European and North American salmon. Lower though still significant levels of divergence exist between allopatric forms in other river systems in Newfoundland (Verspoor, 1994), as well as in the River Namsen in Norway (Vuorinen & Berg, 1989) and the Götälv-Lake Vänern system in Sweden (Ståhl, 1987). Koljonen (1989) found that the nonanadromous Saimaa and Lake Onega salmon differed more from the anadromous Neva salmon than the anadromous stocks in the Baltic differed from each other.

These findings are reinforced by the significant differentiation between the anadromous and non-anadromous (Lake Onega) forms in the Neva system (present paper) as well as between nonanadromous samples within this river system. This supports the view that the Neva river system, consisting of many large lakes and connecting rivers, contains multiple, genetically different non-anadromous populations, as is suggested by the presence of river specific spawning stocks (Kazakov, 1992). The study by Verspoor & Cole (this volume) indicates that, at least in the Gander River system in Newfoundland, some non-anadromous populations are restricted to single lakes and are lake spawners. As non-anadromous fish occur in many of the lakes of Newfoundland, this suggests that large numbers of genetically distinct nonanadromous populations occur in this part of the species range.

TEMPORAL VARIATION AMONG SAMPLES

Allele frequencies in samples from the same location, in general, show temporal stability, supporting the view that the observed spatial patterns are stable and reflect population structuring. Jordan et al. (1992) examined variation across either 2 or 3 year classes of 0+ juveniles collected from 12 separate locations in Scotland for six variable loci. Overall genetic differences among year classes were not significant. This conclusion supported by results from five of the sites which were monitored for a total of 5 years (unpubl. data). The only site studied by Jordan et al. (1992) to show significant heterogeneity was one, the River Polly, which had a history of stocking with various non-native salmon from a number of sources. Indeed, the recent work of Møller (this volume) shows that, at least for some loci, genetic protein variation in wild, undisturbed populations, is likely to be stable over decadal timeframes. His study, of the TF* polymorphism in mixed cohort samples from wild populations in twelve rivers in Atlantic Canada, showed stable allele frequencies in ten of the rivers between 1969/1970 and 1998. The two rivers where changes were seen had either undergone a dramatic decline or stocking had been carried out.
Ponomareva et al. (2002) found significant heterogeneity in the Nil’ma, a small river in the White Sea region of northern Russia, when 5 year classes of juveniles were compared. However, this river also experienced a severe population reduction during the sampling period and had been significantly disturbed by human activities. Temporal heterogeneity has also been reported in other studies where rivers have been maintained by or subject to stocking with hatchery reared fish (Ståhl, 1987; Koljonen, 1989; Morän et al., this volume b). However, in most studies where multiple year classes have been sampled in consecutive years, there has been no evidence of temporal variation (Crozier & Moffett, 1989; Danielsdottir et al., 1997; Blanco et al., this volume; Consuegra et al., this volume; Verspoor, this volume). Yet this may not always be the case, as demonstrated by studies of undisturbed populations of other species (Dempson et al., 1988). With regard to salmon, some loci such as MEP-2* seem to show temporal heterogeneity more than others, perhaps a reflection of the effect of natural selection on these loci. However, the evidence suggests that, in general, temporal homogeneity seems a reasonable assumption where samples are from healthy wild populations and are composed of multiple year classes. Thus the observed patterns of variation described, both among and within river systems, can reasonably be assumed to reflect actual spatial differentiation.

PHENOTYPIC AND ENVIRONMENTAL ASSOCIATIONS

Few early studies on Atlantic salmon investigated the association of protein variation with phenotypic or environmental variation to address the question of whether it might be subject to selection and of adaptive relevance. The exception was Payne (1974) who noted a heterozygote deficit at TF* among male grilse in one coastal fishery not present among female grilse or multi-sea winter fish, though this deficit is not significant if corrected for multiple tests. Later for this same locus, Verspoor (1986) showed allele frequencies in North American populations to correlate with the proportion of grilse in rivers. Other associations with both phenotypic and environmental variation have been reported for MEP-2* and a number of other loci.

Verspoor & Jordan (1989) found a significant latitudinal variation in MEP-2* variation among rivers in both North America and Europe. Latitudinal clines can arise as a result of historical gene flow into newly colonized regions due to gene flow from two differentiated refugial populations, where gene flow is a function of distance. However, this seems unlikely for MEP-2* given the cline occurs in both Europe and North America and, in both areas, is even more strongly correlated with summer temperature (Verspoor & Jordan, 1989). Furthermore, the same association of MEP-2*100 with warmer temperatures has been found within a number of river systems (Verspoor & Jordan, 1989; Verspoor et al., 1991; Gilbey et al., 1999). It has also been found in a more intensive recent analysis by Jordan et al. (this volume) of rivers in the British Isles. Again MEP-2*125 was highly correlated with latitude, January and July temperature, and local river gradient; a significant association with latitude, but not temperature, was also found for AAT-4*50 but not with IDDH-2*, IDHP-3* or MDH-3,4* variation. Genetic differentiation within river systems at MEP-2* is generally (Jordan et al., 1997), though not always (Gilbey et al., 1999), more
MEP-2* variation has also been reported to be associated with phenotypic traits such as mean size at age (Jordan & Youngson, 1991; Gilbey et al., 1999), specific growth rate (Pringle, 1995), and sea age (Jordan et al., 1990; Morán et al., 1994b), and with growth variation within families (Pringle, 1995; Gilbey, 2003). Experimental studies by McGinnity (1997) found associations of MEP-2* genotype with survival and growth in the early life-history stages among planted out fish as well as an association with smolt age and male parr maturation. McCarthy et al. (2003) have found MEP-2* to be associated with growth and age of first feeding in hatchery fish, while Consuegra et al. (this volume) have found evidence of small but significant differences in the MEP-2* allele frequencies between adults sampled in the fishery early in the spring and later in the summer.

At PGM1-r*, a regulatory locus controlling liver expression of PGM1, the frequency of the gene for expression, designated *b, was consistently higher in males than in females in one population and there was a positive correlation with the proportion of early maturing male parr (Pollard et al., 1994). Verspoor & Moyes (this volume) found evidence in hatchery fish of differential mortality in the early juvenile stage among genotypes at this locus within families. In rainbow trout, O. mykiss, individuals with liver expression show faster embryonic development (Allendorf et al., 1982, 1983; Leary et al., 1983; Ferguson & Danzmann, 1985). Growth and development have also been reported to be associated in Atlantic salmon with electrophoretic variation in trypsin, a gut enzyme (Torrissen & Torrissen, 1984; Torrissen, 1991; Torrissen & Shearer, 1992; Torrissen et al., 1993, 1994), which may be related to differences in the ability of fish to absorb free amino acids and digest proteins. However, the genetic control of the variation is unclear (Torrissen et al., 1993).

Some studies (Blanco et al., 1998; McCarthy et al., 2003; Pineda et al., 2003) report that higher multi-locus heterozygosity in juvenile hatchery fish is associated with early first feeding, faster growth, a higher proportion of mature males in the first year of life, and a higher proportion of smolts after 1 year. Higher heterozygosity has also been found to be associated with lower levels of fluctuating asymmetry for bilateral traits such as fin ray number (Blanco et al., 1990; Borrell et al., 2004), but this has not been found in other studies (Vøllestad & Hindar, 1997).

CULTURED POPULATIONS

Protein studies were the first to show, as predicted by theory (Gall, 1987), reduced genetic variability in cultured populations of Atlantic salmon. Studying nine Baltic Sea river systems, Ståhl (1983, 1987) found 10% less genetic variability in hatchery stocks, reared for ranching and supplementation, compared to wild stocks. Though not statistically significant, divergence within drainages, and between drainages within regions, was three times higher for hatchery stocks. Koljonen et al. (1999) found mean heterozygosity (H) to be lower in hatchery (3-6%) as opposed to wild (4-8%) populations in the Baltic Sea area.
These observations are consistent with genetic losses in cultured stocks due to founder effects and genetic drift.

Verspoor (1988b) compared genetic variability in wild and farmed Atlantic salmon in Canada at one regulatory and 19 structural protein loci. The comparison was based on juveniles in samples of first generation hatchery fish from 7 hatchery stocks, and samples of 7 wild stocks in New Brunswick and Nova Scotia. Mean heterozygosity ($H$) differed significantly ($P = 0.013$), with $H = 0.145$ in wild stocks and $H = 0.107$ in the cultured stocks. The wild stocks also had a higher mean number of alleles, although this was not statistically significant. It was concluded that a significant reduction in genetic variability had already taken place in the first generation due to founder effects. Cross & King (1983) analysed six polymorphic loci in the progeny of fifth generation ranched stock in Ireland subject to growth and disease resistance selection and maintained using from 20 to 60 breeders. They showed divergent allele frequencies as well as lower heterozygosity and allelic diversity compared to the wild stock from which the ranched stock was derived. In contrast, in the River Bush in Northern Ireland, where numbers of spawners were in the order of 200 males and females, Crozier & Moffett (1989) found no reductions. However, Crozier (1994) found significant differences between hatchery and wild fish and temporal heterogeneity among different year classes of hatchery fish not seen for wild fish. Crozier (1998) also found 1+ smolts in first generation hatchery fish from the River Bush to differ genetically from non-smolts of the same cohort, something not observed in wild fish and possibly related to an association of the genetic variation studied with growth performance.

Youngson et al. (1989, 1991) examined genetic variation at six variable loci in 12 lines of farmed Scottish salmon in each of two years compared with 18 wild Scottish salmon populations. As well as differences among farmed strains, significant temporal changes were observed across years within strains. Significant differences also occurred in mean allelic frequencies between strains of Scottish and Norwegian origin at two loci (IDHP-3* and MDH-3,4*), between Scottish wild and Norwegian farmed at four loci (AAT-4*, MDH-3,4*, MEP-2*, IDDH-1*), probably reflecting regional differences in their wild source stocks, and between wild Scottish and farmed Scottish at two loci (IDHP-3*, MEP-2*). Significant differences were found in all but one comparison of farmed strains with their wild founder populations and were comparable to the natural differences that were observed among wild salmon in separate river catchments. Genetic variability was not reduced in the farmed strains generally, though a few strains did show markedly less variation. Difference between farm stocks and their source populations were attributed largely to genetic drift, but a directional change in favour of MEP-2* 125 was found when comparing strains to their wild source populations. This supports the view that this locus is subject to selection, and may illustrate a genetic character with a different effect on the fitness of an individual in domesticated and natural environments, where gene flow from escaped salmon would reduce performance of a wild population. Others have reported changes, during hatchery rearing, in allele frequencies at AAT-4*, IDHP-3*, and MEP-2*, attributable to selective mortality (Kazakov et al., 1988; Ofitserov et al., 1989; Artamonova & Kholod, 2004).
Cross & Ni Challanain (1991) compared the five main Irish farm strains used in Ireland with a wild Irish salmon stock at 6 polymorphic enzyme loci. The four Norwegian and one Scottish origin strain were significantly heterogeneous. This was also the case for the four Norwegian strains, most probably due to founder effects. Only two of the minor farm lines showed reduced genetic variability. Mjølnerød et al. (1997) compared a Norwegian farm strain with two large wild populations, and found the farm strain to have 14% fewer alleles but higher heterozygosity, though neither difference was statistically significant. Differentiation, as estimated by $F_{ST}$, was consistent over three classes of markers, with the wild populations being most similar, and the farmed population most divergent. More recently, Skaala et al. (this volume) studied the five main Norwegian farmed strains and four major wild Norwegian populations at eight enzyme coding loci. The mean number of alleles and proportion of polymorphic loci were c. 12% and 14% lower, respectively, in the farm strains, while mean heterozygosity was c. 17% lower. A paired comparison between one farm strain and its wild source populations found a slight reduction in mean number of alleles and in number of polymorphic loci, but no reduction in heterozygosity. However, differentiation was c. 10 times greater than between three large, wild populations. Mean $F_{ST}$ over 8 polymorphic loci was 0.161 among the domesticated strains, with high values at $MEP-2^*$ (0.340), $TPI-3^*$ (0.115) and $MDH-2^*$ (0.111), compared to 0.021 among the four wild stocks. Thus, in 8 generations, the farm strains both lost variation and differentiated significantly from their source stocks. In contrast, a study by Ward et al. (1994) found no differences in variability at allozyme loci between farmed salmon in Tasmania and the original source population in Canada from which the stock was established 30 years earlier.

**DISCUSSION**

**POPULATION STRUCTURE AND GENE FLOW**

Genetic heterogeneity among spatial samples is clearly widespread and provides strong support for the view that Atlantic salmon are highly structured into genetic populations, both within and among rivers. Given the observed temporal stability of the spatial variation, structuring represents the most parsimonious explanation for the heterogeneity and is consistent with homing of anadromous salmon back to natal rivers (Stabell, 1984) as well as to tributaries within larger rivers (Youngson et al., 1994). The absence of detectable differences in some cases most probably reflects the fact that genetic drift is the main factor underlying differentiation and, by chance in some, that differences have not evolved at the small selection of loci screened. This also probably accounts for tributary populations in some rivers, and neighbouring rivers being as different as distant rivers for the loci examined (e.g. as seen for the Hammond and Tobique tributaries of the St John River; Fig. 5).

The probable importance of genetic drift is highlighted by the comparisons made between anadromous and nonanadromous populations. Nonanadromous populations are more divergent and have less variability than geographically adjacent anadromous samples though their overall mean allele frequencies are the same (Verspoor, 1994; Fig 9). This is unlikely if selection is responsible for
differentiation at most of the loci (Lewontin, 1974). However, in some cases differences may reflect variation in the refugial sources of founders at the time of post-glacial colonization, a possibility discussed below.

$F_{ST}$, a measure of genetic divergence, has been used to estimate historical levels of gene flow in a number of studies, based on the assumption that genetic drift is the main factor underlying genetic differentiation and an island model of gene flow. Some studies have also used the private allele method of Barton & Slatkin (1986). Under the island model, $F_{ST}$ will be inversely proportional to the historical effective population size ($N_e$) and the mean historical level of gene flow ($m$) according to the formula: $F_{ST} = 1/[1 + 4N_em(N/N_e - 1)^2]$ (Chakraborty & Leimar, 1987). Estimates of the genetically effective number of migrants between populations ($N_em$) derived from $F_{ST}$ values are generally consistent across studies for a given geographical level when MEP-2* is omitted from consideration. The effect of excluding MEP-2*, a locus for which there is evidence that differentiation is affected by selection, is variable (Verspoor et al., 1991; Elo et al., 1994; Wilson, 1996; Jordan et al., 1997), though in most cases its inclusion does not change the inferences which can be drawn. Estimates of $N_em$ based on private alleles, which are not sensitive to the effects of selection (Barton & Slatkin, 1986), are in general agreement with those obtained using $F_{ST}$, though they are often lower.

Estimates of $N_em$ based on $F_{ST}$ can only place upper limits for historical gene flow, as the model used for their derivation assumes an infinite number of populations, an equal likelihood of migration among all populations, selective neutrality, and that the populations have reached equilibrium between genetic drift and gene flow (O’Connell et al., 1995). Where effective population sizes are large and gene flow is highly restricted, populations will take many generations until differentiation stabilizes and equilibrium is reached between genetic drift and gene flow. This may explain why estimates based on the private alleles method tend to be lower. Thus estimates of $N_em$ based on $F_{ST}$ define an absolute upper limit. Furthermore, such values represent historical averages and it is most likely that gene flow will have been greatest in the early history of populations, when they were first established following the deglaciation of much of the species’ modern range. Thus levels of contemporary gene flow will in reality be much less than suggested by estimates derived from observed levels of differentiation. As such, it is difficult to interpret the observed genetic differentiation, even within rivers, as indicating anything other than the existence of large numbers of distinct genetic populations among which little effective gene flow occurs.

Estimates of $N_em$ derived from $F_{ST}$ between continents of $<1$ (Verspoor, this volume) provide strong evidence that, overall, little gene flow has occurred between North America and Europe, either in recent times or in the past. However, as discussed below, there is evidence to suggest limited intercontinental gene flow occurred in the early post-glacial period (Verspoor et al., 1999; Knox et al., 2002; Makhrov et al., this volume) in both directions. In the case of sympatric anadromous and resident salmon in Little Gull Lake, Newfoundland, $N_em$ estimates from $F_{ST}$ of 0-1 (Verspoor & Cole, 1989) point to their almost complete historical reproductive isolation. In contrast, higher $N_em$ values are found among anadromous fish from different tributaries. Within the Kyle of Sutherland system in northeast Scotland, estimates of $N_em$ obtained

are 15·1 and 2·0, based on \( F_{ST} \) and the private alleles method, respectively (Verspoor et al., 1991), while for the Näätämö and Teno rivers in Finland Elo et al. (1994), \( N_{em} \) values calculated from \( F_{ST} \) are 4·6 to 3·1; the former river is of similar size to the Kyle of Sutherland system and the latter several orders of magnitude larger. Based on \( F_{ST} \), \( N_{em} \) was estimated to be \(~\sim\)8 for three medium sized catchments in Wales (O’Connell et al., 1995).

Wilson (1996) estimated \( N_{em} \) to be 20·4 among tributaries in the river Dee in Scotland using a maximum likelihood method. He found no evidence that differentiation was a function of geographic separation in the River Dee, River Teno, or in other rivers examined, though the test was relatively insensitive given the low levels of heterozygosity observed. However, in the Teno he found a correlation with branching, suggesting that the physical structure of a river system has influenced historical gene flow. Estimates of gene flow among rivers are generally lower. For Baltic rivers, estimates based on \( F_{ST} \) ranged from 0 to 10 migrants (Koljonen et al., 1999). The average among river value for North America, (based on the combined data sets analysed above but excluding replicate St John River samples) with an overall \( F_{ST} \) of 0·088, is 2·5 with an average based on private alleles of 0·94 (Genepop Version 3.1d – Raymond & Rousset, 1995).

Actual effective population sizes (i.e. \( N_e \)) for Atlantic salmon in most rivers are unknown, making it difficult to translate \( N_{em} \) into maximum straying rates, though, in most cases, populations are likely to number in the hundreds or thousands rather than the tens. This suggests maximum historical genetically effective straying rates of a few per cent per generation, lower than postulated contemporary rates of straying (Stabell, 1984; Youngson et al., 1994). However, modelling of genetic divergence (\( F_{ST} \)) as a function of genetic drift and migration, using estimates of local homing and population sizes as input, suggests that salmon populations are genetically more different at enzyme loci than would be expected from observed migration patterns (Hindar & Tufto, 1995). This indicates that a substantial proportion of straying does not lead to gene flow.

**PHYLOGEOGRAPHY**

The large and consistent differentiation of European and North American salmon seen at allozyme loci suggests populations on the two continents belong to different lineages with separate origins on these continents. It also suggests these lineages dominated in refuges on the two continents during the Pleistocene. Atlantic salmon vertebrae from caves in northern Spain confirm that rivers along the Atlantic coast to the south of the ice sheet served as a refugial area in Europe for the European type (Consuegra et al., 2002). This area probably extended north and east to the ice edges in southern Britain and western Denmark, given that modern populations in West Greenland and Iceland are found within kilometres of ice fields (MacCrimmon & Gots, 1979). The location(s) of refugia in North America remains speculative as all modern populations are found in deglaciated areas (Verspoor et al., 2002).

Benhke (1980) hypothesised that both Atlantic salmon and Arctic charr *Salvelinus alpinus* (L.) from Europe as well as North America colonized Newfoundland. This view is supported for Atlantic salmon by the occurrence
of rare mtDNA haplotypes typical of Europe in Newfoundland and Labrador (Birt et al., 1991; Knox et al., 2002), and it could explain some aspects of the distribution of protein variation. The TF*1 variant, fixed in most European populations (Payne et al., 1971; Child et al., 1976), is more common in Newfoundland and Labrador than in other parts of the North American range (Verspoor, 1986), something also seen with respect to European variants at AAT-4*, IDDH-1*, and IDDH-2* (Verspoor, this volume), and at ESTD* (Skaala et al., 1998; Verspoor, unpubl.).

Post-glacial colonization of parts of Europe by salmon from North America was first suggested by the occurrence in northern Russia of low levels of mtDNA types typical of North America, variants not detected elsewhere in Europe (Verspoor et al., 1999; Nilsson et al., 2001; Consuegra et al., 2002). This view is also consistent with protein studies which show that ESTD*80, apparently fixed in most North American populations, also occurs in this region but not elsewhere in Europe (Bourke et al., 1997; Makhrov et al., 1998b). This may also explain why the closest genetic relationship of North American populations with those in Europe is with populations in northern Russia (Verspoor, this volume). This may also account, in part, for the differentiation of northern Russian populations from those elsewhere in Europe and for the higher levels of genetic diversity seen in this region compared to southern Norway and Iceland and the Baltic Sea (Table VI). It has been proposed that the region was colonized by European salmon from both the Atlantic and Baltic Sea drainages (Kazakov & Titov, 1991; Makhrov et al., this volume). If true, this would have further enhanced the differentiation among populations in this region. More detailed information on allozyme variation in Norway would allow these questions to be considered more critically.

Populations in southern Britain and Northern France lie at the centre of the genetic distribution of allozyme variation for Europe (Fig. 9), with allele frequencies similar to the overall European mean and the highest genetic diversity (Table VI). This is as expected if the area was the main source of incoming migrants in newly colonized areas (Makhrov et al., 2003). This assumes genetic differentiation is dominated by gene flow and genetic drift, and colonisation occurs by a combination of local range expansion and long distance dispersal out of refugia, with the latter followed by further local range expansion (Verspoor et al., 2002; Verspoor, this volume). If so, the mean genetic character of salmon across its range would remain more or less the same with just the distribution of variation among populations changing, and the mean genetic character today should be the same as the refugial ‘prepopulation’. Under this scenario, the observed associations of genetic and geographic distance would be expected, as well as the restricted regional distributions observed for many of the rare variants. In North America, the centre of the distribution of allozyme variation is associated with populations from the southern part of the species modern range (Fig. 6) closest to where refugial areas are likely to have been located.

Aspects of the distribution of protein variants suggest a second refugial area may have existed in Europe. The distribution of TF*2 led Child et al. (1976) to propose that there were two geographically distinct salmon races in the British Isles, derived from different glacial refugia; these areas also show a general differentiation across allozyme loci (Jordan et al., this volume). The existence of
a second refuge is also suggested by the overall distribution of \( AAT-4^*25 \), the main variant in the Baltic Sea area and many Icelandic populations, as well as being more common in the northern British Isles and North Sea rivers than to the south (Table III; Jordan et al., this volume). The allele’s virtual absence from rivers in the main refuge area suggests this is unlikely to have been its origin. Rather its wide northerly and easterly distribution suggests an origin in a second more easterly refuge. This could have been in the southern North Sea (Verspoor et al., 1999), centred on the ancient Elbe drainage which was largely unglaciated and drained northward into the Skagerrak, isolating it from Atlantic drainages, or further east in a proglacial ice lake. The latter has been suggested by Koljonen et al. (1999) to explain regional differentiation within the Baltic Sea drainage area. They postulate colonization of the Baltic Sea area from an ice lake refuge in eastern Russia to the south of the ice sheet, followed by a later secondary colonization of the northern Baltic Sea from the west, when the Baltic Sea was briefly connected to the North Sea (Koljonen et al., 1999).

The lack of consistent differentiation between anadromous and nonanadromous forms of Atlantic salmon (Verspoor, 1994) suggests the latter to be polyphyletic and to have evolved independently from anadromous forms in different river systems. This view is consistent with their occurrence only in areas covered by the Pleistocene ice sheets. Most probably nonanadromous forms evolved early on when initial populations in rivers were confronted by physical barriers to migration soon after they colonized some systems following the retreat of the Pleistocene ice sheets (Power, 1958; Berg, 1985). Alternatively, ‘lake salmon’ in deglaciated areas of Northern Europe may have evolved from resident populations in proglacial lakes at the edge of the ice sheets, a scenario which could also explain their lack of marine migration.

Despite possible post-glacial gene flow between the two continents, the major divergence between North America and Europe seen at many protein loci supports the view of Payne et al. (1971), based on the distribution of \( TF^* \) variation, that salmon from the two continents should be designated as distinct sub-species. \( I \), a measure of genetic similarity (Chakraborty & Leimar, 1987), is 0.919 between the St John River in Canada and the North Esk in Scotland (Bourke et al., 1997), two rivers whose divergence is typical of that seen between Europe and North America (Verspoor, this volume). This is of the same order as seen between cutthroat trout (\( O. clarkii \)) sub-species (Loudenslager & Gall, 1980) and equates to a distance \( D \) of 0.085 (−ln \( I \)), well above the value of 0.05 cited by Nei (1987) as being typical of sub-species. This divergence is also seen for satellite and mitochondria DNA variation (King et al., 2000, 2001; Koljonen et al., 2002) as well as numbers of chromosomes (North America \( 2N = 56 \), Europe \( 2N = 58 \)) and chromosome arms (Hartley, 1987).

### POPULATION ADAPTATION

It has been recognized since Lewontin & Hubby (1966) and Harris (1966) first revealed the high levels of protein polymorphism in wild populations that some of this variation may be selectively maintained. In principle, at least, variation that alters the charge state or conformation of a protein can affect protein function. This can in turn influence an individual’s survival chances and...
reproductive success. However, it is difficult to demonstrate that a polymorphism is subject to selection, and therefore reflects adaptive variation, without detailed experimental and genetic analysis (Endler, 1986). This has led to considerable debate regarding the extent to which studies of electrophoretic protein variation can inform the debate (Lewontin, 1974; Allendorf & Utter, 1979; Nevo, 1983).

Few studies have specifically addressed whether selection is acting on the protein polymorphisms in Atlantic salmon. There is no clear expectation of what effect selection will have on spatial distributions of variation, making it difficult to reject the null hypothesis of no selection (i.e. neutrality). Thus the evidence for selection which exists is largely circumstantial and is restricted largely to observed associations of genotypic and allelic variation with environmental variation and phenotypic performance at the MEP-2* locus.

The possibility that selection might be affecting MEP-2*, first mooted by Verspoor & Jordan (1989), is based on detection of a significant latitudinal cline in MEP-2* variation among rivers in both North America and Europe, as well as within two rivers in the British Isles. Though clines may arise without selection (Endler, 1977), this seems unlikely in the case of MEP-2* given that they occur independently in different areas and at different scales and, in all cases, variation is even more strongly correlated with temperature than with latitude (Verspoor & Jordan, 1989; Verspoor et al., 1991; Gilbey et al., 1999; Jordan et al., this volume). This view is supported by the commonly observed associations of MEP-2* variation with performance measures such as mean size at age, specific growth rate, and sea age, previously described, and with survival (McGinnity, 1997). Such consistent associations have not been detected for other polymorphisms. If selection is operating, this could explain why population differentiation at the locus is generally more constrained within rivers but more marked at the continental scale than observed for other loci (Jordan et al., 1997).

From studies to date, it cannot be concluded whether selection is acting directly on MEP-2* or on linked variation. The enzyme has an important metabolic role and, consistent with this, mitochondria from different MEP* genotypes from the same population show different rates of oxygen utilization (Godolphin, 1990). Furthermore, MEP-2* was found to be the molecular marker most strongly linked to growth in a recent study identifying markers for quantitative trait loci (QTL) (Gilbey, 2003). Though selection appears to be mediated through the effects of variation at the locus on growth, further work is needed to determine the mechanism underlying any selection.

Selection may be acting on other polymorphisms studied, but the evidence for this is more limited. There is some support for selection acting on PGM1-β*, the regulatory locus controlling liver expression of PGM1, given evidence relating to the same locus in rainbow trout, O. mykiss, with variation in embryonic development (Allendorf et al., 1982, 1983; Leary et al., 1983; Ferguson & Danzmann, 1985). Further insight into the adaptive relevance of variation at this locus and MEP-2* will come from further examination of the polymorphisms at the DNA level and examining gene expression and function in more detail at the molecular level. Certainly recent studies, many DNA based, suggest that selection does act on many protein polymorphisms (Powers et al., 1991; Verrelli & Eanes, 2000; Saunders et al., 2002; Sezgin et al., 2004).
INDIVIDUAL AND STOCK DISCRIMINATION

Research into protein variation from the very start searched for variation which could discriminate individuals and stocks of different origins, for example, in high seas fisheries. Atlantic salmon are subject to a number of offshore or coastal fisheries targeting populations from different rivers. The fishery in West Greenland, perhaps the best known, presented a major challenge for fisheries management and stock identification since the mid-1960s (Reddin & Friedland, 1999), as it is known to include salmon from Canadian, US, Scottish, English, Irish, French and Spanish rivers, as well as possibly other areas (Shearer, 1992). Early on, work showed differences in allele frequencies between continents (Nyman & Pippy, 1972; Payne, 1980; Payne & Cross, unpubl.; Cross & Healy, 1983; Ståhl, 1987; Verspoor, 1988c), and more were identified later (Bourke et al., 1997; Verspoor & McCarthy, 1997; Makhrov et al., 1998b). Initially, however, baseline genetic data was limited, and identification of fish to continent of origin focused on scale pattern recognition techniques (Reddin, 1986). Once more baseline information was available, it was possible to use proteins to identify fish to continent of origin with a high degree of accuracy (Verspoor, 1988c), and, for a brief period, allozyme analyses were used in combination of scale characteristics (Reddin et al., 1990). However, in the end, the use of proteins proved problematic due to difficulties in obtaining the fresh or frozen tissue needed for genotyping.

Allozyme based mixed-stock analysis has been most extensively applied in Atlantic salmon to estimating stock group proportions in the Baltic Sea (Koljonen, 1995; Koljonen & McKinnell, 1996). With allozyme data from seven variable loci it was possible to identify the occurrence of six different groups of genetically similar populations (Koljonen & McKinnell, 1996). One of those comprised the three major riverine populations (Tornionjoki, Kalixälven and Simojoki) and could thus serve as an index of the proportion of wild stocks in the catches. By combining the allele frequency information with freshwater age distribution of the stocks, it was possible to group all six wild baseline rivers into one group and get a better estimate of the total wild production in the catches. The pooling of allozyme and freshwater age information also increased the accuracy and precision of the estimate. The proportion of the wild group varied seasonally and at the Finnish coast was at its highest in May, when it made up over 40% of total catch (Koljonen & Pella, 1997).

Allozyme variation has also been used as a marker in other applications. Hutchings & Myers (1988) used variation at the MDH-3,4* locus to mark spawners in tanks and demonstrate the successful fertilization of anadromous females by mature male parr in competition with large anadromous males. Subsequently, Jordan & Youngson (1992) used MEP-2* variation in the same way. In their study of the participation of mature male parr in spawning under semi-natural conditions, the percentage of eggs fertilized by mature parr varied from 0-9 to 27-7% (mean 10-85) in the five redds examined. The study provided the first evidence that mature male parr in the wild breed successfully and led to considerable speculation on the evolutionary forces underlying the variation in the proportions of such parr seen among populations (Myers, 1984).

García de Leãñiz et al. (1989) used differences in MEP-2* between native and stocked fish to estimate the contribution of stocked northern European salmon...
to runs of returning adults in two rivers in northern Spain. They were able to show that non-native fish stocked as eyed eggs or fry were <20% as successful in returning as adults as native fish, something also found in subsequent years (Verspoor & García de Leániz, 1997). This view is supported by the observation of continuing major differences between the native populations in northern Spain and salmon in northern Europe (Vázquez et al., 1993; Morán et al., 1994c, this volume b). Crozier & Moffett (1995) used a rare variant at the GPI-3* locus to mark a hatchery reared native strain whose performance in the wild was then measured in juveniles post-stocking based on the changes in the frequency of the GPI-3* variant and in a MDH-1* allele, which was serendipitously found to also differ between stocked and wild fish. Fleming et al. (2000) used alternate homozygotes at MEP-2* to demonstrate that farmed salmon have lower lifetime reproductive success than native fish when released as spawners into a natural stream in Norway.

Allozyme variation has also been utilized to detect mature escaped farm salmon in wild spawning populations (Lund & Hansen, 1991) and demonstrate their spawning success (Lura & Saegrov, 1991). One opportunistic study (Crozier, 1993) was carried out in the Glenarm River to investigate the actual gene flow from escaped farmed salmon to wild stocks, following escapement from a nearby fish farm. Samples of ascending salmon (presumed escapes) and wild baseline samples were highly significantly heterogeneous across loci, and two alleles (MDH-B1,2*85 and SOD*120) absent in native fish were found in the farm salmon. This indicated the farm fish differed significantly from the native population. Samples of F1 0+ fry from the following summer showed significant linkage disequilibrium and shifts from the baseline native samples at four of seven loci (AAT-3*, IDDH-1*, MDH-B1,2* and MEP-2*) in the direction of the farmed strain. This was consistent with occurrence of interbreeding, but whether further introgression into the wild population occurred is unclear.

MANAGEMENT IMPLICATIONS

Protein studies have made a significant contribution to Atlantic salmon management and conservation. They have helped resolve the evolutionary relatedness and systematics of the genus Salmo and made clear its distinctiveness from rainbow trout. The work has also shown the extent of natural hybridization with brown trout and demonstrated that it can be increased by factors such as incursions of farm escapes and stocking. However, most significantly, the work has provided important scientific evidence which has led to the widespread acceptance of the view that the species is composed of a large number of river, and often tributary, specific populations. Allozyme work has reinforced the more limited behavioural evidence for this view, based on studies of natal homing, leading to the move towards the more population centred management seen today. At the same time, studies of cultured fish have drawn attention to the potential impact of stocking and farm escapes on wild populations and the need to use large numbers of broodstock to maintain genetic variability in captive populations. These issues are now being explored in more detail with DNA based analyses (McGinnity et al., 2003).
The contribution of protein studies to the understanding of local adaptation has been more limited, but the evidence from MEP-2* provides genetic support for the widespread occurrence of local adaptation and has contributed to a wider acceptance of this view. That further evidence has not emerged is not surprising. In only a few cases have convincing cases for the adaptive significance of protein variation in fish species been put together (e.g. LDH-B* in Fundulus heteroclitus L., Powers et al., 1991). In this regard, quantitative studies of phenotypic variation, supported by breeding work, have generally proven to be just as insightful as allozyme work (Riddell & Leggett, 1981; Riddell et al., 1981; Stewart et al., 2002).

Allozyme based mixed-stock analysis (MSA) has been used extensively in Pacific salmonid fisheries to provide information to regulate fishery openings and closures, provide harvest benefits or meet conservation needs, address catch allocation and equity issues among user groups and between countries, provide data for in-season run-size updates, and investigate migration patterns and timing (Shaklee et al., 1999). This has been applied to pink salmon Oncorhynchus gorbuscha (Walbaum) (Beacham et al., 1985a), sockeye salmon O. nerka (Walbaum) (Grant et al., 1980; Wood et al., 1989), chum salmon O. keta (Walbaum) (Beacham et al., 1985b; Shaklee et al. 1990b), chinook salmon O. tshawytscha (Walbaum) (Brodziak et al., 1992). In contrast, its use with regard to Atlantic salmon has been more limited, in part because Atlantic salmon show generally lower levels of protein polymorphism and spatial differentiation than Pacific species (Allendorf & Utter, 1979; Gyllensten, 1985). Also, unlike Pacific salmon, there was no co-ordinated programme developed early on to provide the required baseline data. Protein markers discriminate continental stock groups of Atlantic salmon, but technical problems getting fresh or frozen tissue samples and the slow development of baseline data on North American and European stocks limited its application to the West Greenland fishery. In contrast, DNA variation diagnostic of continent of origin identified (Bermingham et al., 1991; McConnell et al., 1995; Taggart et al., 1995; King et al., 2000, 2001; Koljonen et al., 2002) is now being applied using ethanol stored tissue samples and has allowed the rapid development of sufficient baseline data (T. King, pers. comm.). However, in the Baltic Sea, protein analyses have provided a practical link between the genetic stock structure information and fisheries management (Koljonen, 2001). It has also been used successfully to demonstrate the failure of stocking of foreign fish in northern Spain, leading to the ban of the stocking in the early 1990s (García de Leániz et al., 1989; Verspoor & García de Leániz, 1997).

CONCLUSION

Protein studies have made a valuable contribution over the last 40 years to our understanding of the Atlantic salmon. Their biggest contribution has been to provide the first genetic evidence that the species was highly structured into populations among and within rivers. The work has had a major impact on management of the species, particularly in underpinning the development of more population centred approaches and in helping to understand the potential genetic implications of population reductions, stocking and farm escapes. This evidence has for the most part only been reinforced and embellished over the last
decade by results from DNA studies. However, the future use of protein electrophoresis is likely to become increasingly limited as direct analysis of DNA, based on PCR (polymerase chain reaction) methods (Avise, 1994), is generally more cost-effective and opens up a way to analyse the entire genome. With the development of DNA methods, large numbers of loci with high allelic diversity can be targeted, giving greater discriminatory power and allowing individual genetic relatedness to be assessed (Beaumont, 2001). This should allow many aspects of the genetics of Atlantic salmon populations, such as defining the spatial boundaries of salmon populations within river systems, to be addressed for the first time. Some of the advantages of DNA analysis have recently been demonstrated by McGinnity et al. (2003) in their experimental field study of the impacts of farm escapes on wild populations using DNA markers to discriminate families.

Though protein electrophoresis itself has been largely superseded, data on protein variation are likely to continue to provide a valuable source of genetic insight and to serve as an invaluable framework for guiding DNA based work. Existing data sets will continue to be important for the formulation and testing of new hypotheses as additional approaches to population analysis are developed. Extending work with DNA based markers for protein polymorphisms (McMeel et al., 2001) will be particularly valuable for known phylogenetic markers (e.g. EST-D*) or selected loci (e.g. MEP-2*). In many respects, it will be more productive to extend and expand existing protein data bases than to begin new ones. Furthermore, given the general scarcity of archival biological material such as scales, developing DNA methods of genotyping protein polymorphisms will allow the unique historical advantages of the baseline protein data to be exploited in the future. This will allow investigations of temporal genetic change in Atlantic salmon populations over timeframes and geographical scales which could not easily be achieved, if at all, using other more recently identified DNA polymorphisms. The long time series of data would give a more powerful tool for assessing long-term genetic change, as evidenced by the work of Møller (this volume), particularly in response to factors such as global climate change and declines in abundance. Protein data will continue to be the main source of genetic insight into a 30 year period when Atlantic salmon populations have been seriously affected by a variety of adverse factors. For some now extinct populations protein data offer the only genetic insight.

This review is a part of the EU funded SALGEN project (Contract Q5AM-2001-00200). The manuscript has been much improved by the input from the referees’ and discussions with A Youngson. The authors are also grateful to L. Stradmeyer and A. Youngson for their editing of the manuscript.

References


Johnson, K. R. & Wright, J. E., Jr. (1986). Female brown trout × Atlantic salmon hybrids produce gynogens and triploids when backcrossed to male Atlantic salmon. Aquaculture 57, 345–358.


Lewontin, R. C. & Hubby, J. L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of


