# THE WATERFORD ESTUARY SALMON GENETICS PROJECT - 2010

## CLIENT CONFIDENTIAL FINAL REPORT TO INLAND FISHERIES IRELAND

**School of Biological, Earth & Environmental Sciences / Aquaculture & Fisheries Development Centre, University College Cork**

**1st August 2011**

**Executive Summary**

A Section 18 authorisation was issued to fishermen to undertake an experimental fishery for Atlantic salmon from July to October 2010 in the Waterford estuary to enable an assessment of river of origin of the fish captured and to determine the presence if any of salmon from other countries by genetic analysis. Subsequently a sample of 460 tissue samples collected from individual salmon captured in Waterford estuary during July to October 2010 were provided to UCC for genetic analysis (452 of these revealed sufficient quality and quantity of DNA for screening).

It has not been possible previously to discriminate among salmon in a mixed sample originating in the Nore, Barrow, Suir and Munster Blackwater rivers, and to some extent from the Slaney river, using existing genetic methods. However, a simulation analysis based on the application of newly available genetic markers (SNPs) undertaken in 2009 suggested that it might now be possible to tell fish from the various rivers in the region apart.

This genetic stock identification project using additional SNP markers suggests the following:

Two fish out of the 452 sampled originated from a river which was not an Irish river. The analysis of genotypic data for these fish showed that these had identical genotypes and we consider that the samples are in fact from the same fish sampled twice. The estimated proportion of Blackwater fish in the fishery is less than 1% (in fact only 1 fish could be assigned with confidence to this river and that the Three Sisters populations contributed at least 98% to the total fishery. No fish could be reliably assigned to the Slaney although two fish showed high affinity to the Boyne. One fish was detected which appeared to assign best to rivers in north-west Ireland and it is speculated that this fish maybe the progeny of releases of ranched fish into the Suir in 2005. The above assignments are provided with a high degree of confidence.

Using a combination of all the markers available to us, the genetic analysis undertaken for fish captured in the experimental fishery suggests that any estimated stock proportions for the Barrow, Nore and Suir individually, have very low confidence due to a failure of assignment replications to converge. Because of this, we do not present stock proportion estimates for these rivers. We have recommended below a number of strategies for increasing the necessary statistical support for stock estimates.

Contributors from the Beaufort Fish Genetics Research Group to this report were as follows: Post Doctoral Fellow: Dr E Dillane (Laboratory analysis), Research assistant: Ms. Mary Cross (Laboratory analysis), Senior Researcher: Dr J. Coughlan (Statistical Analysis and initial report drafting), Principal Investigator: Dr P McGinnity (Report drafting and editing), Scientific co-ordinator: Professor TF Cross (Report drafting and editing). Only Dr Dillane was directly funded by IFI.

**Background and Aims**

The NGSI baseline for Ireland was constructed using 15 microsatellite loci as described in other reports. These loci revealed significant but low genetic differentiation between rivers in the Waterford and Lismore districts of Ireland and in particular between the Blackwater/Bride and the Three Sisters (Barrow, Nore and Suir). This resulted in a failure to discriminate between these rivers or indeed between the rivers in the Waterford and Lismore fisheries districts using available mixed stock statistical techniques. Consequently for analysing and assessing stock contributions in Ireland offshore mixed stock fisheries, these rivers were treated as a single reporting group and was then referred to as the south-eastern population complex (SEPC). In this context, stock proportions for Blackwater, Bride, Barrow, Nore and Suir in Irish fisheries were combined and reported as single stock unit. While this was acceptable for the purposes of the NGSI offshore drift nets programme, no reliable information could be inferred with respect to the individual stock contributions of these rivers.

There are a number of reasons why this situation has occurred and probably not all have been identified. The most likely explanation is that low differentiation may be result of common ancestry of these populations and where consistently large populations have not allowed genetic drift to act in a way to change allele frequencies between populations. The populations that inhabit the Blackwater, Barrow, Nore and Suir are characterised by large population sizes and display high allelic variability at microsatellite loci (as described in the Biodiversity report). This may mean that we are not adequately sampling allele frequencies for these populations (where a large number of alleles occur at different loci the likelihood of sampling all alleles in correct frequencies is severely reduced, especially where allele frequency differences between populations is low,). It may also be that we have not used the optimal set of genetic markers for identifying differentiation between these particular rivers (although the set of 15 microsatellite loci appear to work well for other populations and rivers around Ireland) and that new or different markers will reveal increase discrimination between these rivers and populations.

New markers in terms of five additional microsatellite loci and approximately 210 SNP (single nucleotide polymorphisms) were screened and tested in this respect in a project for the SRFB (Report to the SRFB, 2009). Using simulated samples it was demonstrated that the addition of extra markers could improve discrimination which would be apparent as improved estimates of stock contributions in any future actual experimental fisheries. However, the data presented in the 2009 report were solely based on simulations which always tend to be optimistic in terms of demonstrating discrimination and in particular where differentiation between populations is low.

For the current project we have an opportunity to test the utility of these new markers (in addition to existing loci) for identifying stock contributions in the 2010 inshore Waterford estuary fishery. The proposal of this project "is to undertake research to determine the river of origin of salmon captured in Waterford estuary and to also determine the presence, if any, of salmon from other countries in the estuary sample". Our previous analysis of Waterford estuary samples (2006 for the NGSI project) appeared to identify a significant proportion of fish from the Blackwater (approximately 8.3%) in the fishery. However, given our previous difficulties in discriminating between rivers in the Waterford and Lismore area, it was unsure whether these represented a true contribution or were an artefact of the low differentiation in this area. Therefore, a further aim of this project was to investigate whether new markers will result in separation of Blackwater from the Three Sisters and hence, remove uncertainty with respect to the presence of Blackwater fish in Waterford estuary.

## Methods

### Molecular markers and genetic stock identification baselines

To complement the existing set of 15 microsatellite loci which have been widely screened in populations throughout Ireland (see Biodiversity report), a subset of the Irish baseline has been screened for variation at an additional five microsatellite loci (full details available on request). Approximately 40 rivers sampled in multiple sites have be screened so far and include two samples from the Blackwater, one from the Bride, three from the Suir, one from the Nore, two from the Barrow and two from the Slaney (in fact two samples were screened from the Nore but allele frequencies from the Nore Kings tributary are considered to be unreliable due to a high incidence of relatedness between individuals (see Temporal report)). Other samples from Ireland are variously located around the country and typically target the largest salmon producing rivers.

In addition, the same samples have also been screened (in collaboration with CIGENE, Norway) for variation at approximately 210 SNP loci. These loci were selected from a panel of 388 markers which have been screened in a small number of samples taken from throughout Europe. The 210 loci were selected from the original panel of 388 loci on the basis of being variable (polymorphic) in Irish salmon. To reduce costs, a subset of 92 of these 210 SNPs, were selected for screening of Waterford estuary samples. We attempted to maximise the discriminatory power of the 92 SNPs by exploiting ascertainment bias and with this in mind, loci were chosen based on highest global FST values revealed among Blackwater/Bride, Barrow, Nore and Suir samples.

For the SRFB project in 2009 a number of small rivers of direct interest to Waterford estuary were also screened for variation at the 15 loci included in the NGSI programme. These included the Pollmounty (which has also been screened for temporal variation (see Temporal Stability report 2011)), the Suir Blackwater, the Lingaun and the Clodiagh. The were combined with the rivers from the Three sisters to define reporting groups as Barrow/Pollmounty, Nore (by itself), Suir Blackwater (by itself) and Suir/Lingaun/Clodiagh as defined by SSC management units. It should also be noted that very low levels of differentiation observed in particular between the Pollmounty and

Barrow and between the Suir and Lingaun/Clodiagh, seem to justify the management units.

In all, five baselines were constructed and used for the present analysis. The first of these is the original NGSI baseline which is composed of over 8000 individuals from 84 rivers located throughout Ireland. Secondly, a new baseline was constructed using 79 of the SNP markers and all 20 microsatellite loci. This new baseline was comprised of over 1900 individuals from 23 rivers located throughout Ireland. These baselines were used to compare estimates of stock proportions in the Waterford estuary fishery 2010 with particular reference to determining the presence of Blackwater and Slaney fish in this area as were potentially indicated in the 2006 estuary sample. The specific aim here was to see if the use of additional genetic markers can resolve the difficulties currently apparent in discriminating Blackwater fish in particular, from the Three Sisters.

Three additional baselines were also constructed and used. These were Three Sisters specific and were intended to estimate stock proportions from the rivers draining in Waterford estuary if fishery samples could be reliably determined as coming from these rivers. The first of these was constructed with approximately 1190 individuals and included genotypes from 15 microsatellite loci in samples taken from the Pollmounty, Barrow, Nore, Suir Blackwater, Lingaun, Suir and Clodiagh. The next two baselines were constructed from approximately 260 samples taken from Barrow, Nore and Suir only and included 99 loci (20 microsatellites and 79 SNPs) and SNP loci only, respectively.

**Fishery samples and genetic screening**

Fishing area and date of capture were provided for tissues samples from 460 tissues samples from fish collected in the Waterford estuary 2010 fishery. DNA was extracted from all samples as follows: Individual fin clips were blotted on tissue paper to remove alcohol and placed in a 1.5ml Eppendorf tube. Approximately 300µl of Cell Lysis Solution and 4µl of Proteinase K (20mg/ml) were added and the tube placed in incubator at 37ºC overnight. The tube was cooled to room temperature and 100µl of Protein

Precipitation Solution was added. It was then vortexed for 20 seconds and placed on ice for 10 minutes before centrifuging at 13,000rpm for 5 minutes. A new 1.5ml Eppendorf tube was prepared for each sample containing 300µl of 100% isoproponal (2-proponal). The supernatant of the spun sample (the pellet is protein to be discarded) was removed and added to the isopropanol containing Eppendorf. The DNA pellet was precipitated by inverting the tube. The isopropanol was then decanted off and the pellet was washed in 70% ethanol for 2 hours. The 70% ethanol was then poured off and the pellet was allowed to air dry for 30 minutes. 50-100µl of 0.1X TE buffer was added to re-suspend the DNA. DNA was quantified using a nanodrop spectrophotometer and diluted to a concentration of 100ng/ul for PCR.

Twenty microsatellite loci (the same as those used in the recent Irish National Genetic Stock Identification Programme plus an additional five loci) were used to screen each DNA sample. Amplifications for each locus were carried out in a total volume of 10uL and contained 1uL of DNA, 1X reaction buffer with 0.5U Taq polymerase (Promega), 250uM dNTPs and 1uM forward and reverse primers (one of which was 5'- end-labelled with IRD800 or IRD700 to enable detection of DNA fragments on a Licor DNA sequencer), total $MgCl_2$ concentration for each locus was 2.0mM. Some loci were combined within PCR reactions to reduce costs, and all amplifications followed a PCR profile of 95ºC for 3m followed by 30 cycles of 95ºC for 30s, 56ºC for 30s and 72ºC for 30s. Amplified fragments were stored at 4ºC until use. All loci were resolved separately or mixed in multiplexes on 18 or 25cm 6% polyacrylamide gels on Licor automated DNA sequencers. Allele sizes and genotypes for each individual at each locus were determined by reference to DNA size markers and allele-standards which were loaded with the samples on each gel. All genotypes were scored by eye and double checked for correctness and Individuals which were scored for fewer than eight of the 15 microsatellites were removed (these were considered to be unreliable samples of poor DNA quality).

For SNP locus screening, each sample was carefully quantified and diluted to 25ng/µl prior to being shipped to CIGENE, Norway for screening on a Sequenom platform. After

a period, genotypes for 92 loci for each sample were resolved and delivered to us for analysis. Scrutiny of the data revealed that a small number of these loci failed to produce genotypes for a significant proportion of the samples so were considered unreliable for the remaining fish. These loci were excluded from further analysis. In addition, a few other markers displayed very low levels of variability which would not help to discriminate populations, these were also removed. In all, 79 reliable SNP loci remained and with the addition of microsatellite loci, resulted in 99 loci in all for analysis. Genotypic data from the 460 samples was also carefully scrutinised and any DNA sample that failed to resolve genotypes for less that 90% of all the loci (probably due to poor quality) was also removed from statistical analysis. There remained 452 individuals for stock assessment.

**Statistical analysis and results**

To determine the country of origin of the 452 Waterford estuary samples, a novel individuals clustering method was implemented using the BAPS package. All individuals bar two clustered strongly with Irish baseline samples (for this analysis population differences between Irish rivers/populations were ignored). It therefore seems likely that these individuals (WD_358 and WD_366) both captured on the 29$^{th}$ September 2010 in Station 4 were of non-Irish origin. A detailed analysis of genotypic data for these samples also showed that these had identical genotypes and we consider that the samples are in fact from the same fish sampled twice. Therefore, we estimate that the proportion of non-Irish fish in the Waterford estuary sample 2010 to be 1/451, however, as full regional structure of Atlantic salmon in Europe is still undergoing resolution; it is difficult to establish at this stage where in terms of country of origin, this fish might have originated. The finding of one foreign fish in the 2010 sample contrasts with the results of the 2006 fishery. However, the 2006 samples were collected from some locations which extended further seaward in Waterford estuary than the 2010 samples. Then, up to six fish of non-Irish origin were identified in a sample of approximately 140 individuals. However, we now appreciate the difficulties in assessing the country of origin for some individuals with respect to country and ambiguous results may occur due to high similarities of allele frequencies between some Irish and UK populations. We now

believe that the proportion of non-Irish fish in the 2006 fishery may have been over estimated.

The computer software cBAYES was used to assess stock proportion in the Waterford estuary sample using both the original NGSI baseline and the new baseline constructed with 99 loci (SNPs and microsatellites). For comparison, stock proportions in the Waterford estuary 2006 fishery as revealed by the original NGSI baseline were also re-calculated (as these samples were not screened for additional loci, it is impossible to compare these against the new 99 locus baseline). Estimated stock proportions for the 2006 and 2010 fishery based on the original NGSI baseline, and for the 2010 fishery based on the new 99 locus baseline are presented in Table 1.

**Table 1** Estimated stock proportion as calculated by cBAYES in the 2006 and 2010 Waterford estuary fishery samples revealed by the original NGSI baseline and for the 2010 fishery as revealed by the new 99 locus baseline (Other refers to rivers outside the Slaney, Three Sisters and Blackwater and includes fish which assign to other jurisdictions, the Boyne, the Srahmore river (particularly in the case of the 2006 fishery) and the sum of small non-significant assignments to other rivers).

| | **Waterford 2006 (NGSI baseline)** | **Waterford 2010 (NGSI baseline)** | **Waterford 2010 (99 locus baseline)** |
|---|---|---|---|
| **Slaney** | 1.67 | 0.013 | **0.03** |
| **Three Sisters** | 75.78 | 89.41 | **98.19** |
| **Blackwater** | 8.30 | 9.96 | **0.067** |
| **Other** | 14.25 | 0.62 | **1.71** |

For the Waterford 2006 fishery, a high proportion of fish from outside the Waterford and Lismore areas where identified. These included samples which assigned to the outside Ireland, the Slaney, Boyne and in particular the Srahmore river in Co. Mayo. The proportion of Boyne fish (1 fish) was low overall but was highly significant given the high degree of genetic differentiation between rivers in Waterford/Lismore and those

from eastern Ireland. The presence of Slaney fish was also not entirely unexpected given the geographical proximity of the Slaney to Waterford estuary although, unlike the Boyne fish, confidence in these was low due to genetic similarity of the Slaney to the Three Sisters. The high proportion of fish that appeared to originate from the Srahmore river was a cause for more concern. These fish assigned with high confidence to this river but were later explained as originating from the experimental release of a large number of Burrishoole ranched micro-tagged smolts into the river Suir in 2005. In addition, a relatively high proportion of Blackwater fish were also identified although these seemed unlikely to be present in significant numbers in Waterford estuary and are assumed to be the result of mis-assignment due to high similarity of Blackwater and Three Sisters samples revealed by the 15 microsatellite loci.

In contrast to 2006, the current (2010) samples when compared to the original NGSI baseline revealed a substantially lower stock proportion for rivers outside Waterford and Lismore (no releases of Burrishoole ranched stocks were undertaken in the previous year) but there remained a similar and high proportion of Blackwater fish. However, when the samples were analysed against the new baseline which consisted of 20 microsatellites and 79 SNP loci, cBAYES revealed that the estimated proportion of Blackwater fish in the fishery to be less than 1% (in fact only 1 fish could be assigned with confidence to this river (WD_061 caught on the 6$^{th}$ August at Station 1) and that the Three Sisters stocks contributed at least 98% to the total fishery. No fish could be reliably assigned to the Slaney although two fish (WD_131 and WD_206 caught on the 18$^{th}$ August in Station 2 and the 3$^{rd}$ September in Station 3, respectively) showed high affinity to the Boyne. One fish was detected which appeared to assign best to rivers in north-west Ireland (WD_397 caught on 6$^{th}$ October in Station 3) and it is speculated that this fish maybe the progeny of releases of ranched fish into the Suir (although the Srahmore river has not as yet been included in the 99 locus baseline).

The provision of additional markers, and in particular SNP loci, has been a very successful development for the discrimination of fish from the Blackwater or Three Sisters which was previously ambiguous with 15 microsatellite markers alone.

Unfortunately, the stock proportions of individual three sisters rivers as revealed by cBAYES remains problematic. The algorithms implemented by this software fail to converge individual estimates for these rivers and in different run replications stock estimates for the Barrow, Nore and Suir vary between 0 and 99% each which results low statistical support for these estimates and in any case, extremely high and intolerable 95% confidence intervals (1.4 to 92%) for each stock estimate. Where allele frequency differences are low between contributing populations, as is the case with Three Sisters rivers, changing assignments of individual fish to the different populations in each replication which is updated by other individuals in the sample, results in low confidence of assignments and consequently unreliable stock estimates. We do not, therefore, present individual stock proportions for these rivers as these may be inaccurate and mis-leading.

**Conclusions and Recommendations**

It terms of discriminating between the Blackwater and Three Sisters rivers, this project seems to have been successful as had been suggested from earlier simulations conducted for the SRFB. The Waterford Estuary fishery is intercepting fish originating from rivers draining into this estuary in at least 98% of cases. Therefore, although this fishery must be considered to be a mixed-stock fishery, only Three Sisters rivers should be included in calculations of quotas for future fishing here. The use of additional loci and particular SNPs has been responsible for this result but we recommend that additional blind-test samples (of know origin with respect to Blackwater or Three Sisters (and perhaps elsewhere)) be assembled and screened for these loci to confirm these findings.

The failure to resolve reliable stock estimates for the Three Sisters, as given by microsatellites and SNPs, may also be the result of sampling bias given that these baselines are dominated by the Suir (50%) with the Barrow (33%) and Nore (17%) with lesser representation. Increased resolution may be achieved through the screening of additional samples from these rivers to gain improved allele frequency estimates and equalise baseline sample sizes to reduce assignment bias.

We also recommend that a large number of samples (up to 1000 individuals) of known provenance (large and small populations) be assembled and screened SNPS to test discrepancies between the number of alleles sampled and allele frequencies in different numbers of individuals. This may also help to improve allele frequency estimates for problem rivers and inform us about numbers of individuals required for future baseline construction in respect to lowly differentiated stocks.

We must recommend that, given the success of the application of SNP loci for discriminating between the Blackwater and Three Sisters, additional SNP loci be screened for the Three Sisters rivers in particular, to identify a suite of most appropriate loci for stock discrimination in this area. Over 5,000 SNP markers are currently available for Atlantic salmon. It is envisaged that screening of these loci on population samples

from the Barrow, Nore and Suir will identify a suite of markers which will aid discrimination between these rivers.