Experiment 5.2
To investigate the accuracy of estimates on an inter-drainage basis. sites within rivers were pooled to calculate baseline frequencies between rivers. Simulations were run with and without the Dyfi and Rhyd-hir stocks which are represented by relatively small sample sizes.

Results

	MIX	With 4 stocks n = 20	With 6 stocks n = 20
Teifi	100%	0.9999	0.5426
Wye		0.0001	0.0000
Usk		0.0000	0.0000
Dee		0.0000	0.0000
Dyfi		not included	0.0002
Rhyd-hir		not included	0.4572

Comment
When only four stocks are included. the GSI estimates are highly accurate. However. the addition of the Rhyd-hir to the baseline stocks generates massive error. This illustrates the potential errors involved in attempting to use a single locus as a diagnostic marker even when resampling techniques are employed.

Genetic Stock Identification (GSI) in Atlantic Salmon

Dr. Caroline Thompson & Professor Andrew Ferguson

Research Contractor
School of Biology and Biochemistry
Queen's University, Belfast

Environment Agency
Rio House
Waterside Drive
Aztec West
Bristol
BS12 4UD

R&D Technical Report W90

Environment Agency
Rio House
Waterside Drive
Aztec West
Bristol
BS12 4UD Tel: 01454 24400 Fax: 01454 624409

Dissemination Status
Internal: Released to Regions
External: Released to Public Domain

Statement of Use
This Technical Report assesses the feasibility of applying modern genetic techniques to the discrimination of migratory salmonid stock components. It is intended for use by fisheries scientists to demonstrate the viability of such techniques to improve the basis of stock management.

Research Contractor
This document was produced under R&D Project W2007 by:

School of Biology and Biochemistry
Queen's University of Belfast
Medical Biology Centre
97 Lisburn Road
Belfast
BT9 7BL Tel: 01232 245133 Fax: 01232 314591

Environment Agency Project Manager
The Agency's Project Manager for R&D Project W2007 was:

Mr Peter Gough, Welsh Region

Additional Copies
Further copies of this document may be obtained internally from Regional R&D Management Support Officers and externally from Foundation for Water Research (Tel: 01628 891589 Fax: 01628 472711)

R&D Technical Report W90

CONTENTS

ACKNOWLEDGEMENTS

The authors would like to thank Michele Masuda for providing the GIRLSEM computer package, and for altering the program to accommodate the parameters of the data used here in modelling.
We would like to thank Stephen Clifford, Alan Pickering, John Taggart, and Eric Verspoor for providing unpublished, or restricted, genetic data for modelling.
We are also grateful to John Taggart for his advice and discussion on the modelling of DNA profiling.

1. EXECUTIVE SUMMARY

1. Genetic Stock Identification (GSI) involves the determination of individual population (stock) proportions in a mixed stock sample and is based on a mathematical comparison of the genetic composition of the mixed stock sample with baseline data for each of the contributing stocks. Until recently, studies of genetic variation in salmonid species have been based on indirect studies of protein variation as determined by gel electrophoresis. In Pacific salmonids, where there is a high level of protein variation, it is possible to carry out GSI successfully using protein markers. However, a low level of protein variation in Atlantic salmon has limited the application of GSI to mixed stock analysis in this species. The development of direct DNA-based techniques is now facilitating many new approaches to salmonid genetics using both nuclear and maternally inherited mitochondrial DNA (mtDNA). In particular, tandemly repetitive nuclear DNA regions known as microsatellites and minisatellites are highly variable and provide genetic markers particularly suited to GSI analysis.

2. Data from minisatellite, microsatellite and mitochondrial DNA studies of Atlantic salmon were obtained from the literature and unpublished sources. These comprised 15 studies (13 from Britain and Ireland) involving 55 drainages. Five data sets were used for modelling of GSI. These were selected as raw data were available either from more than one tributary of a river to model intra-drainage GSI, or from two or more rivers to model inter-drainage GSI. All models used minisatellite or minisatellite and mtDNA data as few complete microsatellite studies have been undertaken to date.

3. Data from minisatellite, microsatellite and mitochondrial DNA studies of potentially anadromous brown trout (sea trout) were obtained from the literature and unpublished sources. These comprised nine studies (five from Britain and Ireland) involving 44 drainages. Two appropriate data sets were used for modelling of GSI, with selection of data sets involving similar criteria to those used for Atlantic salmon.

4. Modelling demonstrated that the availability of novel highly variable minisatellite and microsatellite genetic markers now enables GSI to be undertaken successfully for Atlantic salmon and brown trout. Levels of accuracy (difference between estimated and real values) and precision (confidence limits on estimate), appropriate to management needs, can be obtained for both inter-drainage and intra-drainage stock mixtures providing appropriate baseline data sets (as outlined below) are available. GSI modelling showed that such stock composition estimates can be made with an accuracy within 10% of the actual value when populations showed typical levels of genetic divergence. The generally higher genetic differentiation among sea trout populations led to greater accuracy in GSI estimates involving this species.

5. The main limitation to undertaking GSI with salmonids is temporal variability among cohorts, within populations, in genetic make-up as a result of relatively low numbers of spawners (i.e. small effective population sizes) and overlapping generations. At present such temporal variation has not been fully defined for Atlantic salmon and brown trout as previous investigations have been restricted to a few cohorts. However, temporal stability can be circumvented by ensuring that all cohorts present in the mixed sample are present in the baseline data set. In some situations this could be estimated retrospectively by

sampling the progeny of equivalent adults to those in the mixed sample. Until such times as temporal stability has been fully established for particular stocks being used for GSI, baseline data sets need to be constantly updated for successive cohorts.

6. Baseline data sets with acceptable levels of sampling error can be obtained using samples of 75 juveniles from each stock. Sampling needs to be designed to ensure that these are obtained from throughout the spawning area of the stock unit and to minimise the sampling of siblings.

7. The minimum representation needed for each contributing population in a mixed stock sample was found to be 20. The presence in the mixed stock sample of fish from populations not included in the baseline data set did not cause significant error in accuracy until such extraneous fish comprised more than 25% of the sample. Elevated estimates tended to be limited to stocks most genetically similar to the extraneous fish, and the level of bias increased with the contribution of the unsampled stock.

8. More accurate and precise GSI estimates can be obtained by using gene loci with a greater number of alleles and, to a lesser extent, by increasing the number of loci. In most cases, use of five loci, each with ten or more alleles, gives estimates with appropriate levels of accuracy and precision. The modelling was undertaken using minisatellite markers, as no microsatellite data sets suitable for modelling inter- and intra-drainage systems are yet available. However, due to recent technical advances, future empirical studies would be undertaken using microsatellites, and it is anticipated that these would be equally, or more, appropriate for GSI analysis. Microsatellites have the advantage that non sacrificial sampling can be carried out with biopsy samples (e.g. adipose fin clips) being obtainable from the 0+ parr onwards.

9. Accuracy and precision of GSI estimates for Atlantic salmon based on minisatellite data are similar to, or greater than, those obtained in empirical GSI studies with Pacific salmonids. As the technique has been validated by independent physical tagging estimates for these latter species, no further non-genetic validation is required for Atlantic salmon and brown trout.

10. It is recommended that four empirical studies should be undertaken for both intra-drainage and inter-drainage situations for Atlantic salmon and sea trout. As each of these GSI trials would involve about 1,000 individuals, each would require about three months to carry out and cost (excluding field sampling) some £15,000 (at 1997 costs). Intra-drainage studies should be combined with DNA profiling of parent-offspring combinations to estimation location and spawning success of different sea age and run time components. Costs would be similar to those for GSI trials.

2. INTRODUCTION

2.1 Objectives of study

To assess the extent of geographical and temporal variability in currently available genetic markers for Atlantic salmon and anadromous brown trout (sea trout), and to determine, by modelling, the precision and accuracy with which genetic stock identification (GSI) could be undertaken at the catchment and sub-catchment levels in the two species, and to produce a technical report on this assessment.

This was to be approached as follows:

1. Compile a database, and produce a written summary, of published and unpublished minisatellite, microsatellite, mitochondrial DNA and MHC geographical and temporal variation in Atlantic salmon stocks in Europe and North America.
2. Compile a database, and produce a written summary, of published and unpublished minisatellite, microsatellite, and mitochondrial DNA geographical and temporal variation in sea trout populations in Europe.
3. In consultation with the project manager, define minimum acceptable levels of accuracy and precision of GSI relevant to management requirements on catchment and sub-catchment scales for both Atlantic salmon and sea trout. Determine minimum level of spatial genetic heterogeneity, and maximum permitted level of temporal variability, to achieve this precision and accuracy. Sampling error, which is not adequately considered in most Atlantic salmon and sea trout studies, requires particular attention in this respect.
4. On the basis of (3) determine the most appropriate marker, or combination of markers, for GSI in Atlantic salmon and sea trout.
5. Identify target catchments in England and Wales for genetic study of Atlantic salmon and sea trout on the basis of management and other requirements with advice and information from the Agency.
6. On the basis of (3), (4) and (5), identify typical catchments suitable for modelling GSI in Atlantic salmon. Model GSI in this catchment using different values of stock admixture, temporal variability, sampling error etc. Use bootstrap / jacknife estimates to determine confidence limits for various approaches.
7. On the basis of (3), (4) and (4), identify typical catchments suitable for modelling GSI in sea trout. Model GSI in this catchment using different values of stock admixture, temporal variability, sampling error etc. Use bootstrap / jacknife estimates to determine confidence limits for various approaches.
8. Determine on the basis of minisatellite / microsatellite data collected in (1), and by modelling, the extent to which DNA profiling ("fingerprinting") for parentage identification could be used to identify within-catchment origin of specific stock components, e.g. spring salmon.
9. Evaluate GSI in the context of management requirements for Atlantic salmon and sea trout.
10. Prepare outline PIDs detailing future approaches for Atlantic salmon and sea trout.

2.2 The purpose of genetic stock identity analysis in salmonids

Fisheries management and conservation are commonly hampered by the occurrence of mixed stock fisheries, either coastal harvesting, offshore nets, or in rivers through angling and commercial fishing. The nature of anadromy means that fish from different rivers migrate into mixed stock areas where they exist together, before returning to natal rivers to spawn.

While attempts can be made to manage the freshwater environment, for a great part of their lives anadromous fish are outwith the control of fisheries managers.

The need to be able to identify individuals from stocks through all stages of their life has traditionally been approached by physical markers. However, tagging by various artificial means (clipping, branding, or microtagging) incurs high costs such as increased mortality through handling, minimum size requirements for marking, loss of marks, and the limitation of the mark to the lifetime of the individual. The use of naturally occurring biological markers (e.g. morphology, scale patterns, meristic counts and parasitic infections) in a phenotypic approach to stock discrimination has also been of limited success, and application to Atlantic salmon has been confined to identifying the continent of origin.

Genetic stock identification (GSI) uses the natural genetic variability of a species to define stocks and assign individuals to them on a probabilistic basis determined by their genetic constitution. GSI may be used to assign proportions of a mixed stock fishery to certain rivers, or even to tributaries within a river. This may be important in terms of commercial factors, such as the buy out of fisheries, or in conservation objectives, being able to identify the most productive or important rivers or streams. GSI could also be used to focus on a component of the fishery, for example, to identify stock components and to manage their source appropriately.

In recent years there has been a growing awareness of the contribution that genetics can make to fisheries management of exploited species. Practices such as the inappropriate use of hatchery fish, and stocking geographically remote (and likely genetically distant) fish, whilst bolstering numbers, may lead to the breakdown of local co-adapted traits and gene complexes (e.g. Altukov & Salmenkova, 1987; Hindar, 1991; Donaghy & Verspoor, 1997). Unique phenotypes and genotypes may be lost by forced mixing or redistribution of geographically separate stocks. Genetic studies can be used to develop more propitious long-term management plans.

2.3 Development of genetic markers

The past three decades have seen an expansion in the number and type of genetic markers that can be applied in population studies. The first applicable genetic marker system developed was protein gel electrophoresis which for the first time allowed examination of variation at single gene loci. The products of protein coding loci are separated according to structural characteristics allowing the frequencies of variant alleles to be calculated for populations.

In recent years, the development of new techniques in molecular biology have made it possible to examine variation in the DNA itself with sufficient ease to make population studies, involving substantial sample sizes, viable. As DNA elements evolve at different rates (for example mitochondrial genes evolve faster than their nuclear counterparts, repetitive DNA evolves faster than non repetitive) it is possible to select appropriate markers, and the small amount of tissue often required makes non-destructive sampling possible. Of particular interest in fisheries is the development of amplification techniques that have made it possible to use dried scale samples for genetic analysis.

Most population genetic studies have used one of two types of DNA marker, a) mitochondrial DNA restriction fragment length polymorphism or b) variable number tandem repeat (VNTR) DNA.

a) The mitochondrial genome is inherited maternally. Several mitochondrial genes can be amplified by the polymerase chain reaction (PCR). Polymorphisms in these regions are detected due the occurrence of a restriction site producing a characteristic DNA fragment pattern. Polymorphisms occur through base changes resulting in the addition or loss of such sites.

b) The detection of variable number of tandem repeats in the nuclear genome has allowed many new developments in genetic marker systems. VNTR loci consist of concatenated repeats of nucleotide sequences, generally less than 5 base pairs (bp) long in microsatellites, and generally 10 - 100 bp in minisatellites. These loci are highly variable with commonly 20 alleles or more, and heterozygosity levels typically several times larger than protein coding loci.

2.4 Genetic principles underlying GSI

The success of GSI depends on the species and geographical location involved meeting the following criteria.

2.4.1 The mixed stock sample is made up of two or more stocks, each of which are identified as being likely contributors. Within those stocks individuals mate at random, and thus the population conforms to Hardy-Weinberg expectations of genotype proportions. The stock unit, or population, is difficult to define. Some species consist of a series of marginally overlapping units, while others may be rigidly defined by areas that are suitable for reproduction. While salmonids may tend towards the former during juvenile stages, environmental conditions and requirements may delimit spawning units.

2.4.2 All individuals in the mixed stock come from the sampled baseline stocks. From genetic population studies and from non-genetic tagging and radiotagging, it is known that reproductive isolation in freshwater is not complete. While homing may be generally accurate, a proportion of fish stray into non-natal streams, and may successfully spawn. Genetic divergence generally increases with geographical distance, suggesting that the level of gene flow is higher between geographically proximate stocks. Straying fish may effect the GSI estimate by a) being present in the baseline samples or b) being present in the mixture sample.

a) If the baseline survey is based on large enough numbers of juveniles, and done so as not to over-sample single families, then straying fish are unlikely to cause significant errors in the baseline allele frequencies.

b) It is probable that non-native fish will be found in any mixed stock sample either in a river, or in coastal fisheries. Sea trout in particular are known to make excursions into non-natal rivers, for periods of time not necessarily associated with spawning (Berg & Berg, 1987). Single tagged salmon, from the Dee, Wales, and Iceland have been found in the River Don (A. Youngson, pers. comm.). While mathematical models suggest that such a rate of immigration can have a homogenising effect on small populations over the course of

years, considering the observed heterogeneity of salmonid populations suggests this is not so for these species. Using standard GSI techniques, individuals that have alleles not found in any baseline stock cannot be analysed. However, alternative analyses using the unconditional likelihood method allows such individuals to be assigned to unsampled baseline stocks, and assumes novel alleles to occur at low frequencies in the sampled baseline stocks (Smouse *et al.*. 1990).

2.4.3 Contributing baseline stocks are genetically different from each other, and that genetic differentiation can be measured. Sampling error will arise when more than one discreet stock is included. However, this can be detected as genotype proportions will no longer conform to Hardy-Weinberg expectations for a randomly mating population. Salmonid populations show a hierarchical distribution of variation, and genetic distance between stocks increases with geographical distance. Thus stocks within a river are generally more genetically similar than stocks from different rivers, and adjacent rivers are more similar than remote rivers. If this is so, then sampling could also be carried out on a hierarchical basis depending on the level of sensitivity required. For example, if the analysis is to determine the river of origin, then the river may be typified from sampling from a restricted number of locations. If however, the catch is to be assigned to a location within a river, then more thorough sampling of baseline stocks would be required.

2.4.4 Genetic data about the baseline stocks and the mixed stock sample is reliable, and that sampling error can be minimised. As with all surveys of natural populations where siblings may be concentrated around the redd site, care should be taken to avoid over sampling from families. An over representation from few families may substantially effect allele frequencies (Allendorf & Phelps, 1981; Webb *et* al., 1993; Hansen *et al.*, 1997). Field sampling strategies can be statistically worked out with some prior knowledge of the distribution of the species and topology of the sampling area.

2.4.5 Where baseline data are not available for each cohort represented in the mixed stock sample, genetic markers must be temporally stable in baseline stocks. If each cohort of the mixed stock sample cannot be represented in the baseline data, allele frequencies must be stable in each cohort so the baseline is fully representative. Stability of allele frequencies is required for GSI as in most situations baseline stocks will consist of several age classes, as might the mixture sample. Jordan (1992) in a review of isozyme variation in salmon found genetic differences between year classes to be less significant than variation between locations within or among rivers. Similar findings have been demonstrated in several isozyme studies of salmonids (e.g. Ferguson & Fleming, 1983; Crozier & Moffett, 1989; Hurrell & Price, 1993; Hansen *et al.*, 1993) and more recently for VNTR studies (Taylor & Beacham, 1994; Galvin *et al.*. 1996). However, temporal stability cannot be assumed, especially for highly variable VNTR markers where the intermittent sampling of rare alleles could be mistaken for sampling error. Where GSI is carried out on individual cohorts, i.e. the baseline and mixed stock sample are from the same cohort(s), temporal stability is not required.

2.5 Procedures for applying GSI

The simplest situation is a mixture involving two or more groups that do not share alleles at a locus. Such fixation of alleles allows individuals to be assigned to a specific group. This rarely occurs among individuals of a single species, and is more typical of variation between

species. However, Taggart *et al.* (1995b) found almost complete fixation of alternate minisatellite alleles in Atlantic salmon populations from North America and Europe. Sympatric freshwater brown trout stocks have also been detected with alternate fixed alleles at a protein coding locus (Ryman *et al.*, 1979). While such polymorphisms allow individuals to be assigned to one or other stock, this is not commonly seen in regional or inter-drainage genetic variation. GSI can assign proportions of a mixture to stocks, but without such diagnostic alleles cannot assign individuals.

A more typical situation is the occurrence of alleles at different frequencies in populations. For example if two stocks contribute to a mixture sample, and have the following allele frequencies at a locus:

Stock 1	A = 0.8	B = 0.2
Stock 2	A = 0.2	B = 0.8

and the allele frequency in the mixture is:

Mixture	A = 0.5	B = 0.5

then assuming these are the only two stocks in the mixture, the contribution of Stock 1 (f_1) can be found from the frequency of the allele A in the mixture (p_m) and the frequency of the allele A in Stock 1 (p_1) and Stock 2 (p_2) by the expression

$$f_1 = (p_m - p_2) / (p_1 - p_2)$$

Therefore the contribution of Stock 1 is 50% (0.5), and if only two stocks are involved, the contribution of Stock 2 is also 50%.

This is a very simplified example making several assumptions about stocks. Firstly, there are few situations when it is certain only two stocks are involved. In most fisheries there is at least the potential for several stocks to contribute to the catch. In many situations the number of stocks contributing is unknown. Secondly, allele frequency differences between salmonid stocks are rarely of that magnitude. While it is possible that geographically distinct stocks do have frequencies differing by 0.6, there will be many other stocks with allele frequencies ranging in-between. While no single locus will show such large allele frequency differences, a combination of loci will give a better estimate that any one alone. With additional stocks and loci, the calculation of estimates becomes a complex mathematical problem. Lastly, the allele frequencies of the stocks and the mixture are taken as absolute. In a realistic situation, baseline stock data is gathered from a sample of the population which is subject to sampling error. Likewise, while the GSI estimate may be absolute for those individuals used in the mixture, that sample may not adequately represent the fishery catch. Therefore sampling error for each estimate must also be calculated.

The success of GSI can be defined in terms of the accuracy of the estimate and the precision or variance of an accurate estimate (Pella & Milner, 1987). Accuracy - the difference between the estimated value of contribution and the true value. This can be particularly effected by the choice of baseline stocks, i.e. if too many non-contributing stocks are used, or a contributing stock is excluded, then the estimates can be erroneous. Accuracy also depends on the level of differentiation between stocks. The contribution of stocks that are not sufficiently different from each other may not be accurately estimated. Precision - the variance associated with the estimate (standard deviation or 95% confidence limits). The precision of an estimate depends

on the sampling error of the baseline stocks, i.e. the baseline allele frequencies must typify the actual population. The sample size of the mixture and the number of individuals from each contributing stock also effects the precision.

2.6 Application of GSI to Pacific salmonids

To date the majority of GSI studies have been on the species of the genus *Oncorhynchus*, and primarily using isozyme loci. Two main applications have emerged, firstly to protect over exploited populations, and secondly, to allocate fisheries among different users and resource managers. The technique has been widely applied in stock composition estimates including aspects of seasonal variation (Milner *et al., 1985;* Beacham *et al.*, 1985, 1987; Seeb *et al.*, 1986; Shaklee *et al.*, 1991). Biological markers (parasites) have also been combined with genetic data (Wood *et al.*, 1989). The accuracy of GSI analysis was tested using coded wire tagged fish (Brodziak *et al.*, 1992). Using 35 polymorphic isozyme loci, and 37 baseline stocks in a blind test, accurate estimates were made for 220 tagged fish from five major stock groups. Between one and three minisatellite loci have also been used in GSI analysis of oncorhynchids (Taylor & Beacham, 1994; Beacham *et al.*, 1995; Miller *et al.*, 1996). Although minisatellite loci were used in these studies, most loci produced complex patterns of more than two alleles in an individual, and analysis techniques were highly modified for these data.

2.7 Application of GSI to Atlantic salmon

Three studies on GSI analysis of Atlantic salmon have been published to date, two using isozyme loci and one minisatellite data. The first paper (Koljonen. 1995) examined catches off the Finnish west coast. Stocks here consist of two groups: 'local' stocks that migrate and remain in coastal waters; and, stocks from five rivers in the north of Finland that move into the main Baltic Sea basin. While the majority of rivers are maintained by hatchery stocking, two of the northern group are essentially natural. Four polymorphic isozyme loci were examined in approximately 2200 individuals in the baseline samples. Catch samples were taken from commercial fisheries for a 17 day period in April / May. It was found that northern stocks constituted over half the catch. While one of the northern stocks appeared to dominate, the five rivers were so genetically similar estimates could not be relied upon. Swedish rivers were then included in the baseline, but despite 10% of fish caught being tagged Swedish salmon. no contribution was detected using GSI .

Studies in the Baltic were extended (Koljonen & McKinnell, 1996) by including 16 Swedish, Finnish. and Russian rivers in the baseline. Seven isozyme polymorphic loci were analysed in approximately 4200 baseline stock salmon, in which five of the loci had two alleles. Catch samples were taken from commercial fisheries during January - June. and from September - November. Clear seasonal differences were observed with the Russian stock contribution varying between 0% - 57%. Standard deviations greater than 10% occurred in estimates for the most genetically similar stocks. However, the total catch proportion based on GSI were realistic when compared with known smolt production proportions for stock groupings. Individual stock contributions of less than 40 fish were subject to high levels of error.

Atlantic salmon show little variation at isozyme loci, and studies have been restricted to a small number of polymorphic loci. To date the only GSI study of Atlantic salmon with minisatellite loci has been carried out on the Shannon in western Ireland (Galvin *et al.*, 1995). The Shannon covers a catchment area of approximately 10,000 km^2. and has been greatly

effected by a hydroelectric scheme for many decades. Baseline data were obtained by sampling approximately 100 juveniles from each of nine tributaries and hatcheries. Mixture samples were taken from 250 ascending adult salmon at the hydroelectric dam trap. Non-destructive sampling techniques were used. Fifty adults were judged by length to be multi-sea-winter salmon. All individuals were screened at the minisatellite loci *Ssa*-A45/1, *Ssa*-A45/2 and *Str*-A9 (Taggart *et al.*, 1995a; Prodöhl *et al.*, 1995) detecting nine, ten and eight alleles respectively. In addition baseline samples were screened at eight allozyme loci. Simulations were done to assess the precision and accuracy of GSI estimates. It was found that the level of error generated using allozyme loci was consistently higher than with minisatellite data. This was maintained throughout the range of sample sizes, although for both sets of data error decreased as the number of individuals contributed from each stock increased from 5 to 100. Much lower bias was found using minisatellite data. It was estimated that approximately one-third of grilse, and half the salmon in the mixture sample originated from hatchery stocks. The remainder were predominantly from tributaries in the lower stretches of the Shannon with the exception a grilse producing tributary in the upper river.

2.8 DNA profiling to identify parent, or parents, of an individual

The high levels of polymorphism observed at VNTR loci make them suitable for determining pedigree and the relationship between individuals. The high numbers of alleles segregating at each locus, and the corresponding high numbers of genotypes makes these markers particularly suitable for identifying the progeny of individuals in naturally mating populations. This may be utilised in determining the intra-drainage origins of important stock components, e.g. spring fish, based on family and / or individual identification rather than total stock identification in the usual context of GSI. Every individual salmon (as with every human other than identical twins) can be defined by its genotypic profile at a number of VNTR loci, thus allowing offspring to be identified to one or more parents whose genotype is known. This has been shown to work in salmon in identifying the parents of individuals from a mixed progeny sample from 60-80 families when the DNA profiles of both parents are known (Ferguson *et al.*, 1995). The technique can be extended where only one parent is known by increasing the number of loci used (J. Taggart, pers. comm.). Where the stock component is predominantly female, as in the case of spring salmon or sea trout, the technique would identify the female parent. This is particularly important considering the often significant contribution by precocious parr (e.g. Jordan *et al.*, 1992; Morán *et al.*, 1996). As the approach is based on the DNA profiling of individuals, rather than stocks, it is not subject to inaccuracy caused by sampling error of baseline data, straying, or temporal instability.

The technique assumes that adult trapping systems, or other catch and release facilities are available to non-destructively sample ascending adults. Subsequent samples of ova/fry/parr, smolts or adults can then be related to specific parents. This approach is particularly pertinent to investigating the spatial and temporal segregation of specific year class stock components e.g. spring salmon, or investigating the distribution of sea trout redds.

3. MATERIALS AND METHODS

3.1 Data sets for modelling

The GSI technique makes two important assumptions about genetic data. Firstly, the populations should be in Hardy-Weinberg equilibrium. This is necessary to ensure that the sample can be adequately described by allele frequencies. Secondly, that all baseline stocks contributing to the mixture sample of unknown origin, or MIX, are sampled. To ensure these conditions were satisfied, only data sets where the genotypes of individuals were available were used. Allele frequencies were calculated from those individuals, and the samples confirmed to be in HWE. Individual genotypes could then be used as components of the MIX sample. Data for all loci had to be available for an individual in a MIX sample. Thus all stocks contributing to the MIX are by definition sampled. However, in some of the experimental models, individuals from a population not sampled as a baseline stock were included in the MIX, to investigate the effect of 'straying' fish on GSI estimates.

3.1.1 Considerations for selecting data sets for modelling

To select data sets for GSI modelling, a number of theoretical and practical considerations had to be made.

1. One of the purposes of modelling data was to estimate the level of divergence that is required between baseline stocks in order to produce reliable GSI estimates. For example if the level of genetic differentiation required is typical of that found between sea trout from different regions, it is unlikely that an intra-drainage study will produce accurate estimates. GSI studies on Pacific salmonids assume significant genetic differentiation of populations (e.g. Wood, 1989; Wood *et al.*, 1989; Smouse *et al.*, 1990, Marshall *et al.*, 1991). However, it was intended to also investigate models where there was no, or minimal, differentiation as a 'worst case scenario'.

2. Size of the data set. It is unlikely that many fisheries of interest exploit only one or two stocks, whether these are within a drainage, or from several. Therefore it was desirable to select data sets with several baseline stocks. Also with several stocks, MIX samples could be manipulated to be more or less skewed. However, limitations of computing power and time would preclude the frequent modelling of large numbers of populations with several loci screened.

For the minisatellite data in particular, while many laboratories have examined variation in the same loci, there is very little inter-laboratory sample exchange, or side-by-side gel comparisons. Resolution of alleles also depends on the electrophoretic conditions, which may vary between laboratories. This has also been a consideration in allozyme studies, but the much higher levels of polymorphism found in DNA based techniques exaggerates this effect. In essence every laboratory is scoring variation according to their own system, and generally comparisons between published data sets from different laboratories / individuals cannot be made (the same is also true of mitochondrial DNA studies). Without deliberate standardisation of techniques and direct contrasts made between each polymorphism found in different studies, only conservative comparisons can be made. Due to the small size of microsatellite alleles (which can be accurately sized), and the more automated techniques that are evolving, this will not be an important consideration in the future. However, at present this limits the data bases to the work of individuals / laboratories.

3. Type of polymorphism. Although mitochondrial, major histocompatibility complex (MHC), and microsatellite variation is reported, locus specific minisatellite screening was used in all but one of the experimental data sets. This was due primarily to the suitability of the populations, and the amount of data available. No comparable data set is available yet for microsatellites. Literature reporting preliminary investigations using microsatellites suggest that both the numbers of alleles at microsatellite loci, and heterozygosity levels are similar to the values typical of minisatellite population studies of salmonids. It can therefore be assumed that the conclusions drawn from minisatellite data will also be borne out by microsatellite data. Mitochondrial DNA data has also been used for modelling. Mitochondrial DNA is inherited maternally in a clonal fashion, and individuals inherit only one variant or haplotype from the female parent. (It should be noted that many GSI studies of Pacific salmonids also include non genetic data e.g. morphometric / meristic characters, scale patterns, parasitic infection, or any character that can be digitally encoded.)

4. Other consideration of the polymorphisms used are the number of loci examined and the levels of polymorphism at those loci. One of the questions of interest is whether using a large number of loci provides more accurate estimates. GSI studies of Pacific salmon commonly use 10 - 30 allozyme loci. However, the majority of these loci are not highly polymorphic and have less than four alleles over the whole range of the species. This leads to the question of whether loci with higher levels of polymorphism are more useful that moderately polymorphic loci. In order to investigate these questions, it is desirable that data sets use more than three loci, and that the loci range from moderately to highly polymorphic.

5. Evidence for temporal stability. Although nearly complete temporal stability has been repeatedly demonstrated for allozyme data in salmon and trout (e.g. Crozier & Moffett, 1989; Ferguson & Fleming 1983, Morán *et al.*. 1995) and for VNTR variation (Taylor & Beacham, 1994), these studies have examined periods of less than five years, and temporal stability cannot be universally assumed. In order to investigate the effects of temporal variation, it is desirable that data from multiple years of sampling are included in the baseline. However, these data are not generally available and the majority of modelling data sets use baseline stocks of mixed cohorts. Temporal instability is associated with significant changes in the effective population size, therefore may be more apparent in numerically unstable populations.

2.1.2 Data sets for Experiments 1 & 2

These come from an unpublished study (C. Thompson, 1995). Experiment 1 used variation at six minisatellite loci in a total of 284 juveniles from tributaries of anadromous sea trout from the Aberdeenshire Dee and Don. Individuals from a river on the Scottish north-west coast (Laxford) were used to simulate 'straying' i.e. fish in the MIX sample that do not come from a sampled baseline stock. For the larger model (Experiment 2) these samples were pooled to form the Dee and Don baseline stocks. Baseline stocks (juvenile samples) from the Tweed (made up from a number of tributaries in the lower river) and the Spey were added to make a four stock baseline covering approximately 370km of the Scottish east coast. Mitochondrial DNA data were also added to this model.

2.1.3 Data sets for Experiment 3

This extensive data set comes from an unpublished ongoing study into life history fitness of Atlantic salmon in the Aberdeenshire Dee (John Taggart). Trapping facilities are seasonally

operated in the Girnock and Baddoch tributaries. Downstream migrating smolts are tagged and a number of these return as adults moving upstream to spawn which can be non-destructively sampled. The Girnock is unique in that redd marking and egg sampling is carried out above the trap. Thus adults and their progeny are sampled in a randomly mating natural population.

For the purposes of GSI modelling, baseline stock frequencies were calculated from adult samples from each year, and the progeny of that year was taken as the MIX sample. This provides a highly realistic method of modelling the MIX sample. Genes in the adults baseline frequency are mixed in random mating events, and individuals may have genotypes not found in the adult sample. This data set allowed two years of comparisons like this. Due to the large data set random subsamples of the progeny groups had to be made.

3.1.4 Data sets for Experiment 4

This experiment combines data from the Dee sites above with published data from the Windermere catchment (S.E. Hartley & A.D. Pickering, 1994). Data from three minisatellite loci could be combined to form a data set with 850 juveniles in the baseline stocks.

3.1.5 Data sets for Experiment 5

Published mitochondrial DNA data (M. O'Connell *et al.*, 1995: 1996) was used to model the effectiveness of using mtDNA as a single marker in inter- and intra- drainage GSI studies. Juveniles were sampled throughout.

3.1.6 Data sets for Experiment 6

This experiment uses unpublished data (Eric Verspoor, 1996) from a minisatellite and mtDNA study of the Tweed. Juvenile samples from three tributaries (Ettrick, Upper River, and Gala) were screened at four minisatellite loci and at the mitochondrial DNA ND1 region.

3.1.7 Data sets for Experiment 7

This experiment uses unpublished data (S. Clifford, 1996). Juvenile samples from several rivers from Co. Donegal (Ireland) were screened at minisatellite loci. Several populations shown to have been effected by fish farm escapes were also screened, but were not used in this model.

3.2 Statistical analysis

Basic genetic information such as allele frequencies, heterozygosity values and F statistics, were calculated using GENEPOP 2.0 program (Raymond & Rousset, 1995). There are four recognised GSI analysis packages world-wide, all originated for the analysis of Pacific salmonids. GIRLSEM was obtained from its authors (M. Masuda *et al.*, 1991) and used throughout this study. GIRLSEM is the most commonly used package in published papers on GSI, and is one of the two PC based packages. GIRLSEM is a FORTRAN-77 coded program, and its dimensions were redefined by the authors to allow for the large number of alleles and genotypes used in this modelling study. The program is based in the work of Pella and Milner (1987) by which a conditional maximum likelihood estimate (CMLE) of stock composition can be computed.

Baseline stocks information is given in the form of allele frequencies and sample sizes. The program finds all possible genotype combinations from the allele frequencies using Hardy-

Weinberg expectations. The MIX sample is input as individual records of genotypic data. The program initially uses the expectation maximisation (EM) algorithm to determine initial estimates of stock composition that are highly supported. As the CMLE is approached, a second algorithm, iteratively reweighted least square (IRLS), continues to search for the best fit estimate to the accuracy and tolerance limits within the program (Pella *et al.*, 1996). Bootstrapping is a method of numerical resampling by replacement. The program can bootstrap the baseline frequencies and the MIX sample, creating empirical distributions of composition estimates that may result from sampling errors for the stocks and MIX.

The same conditions were used in all experiments. The search algorithms were allowed to cycle up to 500 times to find a CMLE. One hundred bootstraps were used to estimate the 95% confidence limits. All other parameters were set by default. If an individual is found in the MIX sample that cannot have come from any of the baseline stocks, no attempt is made to analyse it. This can arise when an individual has an allele not present in the baseline stocks, or if it is heterozygous with one allele being unique to one baseline stock, and a second allele unique to another. If the genotype cannot be accommodated by permutations of the alleles given in the baseline stock, it cannot be assigned.

3.3 DNA profiling

The ability to identify a parent of a fish from a mixed sample of adults, depends on the DNA profile of that adult being unique within the drainage or tributary screened. The likelihood of individuals sharing the same genotype depends on the frequency of the most common allele. Using the frequency of the most common allele, the proportion of offspring with that allele is calculated from Hardy-Weinberg expectations. This is repeated for several loci and the probabilities combined. The number of individuals with the same genotype was modelled using the frequency of the most common allele at one to ten loci. Statistics were modelled in an Excel spreadsheet.

4. RESULTS OF GSI SIMULATIONS

This section details the results of GSI simulations using various data sets derived from published or unpublished studies of Atlantic salmon or sea trout. The data sets are described in terms of sample size, number of alleles / haplotypes, and heterozygosity. Fst values, indicating the level of genetic divergence between samples, are also given. Those stocks with the lowest Fst value in the matrix, are therefore the most genetically similar to each other. Those with Fst values over 0.05 are deemed significantly different from each other at the 5% significance level. Each simulation, or experiment, is described under a sub-heading. The aims of each experiment are described, and the results commented upon.

The actual composition of the mixture sample (MIX), constructed using the genotypes of the baseline stocks in varying proportions, is given in percentage form. The estimates produced by the simulations are given as a proportion. The two values are given beside each other to allow comparison. Mean standard deviation (SD) is the average SD for that estimate. Bootstrapped estimates are calculated using 100 re sampled sets, and the mean and 95% confidence limits calculated from those estimates.

4.1 Experiment 1. Assessment of intra-drainage genetic variation (minisatellite DNA) in sea trout and its use in GSI

Data from the screening of juveniles from three anadromous brown trout tributaries of the Aberdeenshire Dee (Beltie, Lumphanan and Sheeoch) and three from Don (Buchat, Fintray and Leochel) were used to model intra-drainage GSI estimates. Although the sites are from two rivers, the mouths of the Dee and Don are less than 3km apart and all samples came from sea trout spawning areas. The level of genetic divergence among samples are typical of intra-drainage systems. Hereafter they will be referred to as sites 1 - 6. Samples from a river on the west coast of Scotland were used in the model as 'strayers', i.e. the sample was not among the baseline stocks, but individuals were added to the MIX in various simulations. The samples were screened at six minisatellite loci. Numbers of alleles per locus, heterozygosities and pairwise Fst values are given. As can be seen from the table the numbers of alleles at each locus, and the average heterozygosities of the sample are approximately equal, indicating a similar amount of variation in each sample. Of the twenty-one pairwise Fst comparisons, four are less than 0.010 indicating a higher level of gene flow between these populations. The lowest pairwise values are found between tributaries of the same river. The sample of 'strayers' is most genetically similar to Site 4 and Site 1. Over all samples the Fst is significantly different from zero indicating significant population structuring.

Sample size, number of alleles per locus and average heterozygosity (H_{avg}) of samples used

	Sample size[a]	*Str*-A3	*Str*-A5	*Str*-A22/1	*Str*-A22/2	*Ssa*-A45/1	*Ssa*-A45/2	Havg
Site 1	43-48	10	6	16	6	13	4	0.704
Site 2	43-49	11	7	16	4	9	4	0.664
Site 3	46-48	11	7	13	6	9	3	0.694
Site 4	45-41	10	7	12	5	11	6	0.727
Site 5	45-44	10	7	12	4	12	6	0.732
Site 6	41-44	12	7	14	6	12	2	0.715

[a] Variation in sample size is due to different number at each loci.

Pairwise Fst values

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Site 2	0.009	—				
Site 3	0.007	0.022	—			
Site 4	0.018	0.037	0.006	—		
Site 5	0.022	0.036	0.022	0.019	—	
Site 6	0.015	0.029	0.017	0.006	0.019	—
('Strayers'	0.028	0.031	0.031	0.026	0.039	0.035)

Fst over sites 1 - 6 = 0.019 (95% CI 0.012 - 0.029)

Experiment 1.1

The aim of this experiment was a) to obtain stock composition estimates under standard conditions, and b) to investigate the effect of the size of the MIX sample on the precision of the estimate. Equal proportions (16.7%) from each stock were entered into the MIX. Firstly a MIX of n = 60 was constructed using 10 individuals picked at random from each stock. The simulation was rerun with a MIX of n = 180 using 30 from each stock.

Results

	MIX	n =60	n =180
Site 1	16.7%	0.1608	0.1246
Site 2	16.7%	0.2025	0.2076
Site 3	16.7%	0.1729	0.1499
Site 4	16.7%	0.1457	0.1759
Site 5	16.7%	0.1515	0.1797
Site 6	16.7%	0.1665	0.1623
Mean SD		0.0683	0.0380

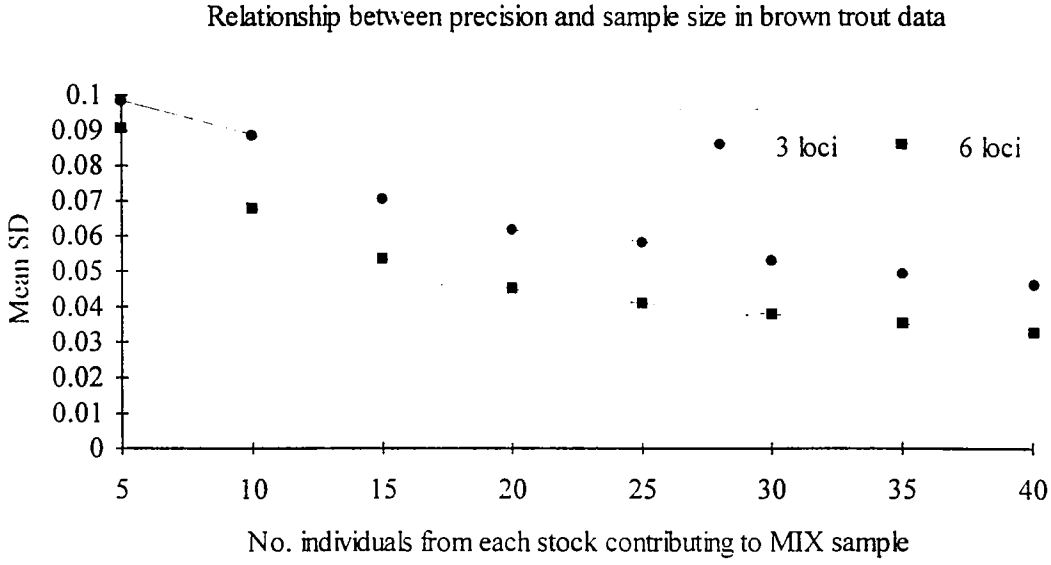
Comment

The effects of a larger MIX size on the stock composition estimate was negligible with an average error of less than 1.5% rising to 2.1% with the larger MIX size. The variance of the estimates however. decreased.

Experiment 1.2

The aim of this experiment was to a) further investigate the effects of MIX sample size on the precision of stock composition estimates, and b) to investigate the effect of using a smaller number of loci. Simulations were run using equal contributions from each stock of 5 to 40 individuals and mean SD of each stcok compositon estimate was plotted. The simulations were rerun using the three more polymorphic loci *Str*-A3 *Str*-A5, *Str*-A22/1, and the SD again plotted.

Results



Comment

From the graph it appears that the precision of the stock composition estimate (as measured by the mean SD) does depend in part on the sample size. Mean SD decreased rapidly as the sample size rose to 20, but after this appeared to change little. Mean SD was consistently higher when only three loci were screened, but followed a similar pattern to the six-loci simulations. It would appear therefore, that 20 or more individuals from each of the contributing stocks, irrespective of what proportion they contribute, are needed in the MIX to optimise the estimate precision. The number of loci screened also appears to contribute to precision, although not to the same extent.

Experiment 1.3
The aim was to investigate the effects of MIX size further, in particular, the effects of MIX size when only one stock contributed. 100% of the MIX was taken from randomly picked individuals from Site 6. The simulation was run with a MIX size n = 10, and again n = 40.

Results

	MIX	n = 10	n = 40
Site 1	0%	0.0000	0.0000
Site 2	0%	0.0000	0.0000
Site 3	0%	0.0000	0.0265
Site 4	0%	0.0000	0.0000
Site 5	0%	0.0000	0.0012
Site 6	100%	0.9999	0.9723
Mean SD		0.0000	0.0012

Comment
In this case the smaller MIX size produced the correct result. The larger MIX size estimate was less accurate and had a slightly higher variance. This is due to the inclusion in the larger data set of fish that cannot be accurately assigned. in this instance alone. The majority of the incorrectly assigned fish were assigned to Site 3, even though the pairwise Fst between these sites was the second highest comparison.

Experiment 1.4

The aim of this experiment was to investigate the effects of using highly polymorphic, or less polymorphic loci in screening samples. Simulations were run with all stocks making equal contributions to the MIX (n = 60) using data from 'less polymorphic' and 'more polymorphic' loci. The level of polymorphism was determined from the numbers of alleles detected. Therefore, *Str*-A3 (15 alleles), *Str*-A22/1 (19 alleles) and *Ssa*-A45/1 (14 alleles) were defined as being more polymorphic than *Str*-A5 (7 alleles), *Str*-A22/2 (6 alleles) and *Ssa*-A45/2 (7 alleles).

Results

	MIX	6 loci	'less' polymorphic	'more' polymorphic
Site 1	16.7%	0.1608	0.2926	0.1656
Site 2	16.7%	0.2025	0.2172	0.1704
Site 3	16.7%	0.1729	0.1239	0.1963
Site 4	16.7%	0.1457	0.2053	0.1883
Site 5	16.7%	0.1515	0.0044	0.1464
Site 6	16.7%	0.1665	0.1566	0.1328
Mean SD		0.0683	0.1126	0.0761

Comment

The less variable loci gave much more inaccurate results than either using six, or the three more variable loci. In particular less than 1% contribution was estimated from Site 5, which could lead to erroneous conclusion that Site 5 did not contribute to the MIX. The mean SD was also particularly high at over 11%. The more variable loci produced results as accurate at least than those obtained from all six loci.

Experiment 1.5

The aim was to investigate if the degree of accuracy obtained in Experiment 1.1 was maintained when the proportions from the contributing stocks was not equal. Two simulations were run with a MIX of n = 80, using data from all six loci. Individuals in the MIX sample were randomly picked from the various contributing stocks.

Results

	i. MIX		ii. MIX	
Site 1	50%	0.4385	0%	0.0015
Site 2	20%	0.2513	16.7%	0.0867
Site 3	0%	0.0107	0%	0.0004
Site 4	10%	0.0926	8.3%	0.0792
Site 5	10%	0.1194	8.3%	0.1519
Site 6	10%	0.0874	66.7%	0.6801
Mean SD		0.0493		0.0374

Comment

In the first example only five stock contributed, and Site 1 contributed 50%. The estimates produced reflected this, although the proportion from the largest contributor was underestimated. Importantly, the estimate from Site 3 was negligible. In the second case the actual proportions were even more skewed, a single site contributed over 65% of the MIX, and Sites 1 and 3 did not contribute at all. Again estimates for the non-contributing sites were accurate. The largest inaccuracy was in the estimation of contribution from Site 5. The mean SD in both these examples is less than 5%.

Experiment 1.6

So far all simulations have assumed that the baseline stocks make up 100% of the MIX sample. In reality this may not be the case as only a finite number of possible baseline stocks can be sampled. The aim of this experiment is to investigate the effect of stray fish in a situation where the proportions of stocks are equal, and one where the proportions are highly skewed. Simulations were run using a MIX from the baseline stocks and including samples from a west coast river as 'strayers'.

Results

	i. MIX	n = 80	ii. MIX	n = 140	iii. MIX	N = 87
Site 1	14.3%	0.1733	17.8%	0.1789	8%	0.0741
Site 2	14.3%	0.2262	17.8%	0.2777	17.2%	0.3287
Site 3	14.3%	0.1689	25%	0.2357	17.2%	0.1946
Site 4	14.3%	0.1523	14.3%	0.2911	17.2%	0.3626
Site 5	14.3%	0.1413		0.0001		0.0002
Site 6	14.3%	0.1380		0.0165		0.0397
'Strayers'	14.3%		25%		40%	
Mean SD		0.0631		0.0699		0.0685

	iv. MIX	N = 80	v. MIX	N = 48
Site 1	14.3%	0.1996	8.3%	0.1753
Site 2	7.1%	0.1992	4.2%	0.2252
Site 3		0.0937		0.1018
Site 4	7.1%	0.2609	4.2%	0.3797
Site 5		0.0004		0.0000
Site 6	21.5%	0.2461	8.3%	0.1179
'Strayers'	50%		75%	
Mean SD		0.0941		0.1210

Comment

Although the 'Strayers' sample does not have any alleles that do not also occur in Sites 1-6, an individual in MIX may be heterozygous for alleles that do not occur in the same sample population, therefore the individual is excluded. One individual of this kind was found in case i., and two in each of the remainder. In the first example all stocks, including strayers, contribute equal numbers. The contribution of the strayers has inflated estimates from various sites, but only Site 2 shows any significant increase over the real value. As the proportions become more skewed the level of error increases. With regard to the non-contributing stocks, the estimates for Site 5 remain accurate throughout, but Site 2 shows an elevation to 10%. The largest error occurs in estimates for Site 4. When strayers make up 40% of the mixture the estimate is more than twice the true value. This rises to almost ten times when strayers make up the majority of the mixture. It is notable that Site 4 and the strayers have the lowest pairwise Fst value in this experiment. It would appear that the straying fish are preferentially assigned to the population least different to them, and that the bias of the estimate increases with the contribution of strayers. However even when strayers equal 40%, estimates of the other stocks are realistic.

Experiment 1.7

The aim of this experiment was to estimate the stock specific bias of the values produced i.e. the proportion of stock A that is falsely attributed to another stock, even though stock A made up 100% of the mixture. As a measure of the precision of the estimate, 95% confidence limits are calculated. To do this separate simulations were run with MIX samples consisting of 40 individuals from each baseline stock in turn. Bootstrapping was performed with 100 iterations and from these the mean value. and 95% CI were calculated.

Results

Simulation	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
100% Site 1	**0.7835 (0.7608 - 0.8063**	0.0544 (0.0430 - 0.0658)	0.0183 (0.0116 - 0.0249)	0.0495 (0.0376 - 0.0614)	0.0335 (0.0245 - 0.0426)	0.0607 (0.0481 - 0.0733)
100% Site 2	0.0886 (0.0734 - 0.1038)	**0.8230 (0.8087 - 0.8499)**	0.0247 (0.0156 - 0.0339)	0.0101 (0.0052 - 0.0150)	0.0241 (0.0159 - 0.0322)	0.0231 (0.0166 - 0.0297)
100% Site 3	0.0186 (0.0113 - 0.0259)	0.0217 (0.0139 - 0.0296)	**0.8000 (0.7800 - 0.8196)**	0.0972 (0.0798 - 0.1145)	0.0123 (0.0063 - 0.0181)	0.0504 (0.0388 - 0.0621)
100% Site 4	0.0122 (0.0052 - 0.0192)	0.0420 (0.0334 - 0.0506)	0.0432 (0.0316 - 0.0548)	**0.7943 (0.7684 - 0.8202)**	0.0719 (0.0579 - 0.0859)	0.0363 (0.0240 - 0.0487)
100% Site 5	0.0149 (0.0082 - 0.0216)	0.0386 (0.0271 - 0.0502)	0.0238 (0.0149 - 0.0328)	0.0187 (0.0114 - 0.0260)	**0.8877 (0.8686 - 0.9067)**	0.0162 (0.0102 - 0.0222)
100% Site 6	0.0584 (0.0460 - 0.0708)	0.0236 (0.0158 - 0.0314)	0.0491 (0.0389 - 0.0591)	0.0247 (0.0164 - 0.0329)	0.0536 (0.0404 - 0.0668)	**0.7906 (0.7707 - 0.8104)**

Comment

The stocks estimates are within 11% to 21% of the true value. The mean stock specific bias (i.e. the average percentage wrongly assigned to a stock) ranges between 3.2% (Site 3) to 4.0% (Site 4), or 4.1% and 5.0% if the upper confidence limits are considered. This is similar the average mean error generally calculated in previous experiments.

4.2 Experiment 2. Assessment of inter-drainage genetic variation (minisatellite and mitochondrial DNA) in sea trout and its use in GSI

Data from the screening of several tributaries of three tributaries of the Tweed, three from the Dee (see EXP2), three from the Don (see EXP2), and one from the Spey rivers in the east of Scotland, were used to model inter-drainage GSI estimates. All samples were from anadromous stocks. Hereafter the rivers will be referred to as Rivers 1 - 4. Samples of juveniles were screened at five minisatellite loci, and for mtDNA RFLP variation at 2 regions. Numbers of alleles per locus, heterozygosities, number of composite mitochondrial haplotypes, and pairwise Fst values are given. The heterozygosity value for River 4 is lower than the remainder, however the difference is not significant when the variance of the mean is taken into account. River 3 has by far the largest number of mitochondrial haplotypes, indicating a higher level of mitochondrial DNA diversity. The highest pairwise Fst derived from minisatellite data occurs between River 1 and River 4 (the geographically most distant rivers), however the Rivers 3 and 4 differ most in mitochondrial DNA. For both marker systems the overall genetic divergence is significant indicating population structuring.

Sample size, number of alleles per locus and average heterozygosity (H_{avg}) of samples used

	Sample size[a]	Str-A3	Str-A5	Str-A22/1	Ssa-A45/1	Ssa-A45/2	H_{avg}	Number haplotypes
River 1	28-46	10	6	16	13	3	0.704	5
River 2	47-137	14	7	18	14	6	0.736	5
River 3	44-96	15	7	17	13	7	0.714	10
River 4	44-47	11	7	13	10	3	0.656	3

[a] Variation in sample size is due to different number at each loci. The lower numbers in the River 2 and River 3 refer to mtDNA data

Pairwise Fst values over nuclear loci (and Φst from mtDNA)

	River 1	River 2	River 3
River 2	0.0600 (0.1776)	—	
River 3	0.0794 (0.2522)	0.0150 (0.1294)	—
River 4	0.0825 (0.2314)	0.0229 (0.1866)	0.0224 (0.2688)

Fst over rivers 1- 4 (6 minisatellite loci) = 0.037 (95% CI 0.023 - 0.068)
Φst over rivers 1 - 4 (15 composite haplotypes) = 0.200

Experiment 2.1

The aim of the experiment was to investigate if estimates of stock composition from a mixture of rivers was accurate, and whether the loci that were used in screening had any effect on the estimates. Simulation were run with a skewed proportion of stocks contributing to the MIX (n = 50) using all five minisatellite loci and mitochondrial DNA data, and were rerun excluding one loci at a time.

Results

	MIX	5 loci + mtDNA	Excluding mtDNA	Excluding *Str*-A3	Excluding *Str*-A5	Excluding *Str*-A22/1	Excluding *Ssa*-A45/1	Excluding *Ssa*-A45/2
River 1	25%	0.2526	0.2284	0.2474	0.2483	0.2446	0.2573	0.2574
River 2	25%	0.2385	0.2032	0.2635	0.2556	0.1903	0.2553	0.2174
River 3	50%	0.5088	0.5646	0.4890	0.4960	0.5651	0.4873	0.5252
River 4	0%	0.0000	0.0039	0.0000	0.0000	0.0000	0.0000	0.0000
Mean SD		0.0612	0.0657	0.0624	0.0624	0.0606	0.0615	0.0611

Comment

The estimates derived from all data were accurate to within 1.2%. There was no significant difference in the values, or the mean SD, produced in each simulation. In particular the estimates for River 4 were less than 0.5%. Previous experiments (1.2, and 1.4) suggest that substantially reducing the number of loci adversely effects the estimate, while use of highly polymorphic increases accuracy. However, this experiment proves that no single (nuclear or mitochondrial) loci is more informative than the others used in the models.

4.3 Experiment 3. Assessment of intra-drainage genetic variation (minisatellite DNA) in Atlantic salmon and its use in GSI

Extensive genetic data has been made available for modelling from Atlantic salmon in two tributaries of the Aberdeenshire Dee (John Taggart, unpublished data). The Girnock and Baddoch tributaries, both predominately early running grilse or salmon spawning areas, have trapping facilities which allow sampling of adult fish moving upstream to spawn. Redd marking and the sampling of eggs is also carried out extensively above the trap in the Girnock Burn. This has allowed adults moving into the study area, and eggs spawned by those adults to be screened at minisatellite loci for the purposes of pedigree and behavioural study. Several cohorts of adults and progeny have been screened. For comparison a sample of adults from another tributary of the Dee (Sheeoch) and a Norwegian river have also been included. Hereafter the Dee tributaries will be referred to as Sites 1 - 3.

The level of heterozygosity of each sample is approximately equal, however as expected smaller numbers of alleles are found in Site 3 and the Norwegian samples due to the small sample sizes. The low pairwise Fst values between sites on the Dee are smaller to those found in Experiment 1 where samples of brown trout from the same river were compared, indicating a possible higher level of intra-drainage genetic divergence in brown trout. As expected the highest Fst values are found in comparisons involving the Norwegian sample.

Sample size, number of alleles per locus and average heterozygosity of samples used

		Sample size	*Ssa*-A45/1	*Str*-A9	*Ssa*-A45/2	*Str*-A22/1	*Str*-A5	*Ssa*-A60	Havg
Site 1	adults 91	118	10	11	19	20	23	11	0.728
	adults 92	75							
	adults 93	79							
	adults 94	96							
Site 2	adults 91	95	6	11	17	19	20	12	0.726
	progeny 91	5424							
	adults 92	79							
	progeny 92	341							
	adults 93	42							
Site 3		14	5	6	7	13	13	4	0.762
Norway		30	4	5	8	9	8	4	0.712

Pairwise Fst values over nuclear loci

	Site 1	Site 2	Site 3
Site 2	0.005		
Site 3	0.002	0.002	
Norway	0.042	0.050	0.035

Fst overall (6 minisatellite loci) = 0.012 (95% CI 0.008 - 0.016)

Experiment 3.1

The aim was to investigate if a progeny cohort relate back to their parental group. Progeny were used as the MIX sample, and their parental group and other adults from the same tributary and others were used as baseline stocks. Using the progeny, rather than random individuals from the baseline stock, provides more realistic data for modelling. Through random spawning genotypes are randomly generated from those of the parental group. Two simulations were run using random samples from the Site 2 1991 and 1992 progeny groups.

Results

i.

		MIX	n = 400
Site 1	adults 91		0.0261
	adults 92		0.0011
	adults 93		0.0082
	adults 94		0.0003
Site 2	adults 91	100% progeny	0.8348
	adults 92		0.0665
	adults 93		0.0610
Site 3			0.0027
Norway			0.0000
Mean SD			0.0184

ii

		MIX	n = 294
Site 1	adults 91		0.0284
	adults 92		0.0319
	adults 93		0.0062
	adults 94		0.0009
Site 2	adults 91		0.1936
	adults 92	100% progeny	0.7646
	adults 93		0.0000
Site 3			0.0000
Norway			0.0000
Mean SD			0.0218

Comment

Estimates in both cases showed the true parental groups as being the most important contributing stock to the MIX. More accurate estimates were obtained for the 1991 progeny group which could be a result of the sample size of the baseline stock, and the MIX sample being larger. In the case of the 1992 progeny, almost 20% were indicated as arising from the previous years spawners. However, although in both cases the true parental group was not identified as the only contributing stock, over 95% of progeny were identified to the correct site. This indicates that there is sufficient variability to correctly identify the parental group in the majority of cases, and where there is not, there is sufficient similarity between years to identify progeny to the correct location.

Experiment 3.2
To further investigate the temporal stability of the data sets, this experiment investigated whether the degree of accuracy seen above was maintained when the true parental group was not included in the baseline stocks. To simulate a more real situation, simulations involved parental groups from a Sites 1 and 2 from a single year only. Three simulations were run for each progeny group using adult stocks from the same year. Three hundred randomly picked individuals were used in MIX samples.

Results

Site 2 1991 progeny

MIX = 100% Site 2 1991 progeny

Site 1	adults 91	0.0794	—		—	
	—		adults 92	0.4221	—	
	—		—		adults 93	0.5106
Site 2	adults 91	0.9206 *	—		—	
	—		adults 92	0.5779	—	
	—		—		adults 93	0.4890

Site 2 1992 progeny

MIX = 100% Site 2 1992 progeny

Site 1	adults 91	0.1984	—		—	
	—		adults 92	0.0921	—	
	—		—		adults 93	0.6960
Site 2	adults 91	0.8016	—		—	
	—		adults 92	0.9079 *	—	
	—		—		adults 93	0.3040

* denotes true parental group

Comment
While in the previous experiment adults from the correct location were correctly estimated as the main contributing stock when all year classes were included, the same is not the case here. In the first example the parental stock is estimated with less than 10% error, but in following years half the progeny would be estimated as being derived from Site 1. In the second example, the parental stock is again highly estimated, as is the previous years spawners, but in the final year, the majority of progeny are estimated as Site 1 stock. From this data the temporal stability of sites appears to vary. In 1991 and 1992 the error of the estimate is on average less than 20%, however, in the following year the percentage error trebles. This could in part be due to a drop in the spawning escapement entering Site 2, from 95 in 1991 to 42 in 1993.

Experiment 3.3
In the light of the above findings this experiment aimed to investigate whether the apparent transient instability effected estimates of mixed stock contributions. Simulations were run with mixtures of the progeny groups and/or other stocks.

Results
i.

		MIX	n = 400
Site 1	adults 91		0.0530
	adults 92		0.0033
	adults 93		0.0111
	adults 94		0.0000
Site 2	adults 91	50% progeny	0.5134
	adults 92	50% progeny	0.4043
	adults 93		0.0145
Site 3			0.0000
Norway			0.0000
Mean SD			0.0175

ii

		Mix	n = 400
Site 1	adults 91		0.0043
	adults 92		0.0309
	adults 93		0.0099
	adults 94		0.0019
Site 2	adults 91		0.1642
	adults 92	87% progeny	0.6714
	adults 93		0.0000
Site 3		4%	0.0290
Norway		9%	0.0882
Mean SD			0.0175

Comment
Results were reasonably accurate in simulation i., although the contribution of Site 2 1992 progeny was underestimated. However, none were erroneously assigned to Site 3 or the Norwegian sample. In simulation ii, while the estimates for Site 3 and Norway were reasonably accurate, over 16% of the 1992 progeny were assigned to 1991 parents. This confirms Experiment 3.2 ii, where 1992 progeny showed a high affinity for 1991 adults. Less than 5% were wrongly assigned to Site 1.

4.4 Experiment 4 Assessment of inter-drainage genetic variation (minisatellite DNA) in Atlantic salmon and its use in GSI

In these experiments data from three populations in the Windermere catchment (Sheila Hartley and Alan Pickering 1994) have been included - Troutbeck, Leven, and Rothay. Variation was scored at three minisatellite loci in juvenile samples according to the same system as the data used in the previous section, and are therefore directly comparable. All adult samples from Sites 1 and 2 (from Experiment 3) were pooled to provide baseline stock frequencies.

The heterozygosities form all sites are similar when the variance of the mean is considered, although the Dee sites (1 and 2) have a significantly greater number of alleles. As only adults from these sites have been included (due to computing time) the Fst between these sites is different to that given in Experiment 3. The highest pairwise Fst value occurs between the Norwegian sample and Site 6. All comparisons within the Windermere catchment are highly significant and greater than the intra-drainage comparisons made in Experiments 1 and 3.

Sample size, number of alleles per locus and average heterozygosity (H_{avg}) of samples used

		Sample size	*Ssa*-A45/1	*Str*-A9	*Ssa*-A45/2	Havg
Site 1	Baddoch	368	10	11	19	0.748
Site 2	Girnock	216	6	11	17	0.713
Site 3	Norwegian	30	4	5	8	0.656
Site 4	Troutbeck	45	5	5	9	0.703
Site 5	Leven	86	7	8	12	0.733
Site 6	Rothay	102	5	7	11	0.683

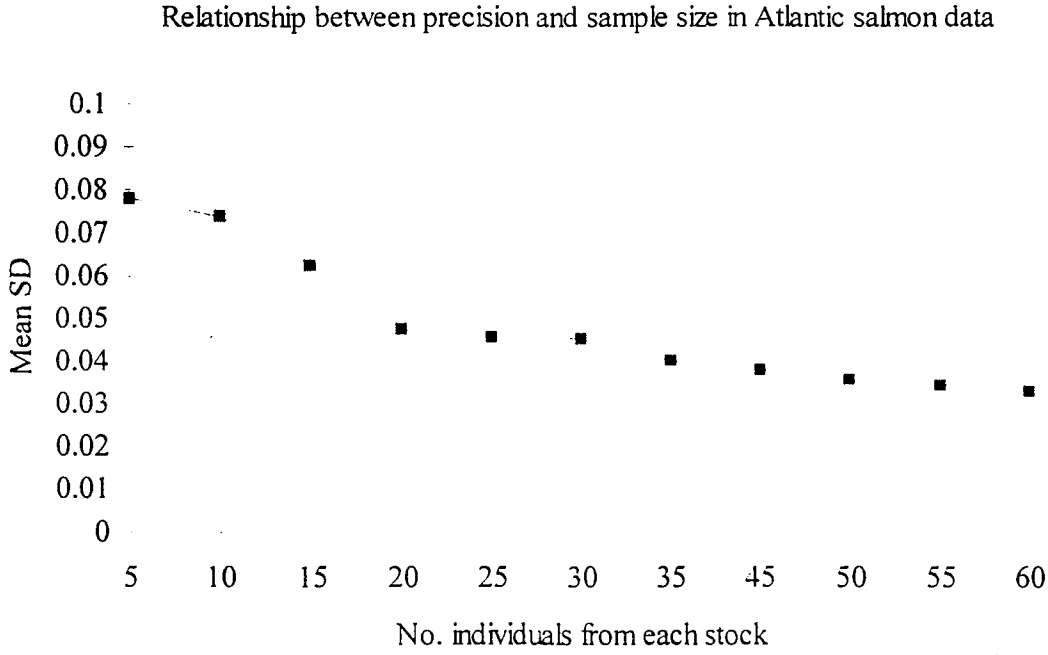
Pairwise Fst values

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Site 2	0.005	—				
Site 3	0.053	0.079	—			
Site 4	0.054	0.057	0.105	—		
Site 5	0.032	0.026	0.094	0.067	—	
Site 6	0.037	0.037	0.112	0.064	0.031	—

Fst over Sites 1 - 6 = 0.032 (95% CI 0.025 - 0.044)

Experiment 4.1

The aim of this experiment was to investigate the effects of MIX sample size on the precision of stock composition estimates, as in EXP 1.2 Simulations were run using equal contributions from each stock, using 5 to 60 individuals randomly picked. Mean SD was calculated and plotted. Data were randomly duplicated in order to make up sample sizes for Sites 3 and 4.



Comment

As in EXP 1.2, the mean SD declines rapidly initially until the mean SD drops below 5% when the MIX contains 20 individuals from each stock. At this point the error values obtained for salmon (0.0469) and trout (0.0451) data sets are similar. Further increase in sample size does not change the error more than 1%. Due to the low number of loci used here, the effect of loci number, or level of polymorphism could not be modelled for salmon.

Experiment 4.2

Stock specific bias of the estimates were investigated by using each stock in turn as 100% of the MIX in order to find the proportion of stock A that is falsely attributed to another stock, even though stock A made up 100% of the mixture. As a measure of the precision of the estimate, 95% confidence limits are calculated.. Separate simulations were run with MIX samples consisting of 40 individuals from each baseline stock in turn. Bootstrapping was performed with 100 iterations and from these the mean value. and 95% CI were calculated.

Results

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
100% Site 1	**0.6653 (0.6320 - 0.6986)**	0.1253 (0.0975 - 0.1533)	0.0098 (0.0049 - 0.0148)	0.1252 (0.1084 - 0.1420)	0.0003 (0.0000 - 0.0008)	0.0739 (0.0619 - 0.0859)
100% Site 2	0.1054 (0.0845 - 0.1263)	**0.6442 (0.6190 - 0.6695)**	0.0081 (0.0021 - 0.0142)	0.1475 (0.1280 - 0.1670)	0.0046 (0.0010 - 0.0081)	0.0901 (0.0709 - 0.1093)
100% Site 3	0.0642 (0.0473 - 0.0810)	0.0242 (0.0147 - 0.0337)	**0.8701 (0.8499 - 0.8914)**	0.0413 (0.0278 - 0.0548)	0.0002 (0.0000 - 0.0007)	0.0000 (0.0000 - 0.0000)
100% Site 4	0.0357 (0.0254 - 0.0459)	0.0000 (0.000 - 0.0001)	0.0181 (0.0119 - 0.0243)	**0.9410 (0.9272 - 0.9541)**	0.0000 (0.0000 - 0.0000)	0.0052 (0.0007 - 0.0096)
100% Site 5	0.0302 (0.0199 - 0.0405)	0.0576 (0.0448 - 0.0704)	0.0007 (0.0002 - 0.0013)	0.1737 (0.1551 - 0.1923)	**0.4285 (0.4052 - 0.4517)**	0.3093 (0.2859 - 0.3327)
100% Site 6	0.0084 (0.0034 - 0.0133)	0.0159 (0.0104 - 0.0215)	0.0003 (0.0000 - 0.0006)	0.0604 (0.0484 - 0.0725)	0.0042 (0.0012 - 0.0070)	**0.9109 (0.8969 - 0.9250)**

Comment

Estimates for the Dee samples (Sites 1 and 2) were poor. although if summed they account for 75% to 80%. Stock specific bias for these sites were 4.9% and 4.5%. or 6.1% and 5.6% if upper confidence limits are considered. The main source of error was Site 4 which had a very high stock specific bias. 11.0% (12.6%). Also Site 5 had a stock specific bias less than 0.5%. but conversely less than half its mixture is assigned to it. although Sites 4. 5 and 6 account for over 91%.

4.5 Experiment 5 Assessment of intra- and inter-drainage genetic variation (mitochondrial DNA) in Atlantic salmon and its use in GSI

Published data from studies of mtDNA variation in Welsh salmon populations (O'Connell, Skibinski, D.O.F, and Beardmore, J.A. 1995, 1996) were used to model GSI estimates. Six drainages were screened, three from the south which were sampled at various sites. The sample sizes of juveniles are low, due in part to the limitations of the technique used in mtDNA analysis. Samples were screened for variation using four polymorphic restriction enzymes detecting 12 haplotypes. The Dee has the highest number of haplotypes detected and the lowest level of genetic divergence between sites (the negative Φst value indicates that genes are more related between sites than within them). The distribution of haplotypes between rivers is not significant, indicating no population structuring except between the Teifi and Wye. Although there was significant heterogeneity within the Teifi and Wye, there was no significant differences between drainages (P = 0.0998).

Drainage	Site	Sample size	Drainage	Site	Sample size
Teifi	1	8	Wye	16	5
	2	2		17	3
	3	6		18	6
	4	8		19	7
	5	8		20	6
	6	4		21	7
Usk	7	2		22	2
	8	1	Dee - Mynach	23	12
	9	8	Ceidiog	24	3
	10	10	Ceirw	25	5
	11	3	Rhyd-hir	26	12
	12	7	Dyfi	27	7
	13	6			
	14	5			
	15	8			

Summary of genetic data

	No. of composite haplotypes	Differentiation within drainages (Φst)
Teifi	8	0.076
Usk	7	0.068
Wye	7	0.071
Dee	9	-0.041
Rhyd-hir	3	—
Dyfi	2	—
overall drainages	12	0.076

Experiment 5.1
The aim was to investigate the accuracy of GSI estimates, and stock specific bias using MIX to which only one stock contributes. Only the sites from which positive contributions are estimated are listed in the results.

Results

	i. MIX	n = 28	ii. MIX	n = 24
Site 2		0.1142		0.0832
Site 3	100%	0.2322		—
Site 4		0.1142		0.0832
Site 8		0.1142		—
Site 12		0.1243		—
Site 15		0.1263		—
Site 20		0.0002		0.0008
Site 21		0.1828	100%	0.7504
Site 24		—		—

Comment
In the first example the stock estimates are highly inaccurate, although the correct sites is estimated as the largest contributor. Also the Teifi three sites (2, 3 and 4) combined are estimated to contribute almost half of the MIX. The estimate given in the second example is more accurate. This is mainly due to a unique haplotype at that site which occurs at a frequency of 0.333, which is reflected in the MIX sample. When estimates are derived from a single locus, a unique haplotype is in fact partially or wholly diagnostic.

Experiment 5.2
To investigate the accuracy of estimates on an inter-drainage basis. sites within rivers were pooled to calculate baseline frequencies between rivers. Simulations were run with and without the Dyfi and Rhyd-hir stocks which are represented by relatively small sample sizes.

Results

	MIX	With 4 stocks n = 20	With 6 stocks n = 20
Teifi	100%	0.9999	0.5426
Wye		0.0001	0.0000
Usk		0.0000	0.0000
Dee		0.0000	0.0000
Dyfi		not included	0.0002
Rhyd-hir		not included	0.4572

Comment
When only four stocks are included. the GSI estimates are highly accurate. However. the addition of the Rhyd-hir to the baseline stocks generates massive error. This illustrates the potential errors involved in attempting to use a single locus as a diagnostic marker even when resampling techniques are employed.

Experiment 5.3
The aim was to investigate the estimates when MIX samples are made up from skewed proportions of the baseline stocks when intra-drainage samples are pooled. The smaller samples from Rhyd-hir and Dyfi were excluded.

Results

	i. MIX	n = 40	ii. MIX	n = 40
Teifi		0.3067	10%	0.4310
Wye	50%	0.2082	65%	0.2443
Usk	50%	0.4509	25%	0.2904
Dee		0.0341		0.0342

Comment
Again estimates are highly inaccurate, particularly in the first example where the Teifi is estimated as contributing over 30%. This is due to the Wye and Teifi sharing haplotype(s) not commonly found elsewhere.

4.6 Experiment 6 Assessment of intra-drainage genetic variation (mitochondrial and minisatellite) in Atlantic salmon and its use in GSI

Unpublished data on juvenile Atlantic salmon from three tributaries of the River Tweed has been made available for modelling (Eric Verspoor, unpublished data 1996). Samples from two upper tributaries (Ettrick and 'Upper') and one slightly further downstream (Gala) were screened at 4 minisatellite loci and the ND1 mtDNA region was screened with 5 polymorphic restriction enzymes. Ettrick, 'Upper' and Gala will hereafter be referred to Sites 1 - 3.

Unlike the data used in previous experiments, there is no genetic differentiation between Tweed sites at either nuclear or mitochondrial loci. Although the heterozygosity values and number of haplotypes do not suggest any lack of variability within sites. The test in this section is to investigate if GSI estimates are accurate and precise when there is no genetic heterogeneity in the baseline stocks.

Sample size, number of alleles per locus, average heterozygosity (Havg), and number of composite mitochondrial haplotypes in samples used

	Sample size[a]	*Str*-A9	*Str*-A22/1	*Ssa*-A34	*Ssa*-A45/2	Havg	No. composite haplotypes
Site 1	86-96	10	8	4	8	0.733	4
Site 2	90-98	10	8	4	7	0.717	5
Site 3	90-96	10	7	4	9	0.733	5

[a] Variation in sample size is due to different number at each loci.

Pairwise Fst values over nuclear loci (and Φst from mtDNA)

	Site 1	Site 2
Site 3	0.0008 (0.0008)	
Site 4	0.0028 (0.0079)	0.0066 (-0.0063)

Fst over sites 1 -3 (4 minisatellite loci) = 0.003 (95% CI -0.000 - 0.010)
Φst over sites 1 - 3 (6 composite haplotypes) = 0.001

Experiment 6.1
Stock specific bias of the estimates were investigated by using each stock in turn as 100% of the MIX. Separate simulations were run with MIX samples consisting of 50 individuals from each baseline stock in turn. Bootstrapping was performed with 100 iterations and from these the mean value, and 95% CI were calculated. Each simulation was rerun excluding mitochondrial DNA data, and these values are given in italics.

Results

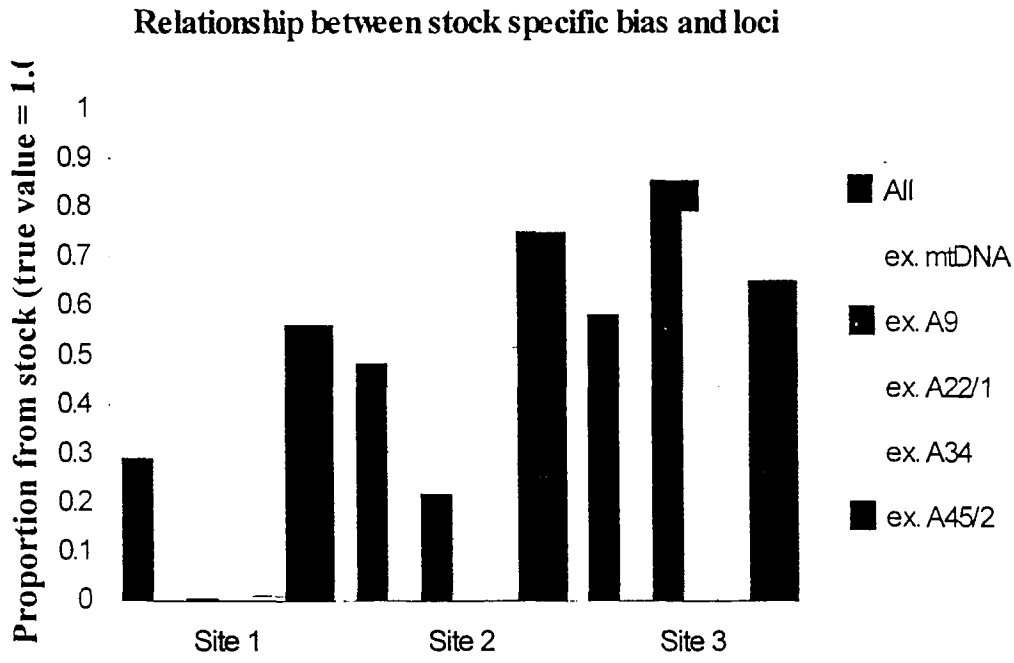
	Site 1		Site 2		Site 3	
	+ mtDNA	- mtDNA	+ mtDNA	- mtDNA	+ mtDNA	- mtDNA
Site 1 (100%)	**0.2930 (0.2494 - 0.3454)**	***0.7915 (0.7417 - 0.8414)***	0.3142 (0.2657 - 0.3628)	*0.1510 (0.1052 - 0.1969)*	0.3918 (0.3423 - 0.4414)	*0.0574 (0.0379 - 0.0770)*
Site 2 (100%)	0.0677 (0.0384 - 0.0971)	*0.2449 (0.1891 - 0.3006)*	**0.4862 (0.4259 - 0.5466)**	***0.6717 (0.6120 - 0.7315)***	0.4460 (0.3902 - 0.5018)	*0.0833 (0.0570 - 0.1097)*
Site 3 (100%)	0.0717 (0.0517 - 0.0916)	*0.1897 (0.1449 - 0.2348)*	0.3445 (0.2972 - 0.3920)	*0.2235 (0.1771 - 0.2699)*	**0.5837 (0.5399 - 0.6276)**	***0.5866 (0.5355 - 0.6378)***

Comment
GSI estimates here showed considerable levels of error with the best estimate being 30% from the true value. It is notable that in two cases (Sites 1 and 2) excluding mtDNA data from the analysis produced significantly different values, nearer to the true ones. In particular a MIX from Site 1 was estimated to consist of only 30% of that stock. The estimate rose to nearly 80% on the basis of nuclear data alone.

Experiment 6.2

As the exclusion of mtDNA data appeared to significantly effect the accuracy of estimates in the previous section, the aim of this experiment was to further investigate the influence of different loci. Simulations were run with a single loci excluded each time and MIX samples consisting of 50 individuals from each stock in turn. The mean estimate for the contributing stock (the true value = 100%) is plotted below.

Results



Comment

At every site excluding at least one loci appears to raise the GSI estimate. At Site 1 exclusion of mtDNA data increases the estimate almost 3 fold, while excluding *Ssa*-A45 2 almost doubles it. Excluding *Str*-A9 reduces the GSI estimate to less than 1%. It is notable that in the sample from Site 1, loci *Str*-A22/1 and *Str*-A9 show a highly significant level of genotypic disequilibrium ($P < 0.0001$) suggesting sampling error at one or more loci. In Sites 2 and 3 no single loci, or pair of loci, has such a pronounced effect. However, mtDNA and *Ssa*-A45 2 have a consistent effect on estimates for all three sites.

Although it cannot be modelled due to the low number of loci, it could be deduced that where there is a lack of differentiation between baseline stocks, a higher number of loci will produce a more accurate estimate. This has been demonstrated for baselines where there is significant heterogeneity, and is likely to be more important in cases where the discriminatory power of each locus is weaker overall.

Experiment 6.3
This experiment examines how accurate stock estimates are when the MIX sample is of skewed or equal proportions. Simulations were run with and without mtDNA data.

Results
i.

		n = 75	
	MIX	+ mtDNA	- mtDNA
Site 1	33%	0.1803	0.4550
Site 2	33%	0.3660	0.3968
Site 3	33%	0.4537	0.1482
Mean SD		0.2321	0.1994

ii.

		n = 60	
	MIX	+ mtDNA	- mtDNA
Site 1	66%	0.1141	0.5111
Site 2	33%	0.5511	0.4401
Site 3		0.3348	0.0487
Mean SD		0.2175	0.1995

iii.

		n = 60	
	MIX	+ mtDNA	- mtDNA
Site 1	5%	0.0117	0.2032
Site 2	15%	0.5322	0.2270
Site 3	80%	0.4560	0.5697
Mean SD		0.2278	0.2242

Comment
The most notable difference between this and previous experiments is the high mean standard errors of the values. These are up to 10 times greater than the values produced when the baseline stocks show significant differentiation. In these three models estimates produced without using mtDNA are demonstrably more accurate than those produced with. This is particularly true for MIX samples where only two stocks contribute. Although the values are inaccurate, the overall ranking baseline stock contribution is correct. The most important contributor to the MIX is correctly identified.

4.7 Experiment 7 Assessment of inter-drainage genetic variation (minisatellite DNA) in Atlantic salmon and its use in GSI

This experiment is based on minisatellite screening of seven rivers in north Co. Donegal (Clifford, 1996). Juveniles from the Owenmarve, Owentochar. Ray, Swilly, Owencarrow, Owenbeagh and Gleannan were sampled by electrofishing. The Ray and Swilly were sampled for three consecutive years. There is no significant difference between the heterozygosity values for each sample. Except for comparisons between Rivers 2 and 4, and Rivers 5 and 7, the pairwise Fst values were highly significant and overall values suggested significant population structuring with values similar to that found in Experiments 2 and 4.

Sample size, number of alleles per locus and average heterozygosity of samples used

	Sample size[a]	*Ssa*-A10	*Str*-A5	*Ssa*-A45/2	Havg
River 1	17-24	5	4	6	0.498 ± 0.171
River 2	21-22	6	7	8	0.470 ± 0.051
River 3 (91,92,93)	53-86	6	9	11	0.547 ± 0.012
River 4 (91,92,93)	54-67	7	8	13	0.517 ± 0.140
River 5	26-41	5	5	8	0.661 ± 0.112
River 6	18-46	5	6	9	0.498 ± 0.085
River 7	17-25	5	4	8	0.554 ± 0.140

[a] Variation in sample size is due to different number at each loci.

Pairwise Fst values

	River 1	River 2	River 3	River 4	River 5	River 6
River 2	0.028	—				
River 3	0.010	0.015	—			
River 4	0.033	0.008	0.016	—		
River 5	0.022	0.055	0.011	0.040	—	
River 6	0.034	0.068	0.042	0.030	0.028	—
River 7	0.052	0.074	0.027	0.081	0.006	0.068

Fst over Rivers 1 -7 = 0.032 (95% CI 0.024 - 0.040)

Experiment 7.1

Stock specific bias of the estimates were investigated by using each stock in turn as 100% of the MIX. Separate simulations were run with MIX samples consisting of 50 individuals from each baseline stock in turn. Bootstrapping was performed with 100 iterations and from these the mean value, and 95% CI were calculated.

Simulation	River 1	River 2	River 3	River 4	River 5	River 6	River 7
River 1 (100%)	**0.8276 (0.7979 - 0.8573)**	0.0113 (0.0048 - 0.0177)	0.0654 (0.0419 - 0.0889)	0.0324 (0.0203 - 0.0445)	0.0205 (0.0090 - 0.0321)	0.0352 (0.0240 - 0.0465)	0.0074 (0.0 - 0.0145)
River 2 (100%)	0.000 (-0.0000 - 0.0145)	**0.6704 (0.6369 - 0.7040)**	0.0712 (0.0458 - 0.0967)	0.1915 (0.1640 - 0.2191)	0.0307 (0.0168 - 0.0446)	0.0010 (-0.0000 - 0.0025)	0.0351 (0.0218 - 0.0484)
River 3 (100%)	0.0224 (0.0129 - 0.0319)	0.0773 (0.0610 - 0.0937)	**0.6214 (0.5860 - 0.6567)**	0.0620 (0.0429 - 0.0811)	0.0309 (0.0196 - 0.0424)	0.0874 (0.0704 - 0.1044)	0.0984 (0.0753 - 0.1215)
River 4 (100%)	0.0213 (0.0140 - 0.0285)	0.0570 (0.0442 - 0.0698)	0.0854 (0.0668 - 0.1040)	**0.7924 (0.7712 - 0.8137)**	0.0088 (0.0033 - 0.0142)	0.0139 (0.0089 - 0.0189)	0.0212 (0.0138 - 0.0286)
River 5 (100%)	0.0594 (0.0433 - 0.0754)	0.0028 (0.0000 - 0.0057)	0.0296 (0.0159 - 0.0432)	0.0011 (0.0000 - 0.0029)	**0.7525 (0.7179 - 0.7871)**	0.0830 (0.0576 - 0.1085)	0.0714 (0.0487 - 0.0942)
River 6 (100%)	0.0132 (0.0046 - 0.0218)	0.0080 (0.0004 - 0.0171)	0.0442 (0.0252 - 0.0632)	0.0112 (0.0024 - 0.0201)	0.0523 (0.0330 - 0.0717)	**0.8248 (0.7925 - 0.8571)**	0.0454 (0.0264 - 0.0644)
River 7 (100%)	0.0182 (0.0070 - 0.0293)	0.0058 (0.0000 - 0.0120)	0.0218 (0.0110 - 0.0325)	0.0340 (0.0213 - 0.0468)	0.1287 (0.0965 - 0.1609)	0.0686 (0.0525 - 0.0846)	**0.7229 (0.6854 - 0.7604)**

Comment

The stock estimates are within 17.2% and 37.9% of the true value. The average level of error (i.e. percentage of stock wrongly assigned to each river) ranges between 2.9% (River 6) and 6.3% (River 3). The highest levels of error occur between Rivers 2 and 4, and Rivers 5 and 7. These samples also have the lowest pairwise Fst values.

Experiment 7.2
This aim of this experiment was to investigate how accurate GSI estimates were for temporally spaced samples. Samples from the Ray and Swilly were pooled in the MIX sample, but the simulations were run with baseline data for each individual year. This would show if the MIX sample was broken down into sampling cohorts.

Results
i.

		MIX	n = 35
River 1			0.0000
River 2			0.0494
River 3	1991	33%	0.3495
	1992	36%	0.3287
	1993	30%	0.2550
River 4	1991		0.0000
	1992		0.0000
	1993		0.0001
River 5			0.0125
River 6			0.0000
River 7			0.0000
Mean SD			0.0483

ii

		MIX	n = 50
River 1			0.0002
River 2			0.0123
River 3	1991		0.0001
	1992		0.0007
	1993		0.0000
River 4	1991	39%	0.3574
	1992	12%	0.1259
	1993	49%	0.4379
River 5			0.0654
River 6			0.0000
River 7			0.0000
Mean SD			0.0299

Comment
The result show a high degree of accuracy and there is no significant difference between the true and estimates values for each year of sampling. Over 93% of River 3, and over 92% of River 4 MIX samples are assigned to the correct rivers.

Experiment 7.3

This experiment further investigates temporal stability of samples in River 4, which has a large enough single year class sample in the MIX to test (n = 25). River 3 year classes were pooled and the River 4 1993 sample was used as a MIX sample in all simulations. Simulations were run using i.) all years in the stock baseline to see if the 1993 sample was estimated correctly, and ii) each year in turn in the baseline. This was to investigate if the sample from which the MIX came was not present in the baseline, could it be related back to other cohorts from the same location.

Results

i.

		MIX	
River 1			
River 2			
River 3			
River 4	1991		
	1992		
	1993	100%	0.9999
River 5			
River 6			
River 7			
Mean SD			0.0002

ii.

		MIX = River 4 1993 (n = 25)						
River 1		0.1708	River 1		0.2251	River 1		0.0000
River 2		0.1623	River 2		0.1999	River 2		0.0000
River 3		0.3619	River 3		0.3539	River 3		0.0000
River 4	1991	0.1891	River 4	—	—	River 4	—	—
	—	—		1992	0.0000		—	—
	—	—		—	—		1993	0.9999
River 5		0.0034	River 5		0.3539	River 5		0.0000
River 6		0.1122	River 6		0.1461	River 6		0.0000
River 7		0.0000	River 7		0.0000	River 7		0.0000
Mean SD		0.0971			0.1010			0.0002

Comment

The River 4 1993 sample is related back to the 1993 baseline stock very accurately, whether or not the other River 4 baseline stocks are included. However, when the 1993 baseline stock is excluded, the estimates are completely erroneous. Less than 20% of the sample is estimated to have come from the River 4 1991 stock, and none from 1992. It should be noted that the River 4 1992 baseline frequencies were based on only 6 individuals and as a results will be highly biased. A large proportion of the MIX is estimated to have come from River 3, with which River 4 has one of the lowest pairwise Fst values. Thus it appears that temporal instability, and unequal sampling errors can combine to lead to highly inaccurate GSI estimates.

Experiment 7.4

The aim of this experiment was to investigate how accurate GSI estimates were when equal or skewed proportions of the baseline stocks were included in the MIX sample.

Results

	i. MIX	n = 70	ii. MIX	n = 100
River 1	14.3%	0.1590	10%	0.1416
River 2	14.3%	0.1080	10%	0.1053
River 3	14.3%	0.1178	10%	0.0095
River 4	14.3%	0.1056	50%	0.4902
River 5	14.3%	0.1760	5%	0.0703
River 6	14.3%	0.2044	5%	0.0851
River 7	14.3%	0.1290	5%	0.1066
Mean SD		0.0766		0.0499

	iii. MIX	n = 40	iv. MIX	n = 66
River 1		0.0000	6%	0.0530
River 2		0.0000		0.0019
River 3		0.0000	12%	0.0099
River 4		0.0000	76%	0.8089
River 5	15%	0.1775		0.0055
River 6	10%	0.0994		0.0004
River 7	75%	0.7230	6%	0.1202
Mean SD		0.0459		0.0384

Comment

GSI estimates were generally accurate and mean SD small. overall the average error was less than 3.5%. In the first case where all baseline stocks are represented equally. the estimate was generally accurate with only the estimates of River 6 being significantly elevated. In the second example. all baseline stocks were represented. but the proportions skewed. The most serious error here is the underestimation of River 3. where the proportion was less than a tenth of the actual value. The third example. where only three baseline stocks were represented produced accurate estimates with the largest error being less than 3% and the non-contributing stocks are estimated as such. Again in the fourth case. non-contributing stocks are estimated at less than 1%,. However, the contribution from River 3 is again seriously underestimated.

4.8 DNA profiling

The aim of this model was to investigate the utility of DNA profiling in identifying a single parent of a mixed progeny sample. Assuming the maternal genotype, or one of several possible maternal genotypes can be matched to a known female the chance of a second unsampled female having the same genotype can be calculated.

Allele frequency		Probability of more than one identical genotype depending on no. loci screened									
Most common	All others pooled	1 locus	2 loci	3 loci	4 loci	5 loci	6 loci	7 loci	8 loci	9 loci	10 loci
0.95	0.05	0.9975	0.9950	0.9925	0.9900	0.9876	0.9851	0.9826	0.9802	0.9777	0.9753
0.9	0.1	0.9900	0.9801	0.9703	0.9606	0.9510	0.9415	0.9321	0.9227	0.9135	0.9044
0.85	0.15	0.9775	0.9555	0.9340	0.9130	0.8925	0.8724	0.8527	0.8336	0.8148	0.7965
0.8	0.2	0.9600	0.9216	0.8847	0.8493	0.8154	0.7828	0.7514	0.7214	0.6925	0.6648
0.75	0.25	0.9375	0.8789	0.8240	0.7725	0.7242	0.6789	0.6365	0.5967	0.5594	0.5245
0.7	0.3	0.9100	0.8281	0.7536	0.6858	0.6240	0.5679	0.5168	0.4703	0.4279	0.3894
0.65	0.35	0.8775	0.7700	0.6757	0.5929	0.5203	0.4565	0.4006	0.3515	0.3085	0.2707
0.6	0.4	0.8400	0.7056	0.5927	0.4979	0.4182	0.3513	0.2951	0.2479	0.2082	0.1749
0.55	0.45	0.7975	0.6360	0.5072	0.4045	0.3226	0.2573	0.2052	0.1636	0.1305	0.1041
0.5	0.5	0.7500	0.5625	0.4219	0.3164	0.2373	0.1780	0.1335	0.1001	0.0751	0.0563
0.45	0.55	0.6975	0.4865	0.3393	0.2367	0.1651	0.1152	0.0803	0.0560	0.0391	0.0273
0.4	0.6	0.6400	0.4096	0.2621	0.1678	0.1074	0.0687	0.0440	0.0281	0.0180	0.0115
0.35	0.65	0.5775	0.3335	0.1926	0.1112	0.0642	0.0371	0.0214	0.0124	0.0071	0.0041
0.3	0.7	0.5100	0.2601	0.1327	0.0677	0.0345	0.0176	0.0090	0.0046	0.0023	0.0012
0.25	0.75	0.4375	0.1914	0.0837	0.0366	0.0160	0.0070	0.0031	0.0013	0.0006	0.0003
0.2	0.8	0.3600	0.1296	0.0467	0.0168	0.0060	0.0022	0.0008	0.0003	0.0001	0.0000
0.15	0.85	0.2775	0.0770	0.0214	0.0059	0.0016	0.0005	0.0001	0.0000	0.0000	0.0000
0.1	0.9	0.1900	0.0361	0.0069	0.0013	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
0.05	0.95	0.0975	0.0095	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Comment

The likelihood of finding more than one identical genotype depends on the number of loci screened, and the frequency of the most common allele. The more polymorphic a loci is, and the more alleles that are present in a population, the lower the frequency of the most common allele. With minisatellite loci of more than 10 alleles, the most common tends to occur at a frequency of less than 0.4. At this level the probability of finding another individual with the same genotype is 0.64. If 10 loci are examined this decreases to a one in a hundred chance. This model provides a conservative estimate as it assumes that the individual has only the most common alleles at each loci, whereas in screening a number of loci this is unlikely to be the case.

4.9 SUMMARY OF RESULTS

1. The relationship between the precision of estimates and the mixture sample size was investigated by comparing the mean error of an estimate with increasing numbers of individuals from each stock contributing to the mixture sample. Simulations were run for salmon (Exp. 4.1) using 5 - 60 individuals per stock, and for sea trout (Exp. 1.2) using 5 - 40 individuals.

   The results from both data sets were similar. Mean error decreased with increasing mixture size, from 7.8 to 3.3% in salmon, and from 9.1% to 3.3% in sea trout. At a sample size of 20 individuals per stock, the mean level of error drops below 5% in both data sets.

2. Several experiments were designed to investigate the effect that choice of loci has on GSI estimates. The number of loci in Experiment 1, a sea trout intra-drainage model, was halved, and simulations run with increasing mixture sample size (Exp. 1.2). Although the level of error dropped with increasing sample size, as it did when six loci were used, error was consistently higher by 1.0 to 1.5%.

   As well as examining the number of loci, the effect of using three highly polymorphic (14 - 19 alleles per locus) loci was compared to using three moderately polymorphic loci (6 - 7 alleles) in Exp. 1.4. The more polymorphic loci consistently gave more accurate estimates, with an average error of 1.8% compared to 6.9%. Levels of error were only marginally higher than obtained when using all loci (1.4%).

   To investigate the effect of individual loci, simulations were carried out on a sea trout inter-drainage model, with data from five minisatellite loci, and mtDNA haplotype frequencies (EXP 2). Each locus was excluded in turn and compared to original estimates. Combining all data produced a highly accurate estimate of stock composition with errors less than 1.5%. The exclusion of any single locus appeared to have no significant effect in improving, or impairing the estimate. There was no evidence that mtDNA data were any more useful than any single nuclear locus.

   To further investigate the effect of using mtDNA, salmon inter- and intra- drainage models were simulated using mtDNA as the only genetic data (EXP 5). GSI estimates were realistic if the stock had a high frequency of a unique haplotype. Otherwise estimates were highly inaccurate and particularly subject to the choice of baseline stocks that were included, with errors up to 45%.

3. Stock specific bias was investigated using three data sets. The sea trout intra-drainage model (Exp. 1.7) showed low levels of bias (< 5%), typical of the mean error obtained using this data. Estimates of stock error were within 21% of the true value.

   Estimates produced by the salmon intra-drainage model (Exp. 4.2) were less accurate with stock estimates being as low as 40% of the true value. However, if the estimate across all sites of a drainage were summed, values were more realistic. Error appeared to be caused by a single sample showing high levels of bias.

The salmon inter-drainage model (Exp. 7.1) produced more consistent results with stock estimates ranging between 62% and 83% of the true value. High values of stock specific bias were caused by atypically low pairwise Fst values.

4. Individuals from stocks not included in the baseline were added to the mixture samples of Exp. 1.6, to simulate the effects of straying. As the proportion of strayers increased from 14.3% (equal proportions) to 75% of the mixture, the level of error almost doubled. Estimates became highly inaccurate when strays made up the majority of the mixture. However, even when strayers made up 25% of the mixture the maximum error was 11.5%. This doubled when strayers comprised 40%. Even when proportions were highly skewed and mixture comprised 75% strays, two non-contributing stocks were estimated as only 10% and 0%. Stocks that were most genetically similar were overestimated.

5. Temporal instability was investigated in two data sets. In Experiment 3 it was possible to use year classes of adults as baseline stocks, and their progeny as mixture samples. When multiple year classes were included in the baseline, estimates were reasonably accurate, and progeny groups were assigned to their parental year group, or to other adult year classes in the same location. Similar results were found in Exp. 7.2. However, inaccuracy increased when the parental year class was excluded from the baseline. In these simulations from 20% to 70% of progeny in Exp. 3.2 were assigned to another location. In both experiments the most accurate estimates were obtained by pooling all year classes in the baseline sample.

6. Various levels of genetic divergence were modelled. In Experiment 1, Fst values for sea trout populations were typical of intra-drainage populations. Estimated proportions were usually less than 5% away from the true value. Even when proportions were skewed, estimates were within 10% of the real value. Only when significant numbers of strayers were introduced did errors become significant.

   The sea trout rivers in Experiment 2 showed typical inter-drainage levels of divergence at nuclear loci and mtDNA haplotype frequencies. Skewed proportions in the mixture resulted in absolute errors of less than 1.2%.

   Genetic divergence among salmon samples in Experiment 7 were also typical of inter-drainage divergence, although some pairwise Fst values were only marginally significant. In this experiment all estimates were within 12% of the true value, and most errors were less than 7%.

   Unlike the previous models, the salmon stocks used in Experiment 6 showed no genetic divergence at four nuclear loci. Estimates were within 26% of the true value, and the majority less than 16%. Where mixture proportions were skewed and estimates erroneous, the ranking order of contributing was still correct. Mean errors were several times higher than those found in other experiments.

5. DISCUSSION

5.1 Effect of mixture sample size and the number of individuals from each contributing stock.

It has been noted that the accuracy and precision of GSI estimates depends in part on their being enough individuals from each contributing stock in the mixture for the proportion to be based on (Pella & Milner, 1987). If a particular stock contributes 10% to a mixture sample, it will be harder to estimate that 10% accurately if the total mixture size is 20, rather than 200.

To examine this, mixture samples made up of equal proportions and increasing sample size were simulated (Experiments 1.2 and 4.1). The precision increased rapidly as the sample size increased from 5 to 20 individuals per stock when the mean error drops below 5%. At sample size 20, the levels of precision in both salmon and trout models were almost identical. Beyond this point however, for both species, increasing the sample size had little effect on the level of error. Trebling the sample size of salmon from 20 to 60 individuals reduced the mean error from 4.69% to 3.30%. Therefore a stock represented by 20 individuals in the mixture will be adequately estimated.

Simulations of data from the River Shannon (Galvin *et al.*, 1995) showed a similar pattern. Error rates were up to three times higher that in this study (which maybe an effect of the substantial stocking in the drainage), and decreased from approximately 30% with five individuals from each stock, to 10% when that figure was increased to 100.

A direct comparison was made between a mixture of 10 , and of 30 individuals per stock (Experiment 1.1). The difference between the accuracy and the precision of each estimate was negligible, although the mean error of the larger sample size was lower. Likewise when only one stock contributes (Experiment 1.3), the estimate given for a total sample size of 10 was less than 0.01% from the true value.

Some studies of Pacific salmonids, and mathematical modelling of GSI have shown similar results. Wood *et al.* (1987) suggested that with 25 - 50 individuals per stock, depending on the number of loci used, could be accurately estimated. Beacham *et al.* (1995) using minisatellite DNA probes in stock composition estimates of sockeye salmon, found that larger mixture sample sizes had little effect on the accuracy of the estimate, although precision did increase.

In contrast to this, Koljonen & McKinnell (1996) found that in GSI estimates of Baltic Atlantic salmon, stocks had to contribute at least 40 individuals to reduce error sizes to less than 15%. However, this study used seven isozyme loci, five of which had two alleles and the remainder three. One of the main influences on the required sample size may be the number of observable possible genotypes. Thus using more variable loci, with a greater number of alleles, may reduce the sample size needed. This may also be true of the baseline stock sample size, although this was not modelled in this study. However, sample sizes less than 50, which were commonly used in this study appear adequate.

5.2 The effect of the number of loci and level of polymorphism on GSI estimates

While there have been many new developments in molecular population genetics in the past decade, very few GSI fisheries studies have utilised DNA based marker systems. These have all used one to three minisatellite markers in the study of chum salmon (Taylor *et al.*, 1994), sockeye salmon (Beacham *et al.*, 1995), and coho salmon (Miller *et al.*, 1996). No study has considered the relative benefits of marker systems, or the optimal levels of polymorphism.

All of the models in this study have used locus specific minisatellite probes and / or mitochondrial DNA variation. The minisatellite loci have all been isolated from Atlantic salmon or brown trout (Prodöhl *et al.*, 1995; Taggart *et al.*, 1995a), and have been characterised in a number of population genetic, and pedigree analysis studies. Several models in this study investigated the effects of the number of loci used and the optimum levels of polymorphism.

For example, in Experiment 1.2, simulations are rerun with six loci and three loci for increasing mixture sample sizes. The mean error is approximately 1.5% greater using three loci. While the error decreases with increasing sample size in both models, the difference between them is maintained. The mean error for the smaller locus set drops below 5% when the sample size is approximately 35. Brodziak *et al.* (1992) in a GSI study of chinook salmon compared results using 11 and 15 isozyme loci. The error was less for the larger number of loci, but the inclusion of more loci did not uniformly produce better results, and for one stock the estimate was incorrectly elevated. Likewise Shaklee *et al.* (1991) found in a GSI analysis of pink salmon, that doubling the number of isozyme loci from 14 to 28 reduced the required mixture sample size by half.

VNTR markers (mini- and microsatellite) are characterised by large numbers of alleles segregating in relatively small sample sizes. This is thought to compromise classical population genetic statistical methods (for example Weir 1992). It has been calculated that in a population with 10 alleles at a locus a sample size of over 500 would be needed to represent all possible genotypes (Chakraborty, 1992). However GSI techniques may benefit from high numbers of alleles. It appears from this study and others that a larger number of possible genotypes can reduce the sample size required (see above). Private, or semi-private alleles (that occur in a single or small group of stocks) would go some way to providing a diagnostic stock specific marker.

This was further investigated in Experiment 1.4 where simulations were run with a set of three highly polymorphic minisatellite loci (average = 16 alleles / loci), and three less polymorphic(average = 6.7 alleles / loci). Using three less polymorphic loci gave much less accurate and less precise results than using either three highly polymorphic loci or all six loci, with a mean error twice as high.

The use of mitochondrial DNA variation was also modelled in this study. The mitochondrial genome is maternally inherited in a clonal fashion down the female lineage. There is no recombination or mixing of alleles between generations, and all the progeny will inherited an identical maternal mtDNA genotype. With the advent of PCR techniques and heterologous primers that can be used in several species, there have been many population genetic studies

of salmonids using mtDNA. Certainly the number of mtDNA studies among published literature at least equals the number on nuclear DNA loci.

In Experiment 5 mitochondrial haplotype frequencies were used as the only data for baseline stocks, but it was only successful where a baseline stock had a high frequency of a unique haplotype acting in effect as a stock specific diagnostic marker. For example a mixture of individuals from the Teifi could be identified with 99.99% accuracy when the Rhyd-hir was excluded. With the Rhyd-hir in the baseline, only 55% were assigned to the Teifi. However when combined with other data, mtDNA data can be at least as effective as any other locus (Experiment 2.1). Mitochondrial data has been used for stock identification in species where there is little nuclear variation e.g. in striped bass (Wirgin *et al.*, 1993).

Modelling has demonstrated some of the advantages of minisatellite loci in GSI, compared to protein coding loci. Use of highly polymorphic loci allows smaller mixture samples to be accurately analysed with lower levels of error than those observed in GSI of Atlantic salmon using protein coding loci (Koljonen, 1995; Koljonen & McKinnell. 1996). It has been shown previously that minisatellites and microsatellites share many of the same characteristics in terms of levels of polymorphism, high heterozygosity and conformation to Hardy-Weinberg expectations. In addition, recent advances have made it possible to automate microsatellite analysis (allowing more samples to be analysed) and the use of PCR has made non-destructive sampling feasible. For example it is possible to use the adipose fin from a 0+ fry after summer growth for analysis. On the basis of this evidence, microsatellite techniques appear the most appropriate for GSI analysis of salmonids in future empirical studies.

5.3 Stock specific bias

Stock specific bias is here defined as the average percentage wrongly assigned to a baseline stock. It was evaluated by simulating mixtures of each stock in turn, and resampling the baseline and mixture data 100 times to calculate 95% confidence limits for the estimate. The values obtained by this method are generally less accurate than using the absolute values, as both baseline stocks and mixture samples are resampled.

In Experiment 1.7 the average stock specific bias ranged between 3.2% and 4.0% (or 4.1% and 5.0% using the upper confidence limits), typically in the range of mean errors obtained with this data set. The GSI estimate for the mixture ranged within 11% to 21% of the true value. Experiment 7.1 produced stock estimates within 38% of the true value. Inaccurate estimates were produced by low pairwise Fst values between certain sites. It is notable that low pairwise Fst produced inaccurate estimates in one direction, but not in the other. For example when a sample from River 2 was used as the mixture, almost 20% of the mixture was assigned to River 4. These two rivers have an atypically low Fst value of 0.0081. However, when the River 4 was used as the mixture, less than 6% was assigned to River 2. This may indicate the 'direction' of gene flow. A similar situation is apparent in Rivers 5 and 7. Even where there is low levels of genetic differentiation between baseline stocks and estimates are imprecise, useful information may be obtained by modelling real data.

Experiment 4.2 produced some anomalous results for the Lake Windermere samples. Mean estimates for the Dee samples are poor, although together they account for up to 80% of the mixture samples. The maximum stock specific bias for the Dee samples is less than 5%, but in the three Windermere samples values range from 0.18% for the Leven to 11.0% for

Troutbeck. In particular less than a half of the Leven mixture is estimated to have come from that baseline. Yet this population has more alleles than the other Windermere sites, and does not show any irregular statistical characteristics (e.g. deviation from HWE). However, the three Windermere sites together account for 91% of the mixture.

The error of estimation obviously depends in part on which stock groups are included in the baseline and the inclusion of too many non-contributing stocks can lead to serious errors in estimation (Brodziak *et al.*, 1992). Low stock specific bias estimates may be of more importance in a fisheries management context than the actual accuracy of GSI estimates. Stock specific bias could lead to a significant proportion of the mixture being assigned to a stock, that in fact does not contribute anything.

5.4 Effect of unsampled stocks contributing to the mixture

The fundamental purpose of GSI is to assign mixtures samples to one or more of the baseline stocks. Therefore all stocks that contribute to the mixture need to be sampled. This may be impractical as a) potential contributing stocks may not be known or b) the number of potential stocks is too large to sample.

Experiment 1.6 investigated the effect of individuals from an unsampled baseline stock. With all stocks contributing equal proportions the maximum error is 8.4%. As the proportion of strayers increases to 50%, this rises to 19%, and when the strayers make up 75% of the mixture the estimate for one stock is almost ten times greater than the true value. The mean error also increases with the proportion of strayers, almost doubling. One stock in particular (Buchat - Site 4) shows a particularly high level of error. This is also the population that has the lowest pairwise Fst value. This suggests that strayers are assigned to populations to which they are most genetically similar, and that the size of the bias increases with the contribution of the strayers. However, even when strayers comprise 25% of the mixture, the maximum absolute error is 11.43%, and the average absolute error is only 5.3%. This demonstrates that useful estimates can be made even when this level of error is introduced.

An alternative GSI analysis (Smouse *et al.*, 1990) uses an unconditional likelihood method which, rather than resampling or estimating the error of the baseline and mixture data, treats them as experimental variables. This leads to a convergent stock estimate and permits estimates of the contribution of an unsampled stock in the mixture The larger the contribution and the more genetically divergent the unsampled stock is, the more likely its contribution in the mixture will be detected.

5.5 Effects of temporal instability

Temporal stability is a prerequisite for GSI analysis (Wood, 1989; Shaklee *et al.*, 1991; Brodziak *et al.*, 1992). Experiment 3 using Girnock and Baddoch data is maybe the most realistic model. A run of adult fish was used as a baseline, and random samples of their progeny were used as the mixture. The use of the progeny genotypes as mixture samples ensures the genotypes of the adults are recombined, or mixed, through random mating and subject to random genetic changes. When all year classes are included in the baseline, stock estimates are reasonably accurate, and the majority of individuals not assigned to their own year class of adults are assigned to another year class at the same location. Likewise in Experiment 7.2 and 7.3i, a mixture of all year classes were assigned to their own parental year with a high degree of accuracy, the highest absolute error in any case being 5.2%. In all cases

the most accurate estimates are obtained by pooling all year classes in the baseline sample (95% for Girnock and 99.9% for Swilly).

However, when the parental year class is excluded from the baseline stocks, significant levels of error are recorded. For example in the worst case (Experiment 3.2), if allele frequencies from 1993 adults were the only baseline data available for the Girnock, 70% of Girnock 1992 progeny would be assigned to the Baddoch. It is notable that the number of adults spawning in the Girnock in 1993 was approximately half the number spawning in previous years, and this may cause biased allele frequencies. In Experiment 7.3ii., the Swilly 1993 year class was poorly assigned to 1991 (20.0%) and 1992 (0.0%), although notably the 1992 baseline was based on just six individuals.

The question of temporal stability cannot be easily resolved without empirical testing, as the existence of instability, the level of instability, and the effect it has on GSI estimates, are highly specific to the baseline stock involved. In the long term studies of Pacific salmonids genetic baseline data from important contributing rivers is updated annually, and less important stocks are updated on a three-yearly cycle. The evidence suggests that the levels of temporal instability observed are minimal compared to the amount of genetic divergence between drainages, and GSI analysis of catches from several rivers are unlikely to be biased. However, it is possible that observed genetic differentiation between populations within rivers may be low enough to be subject to such error.

In the absence of long term data collections two possible strategies could be adopted to eliminate, or at least minimise the effects of any possible temporal instability, in an intra-drainage GSI study. The ideal situation would be to perform GSI analysis on the same cohort, or cohorts, of the stock. That is to sample juveniles or smolts in the baseline stock, and to resample the same cohort in the catch mixture. For example, sampling of 2+ juveniles early year would allow GSI analysis of grilse (that have smolted after two years in freshwater) in the autumn of the following year.

A second method of eliminating the effects of instability, is to sample the adult catch before the baseline is surveyed. That is, if a high proportion of ascending adults were sampled non-destructively, the next years cohort, i.e. their progeny, could be sampled, to provide baseline data. This would give a GSI analysis for that cohort of adults that is not subject to the effects of straying, or temporal instability. This approach uses the same sampling strategy as the proposal outlined for DNA profiling and the two approaches could be combined.

5.6 Degree of genetic divergence needed between baseline stocks

The nature of anadromy means that populations of sea trout and Atlantic salmon can be wholly or partially reproductively isolated, leading to and maintaining genetic differentiation. Non-genetic (trapping and tagging) studies have shown Atlantic salmon home back to natal streams within a river. Thus it seems that population structuring occurs on a sub-catchment level. This is borne out by many genetic studies of salmon and trout showing stable genetic differences within and among rivers. The degree of genetic differentiation depends on geographical distance, and on the relative amount of interaction between populations. This study made some attempts to investigate how much differentiation is needed for accurate GSI estimates.

The measure of genetic differentiation commonly used here is Fst, or Φst when mitochondrial genes are being considered. Values range from 0 were there is no genetic differentiation to 1.0 indicating no genetic exchange between populations. Anadromous species fall between two extremes in levels of genetic differentiation (Gyllensten, 1985) as although each population reproduces in physical isolation, adults at sea have the opportunity to stray into non-natal rivers. Landlocked species, or landlocked populations of salmon and trout, have no opportunity to stray, and so typically display high Fst values. Populations of marine species are not subject to any actual physical barriers, apart from distance and hostile environments, and often show little or no genetic differentiation leading to low Fst values. Jordan (1992) reviewed Atlantic salmon isozyme variation in the British Isles. Combining 18 separate studies it was found that Fst values within sites on a river were on average 0.023. The level of differentiation between rivers, in which data from different tributaries are pooled to form a single sample, were on average twice as high (0.046).

In Experiment 1 the over all sea trout samples Fst was significant at 0.019 although the range of values went from near zero (0.006) to near inter-drainage values (0.037), and six minisatellite loci were used. GSI estimates were generally within the 5% absolute error range. Where the proportions in the mixture where highly skewed, the level of error was under 10%. In all but one case, non-contributing stocks were identified as contributing less than 2%. Only when 'strayers' were introduced did error levels become very significant.

The rivers used in Experiment 2 had an overall Fst of 0.037, typical of that between rivers, although pairwise values ranged from 0.015 to 0.082) There was also significant mitochondrial DNA divergence between these samples. (Φst = 0.200). Here skewed mixture proportions resulted in absolute errors at a maximum of 1.2% when all five minisatellite loci and mtDNA data were included in the analysis.

The salmon samples used in Experiment 7 had a significant Fst over all of 0.032, (0.0062 - 0.0744). Here all estimates were within 12% of the true value and most within 7%. The inaccuracies were mostly due to low pairwise Fst values. Three minisatellite loci were used in this model.

The samples used in Experiment 6, from the Tweed, were not genetically different (Fst = 0.003) at the 5% level. MtDNA divergence was also non-significant. Three Tweed samples from the upper and middle Tweed were found could not be differentiated with four minisatellite loci, or with mtDNA. When the mtDNA data was excluded however, all estimates had absolute errors less than 25%, and the majority less than 16%. Importantly, where the mixture proportions were skewed, the estimates at least ranked the stocks correctly, and the most and least important contributors were identified. However, the mean SD for these estimates were up to 10 times greater than in other experiments.

In GSI studies of Pacific salmonids baseline stocks are assumed to be genetically differentiated at the 5% level, if not, samples are pooled (for example Shaklee *et al.*, 1991; Marshall *et al.*, 1991). Models here show that even where there is little or no differentiation GSI can provide some information of interest in fisheries management. Genetically similar tributaries of a river may geographically proximate and form a natural single management unit. With regard to GSI, estimates could be made for such a group of tributaries as opposed to upper more differentiated streams. The amount of grouping, and the amount of genetic

homogeneity that can be tolerated, depends therefore on the sensitivity required. For example, to assign the percentage of a catch to tributaries dominated by spring fish, it is only necessary that those are differentiated from other grilse stocks. It is not required that the grilse tributaries are differentiated from each other. Or, if the relative importance of rivers in a mixed stock fishery is to be assess, it may take the minimum amount of genetic differentiation to produce an accurate ranking order. Likewise, it may be a simple question of whether a stock contributes or not, rather than an absolute value.

5.7 DNA profiling

DNA profiling has been used extensively in forensic science and in the study of parentage and mating success (e.g. Burke, 1989). It is due to the high levels of polymorphism commonly seen at minisatellite and microsatellite loci that it is possible to apply these techniques in wild randomly mating populations. The model here illustrates how DNA profiling could be applied. If ascending adults are wholly or partially sampled, and the offspring can be assigned to one of these on the basis of genotype at ten loci, the probability of that match being correct is high. The likelihood of the match being correct is even higher if the adults are sampled ascending into a tributary, or group of tributaries. This approach is limited to examining one generation at a time, but is not subject to temporal instability, straying or sampling error. It is also more informative and cost effective than radio tracking, and provides information not only on the movements of individuals, but also on reproductive success.

6. CONCLUSIONS

Previous studies on genetic stock identity in Pacific salmonids, and to a limited extent on Atlantic salmon, have suggested that the technique is of use in fishery management of salmonid species. This study supports that view, and has gone some way in establishing that Atlantic salmon and sea trout populations meet the criteria for application of GSI. This study has also shown that an acceptable level of accuracy and precision can be attained under some circumstances, but not in others.

GSI could be empirically applied to Atlantic salmon and sea trout populations by applying standard sampling techniques and DNA-based genetic screening methods. Salmonid populations commonly show hierarchical genetic diversity, with tributaries within a river being more similar than rivers, rivers within a region being more similar than distant rivers etc. Therefore it can be assumed that genetic differentiation exists in possible rivers of stocks of interest, although widespread stocking or enhancement, and population instability must be ruled out, as these may effect the reliability of GSI estimates.

Modelling has shown that screening of Atlantic salmon and sea trout with minisatellite genetic markers meets many of the criteria needed for GSI analysis, and in the case of Atlantic salmon has been shown to be a considerable advance on isozyme based analysis. With recent advances in microsatellite screening, it is anticipated that this would be the genetic marker of choice as larger sample sizes / more loci can be screened, and the characteristics of the polymorphisms are similar to minisatellites. A common feature of minisatellite and microsatellite genetic markers is the high number (>15) of alleles found in population samples, compared to traditional protein markers which are commonly diallelic. Where the sample size is inadequate this may lead to sampling error and the genetic information about a baseline population may not be accurate. The minimum sample size needed to ensure that 20 alleles (which occur at equal frequencies) are detected is 60 (Chakraborty, 1992). Data sets used here for modelling GSI have an average baseline stock sample size of 55 - 60 individuals, and have provided adequate stock composition. However, a sample size of 75 may provide a useful comprise between accuracy and practical considerations. The number of alleles occurring at a particular genetic marker may also effect the minimum number of each stock that can be accurately estimated. Modelling has shown that a minimum of 20 individuals can be accurately estimated in a mixture sample.

Straying is common to salmonid populations. The level of gene flow between populations effects the level of genetic differentiation. Where the genotype of an immigrant cannot be reconciled with the allele distribution of baseline stocks, the individual cannot be assigned. Otherwise 'strayers' will be assigned to the baseline sample to which they are most genetically similar. The effect of 'straying' on baseline samples cannot be estimated without empirical data. In a 'catastrophe' event where all or a major part of a cohort is prevented from homing to spawn, GSI will be comprised. However, the effect depends on the size of the spawning escapement, both in the population of interest, and in the populations they might enter. It also depends on the level of genetic divergence between populations before and after the spawning event, and may require that the stocks are considered as a spawning 'unit' rather than discreet populations. As spawners are likely to enter the nearest 'available' spawning area this may not necessarily compromise management aims.

The question of temporal stability among baseline stocks remains unresolved. The amount of variation that could be caused by instability within a stock is minimal compared to inter-drainage differentiation. However, in comparisons between more genetically similar stocks within drainages, temporal instability could cause bias in stock composition estimates. Modelling has shown that temporal instability in baseline stocks is a limiting factor in obtaining accurate GSI estimates arising from sampling error and inter-cohort instability. While the former can be addressed by experimental design, the latter cannot be estimated without empirical study. The use of preserved samples (e.g. dried scales) that include a considerable number of cohorts could be studied to investigate temporal instability in that river, and to derive more general information that could be applied in fisheries studies. The effects of temporal instability are negated when the baseline and mixture sample are of the same cohort(s). Sampling strategy must ensure adequate identification and sampling of possible contributing stocks, especially in intra-drainage studies which focus on particular stock components. Thus the baseline data must be representative of the population in a tributary or river. Mixed stock sampling must include all stock components considering seasonal variation.

Modelling has shown that GSI estimates with absolute errors of less than 5% for sea trout and less than 12% for Atlantic salmon are possible where significant genetic differentiation exists. Two observations may be of particular interest with regard to fisheries management. Firstly a particular feature of simulations was the generally accurate estimates provided for non-contributing stocks. Secondly, even in the absence of significant genetic differentiation, the ranking of contributions was accurate, although the absolute values may be erroneous.

7. RECOMMENDATIONS

Minisatellite and microsatellite loci are highly suitable for GSI analysis, and with rational and strategic sampling of baseline and mixture samples, can provide accurate estimates of stock composition. The recent advances in microsatellite techniques enabling larger samples to be screened, and the highly polymorphic nature of these loci, suggest that microsatellites should be the loci of choice in future GSI studies in Atlantic salmon or sea trout.

Empirical studies of GSI, and tests using tagged fish have authenticated GSI analyses in Pacific salmonids. Modelling of Atlantic salmon data sets has shown that the precision and accuracy achieved is comparable with that seen in Pacific salmonids studies. Therefore, it is suggested that further validation of GSI in Atlantic salmon using tagging, or tracking, is not required. Sea trout populations are demonstrated to be more amenable to GSI techniques, because of a higher level of genetic variability and divergence. However, it is concluded that GSI can be applied to both species to answer successfully many management related questions.

Conclusions drawn from modelling should be empirically tested on an inter-drainage, and intra-drainage context for both species.

7.1 Empirical inter-drainage GSI analysis of Atlantic salmon

GSI should be empirically tested on an inter-drainage basis examining catches from coastal fisheries where the catch has the potential to originate from up to ten drainages.

The aims of such a study would be to:

a) assign proportions of commercial catches to management areas, and more specifically to drainages.
b) to examine seasonal changes in the catches.
c) enable realistic management exercises in protecting over-exploited stocks.

The study should be carried out by:

a) baseline sampling of ten potentially contributing drainages. Baseline samples would consist of approximately 75 juveniles from each river, representing all of the major tributaries.
b) preliminary screening of samples with five highly polymorphic microsatellite loci.
c) modelling GSI analysis of this data in order to confirm that mixed stock analysis of those populations is possible. On the basis of information gathered and modelling of data, the sampling strategy and / or genetic analysis could be modified, or optimised. For example, more loci could be used to increase resolution.
d) sampling of net fisheries, at regular time intervals.

Possible locations of study:

This study should be carried out in a mixed stock commercial fishery of interest. At least five drainages should potentially contribute to the fishery which produce significant numbers of smolts. Locations sampled within these drainages should be numerically stable populations, and major smolt producing stocks. Possible areas for study included the north-east coast of England straddling the Border regions. In 1994 over 42,000 salmon and grilse were caught off

the Northumbrian coast. Baseline stocks could include the Tay, Forth and Tweed in Scotland, and the Coquet, Tyne, Wear and Esk in England. Another area of interest is the Morecambe Bay fishery.

7.2 Empirical intra-drainage GSI analysis of Atlantic salmon

GSI should be empirically tested on an intra-drainage basis examining the potential exploitation and environmental impact on salmon from a number of major tributaries. The study should also determine catches of different sea-age and run-time components.

The aims of such a study would be to:

a) assign proportions of net or angling catches to tributaries within a drainage.
b) examine seasonal changes in the catch.
c) identify over-exploited populations that may require reductions in exploitation rates or environmental protection.
d) examine any relationship between time of entry into freshwater, and tributary of origin. It would be possible to confirm that early running fish originate throughout the drainage, or were restricted to a small number of tributaries, that should be environmentally protected. The origins of late running fish may also be of interest in relation to competition with early fish.
e) alternatively, or in combination with the above aims, DNA profiling could be used to examine the behaviour of individuals from an early-running stock component. This would require appropriate sampling facilities such as released angled fish, or fixed/temporary trapping facilities, by which adult fish could be non-destructively sampled.

The study should be carried out by:

a) baseline sampling of all potentially contributing tributaries. Samples would consist of approximately 75 juveniles of all age classes from each tributary.
b) preliminary screening of samples with five highly polymorphic microsatellite loci.
c) modelling GSI analysis of the data in order to confirm that mixed stock analysis of those populations is possible. On the basis of information gathered and modelling of data, the sampling strategy and / or genetic analysis could be modified, or optimised. For example, more loci could be used to increase resolution.
d) sampling net and / or angling catches.

An alternative strategy to sampling which would evaluate, and eliminate, baseline temporal instability should be carried out by:

a) a preliminary baseline sampling of all tributaries potentially contributing significant numbers of smolts. Samples would consist of approximately 50 juveniles from a single year class. Samples would be screened at five highly polymorphic microsatellite loci.
b) sampling commercial and / or angling catches throughout the season.
c) sampling the following years cohort as emerging alevins or fry in all major tributaries. Samples would consist of 75 individuals.
d) Screening adult catches and progeny with five highly polymorphic loci and performing GSI analysis to assign catch to baseline stocks.
e) comparing progeny from a) and d) to ascertain of there is temporal stability in those locations, and so establishing if baseline data can be used in future study.

DNA profiling could be done in combination with the above strategy. The GSI analysis of morphologically grouped fish would determine if early-running fish originate throughout the system, or are restricted to particular tributaries. DNA profiling will also allow the study of individual fish. The additional steps required are:

b.1) sampling early-running fish non-destructively in an adult trap, or other catch and release facilities.

d.1) screening early running adults, and progeny from the baseline tributaries which are thought to produce early running fish at an additional 5+ microsatellite loci. The intensity of progeny sampling could only be determined with respect to the circumstances involved, such as the number of adults sampled, and the geographical distribution of sampled adults in the river or tributary. That is, a tributary, or group of tributaries could be sampled intensively, providing precise information about the adults. However, the same level of sampling could not be carried out throughout a drainage.

Possible locations of study:

This study should be carried out in a drainage system that has stock components with a wide range of time of entry into freshwater. The tributaries sampled should be numerically stable, and not been subject to widespread stocking. The sampling of adult catches throughout the season must be representative. Possible rivers for study are the Exe, Eden, Wye, Dee, Lune or Usk.

7.3 Empirical inter-drainage GSI analysis of sea trout

GSI should be empirically tested on an inter-drainage basis examining catches from coastal fisheries where the catch has the potential to originate from five to ten drainages.

The aims of such a study would be to:

a) assign proportions of commercial catches to management areas, and more specifically to drainages.

b) to examine seasonal changes in the catches.

c) enable realistic management exercises in protecting over-exploited stocks.

The study should be carried out by:

a) baseline sampling of ten potentially contributing drainages. Baseline samples would consist of approximately 75 juveniles from each river, representing all of the major tributaries.

b) preliminary screening of samples with five highly polymorphic microsatellite loci.

c) modelling GSI analysis of this data in order to confirm that mixed stock analysis of those populations is possible. On the basis of information gathered and modelling of data, sampling strategy and / or genetic analysis could be modified, or optimised. For example, more loci could be used to increase resolution.

d) sampling of net fisheries, at regular time intervals.

Possible locations of study:

This study should be carried out in a mixed stock net fishery of interest. At least five drainages should potentially contribute to the fishery which produce significant numbers of smolts. Locations sampled within these drainages should be numerically stable populations, and major smolt producing stocks. Possible areas for study included the north-east coast of England straddling the Border regions. In 1994 over 31,000 sea trout were caught off the

Northumbrian coast. Previous tagging studies have shown that a significant proportions of these may be from Scottish rivers. Baseline stocks could include the Tay, Forth and Tweed in Scotland, and the Coquet, Tyne. Wear and Esk in England.

7.4 Empirical intra-drainage GSI analysis of sea trout

GSI should be empirically tested on an intra-drainage basis examining the exploitation and environmental impact on sea trout from a number of major tributaries. The study should also determine catches of different sea-age and run-time components such as whitling.

The aims of such a study would be to:

a) assign proportions of net or angling catches to tributaries within a drainage.
b) examine seasonal changes in the catch.
c) identify over-exploited populations that may require reductions in exploitation rates or environmental protection.
d) examine any relationship between time of entry into freshwater, and tributary of origin.
e) alternatively, or in combination with the above aims, DNA profiling could be used to examine the behaviour of individuals. This would require appropriate sampling facilities such as released angled fish, or fixed/temporary trapping facilities, by which adult fish could be non-destructively sampled.

The study should be carried out by:

a) baseline sampling of all potentially contributing tributaries. Samples would consist of approximately 75 juveniles of all age classes.
b) preliminary screening of samples with five highly polymorphic microsatellite loci.
c) modelling GSI analysis of the data in order to confirm that mixed stock analysis of those populations is possible. On the basis of information gathered and modelling of data, sampling strategy and / or genetic analysis could be modified, or optimised. For example, more loci could be used to increase resolution.
d) sampling net and / or angling catches. GSI analysis of morphologically grouped individuals (e.g. whitling) would determine if whitling originate throughout the system or are restricted to particular tributaries.

An alternative strategy to sampling which would evaluate, and eliminate, baseline temporal instability should be carried out by:

a) a preliminary baseline sampling of all tributaries potentially contributing significant numbers of smolts fish run. Samples would consist of approximately 50 juveniles from a single year class. Samples would be screened at five highly polymorphic microsatellite loci.
b) sampling net and / or angling catches throughout the season.
c) sampling the following years cohort as emerging alevins or fry in all major tributaries. Samples would consist of 75 individuals.
d) Screening adult catches and progeny with five highly polymorphic loci and performing GSI analysis to assign catch to baseline stocks.
e) comparing progeny from a) and d) to ascertain of there is temporal stability in those locations, and so establishing if baseline data can be used in future study.

Possible locations of study:
This study should be carried out in a drainage system that has stock components with a wide range of time of entry into freshwater. The tributaries sampled should be numerically stable, and not been subject to widespread stocking. Tributaries sampled should be numerically stable. The sampling of adult catches throughout the season must be representative. Possible rivers for study are the Teifi, Tywi, or Lune.

7.5 Cost and time scale for empirical studies

The current cost (inclusive of all labour, consumables and indirect costs) of screening for 5 microsatellites is some £15 per individual fish. This cost is likely to decrease over the next couple of years as more automated procedures are developed. However, decreased labour costs will be to some extent negated by increased capital and depreciation costs of the automated equipment and is not likely to be less than £10 per fish. At the moment, an experienced researcher can analyse some 4000 individuals in a working year, and this is likely to increase some two to three fold with automation. As each of the empirical GSI trials recommended above involves about 1,000 individuals, laboratory analysis (excluding fieldwork) of each would require about three months to carry out at a cost of £15,000 (based on 1997 costs). Costs for studies involving DNA profiling are similar as the laboratory analysis is the same for both approaches.

8. REFERENCES

Allendorf, F. W. & Phelps, S. R. (1981). Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences*, **31**, 1507-1514.

Altukov, Y.P. & Salmenkova, E.A. (1987). Stock transfer relative to natural organisation, management, and conservation of fish populations. In *Population Genetics and Fishery Management* (Ryman, N. & Utter, F., eds), pp 333-343. London: University of Washington Press.

Beacham, T.D., Withler, R.E. & Gould, A.P. (1985). Biochemical genetic stock identification of pink salmon (*Oncorhynchus gorbusha*) in southern British Columbia and Puget Sound. *Canadian Journal of Fisheries and Aquatic Sciences*, **42**, 1474-1483.

Beacham, T.D., Withler, R.E., Murray, C.B. & Barner L.W. (1987). Biochemical genetic survey and stock identification of chum salmon (*Oncorhynchus keta*) in British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences*, **44**, 1702-1713.

Beacham, T.D., Withler, R.E. & Wood, C.C. (1995). Stock identification of sockeye salmon by means of minisatellite DNA variation. *North American Journal of Fisheries Management*, **15**, 249-265.

Berg, O.K. & Berg, M. (1987). Migrations of sea trout, *Salmo trutta* L., from the Vardnes River in northern Norway. *Journal of Fish Biology*, **31**, 113-121.

Brodziak, J., Bentley, B., Bartley, D., Gall, G.A.E., Gomulkiewicz, R. & Mangel, M. (1992) Tests of genetic stock identification using coded wire tagged fish. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 1507-1517.

Burke, T. (1989). DNA fingerprinting and other methods for the study of mating success. *Trends in Ecology and Evolution*, **4**, 139-144.

Chakraborty, R. (1992). Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. *Human Biology*, **64**, 141-159.

Clifford, S. (1996). The genetic impact on native Atlantic salmon populations resulting from the escape of farmed salmon. PhD. Thesis. The Queens University of Belfast

Crozier, W. W. & Moffett, I. J. J. (1989). Amount and distribution of biochemical-genetic variation among wild populations and a hatchery stock of Atlantic salmon, *Salmo salar* L., from north-east Ireland. *Journal of Fish Biology*, **35**, 665-677.

Donaghy, M.J. & Verspoor E. (1997). Egg survival and timing of hatch in two Scottish Atlantic salmon stocks. *Journal of Fish Biology*, **51**, 000-000.

Ferguson, A., & Fleming, C. C. (1983). Evolutionary and taxonomic significance of protein variation in the brown trout (*Salmo trutta* L.) and other Salmonid fishes. In G. S. Oxford & D. Rollinson (Eds.), *Protein polymorphism: Adaptive and taxonomic significance* (pp. 86-99). Academic Press. London.

Ferguson, A., Taggart, J.B., Prodöhl, P.A., McMeel, O., Thompson, C., Stone, McGinnity, P. & Hynes, R.A. (1995). The application of molecular markers to the study and conservation of fish populations, with special reference to *Salmo*. *Journal of Fish Biology*, **47**, (Suppl A), 103-126.

Galvin, P., McKinnell, S., Taggart, J.B., Ferguson, A., O'Farrell, M. & Cross, T.F. (1995) Genetic stock identification of Atlantic salmon using single locus minisatellite DNA probes. *Journal of Fish Biology*, **47** (Suppl A) 186-199.

Galvin, P., Taggart, J., Ferguson, A., O'Farrell, M. & Cross, T. (1996). Population genetics of Atlantic salmon (*Salmo salar*) in the River Shannon system in Ireland; An appraisal using single locus minisatellite (VNTR) probes. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 1933-1942.

Gyllensten, U. (1985). The genetic structure of fish: differences in the intraspecific distribution of biochemical genetic variation between marine, anadromous and freshwater species. *Journal of Fish Biology*, **26**, 691-699.

Hansen, M.M., Loeschcke, V., Rasmussen, G., & Simonsen, V. (1993). Genetic differentiation among Danish brown trout (*Salmo trutta*) populations. *Hereditas*, **118**, 177-185.

Hansen, M.M., Nielsen, E.E., & Mensberg, K.-L.D. (1997). The problem of sampling families rather than populations: Relatedness among individuals in samples of juvenile brown trout (*Salmo trutta* L.). *Molecular Ecology*. **6**, 469-474.

Hartley, S.E. & Pickering, A.D. (1994) Genetic differentiation between Atlantic salmon populations in the Windermere catchment. *Institute of Freshwater Ecology* NRA (North West Region)

Hindar, K., Ryman, N., & Utter, F. (1991). Genetic effects of cultured fish on natural fish populations. *Canadian Journal of Fisheries and Aquatic Science*, **48**, 945-957.

Hurrell, R.H. & Price, D.J. (1993) Genetic variation in Atlantic salmon, *Salmo salar* L., within the Tamar catchment on south-west England. *Journal of Fish Biology*, **42**, 153-156.

Jordan, W.C. (1992). Genetic protein variation in Atlantic salmon (*Salmo salar*) from the British Isles. Unpublished MAFF contract report. *Unpublished*, 43 pp.

Jordan, W.C. & Youngson, A.F. (1992). The use of genetic marking to assess the reproductive success of mature male Atlantic salmon parr *Salmo salar* L. under natural spawning conditions. *Journal of Fish Biology*, **41**, 613-618.

Koljonen, M.-L. (1995) Distinguishing between resident and migrating Atlantic salmon (*Salmo salar*) stocks by genetic stock composition analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 665-674.

Koljonen, M.-L. & McKinnell, S. (1996) Assessing seasonal changes in stock composition of Atlantic salmon catches in the Baltic Sea with genetic stock identification. *Journal of Fish Biology*, **49**, 998-1018.

Marshall, A.R., Miller, M., Busack, C. & Phelps, S.R. (1991). *Genetic stock identification analysis of three 1989 Washington Ocean and Strait of Juan de Fuca chinook salmon fisheries*. Washington Dept of Fisheries. GSI Summary Report 91-1.

Masuda, M., Nelson S. & Pella, J. (1991). *User's manual for GIRLSEM, GIRLSYM, and CONSQRT*. The computer programs for computing conditional maximum likelihood estimates of stock composition from discrete characters. Auke Bay: USA-DOC-NOAA-NMFS Auke Bay Laboratory.

Miller, K.M., Withler, R.E. & Beacham, T.D. (1996). Stock identification of coho salmon (*Oncorhynchus kisutch*) using minisatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**, 181-195.

Milner, G.B., Teel, D.J., Utter, F.M. & Winans, G.A. (1985). A genetic method of stock identification in mixed populations of Pacific salmon. *Oncorhynchus* spp. *Marine Fisheries Review*, **47**, 1-8.

Móran, P., Pendás, A.M., Izquierdo, J.I., Loboncervia, J., & García-Vázquez, E. (1995). Temporal stability of isozyme allele frequencies in wild populations of brown trout (*Salmo trutta* L.). *Hereditas*, **123**, 221-225.

Móran, P., Pendás, A.M., Beall, E., & García-Vázquez, E. (1996). Genetic assessment of the reproductive success of Atlantic salmon precocious parr by means of VNTR loci. *Heredity*, **77**,

O'Connell, M., Skibinski, D.O.F. & Beardmore, J.A. (1995). Mitochondrial DNA and allozyme variation in Atlantic salmon (*Salmo salar*) populations in Wales. *Canadian Journal of Fisheries and Aquatic Sciences*, **52**, 171-178.

O'Connell, M., Skibinski, D.O.F. & Beardmore, J.A. (1996). Allozyme and mtDNA divergence between Atlantic salmon populations in North Wales. *Journal of Fish Biology*, **48**, 1023-1026.

Pella, J.J. & Milner, G.G. (1987) Use of genetic marks in stock composition analysis. In *Population Genetics and Fishery Management* (Ryman, N. & Utter, F., eds), pp 246-276. London: University of Washington Press.

Pella, J., Masuda, M., & Nelson, S. (1996). Search algorithms for computing stock composition of a mixture from traits of individuals by maximum likelihood. NAOO Technical Memorandum, **NMFS-AFSC-61**, 67pp.

Prodöhl, P.A., Taggart, J.B. & Ferguson, A. (1995). Single locus inheritance and joint segregation analysis of minisatellite (VNTR) DNA loci in brown trout (*Salmo trutta* L.). *Heredity*, **73**, 556-566.

Raymond, M. & Rousset, F. (1995). GENEPOP (Version 2): A population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.

Ryman, N., Allendorf, F. W., & Ståhl, G. (1979). Reproductive isolation with little genetic divergence in sympatric populations of brown trout (*Salmo trutta*). *Genetics*, 92, 247-262.

Seeb, J.E., Seeb, L.W. & Utter, F.M. (1986). Use of genetic marks to assess stock dynamics and management programs for Chum salmon. *Transactions of the American Fisheries Society*, **115**, 448-454.

Shaklee, J.B., Klaybor, D.C., Young, S. & White, B.A. (1991) Genetic stock structure of pink salmon, *Oncorhynchus gorbuscha* (Walbaum), from Washington and British Columbia and potential mixed-stock fisheries applications. *Journal of Fish Biology*, **39** (Suppl A), 21-34.

Smouse, P.E., Waples, R.S. & Tworek, J.A. (1990). A genetic mixture analysis with incomplete source population data. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**, 620-634.

Taggart, J.B., Prodöhl, P.A. & Ferguson, A. (1995a) Genetic markers for Atlantic salmon (*Salmo salar* L.): single locus inheritance and joint segregation analyses of minisatellite (VNTR) DNA loci. *Animal Genetics* **26**, 13-20.

Taggart, J.B., Verspoor, E., Galvin, P., Morán, P., & Ferguson, A. (1995b) A minisatellite DNA marker for discriminating between European and North American Atlantic salmon (*Salmo salar* L.). *Canadian Journal of Fisheries and Aquatic Science*, **52**, 2305-2311.

Taylor, E.B. & Beacham, T.D. (1994). Population structure and identification of north Pacific Ocean chum salmon (*Oncorhynchus keta*) revealed by an analysis of minisatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1430-1442.

Thompson, C. (1995) Population genetics of anadromous brown trout (*Salmo trutta* L.) in Scotland and Ireland. PhD. Thesis. The Queens University of Belfast

Utter, F. & Ryman, N. (1993). Genetic markers and mixed stock fisheries. *Fisheries*, **18**, no.8, 11-21.

Webb, J.H., Younson, A.F., Thompson, C.E., Hay, D.W., Donaghy, M.J. & McLaren, I.S. (1993). The spawning of escaped farmed salmon (*Salmo salar* L.) in western and northern Scottish rivers. *Aquaculture and Fisheries Management*, **24**, 663-670.

Weir, B. S. (1992). Population genetics in the forensic DNA debate. *Proceeding of the National Academy of Sciences USA*, **89**, 11654-11659.

Wirgin, I., Maceda, L., Waldman, J.R. & Crittenden R.N. (1993). Use of mitochondrial DNA polymorphisms to estimate the relative contributions of the Hudson River and Chesapeake Bay striped bass stocks to the mixed fishery on the Atlantic coast. *Transactions of the American Fisheries Society*, **122**, 669-684.

Wood, C.C., McKinnell, S., Mulligan, T.J. & Fournier, D.A. (1987). Stock identification with the maximum-likelihood mixture model: Sensitivity analysis and application to complex problems. *Canadian Journal of Fisheries and Aquatic Sciences*, **44**, 866-881.

Wood, C.C. (1989). Utility of similarity dendrograms in stock composition analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**, 2121-2128.

Wood, C.C., Rutherford. D.T. & McKinnell, S. (1989). Identification of sockeye salmon (*Oncorhynchus nerka*) stocks in mixed stock fisheries in British Columbia and southeast Alaska using biological markers. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**, 2108-2120.

9. GLOSSARY

Allele - variant of a DNA sequence within a gene or region of DNA. Diploid organisms have two alleles at each gene, one inherited form each parent. The alleles may be of the same size, making the individual **homozygous**, or of different sizes, **heterozygous**. The frequency of the occurrence of each allele can be calculated from all individuals in a sample, and the resulting allele frequency distribution is taken to be approximately representative of that population.
Base pairs / bp - the partnership of nucleotide bases in a DNA sequence, often used as a measurement of the size of a sequence. 1000 base pairs make a kilobase.
Baseline stock - a population that contributes, or is likely to contribute, to the fishery of interest. A baseline stock may contribute any proportion of the mixture sample.
Bootstrapping - A statistical method based on repeated random sampling with replacement from an original sample to provide a collection of new estimates of a particular parameter, from which confidence limits can be calculated.
Electrophoresis - size separation of DNA fragments through a supportive solid matrix.
Fst - the correlation of genes of different individuals in the same population, relative to random genes from the whole species, thus providing a measure of the level of inbreeding within subdivisions relative to the total sample. Fst values range between 0 (indicating complete panmixia) and 1.0 (indicating complete absence of current or past gene flow). The significance of Fst values can be confirmed if their 95% confidence interval (CI) does not include zero. The equivalent value for mitochondrial DNA is **Φst**, although due to the maternal inheritance of mtDNA, the two values are not directly comparable.
Genotype - the genetic constitution of an organism with may partly or wholly be expressed in the phenotype.
Haplotype - the mitochondrial genome equivalent of a genotype, which is maternally inherited and not subject to recombination.
Hardy-Weinberg equilibrium / HWE - an equilibrium of genotypes achieved in a population of infinite size, mating at random, and in which there is no immigration, emigration, or selection acting on genotypes. With two alleles A and B of frequency p and q the HWE proportions of the possible genotypes AA, AB and BB will be p^2, $2pq$ and q^2 receptively.
Heterozygous - individual with different sized alleles at a particular gene. The heterozygosity of a population is a measure of how many individuals are heterozygous at one or more loci, and can taken as a measure of the relative level of variability at that locus in that population. Heterozygosity values are given as a proportion ranging from 0 to 1.0.
Homozygous - individual with identical alleles at a particular gene.
Locus (pl. loci) - site or region on the chromosome where a particular gene is located.
Major histocompatibility complex (MHC) - the MHC genes produce a molecule of great importance in the immune response. Two classes of MHC gene (Class I and II) have been found to be highly polymorphic in the mammalian genome.
Microsatellite - VNTRs consisting of short (< 5 bp) repeat sequences that can form tandem repeats to create alleles of up to 300bp or larger. The loci are highly variable and, due to their small size, are amenable to PCR amplification.
Minisatellite - tandemly repeated sequences of DNA forming **VNTR** loci. Minisatellite DNA alleles consists of repeat units of DNA sequences up to several hundred base pairs, and are highly variable. Forensic DNA fingerprinting/profiling was developed using these loci.

Mitochondrial DNA / mtDNA - DNA that occurs in a series of genes in the mitochondria of the cell, which are inherited from the female parent only. The mtDNA genome is therefore unlike diploid genes, in that there is only one copy inherited.
Mixture samples / MIX - a sample of individuals of unknown baseline stock composition origin for example the catch from a mixed stock fishery. Any number of baseline stocks may contribute to the mixture, or it may consist partially or wholly of unknown stocks not sampled among the baseline.
Nucleotides - the components of DNA, the sequence of which determines the genotype of an individual.
PCR / polymerase chain reaction - the technique by which particular regions of the DNA can be examined for polymorphisms. Primers (short known sequences of DNA) are designed to complement the region of the genome of interest, and using cycles of heating and annealing, the target DNA is amplified to a point where **Electrophoresis** or **RFLP** analysis can be performed. The development of PCR techniques allows the examination of small (from non-destructive sampling) or preserved, tissue samples.
Phenotype - expressed characteristics of an organism, derived from the genotype and the interactions between genotype and environment.
Private / unique allele - refers to a polymorphism, or an allelic / RFLP variant that occurs in a single population sample.
Polymorphism - simultaneous occurrence in a population of genomes showing two or more variations in DNA sequence leading to changes in and RFLP pattern, alternatively two or more alleles at a VNTR locus. A genome / locus is said to be polymorphic if more than one genotype is present in a population.
Restriction enzymes - biochemicals that fragment DNA at nucleotide sequences specific to that enzyme.
RFLP / restriction fragment length polymorphism - the occurrence of a restriction site in a region of DNA produces a characteristic DNA fragment pattern. Polymorphisms occur through the addition or loss of such sites.
Tandem repeats / VNTR - variable number tandem repeats of DNA sequences that make up **microsatellites** and **minisatellites**.

10. APPENDIX - SUMMARY OF DATA AVAILABLE ON GENETIC VARIATION IN ATLANTIC SALMON AND SEA TROUT

Notes:

1. Only studies that have used minisatellite, microsatellite, MHC, or mitochondrial DNA have been included in the summary, thus excluding two main areas of work on protein coding loci, and DNA sequencing. If a study has used a mixture of genetic marker systems, then only results from the former have been included here.

2. No data from unpublished studies, or studies with restricted access, is included. Papers that have been accepted for publication are included.

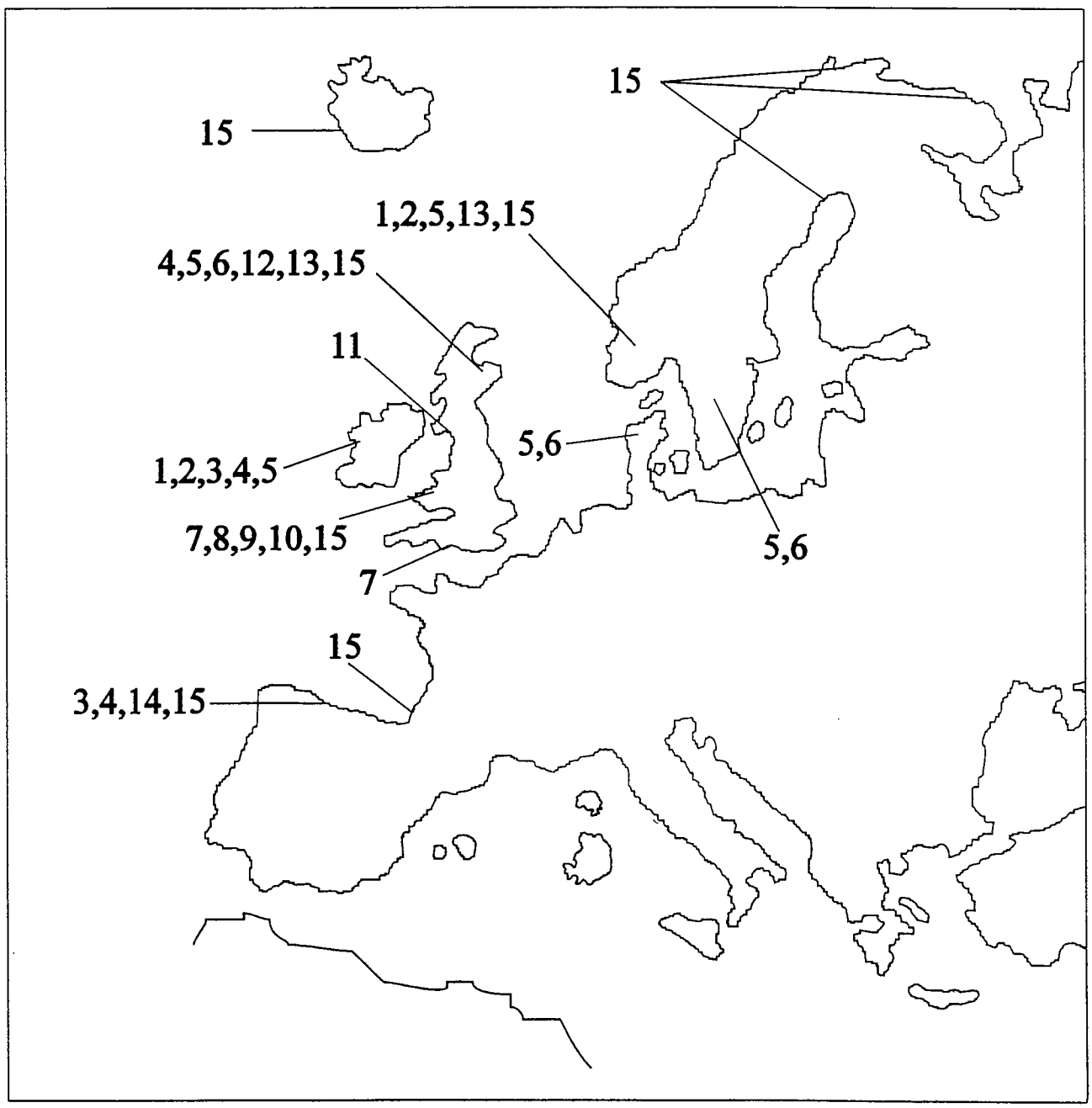
3. Studies of genetic variation in hatchery, or farmed stocks have been excluded as not being relevant in genetic stock identity. Studies of drainages where stocking is noted to have taken place, are however, included.

4. Studies of brown trout populations that are landlocked and are inaccessible from the sea have been excluded. If a paper examines such populations amongst other accessible populations, then the former are excluded from this summary. The majority, if not all, brown trout stocks in coastal drainages will have a mixture of anadromous and nonanadromous individuals, whose relative proportions are unknown. Therefore for the purposes of this summary, populations with unrestricted access to the sea are included, whether or not anadromy has been noted by the authors.

5. Likewise with Atlantic salmon, those populations noted as landlocked have been excluded.

List of papers and unpublished studies on the population genetics of Atlantic salmon using minisatellite, microsatellite, MHC, and mitochondrial DNA techniques.

Map showing approximate sampling areas discussed. Numbers refer to reference list.



10.1.1 Polymorphic microsatellite loci from Atlantic salmon (*Salmo salar* L.): genetic differentiation of North American and European populations

S.K. McConnell, P. O'Reilly, L. Hamilton, J.M. Wright and P. Bentzen
***Canadian Journal of Fisheries and Aquatic Sciences* (1995a)**
52, 1863-1872

Country	Region	Drainage	Life stage	Sample size
Canada	Nova Scotia	Gold	parr	35
		Salmon	parr	20
		Stewiackie	parr	45
Norway	south-west	Dale	adults	10* Pooled for
		Onarheim	adults	10* analysis
Ireland	west	Corrib		26

Three microsatellite loci isolated from Atlantic salmon were screened by PCR amplification and ^{32}P labelling.

Summary of genetic data

		Microsatellite loci		
		Ssa4	Ssa14	Ssa289
Gold	no. alleles	11	2	4
	H_{obs}	0.89	0.36	0.38
Salmon	no. alleles	8	2	2
	H_{obs}	0.60	0.35	0.30
Stewiackie	no. alleles	11	3	3
	H_{obs}	0.65	0.56	0.59
Ireland	no. alleles	13	3	3
	H_{obs}	0.69	0.46	0.44
Norway	no. alleles	19	3	6
	H_{obs}	0.65	0.53	0.75

This study examines intercontinental, inter-drainage (Nova Scotia) and to a lesser extent inter-regional (Europe). Nova Scotia rivers are separated by distances of 300 to 500 km. The allele frequency distributions at all loci showed significant heterogeneity. There were large discrepancies between North America and European samples at two loci.
The locus Ssa4 had 10 to 13 alleles in the north American samples, with a maximum frequency difference of 0.17. Four alleles present in these samples with an overall frequency of 0.54 to 0.61, were totally absent in the Irish or Norwegian samples. Thirteen to 16 alleles occurred in the European samples with four private alleles occurring up to a frequency of 0.30. No more than three alleles occurred at the locus Ssa14 in any population, and there were no private alleles above a frequency of 0.05. The largest allele frequency difference occurred between Ireland and Norway (0.21).
At the locus Ssa289 no more than four alleles were present in any Nova Scotia sample, with a maximum frequency difference of 0.36. Up to six alleles occurred in European samples, with a private allele of frequency 0.22.

10.1.2 Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian east coast stocks of Atlantic salmon
S.McConnell, L.Hamilton, D.Morris, D.Cook, D.Paquet, P.Bentzen and J.Wright
***Aquaculture* 1995b)**
137, 19-30

Country	Region	Drainage	Life stage	Sample size	
Canada	Nova Scotia	Gold	parr	35	
		Salmon	parr	20	
		Stewiackie	parr	45	
		LaHave	parr	35	
		Isaac's Harbour	parr	27	
Norway	south-west	Dale	adults	10*	Pooled
		Oeyreselv	smolts	20*	for
		Onarheim	adults	10*	analysis
Ireland	west	Corrib		26	

Four microsatellite loci were screened. Two loci isolated from Atlantic salmon (as described in McConnell *et al.*, 1995a) and two previously isolated from *Oncorhynchus mykiss*. Microsatellite were screened with PCR amplification and ^{32}P labelling.

Summary of genetic data

		Microsatellite loci			
		Ssa4	Ssa14	Omy27	Omy38
Gold	no. alleles	11	2	10	15
	H_{obs}	0.89	0.36	0.74	0.79
LaHave	no. alleles	13	2	12	24
	H_{obs}	0.80	0.36	0.79	0.83
Stewiackie	no. alleles	11	3	17	21
	H_{obs}	0.65	0.56	0.82	0.76
Salmon	no. alleles	8	2	7	11
	H_{obs}	0.60	0.35	0.70	0.65
Isaac's Hbr.	no. alleles	11	2	7	16
	H_{obs}	0.52	0.57	0.48	0.82
Ireland	no. alleles	13	3	3	25
	H_{obs}	0.69	0.46	0.08	0.92
Norway	no. alleles	19	3	4	31
	H_{obs}	0.65	0.53	0.10	0.84

This study examines intercontinental, inter-drainage (Nova Scotia) and to a lesser extent inter-regional (Europe). Much of the data analysed was reported in McConnell *et al.* (1995a). Four microsatellite loci were screened, two from Atlantic salmon (Ssa4 and Ssa14) and two previously isolated from rainbow trout (Omy27 and Omy38). Allele frequencies are given in graph form. All allele frequencies showed significant heterogeneity (Gst = 0.11), and three showed intercontinental differentiation.

10.1.3 Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic salmon)

J.A. Sánchez, C. Clabby, D. Ramos, G. Blanco, F. Flavin, E. Vázquez and R. Powell
***Heredity* (1996)**
77, 423-432

Country	Region	Drainage	Tributary	Life stage	Sample size
Ireland	west	Moy	Bunree	parr	19
		Shannon		parr	19
	east	Slaney	Moy	parr	22
Spain	north	Narcea		Adults	9
		Bidasoa		Adults	16
		Sella		Adults	14
		Esva		Adults	12

Three microsatellites isolated from Atlantic salmon were screened by PCR amplification and visualised by silver staining of acrylamide gels.

Summary of genetic data

		N	Average observed heterozygosity (± SE)
Ireland	Moy	19	0.579 (± 0.079)
	Shannon	19	0.529 (± 0.187)
	Slaney	22	0.532 (± 0.202)
Spain	Narcea	9	0.370 (± 0.225)
	Bidasoa	16	0.420 (± 0.212)
	Sella	14	0.310 (± 0.156)
	Esva	12	0.518 (± 0.224)

The paper examines inter and inter-drainage genetic variation using microsatellite loci isolated by the authors. There were significantly different allele frequency distributions among Irish samples, and between Irish and Spanish samples. However there was no heterogeneity in Spanish samples which were also less heterozygous ($P < 0.05$). Nei's genetic distance between Ireland and Spain was 0.10, reflected in two clusters in a dendrogram.
Four alleles were detected in all samples at the locus μ20.19. The maximum allele frequency difference occurs between two Spanish populations (0.168)
Four alleles unique to either Irish or Spanish samples were detected at the locus μF-43. However, all of these occur at low frequency (up to 0.068) and are not diagnostic. Four to five alleles occurred in all samples. The maximum frequency difference arose between two Spanish samples (0.292).
Three Spanish populations were monomorphic at the locus μ-D-30, while the fourth was not polymorphic at the 99% level. These populations are fixed/approaching fixation for the most common Irish allele. Between two and five alleles occurred in the Irish samples, one of which has a private allele at a frequency of 0.139.

10.1.4 Single locus minisatellite DNA variation in European populations of Atlantic salmon (*Salmo salar* L.).

C.E. Stone, J.B. Taggart and A. Ferguson

in press, Hereditas

Country	Region	Drainage	Life stage	Year	Sample size
Scotland	Sutherland	Shin	fry	1992	50
Ireland	Kerry	Blackwater	fry	1992	50
				1991	44
	Mayo	Burrishoole	parr	1991	50
Spain	Asturias	Esva	fry	1993	39
				1992	50

Four locus specific minisatellite probes, previously isolated from Atlantic salmon and brown trout, were used to screen 283 individuals. Three samples of farmed fish (n = 50 each) were also screened, but are not included here. Radiolabelled probe DNA was hybridised to Southern blots of *Hae*III digested genomic DNA, and the fragments visualised by autoradiography.

Summary of genetic data

			N	Single locus heterozygosity *Ssa*-A45/1	*Str*-A9	*Ssa*-A60	*Ssa*-A45/2	Average observed heterozygosity (± se)
Scotland	Shin		1992	0.735	0.669	0.396	0.650	0.612 (± 0.075)
Ireland	Blackwater		1991	0.700	0.792	0.321	0.634	0.612 (± 0.102)
			1992	0.633	0.804	0.250	0.698	0.596 (± 0.121)
	Burrishoole		1991	0.584	0.771	0.206	0.817	0.595 (± 0.139)
Spain	Esva		1992	0.670	0.664	0.572	0.653	0.643 (± 0.025)
			1993	0.685	0.673	0.575	0.706	0.656 (± 0.028)

Four minisatellite loci, previously isolated from salmon and brown trout, were used to examine inter and intra-regional variation. As there was no heterogeneity between years in Esva and Blackwater, temporal stability was proved. Over 85% of total variation was found within samples. The Fst value within Scottish and Irish samples was highly significant, Fst = 0.057 (95% CI 0.036 - 0.072), as was the comparison between these and the Spanish river, Fst = 0.145 (95% CI 0.091 - 0.200). This is reflected in the Spanish samples clustering away from the others in a dendrogram.

Four to seven alleles were found at the locus *Ssa*-A45/1, although there were no private alleles. The maximum frequency difference was 0.34. Neither did *Str*-A9 detect any private alleles. The largest frequency difference (0.19) occurred between Ireland and Scotland. The largest allele frequency difference at the locus *Ssa*-A60 also occurred here (0.19). *Ssa*-A45/2 was the most discriminatory locus with almost 25% of variation at this locus being distributed between samples. This locus also had the largest number of alleles segregating.

10.1.5 Genetic structure of European populations of *Salmo salar* L. (Atlantic salmon) inferred from mitochondrial DNA

E.E. Nielsen, M.M. Hansen and V. Loeschcke
***Heredity* (1995)**
77, 351-358

Country	Region	Drainage	Life stage	Sample size
Denmark	Jutland	Skjern	parr	33
Sweden	south-west	Ätran	smolts	27
		Lagan	parr	37
Norway	south-east	Numedalslågen	parr	26
Scotland	north-east	Conon	parr	34
		Dee	adults	111
Ireland	west	Burrishoole	parr	32
		Corrib	parr	27

Total DNA was PCR amplified at the mitochondrial ND1 segment and RFLPs were directly observed using ethidium bromide staining of agarose gels. Four restriction enzymes (*Ava*II, *Hae*III, *Hin*f1 and *Rsa*I) were found to be commonly polymorphic resulting in six haplotypes detected in a total of 327 individuals.

Results of mtDNA screening

		Haplotypes					
		1	2	3	4	5	6
Denmark	Skjern	0.67	0.30	—	0.03	—	—
Sweden	Ätran	0.11	0.74	0.11	—	—	0.04
	Lagan	—	0.22	0.68	0.10	—	—
Norway	Numedalslågen	—	0.65	0.23	0.12	—	—
Scotland	Conon	0.24	0.24	0.09	0.44	—	—
	Dee	0.09	0.54	0.07	0.25	0.05	—
Ireland	Burrishoole	0.22	0.44	—	0.28	—	0.06
	Corrib	—	0.41	0.15	0.44	—	—

This paper examines inter-regional, and to a smaller extent, inter-drainage variation at the mitochondrial ND1 region. Significant differences in the frequency of haplotypes were found in most pairwise comparisons.
Over all 79.5% of variation was within samples, and 0.2% among groups with a significant Φ st of 0.335. However χ^2 between and among Scottish/Irish samples were not significant. Only the Dee has a private haplotype at a frequency of 0.05.
There was no correlation between haplotypes and geographical distance, with the three most common haplotypes found everywhere.

10.1.6 Analysis of microsatellite DNA from old scale samples of Atlantic salmon (*Salmo salar*): A comparison of genetic composition over 60 years

E.E. Nielsen, M.M. Hansen and V. Loeschcke
***Molecular Ecology* (1997)**
6, 487-492

This paper examines microsatellite variation in historical and extant populations. Only the data set of current populations is considered here.

Country	Region	Drainage	Sample size
Denmark	Jutland	Skjern	36
Scotland	north-east	Conon	26
Sweden	south	Ätran	27

Four microsatellite loci, previously isolated from Atlantic salmon, were screened. Loci were PCR amplified and analysed by automated sequencer.

Summary of genetic data

	Skjern		Conon		Ätran	
	no. alleles	H_{obs}	no. alleles	H_{obs}	no. alleles	H_{obs}
SsoSL 85	6	0.74	9	0.74	12	0.87
SsoSL 311	10	0.88	14	0.91	19	0.95
SsoSL 417	4	0.55	12	0.86	13	0.86
SsoSL 438	4	0.65	6	0.79	8	0.82
Mean	6 ± 2.8	0.70 ± 0.14	10.2 ± 3.5	0.82 ± 0.08	13 ± 4.5	0.88 ± 0.05

Low frequency private alleles occurred at the loci SSOSL85 and SSOL438 at frequencies of 0.02 and 0.07 respectively. The largest allele frequency differences involved inter-regional comparisons (SSOSL85 max. frequency difference = 0.46 between Scotland and Denmark; SSOSL438 max. frequency difference = 0.32 between Denmark and Sweden).
Eight private alleles were found at the locus SSOSL311. The highest maximum frequency private allele occurred in the Skjern at a frequency of 0.15.
Seven private alleles were detected at locus SSOSL417, the most frequent of which occurred in the Ätran at a frequency of 0.17.
Significantly fewer alleles were detected in the Skjern (P = 0.0481, P = 0.0106), although heterozygosity values show no similar pattern.

10.1.7 Mitochondrial DNA and allozyme analysis of Atlantic salmon, *Salmo salar* L., in England and Wales
S.J. Hovey, King, D.P.F. King, D. Thompson and A. Scott
***Journal of Fish Biology* (1992)**
35 (Suppl. A) 253-260

10.1.8 Mitochondrial DNA variation in Atlantic salmon, *Salmo salar* L., populations
D.P.F. King, S.J. Hovey, D. Thompson and A. Scotts
***Journal of Fish Biology* (1993)**
42, 25-33

These papers can be summarized together as the latter paper includes and extends the results of the first.

					Sample size			
Country	Region	Drainage	Site	Grid Ref.	1988 parr	1989 smolts	1989 parr	1990 smolts
England	south	Itchen	1	SU 485289	5	0	11	0
			2	SU 481282	5	8	0	0
			3	SU 484249	12	10	6	9
			4	SU 474227	4	13	12	10
			5	SU 466192	14	18	11	10
Wales	north	Conwy	1	SH 702518		10		
			2	SH 785567		12		
			3	SH 806585		9		
			4	SH 802608		10		
			5	SH 768700		10		

Purified mitochondrial DNA was isolated and digested with restriction enzymes. Fragments were visualised by silver staining of polyacrylamide gels. Four restriction enzymes (*Ava*II, *Hae*III, *Hin*f1 and *Mbo*I) were found to be commonly polymorphic resulting in ten haplotypes being detected in a total of 209 individuals.

Results of mtDNA screening

haplotype	Itchen 1988 parr[a]	Itchen 1989 smolts[a]	Itchen 1989 parr[b]	Itchen 1990 smolts[a]	Conwy 1989 parr[a]
1	0.58	0.18	0.38	0.17	0.39
2	0.25	0.08	0.02	—	0.12
3	0.02	0.02	—	—	—
4	0.05	—	0.10	—	—
5	0.02	0.02	—	—	—
7	0.08	0.41	0.10	0.34	0.49
11	—	0.02	—	—	—
15	—	0.06	0.20	0.21	—
16	—	0.02	—	—	—
18	—	0.18	0.20	0.28	—

[a] No significant between site differentiation so data from sites 1-5 are pooled.
[b] Significant between site differentiation ($P = 0.002$)

This is an inter- and intra- drainage study of mitochondrial DNA variation. There was significant intra-drainage differentiation between sites on the Itchen in 1989 parr sample (P = 0.002), although there was none between years. However, there was significant differences between smolts and parr of the same cohort.
Notably only three composite haplotypes were found in the Conwy in more than 50 samples from five sites.

10.1.9 Mitochondrial DNA and allozyme variation in Atlantic salmon (*Salmo salar*) populations in Wales

M. O'Connell, D.O.F. Skibinski and J.A. Beardmore
***Canadian Journal of Fisheries and Aquatic Sciences* (1995)**
52, 171-178

10.1.10 Allozyme and mtDNA divergence between Atlantic salmon populations in North Wales

M. O'Connell, D.O.F. Skibinski and J.A. Beardmore
***Journal of Fish Biology* (1996)**
48, 1023-1026

These papers can be summarized together as the authors have used the same scoring system for mtDNA screening throughout. Haplotypes shown in the later paper are renumbered in this summary in accordance with the earlier work.

Country	Region	Drainage	Site	Life stage	Sample size
Wales	south	Teifi	1	parr	8
			2	parr	2
			3	parr	6
			4	parr	8
			5	parr	8
			6	parr	4
		Usk	1	parr	2
			2	parr	1
			3	parr	8
			4	parr	10
			5	parr	3
			6	parr	7
			7	parr	6
			8	parr	5
			9	parr	8
		Wye	1	parr	5
			2	parr	3
			3	parr	6
			4	parr	7
			5	parr	6
			6	parr	7
			7	parr	2
	north	Dee	Mynach		12
			Ceidiog		3
			Ceirw		5
		Rhyd-hir			12
		Dyfi			7

Isolated mtDNA was digested with four polymorphic restriction enzymes. Fragments were ^{32}P endlabelled and resolved on polyacrylamide gels. The digest patterns were visualised by autoradiography.

Results of mtDNA screening

	Teifi						Wye						
haplotype	1	2	3	4	5	6	1	2	3	4	5	6	7
1	0.375	1.000	0.333	1.000	0.750	0.750	0.600	1.000	0.400	0.714	0.333	0.143	—
2	—	—	—	—	0.125	—	—	—	—	0.143	—	—	0.50(
3	0.125	—	0.167	—	—	—	—	—	—	—	—	—	—
4	—	—	0.167	—	—	—	0.400	—	0.400	—	0.167	0.285	—
5	—	—	0.167	—	—	—	—	—	—	—	—	—	—
6	0.125	—	—	—	—	—	—	—	—	—	0.167	0.143	—
7	—	—	—	—	—	—	—	—	—	—	—	—	—
8	—	—	—	—	—	—	—	—	—	—	—	—	0.50(
9	0.125	—	0.166	—	—	—	—	—	—	—	—	0.429	—
10	0.250	—	—	—	0.125	0.250	—	—	0.200	0.143	0.333	—	—
11	—	—	—	—	—	—	—	—	—	—	—	—	—
12	—	—	—	—	—	—	—	—	—	—	—	—	—

	Usk								
haplotype	1	2	3	4	5	6	7	8	9
1	0.500	—	0.375	0.500	0.333	0.571	0.833	0.600	0.750
2	—	—	—	—	—	—	—	—	0.125
3	—	—	—	—	—	—	—	0.400	—
4	0.500	1.000	0.500	0.400	0.333	0.143	0.167	—	0.125
5	—	—	—	—	—	—	—	—	—
6	—	—	—	0.100	—	—	—	—	—
7	—	—	—	—	—	0.286	—	—	—
8	—	—	—	—	—	—	—	—	—
9	—	—	—	—	—	—	—	—	—
10	—	—	0.125	—	0.333	—	—	—	—
11	—	—	—	—	—	—	—	—	—
12	—	—	—	—	—	—	—	—	—

haplotype	Dee			Dyfi	Rhyd-hir
	Mynach	Ceidiog	Ceirw		
1	0.585	0.334	0.400	0.714	0.667
2	—	—	—	0.286	—
3	0.083	—	—	—	0.083
4	—	—	—	—	—
5	0.083	—	—	—	—
6	—	0.333	0.200	—	—
7	—	—	—	—	—
8	0.083	—	—	—	—
9	—	—	0.200	—	—
10	0.083	—	0.200	—	0.250
11	—	0.333	—	—	—
12	0.083	—	—	—	—

Summary of genetic data

	Sample size	Percentage nucleotide diversity	No. of composite haplotypes
Teifi	36	0.13	8
Usk	50	0.07	7
Wye	36	0.15	7
Dee	20	0.20	9
Rhyd-hir	12	0.17	3
Dyfi	7	0.08	2

This paper examines inter- and intra- drainage mitochondrial DNA variation. Significant differences in haplotype frequencies were found within the Teifi ($\Phi = 0.076$) and Wye ($\Phi = 0.105$) although sample sizes in are general and prone to sampling error. The Usk has a unique haplotype at a frequency of 0.286. The Dee has 2 unique haplotypes at a frequency of 0.083 and 0.333. NB These unique haplotypes within drainages are restricted to single sampling sites, and do not typify the river. Over all drainages χ^2 and Φst ($\Phi = 0.062$) were not significant ($P = 0.998$). The majority of the variation (93.8%) was found within samples.

Unpublished or restricted studies

10.1.11 Genetic differentiation between Atlantic salmon populations in the Windermere catchment
S.E. Hartley and A.D. Pickering
Institute of Freshwater Ecology (1994)
NRA (North West Region)

10.1.12 Minisatellite and mitochondrial DNA analysis of Atlantic salmon populations from the River Tweed
E. Verspoor
unpublished data

10.1.13 Application of DNA markers in the environmental genetic study of Atlantic salmon
J. Taggart
unpublished data

Data from 10.1.11, 10.1.12 and 10.1.13 were used for modelling by kind permission of the authors.

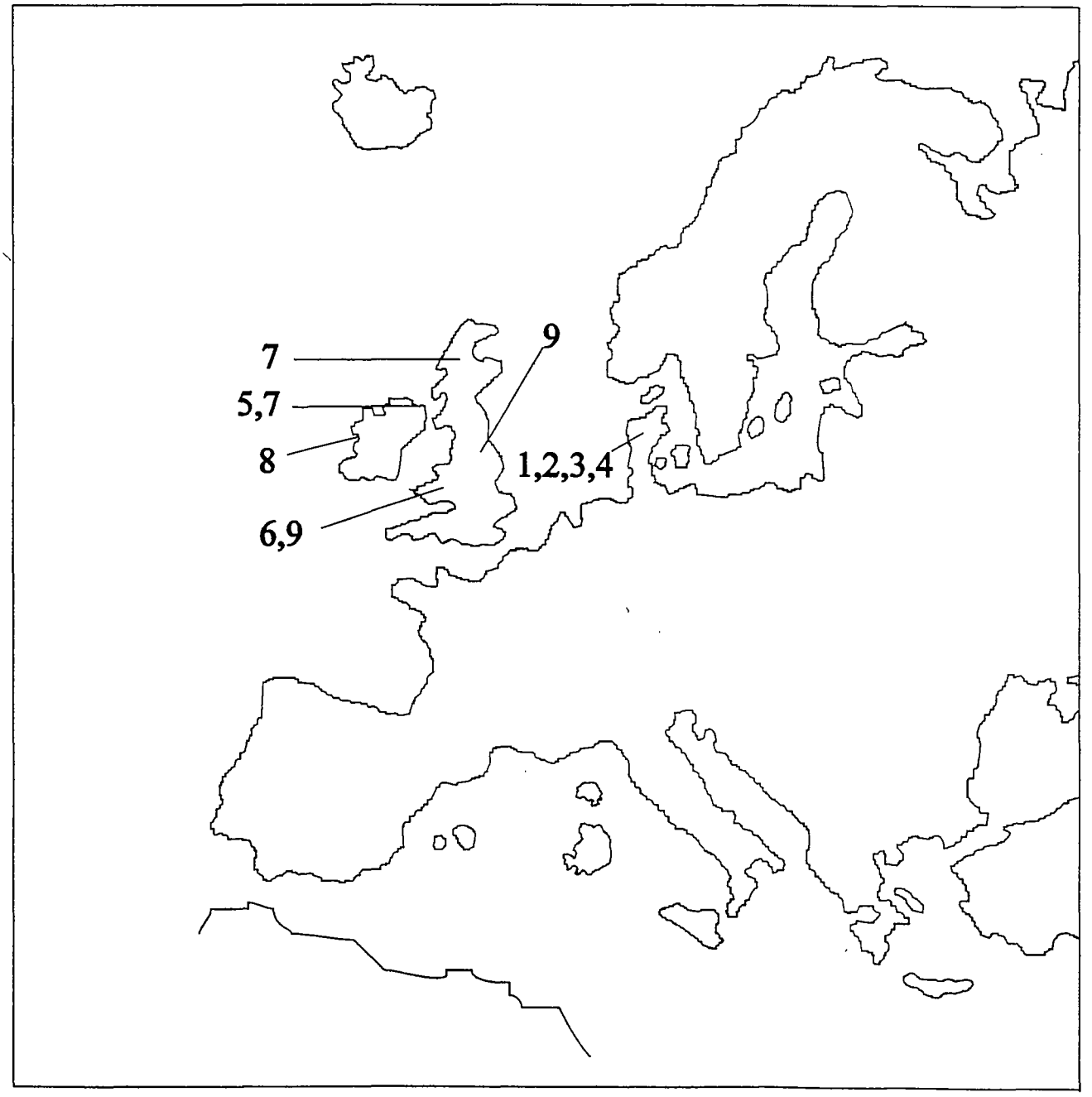
EU Contract AIR1 3033 92 0719
An Assessment of the Genetic Consequences of the Deliberate or Inadvertent Introduction of Non-native Atlantic Salmon into Natural Populations

10.1.14 Component B.1. Analysis of salmon populations from the Galacian and Asturian regions of northern Spain
E. Vazquez *et al.*
unpublished

10.1.15 Component C. Genetic variation of Atlantic salmon populations in western Europe, and North America
T. Cross *et al.*
unpublished

List of papers and unpublished studies on the population genetics of anadromous brown trout using minisatellite, microsatellite, and mitochondrial DNA techniques.

Map showing approximate sampling areas discussed. Numbers refer to reference list.



10.2.1 Assessment of the stocked or wild origin of anadromous brown trout (*Salmo trutta* L.) in a Danish river system, using mitochondrial DNA RFLP analysis

M.M. Hansen, R.A. Hynes, V. Loeschcke and G. Rasmussen
***Molecular Ecology* (1995)**
4, 189-198

Country	Region	Drainage	Tributary	Life stage	Sample size
Denmark	north	Karup	main river	anadromous adults	50
			Sejbæk	juvenile	25
			Haderis	juvenile	28
			Trevad	juvenile	28
	east Jutland	Odder	Stampenmølle	juvenile	20

Total DNA was digested with seven polymorphic restriction enzymes (*Ava*II, *Eco*RV, *Hae*III, *Hin*fI, *Mbo*I, *Rsa*I and *Xba*I). The fragments were resolved on agarose gels, and Southern blots were hybridized with ^{32}P-labelled purified mtDNA which was visualised using autoradiography.

Results of mtDNA screening

	Karup				Odder
haplotype	main	Sejbæk	Haderis	Trevad	Stampmølle
1	0.18	0.04	0.39	0.29	0.10
2	0.58	0.56	0.39	0.39	0.15
3	0.22	0.32	0.14	0.32	0.10
4	0.02	—	—	—	0.30
5	—	0.04	0.07	—	0.10
6	—	—	—	—	0.20
7	—	—	—	—	0.05
8	—	0.04	—	—	—
Nucleon diversity	0.59	0.60	0.69	0.69	0.86

This study examined the extent of stocking success on an intra-drainage basis. Samples from hatchery stocks and landlocked populations were also studies but have been excluded from this summary.
Haplotype 2 was the most common in all Karup populations, and three common haplotypes were found in all samples. The largest frequency difference occurred among Karup samples (0.35). χ^2 among Karup samples was not significant, neither was the Φst value (Φst = 0.029, P = 0.059), indicating little intra-drainage structuring. However significant differences occurred between the Karup and Odder rivers. Two unique haplotypes were found in the Odder, one with a frequency of 0.2. The Φst value (0.166) between the two rivers was significant (P = 0.039), and was the χ^2.

10.2.2 Genetic differentiation among Danish brown trout populations, as detected by RFLP analysis of PCR amplified mitochondrial DNA segments
M.M. Hansen and V. Loeschcke
Journal of Fish Biology **(1996)**
48, 422-436

10.2.3 Founder effects and genetic population structure of brown trout (*Salmo trutta*) in a Danish river system
M.M. Hansen and K.-L.D. Mensberg
Canadian Journal of Fisheries and Aquatic Science **(1997)**
in press

While these papers examine different aspects of the population genetics of brown trout, they can be summarised together as the methods of analysis, and the scoring of variation is the same. The first paper examines mtDNA data primarily from Lake Hald, however, access to this system has been blocked for several centuries and the results are not summarised here. All other drainages sampled are known to contain a mixture of anadromous and nonanadromous brown trout, and all have access to the sea. The latter paper examines data from the Odder River system. Stocks from the Odder are noted as being a mixture of anadromous and nonanadromous. Two hatchery strains were also screened but the results are not included here

Country	Region	Drainage	Tributaries	Life stage	Sample size
Denmark	east Jutland	Gudenaa	Brandstrup	juvenile	30
			Tjærbæk	juvenile	25
		Odder	main river upper	juvenile	40
			main river lower	juvenile	42
			Stampemølle	juvenile	40
			Asbæk	juvenile	39
			Kragebæk	juvenile	39
			Assedrup	juvenile	42
			Fiskbæk	juvenile	42
	Baltic	Bornholm	Tejn	juvenile	23
			Vellens	juvenile	22
			Dyndals	juvenile	23

Total DNA was PCR amplified at the ND1 and ND5/6 mitochondrial regions, and digested with ten polymorphic restriction enzymes (ND1: *Ava*II, *Hin*fI, *Alu*I, *Hae*III, *Hpa*II; ND5/6: *Ava*II, *Hin*fI, *Hae*III, *Taq*I, *Xba*I). The fragments were resolved on agarose gels, and viewed by ethidium bromide staining.

Results of mtDNA screening

	Gudenaa		Bornholm		
haplotype	B'strup	Tjærbæk	Tjen	Vellens	Dyndals
1	0.03	0.04	0.30	0.18	0.13
2	0.30	0.16	0.70	0.64	0.30
3	0.10	—	—	—	—
4	0.20	0.40	—	0.14	0.17
5	0.23	0.16	—	—	0.09
6	0.03	0.08	—	—	—
7	—	0.12	—	0.05	0.04
8	—	0.04	—	—	—
9	0.03	—	—	—	—
10	0.07	—	—	—	—
11	—	—	—	—	0.13
12	—	—	—	—	0.09
13	—	—	—	—	0.04
Nucleon diversity	0.83	0.81	0.47	0.59	0.87

	Odder						
haplotype	upper	lower	S'mølle	Asbæk	Kragebaek	Assedrup	Fisbæk
1	—	—	—	—	—	—	—
2	0.13	0.33	0.10	0.41	0.33	0.26	0.31
3	—	—	—	—	—	—	—
4	0.08	0.17	0.10		0.03	0.02	0.07
5	0.48	—	0.05	0.08	0.10	0.05	0.07
6	—	—	—	—	—	—	—
7	0.13	0.31	0.28	0.15	0.38	0.21	0.35
8	0.03	0.05	0.23	0.05		0.24	0.10
9	—	—	—	—	—	—	—
10	0.18	0.14	0.25	0.31	0.15	0.21	0.10
11	—	—	—	—	—	—	—
12	—	—	—	—	—	—	—
13	—	—	—	—	—	—	—
Nucleon diversity	0.72	0.76	0.81	0.72	0.72	0.80	0.77

This was an inter- and intra- drainage study of mitochondrial DNA variation. Samples from landlocked populations were also studies but have been excluded from this summary. Haplotype 2 is the most common in all populations. Haplotype 1 occurs in the Gudenaa and Bornholm at frequencies up to 0.30, but is absent from the Odder. Three unique haplotypes occur in the Gudenaa at a frequencies up to 0.10, and three in the Bornholm at frequencies up to 0.13. Significant genetic variation is found over all ($\Phi st = 0.150$), and within the Gudenaa (0.108), Bornholm (0.069) and Odder (0.059). Variation between individuals in a sample accounts for 84.9% of the total variation.

10.2.4 The problem of sampling families rather than populations: Relatedness among individuals in samples of juvenile brown trout (*Salmo trutta* L.).

M.M. Hansen, E.E. Nielsen and K.-L.D. Mensberg

Molecular Ecology **(1997)**
6, 469-474

Country	Region	Drainage	Tributaries	Life stage	Sample size
Denmark	east Jutland	Gudenaa	Skibelund	0+	18
				1+	16
			Tjærbæk	0+	19
				1+	33

Total DNA was PCR amplified at the ND1 and ND5/6 mitochondrial regions, and digested with ten polymorphic restriction enzymes (ND1: *Ava*II, *Hin*fI, *Alu*I, *Hae*III, *Hpa*II; ND5/6: *Ava*II, *Hin*fI, *Hae*III, *Taq*I, *Xba*I). The fragments were resolved on agarose gels, and viewed by ethidium bromide staining. The seven microsatellite loci screened were previously isolated from brown trout and Atlantic salmon. Microsatellite loci were PCR amplified and analysed by automatic sequencer.

Results of mtDNA screening

haplotype	Skibelund	Tjærbæk
2	0.03	0.22
3	0.71	0.02
4	0.12	0.06
5	0.09	0.16
6	—	0.04
7	0.03	0.16
8	—	0.06
9	—	0.02
10	0.03	0.24
14	—	0.04
Nucleon diversity	0.53	0.85

Results of microsatellite screening

Loci	Skibelund		Tjærbæk	
	No. alleles	H_{obs}	No. alleles	H_{obs}
Str 15	6	0.61	4	0.70
Str 60	2	0.39	4	0.42
Str 73	2	0.48	4	0.52
SsoSL 85	6	0.36	8	0.50
SsoSL 417	9	0.94	10	0.90
SsoSL 438	5	0.76	5	0.61
SsHaeIII 14.20	10	0.85	12	0.81

This was an inter- and intra- drainage study of genetic variation in mtDNA and microsatellite loci. Heterogeneity was between year classes in both rivers. There was significant haplotype frequency differences between the rivers ($\Phi st = 0.279$), and χ^2 tests indicated heterogeneity at microsatellite loci. Combining the data possible groups of siblings were identified in Skibelund.

10.2.5 Variation in mitochondrial DNA and post-glacial colonisation of NW Europe by brown trout (*Salmo trutta* L.)

R.A. Hynes, A. Ferguson and M.A. McCann
***Journal of Fish Biology* (1996)**
48, 54-67

Thirty-seven samples were analysed consisting of brown trout from nonanadromous, landlocked, anadromous and hatchery stocks. Only those samples noted in the text to have a high proportion of anadromous individuals are included here.

Country	Region	Drainage	Site	Sample size
Ireland	Antrim	Glenariff	1	50
			2	51
		Glynn		102
		Carnlough		28
		Glencloy		15
		Glenarm		39

Total DNA was digested with six restriction enzymes known from previous studies to be polymorphic (*Xba*I, *Eco*RV, *Ava*II *Hin*fI, *Hae*III and *Mbo*I). The fragments were resolved on agarose gels, and Southern blots were hybridized with ^{32}P-labelled purified mtDNA which was visualised using autoradiography.

Results of mtDNA screening

	Glenariff					
haplotype	1	2	Glynn	Carnlough	Glencloy	Glenarm
1	—	0.08	—	—	—	—
6	0.42	0.06	0.92	0.64	0.40	0.51
10	—	0.06	—	—	—	—
12	0.42	0.64	0.01	0.15	0.40	0.10
13	0.16	0.06	—	0.21	0.20	0.39
14	—	—	0.07	—	—	—
Haplotype diversity	0.6343 (± 0.03007)	0.4369 (± 0.08336)	0.2462 (± 0.06860)	0.5397 (± 0.08556)	0.6857 (± 0.06147)	0.5938 (± 0.04301)
Nucleotide diversity	0.009271	0.007196	0.000880	0.005685	0.009818	0.005203

This was part of a much larger European wide study of mtDNA variation. This section is an inter-and to a small extent intra- drainage study. Low frequency haplotypes occur in the Glenariff and Glynn. Three common haplotypes occur in all populations, although at significantly different frequencies. There was significant haplotype frequency differences within (Φst = 0.180) and among rivers (Φst = 0.621).

10.2.6 Mitochondrial DNA variation in River Usk brown trout, *Salmo trutta*

D.G. Bembo, A.J. Weightman, R.J.H. Beverton and R.C. Cresswell

***Journal of Fish Biology* (1994)**
44, 717-723

While the author claims that the populations examined were "essentially" nonanadromous, the locations studied do have open access to the sea, and are therefore included in this summary.

Country	Region	Drainage	Tributaries	Sample size
Wales	south	Usk	Bran	6
			Crai	5
			Gavenny	6
			Grwyne fawr	6
			Grwyne fechan	6
			Menascin lower	6
			Menascin upper	6
			Senni	12
		Wye	Hierant	12

Isolated mtDNA was digested with six restriction enzymes, two of which (^{32}P and resolved on agarose or acrylamide gels. Patterns were visualised by autoradiography.

Results of mtDNA screening

	Usk								Wye
haplotype	Bran	Crai	Gavenny	Grwyne fawr	Grwynef echan	Menasci n lower	Menasci n upper	Senni	Hierant
1	1.00	—	1.00	1.00	1.00	0.83	1.00	0.25	0.58
2	—	1.00	—	—	—	—	—	0.67	0.25
3	—	—	—	—	—	—	—	0.08	—
4	—	—	—	—	—	0.17	—	—	—
5	—	—	—	—	—	—	—	—	0.17
Nucleon diversity	0.00	0.00	0.00	0.00	0.00	0.34	0.00	0.53	0.62

χ^2 and Φst values indicate a high degree of differentiation in the Usk drainage, with Crai being fixed for an alternative haplotype, and a maximum of three haplotypes per location. Only two of the tributaries showed any variation. There was significant variation between drainages as well ($P = 0.019$).

10.2.7 Population genetics of anadromous brown trout (*Salmo trutta* L.) in Scotland and Ireland

C.Thompson
PhD. Thesis (1995)
The Queens University of Belfast

This study primarily examined genetic variation in anadromous populations of sea trout (i.e. samples from locations known to be sea trout spawning areas with open access to the sea). However, samples were also taken from populations above waterfalls (eight samples from the Dee and Glenariff) which are not reported here. Juveniles were sampled at all locations, apart from Burrishoole where smolts and adults were sampled.

Country	Region	Drainage	Tributary	Grid Ref.
Scotland	East	Tweed	Whiteadder	unknown
			Monynut	unknown
			Dye Water	unknown
		South Esk	Quarhity	NO 395 590
		North Esk	Kinnaber Mill Lade	NO 715 621
		Dee	Sheeoch	NO 755 933
			Lumphanan	NJ 566 005
			Beltie	NJ 610 040
		Don	Fintray	NJ 832 161
			Buchat	NJ 401 149
			Leochel	NJ 577 111
			Gadie	NJ 589 245
		Spey	Conglass	NJ 231 156
	north	Thurso	Aclachan	ND 134 528
		Polla	White Burn	NC 544 386
	west	Laxford	Unnamed Burn	NC 263 441
		L. Assynt	River Loanan	NC 248 195
		Coulin	Farm Burn	NH 015 548
		Squod	Main river	NJ 811 895
		Kishorn		unknown
	south-west	Doon	Smithston Burn	NS 413 125
		Nith	River Scaur	NS 759 030
Ireland	north-east	Spences		J 37 22
		Fork		J 33 15
		Burren		J 35 33
		Glenariff	Main river	D 22 21
			Laragh Lodge	D 22 20
	west	Burrishoole	Lough Feeagh	95°55'W 53°55'W

Locus specific minisatellite probes detecting six loci (*Str*-A3, *Str*-A5, *Ssa*-A45/1, *Str*-A22/1, *Ssa*-A45/2 and *Str*-A22/2), previously isolated from Atlantic salmon and brown trout, were used to screen almost 1200 individuals. Radiolabelled probe DNA was hybridised to Southern blots of *Hae*III digested genomic DNA, and the fragments visualised by autoradiography.

Total DNA was PCR amplified at the mitochondrial ND1 and ND5/6 segments and RFLPs were directly observed using ethidium bromide staining of agarose gels. Eight restriction enzymes (ND1: *Ava*II, *Hae*III, *Alu*I and *Hpa*II; ND5/6: *Ava*II. *Xba*I *Hin*f1 *Hae*III, *Alu*I, *Dde*I and *Taq*I) were found to be commonly polymorphic. resulting in thirty-one detectable haplotypes.

Summary of genetic data

Drainage	N	No. alleles per locus (± s.e)	H	± s.e	No. mtDNA haplotypes	Nucleotide diversity
Tweed	46	9.0 ± 4.5	0.704	± 0.088	5	0.006321
S. Esk	47	8.3 ± 4.3	0.707	± 0.090	—	—
N. Esk	48	8.7 ± 3.5	0.700	± 0.090	—	—
Dee	189	10.6 ± 5.4	0.736	± 0.074	5	0.005555
Don	139	10.5 ± 5.1	0.714	± 0.087	10	0.005639
Spey	47	8.3 ± 3.7	0.656	± 0.119	3	0.001725
Thurso	42	6.8 ± 3.2	0.630	± 0.009	—	—
Polla	22	6.8 ± 2.3	0.728	± 0.006	—	—
Laxford	38	7.5 ± 2.8	0.727	± 0.072	—	—
L. Assynt	47	7.6 ± 2.3	0.694	± 0.056	—	—
Coulin	48	7.0 ± 1.8	0.641	± 0.007	—	—
Squod	48	6.2 ± 2.5	0.635	± 0.081	5	0.009344
Kishorn	40	7.0 ± 2.1	0.746	± 0.036	—	—
Doon	48	6.2 ± 2.3	0.690	± 0.072	—	—
Nith	35	7.0 ± 3.0	0.647	± 0.098	—	—
Burrishoole	106	8.3 ± 3.4	0.671	± 0.075	18	0.005145

Summary of intra-drainage genetic data

Tributaries	N	No. alleles per locus (± s.e)	H	± s.e.	No. mtDNA haplotypes	Nucleotide diversity
Dee						
Beltie	47	9.2 ± 4.7	0.704	± 0.090	—	—
Sheeoch	48	8.2 ± 3.6	0.694	± 0.103	—	—
Lumphanan	45	8.5 ± 4.6	0.664	± 0.112	5	0.005555
Don						
Fintray	49	8.5 ± 3.3	0.732	± 0.072	10	0.005639
Buchat	46	8.5 ± 2.9	0.727	± 0.075	—	—
Leochel	44	8.8 ± 4.6	0.715	± 0.104	—	—

Summary of genetic data from populations from the north-east of Ireland

	N	No. alleles per locus (± s.e)	H	± s.e.	No. mtDNA haplotypes	Nucleotide diversity
Spences	35	5.0 ± 2.3	0.522	± 0.121	—	—
Burren	39	7.0 ± 2.8	0.677	± 0.067	—	—
Fork	60	7.4 ± 3.4	0.634	± 0.068	—	—
Glenariff						
Lower	25	5.2 ± 1.9	0.577	± 0.085	5	0.004024
Laragh	25	3.8 ± 1.6	0.468	± 0.144	—	—

This study examines inter- and intra- drainage variation at 6 minisatellite loci, and to a lesser extent, mtDNA. All drainages in Scotland had significantly different allele frequency distributions ($P < 0.0001$) and F_{st} values were significantly different from zero ($F_{st} = 0.059$; 95% CI 0.044 - 0.075). However, the F_{st} value between north and west coast drainages (0.089) was over three times the value between east coast drainages (0.027), where 20% of pairwise comparisons were not significant at the 5% level. There was also divergence between Scottish and north-eastern Irish drainages ($F_{st} = 0.098$). MtDNA showed a similarly high level of divergence between drainages, $\Phi = 0.245$ overall populations and all pairwise comparisons being significant. Intra drainages comparisons, within the Dee, Don and Glenariff, were all significant, $F_{st} = 0.028$, 0.015 and 0.041 respectively.

Unpublished or restricted studies

10.2.8 Stocking of non-native brown trout into an Irish sea trout river.
R. Hynes and A. Ferguson
unpublished

10.2.9 The application of genetic techniques as a tool for studying sea trout populations in England and Wales
H. Hall
NRA (Draft Final Report) 1996