Balancing selection on MHC class I in wild brown trout (*Salmo trutta*)

Authors

B O'Farrell

Microbial Phylogeography, Environmental Research Institute (ERI), University College Cork (UCC), Cork, Ireland

C Dennis

Biodiversity and Ecotoxicology, Environmental Research Institute (ERI), University College Cork (UCC), Cork, Ireland

JA Benzie

Biodiversity and Ecotoxicology, Environmental Research Institute (ERI), University College Cork (UCC), Cork, Ireland

P McGinnity

Aquaculture & Fisheries Development Centre (AFDC), School of Biological, Earth and Environmental Sciences (BEES), UCC, Cork, Ireland & Marine Institute, Newport, Co. Mayo, Ireland

J Carlsson

Aquaculture & Fisheries Development Centre (AFDC), School of Biological, Earth and Environmental Sciences (BEES), UCC, Cork

E de Eyto

Marine Institute, Newport, Co. Mayo, Ireland

J Coughlan

Aquaculture & Fisheries Development Centre (AFDC), School of Biological, Earth and Environmental Sciences (BEES), UCC, Cork

F Igoe

Inland Fisheries Ireland, Limerick, Ireland

R Meehan

Talamh Ireland Ltd, 86 Athlumney Castle, Navan, Co. Meath, Ireland

T Cross

Aquaculture & Fisheries Development Centre (AFDC), School of Biological, Earth and Environmental Sciences (BEES), UCC, Cork

Corresponding author:

B O'Farrell

b.ofarrell@ucc.ie

Tel +353 21 490 1980

Running title: Balancing selection on MHC in trout

Abstract

Evidence is reported for balancing selection acting on variation at Major Histocompatibility Complex (MHC) in wild populations of brown trout (*Salmo trutta*). First, variation at an MHC class I–linked microsatellite locus is retained in small trout populations isolated above waterfalls although variation is lost at neutral microsatellite markers. Second, populations across several catchments are less differentiated at the MHC-linked locus than at neutral markers, as predicted by theory. The population structure of these trout was also elucidated.

Keywords

Major Histocompatibility Complex, variation, fish, genetic diversity, isolated populations

Introduction

Genes which are under selection and which, in turn, may be involved in local adaptation are of clear interest to evolutionary biology (Nielsen, 2005). Many studies have attempted to identify genes under selection by contrasting patterns and levels of variation at these genes with genes which are presumed to conform to neutral expectations (Kimura, 1983; Karl & Avise, 1992; Pogson *et al.*, 1995; Jordan *et al.*, 1997; Vitalis & Couvet, 2001; Ford, 2002; Dufresne *et al.*, 2002; Goldringer & Bataillon, 2004; Dhuyvetter *et al.*, 2004; Beaumont & Balding, 2004; Vasemagi *et al.*, 2005; Consuegra *et al.*, 2011).

A particularly interesting phenomenon is the maintenance of extensive allelic polymorphism at a locus by selection. This is usually termed balancing polymorphism, where some form of rare allele advantage is necessarily implicated in the prevention of a particular allelic lineage predominating (Takahata & Nei, 1990). The Major Histocompatibility Complex (MHC) gene family, which is critical for determining self from non-self in immune system responses in vertebrates, is commonly held to be under this type of balancing selection (Hedrick, 1994; Apanius *et al.*, 1997). Classical MHC Class Ia molecules are found on the surface of all the nucleated cells of the body. These are composed of a heavy and light chain encoded by polymorphic MHC Class Ia genes and the invariant β2-microglobulin gene. MHC Class II genes, in contrast, are only expressed in a reduced set of cells, e.g. the antigen-presenting cells such as dendritic cells, B cells and macrophages. MHC Class I and II function in the presentation of self and non-self peptides derived from endogenously (i.e. mutated, misfolded or viral) and exogenously (e.g. bacterial or macroparasitic) derived proteins, to cytotoxic T lymphocytes (CTL) or

helper T cells (Th), respectively. Variation in residues within the peptide binding region of different MHC alleles allows binding of different antigenic peptides.

There are three primary non-mutually exclusive theories on how pathogen-driven balancing selection occurs. These are generalised overdominance (Doherty & Zinkernagel, 1975; Hughes & Nei, 1988), whereby heterozygotes have improved immune surveillance over homozygotes; negative frequency dependence (Clarke & Kirby, 1966; Slade & McCallum, 1992), whereby rare alleles may incur an advantage through, for example, pathogen adaptation to more common MHC alleles; and fluctuating selection (Spurgin & Richardson, 2010) among time and place, which may lead to localised patterns of MHC polymorphism and adaptation (Hill, 1991; Hedrick, 2002; Bernatchez & Landry, 2003; Loiseau *et al.*, 2011). Selection for heterozygosity implies selective equivalence of different alleles and, by extension, that different alleles are maintained at a similar frequency (1/k, where k is the number of alleles) in a given population (Richman, 2000). However, this does not appear to be the case with significantly uneven distribution of alleles being the norm even where many MHC alleles are maintained at an appreciable frequency (Salamon *et al.*, 1999). Additionally, it has been proposed that there may be a sexual selection component to MHC evolution, arising from mate selection improving the inclusive fitness of offspring through assortative or disassortative mating (Trivers, 1972; Hamilton & Zuk, 1982). Sexual selection on MHC has clearly been demonstrated in salmonids (Landry *et al.*, 2001; Bernatchez & Landry, 2003; Pitcher & Neff, 2006; Neff *et al.*, 2008; Consuegra & Garcia de Leaniz, 2008). Recently, a role for kin association in maintaining MHC polymorphism in salmonids has been posited (O'Farrell *et al.*, 2012).

MHC diversity is considered crucial for the ability of populations to resist disease challenges (O'Brien & Evermann, 1988; Muirhead, 2001; Bernatchez & Landry, 2003; Kurtz *et al.*, 2004). In keeping with this hypothesis, MHC variation has been shown to be vulnerable to genetic erosion arising from bottleneck events in New Zealand robins (Petroicidae) (Miller & Lambert, 2004) and in northern elephant seals (*Mirounga angustirostris*) (Weber *et al.*, 2004). Populations which have lost MHC variation may not be viable in the medium to long-term, being less capable of fending off novel disease challenges (O'Brien & Evermann, 1988).

However, the situation appears to be complicated by balancing selection acting to maintain genetic variation at MHC despite the loss of genetic variation at neutral genetic markers (Hedrick *et al.*, 2000a; Hedrick *et al.*, 2000b; Hedrick, 2003; Richardson & Westerdahl, 2003; Richman *et al.*, 2003; Aguilar *et al.*, 2004; Jarvi *et al.*, 2004; van Oosterhout *et al.*, 2006). The suggestion is that intense balancing selection serves to maintain MHC variation in the face of these demographic factors, acting through pathogenic pressures (Klein & O'Huigin, 1994; Jeffery & Bangham, 2000; Prugnolle *et al.*, 2005). Despite theoretical predictions, a meta analysis of published empirical data has shown a moderate but significantly greater loss of variation at MHC than neutral loci (Sutton *et al.*, 2011). This study mainly included data from MHC class II (94%) due to an apparent publication bias.

Selective pressure on MHC may lead to differential maintenance of variation at MHC and at neutral loci. In isolated populations, variation at MHC may be maintained where variation is lost at neutral markers, despite genetic drift and lack of inward gene flow

(Bernatchez & Landry, 2003; Aguilar & Garza, 2006; van Oosterhout *et al.*, 2006; Oliver *et al.*, 2009); and populations should be less differentiated at MHC than at neutral markers, due to higher effective gene flow and more even allele frequency distributions (Muirhead, 2001). Reasonably, population differentiation at a MHC-linked microsatellite should also be less than that at neutral microsatellites. However, higher population differentiation than neutral expectations is normally observed empirically at MHC loci (Bernatchez & Landry, 2003; Sutton *et al.*, 2011). This has been attributed to fluctuating selection (Spurgin & Richardson, 2010) on MHC alleles arising from differential pathogen pressures (Muirhead, 2001). These predictions of balancing selection at MHC class I in wild brown trout (*Salmo trutta* L.*)* populations are tested. Previous work on MHC in non-model vertebrates has tended to focus on MHC class II (Bernatchez & Landry, 2003; Sutton *et al.*, 2011). MHC class I (*UBA*) and class II (*DAA/DAB*) have only one expressed locus each in salmonids and these loci are not linked (Shum *et al.*, 2001; Grimholt *et al.*, 2002; Stet *et al.*, 2002; Aoyagi *et al.*, 2002). In this respect, these genes are not a “complex”, as in other vertebrates, and are referred to as “Major Histocompatibility” (MH) genes in salmonids. However, the common acronym “MHC” is used to refer to them in this paper. The unlinked nature of the class I and class II loci in salmonids allows independent detection of selection on each locus, something which is confounded in similar studies in most other vertebrates.

Materials and Methods

Samples were taken from eight Irish River catchments; (from North to South, see Figure 1) the Owenmore, Owenduff, Burrishoole (Goulaun and Srahrevagh tributaries), Newport (Skerdagh tributary), Owenwee, Carrowniskey, Erriff and Mulkear) (Table I) which were 25-370km apart. In the case of the Burrishoole (Srahrevagh tributary) and Mulkear Rivers (300km apart), these have populations isolated above waterfalls for over 12,200 years. The ice cleared from the Mulkear waterfall 16,500±300 years ago and isostatic uplifting would have created this waterfall 3,000-4,000 years after that (McCabe, 2007). In the case of the Srahrevagh waterfall, ice would have cleared 16,950±50 years ago and isostatic uplifting would, again, have created the waterfall 3,000-4,000 years later (Ballantyne *et al.*, 2008). These two long-term isolated *S. trutta* populations were compared with their downstream counterparts and offer a unique opportunity to study the maintenance of genetic diversity. The data from the broader geographical area across all eight catchments allowed us to test the relative population differentiation of *S. trutta* populations at neutral loci and MHC.

A total of 964 individuals were screened at eight selectively neutral microsatellite loci: *Str73* (Estoup *et al.*, 1993), *Ssa85* and *Ssa197* (O'Reilly *et al.*, 1996), *Ssa2216* (Paterson *et al.*, 2004), *SsOsl417* and *SsOsl85* (Slettan *et al.*, 1995), *F43* (Sanchez *et al.*, 1996) and *Str543* (Presa & Guyomard, 1996) and a microsatellite locus embedded in the untranslated 3' end of the MHC class I locus (*Satr-UBA*). This dinucleotide microsatellite marker has been demonstrated to be tightly linked to the class I locus in Atlantic salmon

(*Salmo salar* L.) (Grimholt *et al.*, 2002). It has been successfully employed previously in studies in *S. salar* (de Eyto *et al.*, 2007; Consuegra *et al.*, 2011) and *S. trutta* (Coughlan *et al.*, 2006; Hansen *et al.*, 2007; O'Farrell *et al.*, 2012). The pattern of linkage has been discussed in some detail in these publications. Briefly, the marker locus is less variable than *Satr-UBA*, with one marker allele often being linked to more than one *Satr-UBA* allele (Coughlan *et al.*, 2006; O'Farrell *et al.*, 2012). It should be noted that marker alleles do not necessarily reflect functional characteristics of linked *UBA* alleles and the proteins they encode. No similar marker was available for MHC class II in these *S. trutta*.

DNA extractions were conducted by dissecting small pieces of tissue (1-5µg) from the samples and added to 0.5ml tubes containing 300µl of 10% (weight/volume) Chelex™ solution. The mixture was heated at 99ºC for 1 hour. Samples were centrifuged at 3000 rpm for 3 min and then stored at -20ºC.

PCR amplifications were carried out in a 10µl reaction volume under the following conditions: 95℃ 3min; (95℃ for 30s, 56℃ for 30s, and 72℃ for 30s) X 30 cycles. Alleles were resolved on 18cm or 25 cm 6% polyacrylamide gels, using a Li-Cor 4200 DNA sequencer. Allele sizes were determined by reference to a 50-350bp size ladder and locus-specific allele size standards. These allele size standards were constructed in the laboratory using the full complement of allele sizes observed in pilot studies, to enable consistent scoring amongst batches of individuals screened for each locus. When initial genotyping was unclear due to gel electrophoresis problems or weak amplification (~4% of genotypes), *S. trutta* fry samples were re-extracted and re-screened. Large allele dropout was identified as an occasional problem but large alleles could usually be reliably scored after re-screening. Following re-screening, the final estimated error rate was ≤0.5% of

composite genotypes per individual (Coughlan *et al.*, 2006).

The MICROCHECKER application (Van Oosterhout *et al.*, 2004) was used to help identify general problems such as mistyping, typographical and scoring errors together with a number of null allele tests prior to further analysis of these *S. trutta* populations. This helped establish whether particular loci might best be removed from further analysis due to unsatisfactory error rates. MICROCHECKER was run with a maximum expected allele size of 300bp. Unusual observations were checked and a randomisation procedure (1,000 randomisations) with Bonferroni correction was used for all tests. Missing or suspect data were omitted from the analysis. MICROCHECKER analysis uncovered no evidence of loci presenting problems.

Only the two known samples from *S. trutta* isolated above waterfalls in the Mulkear and Srahrevagh were assumed *a priori* to be populations. Instead, an analysis was conducted on all *S. trutta* samples using the software STRUCTURE (Pritchard *et al.*, 2000; Falush *et al.*, 2003) with input files created using CONVERT (Glaubitz, 2004). The range of values for the number of clusters (K) was narrowed using three short runs (10,000 burn-in and 10,000 MCMC iterations thereafter) of STRUCTURE (Pritchard *et al.*, 2000; Falush *et al.*, 2003; Evanno *et al.*, 2005), with an admixture model with correlated allele frequencies, in the range K=2 to K=17. Having narrowed the range suitably and plotted the log likelihood values for these runs, an admixture model with correlated allele frequencies was again used and the appropriate number of clusters (K) identified using three runs each at each value of K between 9 and 14 (Burn-in 100,000 with subsequent 1,000,000 MCMC iterations). The appropriate number of clusters (K) was identified with reference to plots of Ln P(D) values while setting a cut off for K at that level wherein additional clusters had no

obvious explanation in geography and where few, if any, individuals strongly assigned to the additional cluster (Falush *et al.*, 2003). The program DISTRUCT was used to generate high quality graphical outputs of the STRUCTURE results (Rosenberg, 2004).

F_{ST} statistics (Φ) (Weir & Cockerham, 1984) between all populations were calculated for the neutral microsatellite loci and the MHC-linked locus, separately, in GENETIX v4.04 (Belkhir *et al.*, 2004). These Φ values have been tabulated using Python software developed in this study, CREATEMATRIX. A jacknifing approach was used to estimate neutral Φ statistics. Φ statistics were re-calculated with the removal of one of the eight loci each time. This allowed us to calculate the 99.9% confidence limits in neutral Φ statistics for each population pair. It was then assessed whether the Φ statistic for the MHC-linked microsatellite was greater or less than the respective neutral Φ statistic and whether they fell inside or outside the 99.9% confidence limits for the neutral Φ statistic. Hardy-Weinberg exact tests were implemented in GENEPOP (Raymond & Rousset, 1995).

Unless specified, subsequent statistical analyses were implemented in Python scripts using standard approaches. Individual heterozygosity was calculated across the eight neutral loci and, separately, at the MHC class I linked marker. For each pairwise comparison, unpaired t-tests on the binomial data for heterozygosity were conducted for each locus. Paired t-tests were also conducted in SPSS on the proportion of heterozygotes over the eight neutral loci.

Allelic richness (AR) for each population was estimated by a bootstrap procedure which corrected for sample size differences. The smallest sample was the Mulkear BW sample, with data for 27 diploid individuals for one locus (Table I). At each iteration of a bootstrap procedure, allelic richness was estimated for each population by taking a random sample of 54 gene copies (27 X2) (*g*) from the frequency distribution at each locus and counting the number of alleles observed for that locus. Allelic richness for each of the eight neutral loci and MHC was calculated as the average over 100,000 bootstraps. A summary statistic of neutral allelic richness for each population was further calculated as the average AR across the eight neutral microsatellite loci for each bootstrap iteration and then the average of those values across the 100,000 bootstraps.

A bootstrap method (Coughlan *et al.*, 2006) was adapted for significance tests on allelic richness at neutral loci and MHC in each case study. In each pairwise population comparison of variability, and under the null model that the two samples do not differ in variability, an allele frequency distribution for each locus was estimated from pooled genotypic data from the two samples. As per the calculation of allelic richness, two samples of 54 gene copies were drawn, with replacement, from this distribution to create a pair of simulated samples. Neutral and MHC allelic richness statistics were calculated (as above). The absolute difference in neutral allelic richness between the two simulated samples was used as the test statistic. This was repeated over 100,000 bootstraps to provide a null distribution for the test statistic. A null distribution for MHC allelic richness was constructed similarly. The proportion of simulated test statistic values which exceeded the test statistic value for the real samples provided a test for significant differences in variability above and below the waterfalls in the Srahrevagh and Mulkear. Additionally, paired sample t-tests were conducted in SPSS on AR statistics for the eight neutral loci.

Results

STRUCTURE analysis found that K=11 was the most appropriate number of clusters across the eight river catchments sampled. Values of K≥12 were not well supported by Ln P(D) values and additional clusters had no individuals strongly assigning to them nor any explicable geographical basis. The eleven clusters identified tended to have broad agreement with river catchments. For instance, good assignment of most individuals to a particular cluster was found in the Owenwee (Cluster 8; 56.3%), Owenduff (Cluster 11; 70.5%), the Goulaun (Cluster 2; 56.0%), Owenmore (Cluster 9; 45.5%) and Erriff (Cluster 6; 44.0%) (see Table I, Fig. 2).

The two populations isolated above waterfalls in the Srahrevagh and Mulkear demonstrated the least admixture amongst clusters. Virtually all individuals strongly assigned to cluster 4 in the Srahrevagh AW sample and the overall sample assignment was 86.1%. However, individuals could be identified in the Srahrevagh below the waterfall (BW) which strongly assigned to the same cluster, 4, at Q≥0.5 (number of individuals, n=8). These were considered likely to be downstream migrants. This was not unexpected as it is possible that fry may pass down over the waterfall but fish are not able to get up the waterfall. This does suggest the possibility for some unidirectional gene flow from above the waterfall to below and there were a further small number (n=12) of fry in the Srahrevagh (BW) sample which may be admixed, showing intermediate assignment (0.1<Q≤0.5) to the “above waterfall cluster”. However, there was strong differentiation

between the above and below waterfall populations in the Srahrevagh (Φ = 0.077, $p<0.001$). This was not the case in the Mulkear (Φ = 0.007, ns) where the same cluster was found in both samples (n=5) although there was more evidence of admixture below the waterfall (Mulkear AW 90.5%, Mulkear BW 78.9%).

There were some important exceptions to the general trend of agreement between river systems and clustering patterns. The Carrowniskey (Cluster 6 44.7%) and Lough Alisheen (Cluster 1 77.0%, see Fig 2) samples were initially considered to be taken from the same population but preliminary tests, including the use of STRUCTURE, identified strong population differentiation (Φ = 0.074, $p<0.001$). The two were considered separately in subsequent analyses. Cluster 6 was the most common cluster in both the Carrowniskey (44.7%) and Erriff (44.0%). The Carrowniskey and the Erriff have a much lower Φ (0.0194) than do the Carrowniskey and the next nearest neighbour, the Owenwee (Φ = 0.055). There is no obvious reason for this similarity between the Carrowniskey and Erriff, although the mouths of the rivers are reasonably close. Cryptic population structure was identified in the Skerdagh S1 and S4 samples. Both are largely composed of cluster 7 (77.2% and 33.3%, respectively, Fig. 2) but the Skerdagh S4 sample demonstrates far more admixture with cluster 8 (23.5%) (which is mainly found in the Owenwee) and is significantly differentiated from Skerdagh S1 (Φ = 0.055, $p<0.001$). The Skerdagh S4 sample is less differentiated from the Owenwee (Φ =0.038) than the Skerdagh S1 sample is from Owenwee (Φ = 0.079). The Goulaun (56.0%) and Srahrevagh (BW) (17.5%) samples are both in the Burrishoole system, cluster 2 was common in both but the Srahrevagh (BW) demonstrated admixture with two other clusters, 3 (29.2%) and 10 (29.5%), which were much less common in the Goulaun (4.4% and 11.3%, respectively).

STRUCTURE identified clear incidences of cryptic population structure which could be resolved, *post hoc*, to discrete sub-samples of small tributaries or different reaches of the same tributary, as in the Carrowniskey and Skerdagh systems, and where Φ estimates of population differentiation were significant. These were considered as the Skerdagh (S1), Skerdagh (S4), Carrowniskey and Lough Alisheen populations in subsequent analyses. The Mulkear above and below samples did not show significant population differentiation. Consequently, the Mulkear (BW) and Mulkear (AW) samples were combined for the analysis comparing population differentiation at neutral loci and MHC.

Hardy-Weinberg exact tests found no deviations from expectations across loci in any of the populations. It was concluded from STRUCTURE and Φ estimates of population differentiation (Table II) that the overall population structure was best defined by 12 populations (data are presented for the Mulkear (BW) and Mulkear (AW) samples separately in Table I).

Variation at the neutral markers was significantly lower in populations isolated above waterfalls in both rivers, as measured by individual heterozygosity and allelic richness (Table I, Fig. 3, 4). Individual heterozygosity at neutral markers in the Srahrevagh had a median value of 0.750 below the waterfall and 0.625 above (Mann-Whitney test, Z=-6.587, p=0.001). A paired t-test on locus by locus proportions of heterozygotes at neutral loci also showed significantly lower variation above the waterfall (t=-4.812, df=7, p=0.002). In the Mulkear, median individual heterozygosity at neutral markers was 0.571 below the

waterfall and 0.470 above (Mann-Whitney test, Z=-2.976, p=0.003) while the paired t-test on locus by locus proportions of heterozygotes at the neutral loci was also significant (t=-3.185, df=7, p=0.015).

Neutral allelic richness was significantly greater below the waterfall in the Srahrevagh (7.59, CI95%±0.0019) than above the waterfall (5.21, CI95%±0.0012) (Bootstrap test, p=0.000001; Paired sample t-test, t=-3.722, df=7, p=0.007). Neutral allelic richness was also significantly greater below the waterfall in the Mulkear (4.70, CI95%±0.0007) than above the waterfall (3.47, CI95%±0.0013) (Bootstrap test, p=0.046; Paired sample t-test t=-3.087, df=7, p=0.018).

Variation at the MHC class I-linked locus was not significantly reduced for either individual heterozygosity or allelic richness in the populations isolated above waterfalls (Table I, Fig. 3, 4). The proportion of heterozygotes in the Srahrevagh above the waterfall was 0.852 while the proportion below was 0.885 (t-test, df=193, t=0.511, p=0.610). In the Mulkear, the proportion of heterozygotes was not significantly different above (0.778) and below (0.795) the waterfall (t-test, df=100, t=0.188, p=0.851). Allelic richness was not significantly different above (8.58, CI95%±0.0040) and below (7.84, CI95%±0.0062) the waterfall in the Srahrevagh (bootstrap test, p=0.636). However, allelic richness was actually, marginally, significantly higher above the waterfall in the Mulkear (6.54, CI95%±0.0037) than below (5.00, CI95%±0.0000) (bootstrap test, p=0.047).

Population differentiation (Φ) (Table II) was significantly less at the MHC class I-

linked locus (mean 0.078±0.0050) than at neutral loci (mean 0.104±0.0074, Wilcoxon Signed Rank Test, Z=-2.701, p<0.001) amongst the twelve distinct populations identified (the Mulkear was considered one population for this analysis due to the lack of a significant Φ between the two Mulkear samples), with significantly lower MHC class I Φ seen in 43 of 66 population pairs (Jacknife test, CI 99.9% on neutral expectations, Fig 5).

Discussion

This study tested two predictions of balancing selection on MHC class I in wild *S. trutta* and found strong evidence for both. First, variation was significantly lower at neutral genetic markers in isolated populations but was maintained at the MHC class I-linked locus. Balancing selection on MHC is thought to be largely driven by exposure to a high diversity of pathogens (Klein & O'Huigin, 1994; Jeffery & Bangham, 2000; Prugnolle *et al.*, 2005) but it should be noted that the waterfalls pose barriers to many pathogens and novel disease vectoring. One study of *S. salar* gut microfauna found it to be remarkably depauperate and dominated by *Mycoplasma* spp., ordinarily obligate intracellular parasites (Holben *et al.*, 2002). This may not be surprising given these salmonids are found in upland systems which amount to freshwater flow-through systems. There are reasonable grounds to conclude that pathogenic pressures above waterfalls differ markedly from those below the waterfall and are likely to be reduced. Sexual selection may also influence the maintenance of genetic variation at MHC in these salmonids (Landry *et al.*, 2001; Bernatchez & Landry, 2003; Pitcher & Neff, 2006; Neff *et al.*, 2008; Consuegra & Garcia de Leaniz, 2008). O' Farrell *et al.* (2012) found evidence for kin association in *S. trutta* below the waterfall in the Srahrevagh tributary based on the sharing of MHC alleles. They went

on to argue how this phenomenon could lead to a form of kin recognition-driven rare allele advantage (Grafen, 1990) leading to balancing selection on MHC in these *S. trutta*. This could also explain the maintenance of MHC variation above the waterfall in the Srahrevagh.

It is not known whether the suite of MHC alleles maintained by alternate mechanisms above waterfalls provide downstream migrants with good resistance to pathogens encountered below the waterfall. Within the isolated populations, this type of behaviourally-mediated balancing selection is decoupled from broader disease pressures. Behaviourally-mediated balancing selection relies on the finite ability of individuals to identify alleles which are different from their own. MHC alleles which have a large number of amino acid pairwise difference to other alleles, such as recombinant alleles, may tend to be favoured. Interestingly, preliminary MHC class I (*Satr-UBA*) sequence data from a sample of adult *S. trutta* in 2004 (unpublished data) found that fish assigned to the above waterfall population had a significantly more divergent suite of MHC alleles when mean amino acid pairwise distances (0.40±0.024) were looked at, than those found in fish assigned to the below waterfall population (0.32±0.022), Mann-Whitney $U=624.0$, $Z=-3.684$, $P<0.001$.

A suite of MHC alleles with a large mean amino acid sequence distance does not imply a similarly diverse immuno-surveillance capacity. In practice, MHC alleles fall into a smaller number of “supertypes” based on their antigen binding capacity (Sette *et al.*, 2003). Behaviourally-mediated selection on MHC alleles, when divorced from pathogen-driven balancing selection over any considerable length of time, may lead to a form of

runaway selection, wherein alleles with rare or even maladaptive antigen binding capacity are favoured. Consequently, some MHC alleles in isolated populations may prove maladaptive in downstream migrants placing them at a selective disadvantage if they are poorly able to deal with more varied pathogenic pressures downstream. Conversely, some of these exotic MHC alleles may provide migrants with unique capacity to defend against epidemics of novel pathogens in the downstream population.

Second, population differentiation (Φ) was significantly less at MHC class I-linked locus across the study as a whole. This is as predicted for a gene under balancing selection or one closely linked to such a locus (Muirhead, 2001). The opposite has been observed for most studies of MHC (Muirhead, 2001; Landry & Bernatchez, 2001; Bernatchez & Landry, 2003; Aguilar & Garza, 2006; Sutton *et al.*, 2011). However, these have usually compared MHC sequence data-derived F_{ST} values with those from neutral microsatellites. Neutral data in the Sutton *et al.* (2011) meta-analysis paper was largely derived from neutral microsatellites (74%). The approach here, comparing F_{ST} (Φ) at neutral and a MHC-linked microsatellite, avoids a potential bias in comparing F_{ST} derived from different types of genetic marker. For example, if these previous studies had compared their MHC sequence data with neutral SNP data (neutral SNP F_{ST} is nearly three times that at neutral microsatellites in salmonids (Narum *et al.*, 2008)) it is possible, even likely, that they would have found lower population differentiation at MHC. Another possibility may be that there are more issues with homology at the MHC-linked microsatellite locus than the neutral loci, which would depress Φ estimates at the former, although there are no data to support this.

Directional selection can also cause lower differentiation at MHC through exposure to the same pathogen (Teacher *et al.*, 2009; Fraser & Neff, 2010). This occurs because a specific pathogen will select for and against the same MHC alleles in separate populations, causing their allele frequencies to become more similar. No agent of homogenising, directional selection could explain the lower differentiation across the *S. trutta* populations that have been monitored for several decades. However, this issue was not examined directly and this may be an interesting avenue for future research.

However, the significantly higher Φ values for MHC seen in 12 of 66 population comparisons may be explained by directional selection, as each of these comparisons involved the Burrishoole, Erriff and Skerdagh rivers (Muirhead, 2001). *S. trutta* in these rivers have a history of disease exposure associated with *S. salar* aquaculture and perturbations associated with fisheries management, not experienced by the other populations sampled. Localised bursts of directional selection (selective sweeps) may have occurred in the Burrishoole, Erriff and Skerdagh rivers and it is clear that an interplay of directional and balancing selection may occur. In the case of the Skerdagh, where cryptic population structure was observed, the neutral Φ between Skerdagh S1 and S4 was 0.058 while the Φ value at the MHC-linked locus was 0.139. Disease might help explain the cryptic population structure in the Skerdagh. The overall contrast in pattern for disturbed/disease-affected populations versus pristine populations in this study may be noteworthy given the growing interest in using selected markers like MHC for identifying stocks in conservation genetics. This is an interesting anecdotal finding. The aquaculture practices in the Mayo region involve only *S. salar*. As such, there is potential for disease exposure to both native *S. salar* and *S. trutta* but only the potential for gene flow from aquaculture escapes in one species. One possible follow-on study would be to examine

gene flow at neutral and at selected loci such as MHC in native *S. salar* and *S. trutta* populations within the study region.

It is concluded that balancing selection at MHC best explains the overall observations of lower than expected differentiation at MHC, and the maintenance of significantly higher variation at MHC than expected in the isolated populations. This study has presented clear evidence of balancing selection on MHC class I in the wild.

Acknowledgments

HEA PRTLI Cycle 3; Beaufort Marine Research Award: Fish population genetics. Irish Government NDP (2007-2014) administered by the Marine Institute; Inland Fisheries Ireland; Science Foundation Ireland (Microbial Phylogeography 05/FE1/B882); Marine Institute (Newport) field staff

Conflict of Interest

The authors declare no conflict of interest.

518 **References**

519

520

521 Aguilar, A., & J. Garza. (2006). A comparison of variability and population structure for
522 major histocompatibility complex and microsatellite loci in California coastal steelhead
523 (*Oncorhynchus mykiss* Walbaum). *Molecular Ecology* **15,** 923-937

524 Aguilar, A., G. Roemer, S. Debenham, M. Binns, D. Garcelon, & R. K. Wayne. (2004). High
525 MHC diversity maintained by balancing selection in an otherwise genetically monomorphic
526 mammal. *Proceedings of the National Academy of Sciences of the United States of*
527 *America* **101,** 3490-3494

528 Aoyagi, K., J. M. Dijkstra, C. Xia, I. Denda, M. Ototake, K. Hashimoto, & T. Nakanishi.
529 (2002). Classical MHC class I genes composed of highly divergent sequence lineages
530 share a single locus in rainbow trout (*Oncorhynchus mykiss*). *The Journal of Immunology*
531 **168,** 260-273

532 Apanius, V., D. Penn, P. R. Slev, L. R. Ruff, & W. K. Potts. (1997). The nature of selection
533 on the major histocompatibility complex. *Critical Reviews in Immunology* **17,** 179-224

534 Ballantyne, C. K., J. O. Stone, & D. McCarroll. (2008). Dimensions and chronology of the
535 last ice sheet in Western Ireland. *Quaternary Science Reviews* **27,** 185-200

536 Beaumont, M. A., & D. J. Balding. (2004). Identifying adaptive genetic divergence among
537 populations from genome scans. *Molecular Ecology* **13,** 969-980

538 Bernatchez, L., & C. Landry. (2003). MHC studies in nonmodel vertebrates: what have we
539 learned about natural selection in 15 years? *Journal of Evolutionary Biology* **16,** 363-377

540 Clarke, B., & D. Kirby. (1966). Maintenance of histocompatibility polymorphism. *Nature*

211, 999-1000

Consuegra, S., E. de Eyto, P. McGinnity, R. J. M. Stet, & W. C. Jordan. (2011). Contrasting responses to selection in class I and class II alpha major histocompatibility-linked markers in salmon. *Heredity* **107,** 143-154

Consuegra, S., & C. Garcia de Leaniz. (2008). MHC-mediated mate choice increases parasite resistance in salmon. *Proceedings of the Royal Society of London.Series B, Biological Sciences* **275,** 1397-1403

Coughlan, J., P. McGinnity, B. O'Farrell, E. Dillane, O. Diserud, E. de Eyto, K. Farrell, K. Whelan, R. J. M. Stet, & T. F. Cross. (2006). Temporal variation in an immune response gene (MHC I) in anadromous *Salmo trutta* in an Irish river before and during aquaculture activities. *ICES Journal of Marine Science: Journal du Conseil* **63,** 1248-1255

de Eyto, E., P. McGinnity, S. Consuegra, J. Coughlan, J. Tufto, K. Farrell, H. J. Megens, W. C. Jordan, T. F. Cross, & R. J. M. Stet. (2007). Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. *Proceedings of the Royal Society of London.Series B, Biological Sciences* **274,** 861-869

Dhuyvetter, H., E. Gaublomme, & K. Desender. (2004). Genetic differentiation and local adaptation in the salt-marsh beetle Pogonus chalceus: a comparison between allozyme and microsatellite loci. *Mol Ecol.* **13,** 1065-1074

Doherty, P., & R. Zinkernagel. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* **256,** 50-52

Dufresne, F., E. Bourget, & L. Bernatchez. (2002). Differential patterns of spatial divergence in microsatellite and allozyme alleles: further evidence for locus-specific selection in the acorn barnacle, *Semibalanus balanoides*? *Molecular Ecology* **11,** 113-123

564 Estoup, A., P. Presa, F. Krieg, D. Vaiman, & R. Guyomard. (1993). (Ct)(N) and (Gt)(N)
565 Microsatellites - A new class of genetic markers for *Salmo trutta* L. (brown trout). *Heredity*
566 **71,** 488-496

567 Evanno, G., S. Regnaut, & J. Goudet. (2005). Detecting the number of clusters of
568 individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14,**
569 2611-2620

570 Falush, D., M. Stephens, & J. K. Pritchard. (2003). Inference of population structure using
571 multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164,**
572 1567-1587

573 Ford, M. J. (2002). Applications of selective neutrality tests to molecular ecology. *Mol Ecol.*
574 **11,** 1245-1262

575 Fraser, B. A., & B. D. Neff. (2010). Parasite mediated homogenizing selection at the MHC
576 in guppies. *Genetica* **138,** 273-278

577 Glaubitz, J. C. (2004). CONVERT: A user-friendly program to reformat diploid genotypic
578 data for commonly used population genetic software packages. *Molecular Ecology Notes*
579 **4,** 309-310

580 Goldringer, I., & T. Bataillon. (2004). On the distribution of temporal variations in allele
581 frequency: consequences for the estimation of effective population size and the detection
582 of loci undergoing selection. *Genetics* **168,** 563-568

583 Grafen, A. (1990). Do animals really recognize kin? *Animal Behaviour* **39,** 42-54

584 Grimholt, U., F. Drablos, S. M. Jorgensen, B. Hoyheim, & R. J. M. Stet. (2002). The major
585 histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage
586 analysis and protein modelling. *Immunogenetics* **54,** 570-581

587 Hamilton, W., & M. Zuk. (1982). Heritable true fitness and bright birds: a role for parasites?
588 *Science* **218,** 384-387

589 Hansen, M., O. Skaala, L. Jensen, D. Bekkevold, & K.-L. Mensberg. (2007). Gene flow,
590 effective population size and selection at major histocompatibility complex genes: brown
591 trout in the Hardanger Fjord, Norway. *Molecular Ecology* **16,** 1413-1425

592 Hedrick, P. W. (1994). Evolutionary Genetics of the Major Histocompatibility Complex.
593 *American Naturalist* **143,** 945-964

594 Hedrick, P. W. (2002). Pathogen resistance and genetic variation at MHC loci. *Evolution*
595 **56,** 1902-1908

596 Hedrick, P. W. (2003). The major histocompatibility complex (MHC) in declining
597 populations: an example of additive variation. Pages 94-111 In *Reproduction Science and*
598 *Integrated Conservation* W. Holt, A. Pickard, J. Rodger, and D. Wildt editors. Cambridge,
599 UK: Cambridge University Press.

600 Hedrick, P. W., R. N. Lee, & K. M. Parker. (2000a). Major histocompatibility complex
601 (MHC) variation in the endangered Mexican wolf and related canids. *Heredity* **85,** 617-624

602 Hedrick, P. W., K. M. Parker, G. A. Gutierrez-Espeleta, A. Rattink, & K. Lievers. (2000b).
603 Major histocompatibility complex variation in the Arabian oryx. *Evolution* **54,** 2145-2151

604 Hill, A. V. (1991). HLA assocations with malaria in Africa: some implications for MHC
605 evolution. Pages 403-420 In *Molecular Evolution of the Major Histocompatibility Complex*
606 J. Klein, and D. Klein editors. Berlin, Germany: Springer.

607 Holben, W. E., P. Williams, M. Saarinen, L. K. Sarkilahti, & J. H. A. Apajalahti. (2002).
608 Phylogenetic analysis of intestinal microflora indicates a novel *Mycoplasma* phylotype in
609 farmed and wild salmon. *Microbial Ecology* **44,** 175-185

610 Hughes, A. L., & M. Nei. (1988). Pattern of nucleotide substitution at major
611 histocompatibility complex class I loci reveals overdominant selection. *Nature* **335,** 167-
612 170

613 Jarvi, S. I., C. L. Tarr, C. E. McIntosh, C. T. Atkinson, & R. C. Fleischer. (2004). Natural
614 selection of the major histocompatibility complex (MHC) in Hawaiian honeycreepers
615 (Drepanidinae). *Molecular Ecology* **13,** 2157-2168

616 Jeffery, K. J. M., & C. R. M. Bangham. (2000). Do infectious diseases drive MHC diversity?
617 *Microbes and Infection* **2,** 1335-1341

618 Jordan, W. C., E. Verspoor, & A. F. Youngson. (1997). The effect of natural selection on
619 estimates of genetic divergence among populations of the Atlantic salmon. *Journal of Fish*
620 *Biology* **51,** 546-560

621 Karl, S. A., & J. C. Avise. (1992). Balancing selection at allozyme loci in oysters:
622 implications from nuclear RFLPs. *Science* **256,** 100-102

623 Kimura M. (1983). *The neutral theory of molecular evolution*. Cambridge: Cambridge
624 University Press .

625 Klein, J., & C. O'Huigin. (1994). MHC Polymorphism and Parasites. *Philosophical*
626 *Transactions of the Royal Society of London Series B-Biological Sciences* **346,** 351-358

627 Kurtz, J., M. Kalbe, P. B. Aeschlimann, M. A. Haberli, K. M. Wegner, T. B. Reusch, & M.
628 Milinski. (2004). Major histocompatibility complex diversity influences parasite resistance
629 and innate immunity in sticklebacks. *Proceedings of the Royal Society B: Biological*
630 *Sciences* **271,** 197-204

631 Landry, C., & L. Bernatchez. (2001). Comparative analysis of population structure across
632 environments and geographical scales at major histocompatibility complex and

633 microsatellite loci in Atlantic salmon (*Salmo salar*). *Molecular Ecology* **10,** 2525-2539

634 Landry, C., D. Garant, P. Duchesne, & L. Bernatchez. (2001). 'Good genes as
635 heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon
636 (*Salmo salar*). *Proceedings of the Royal Society of London Series B-Biological Sciences*
637 **268,** 1279-1285

638 Loiseau, C., R. Zoorob, A. Robert, O. Chastel, R. Julliard, & G. Sorci. (2011). Plasmodium
639 relictum infection and MHC diversity in the house sparrow (*Passer domesticus*).
640 *Proceedings of the Royal Society of London Series B-Biological Sciences* **278,** 1264-1272

641 McCabe A. M. (2007). *Glacial Geology and Geomorphology: The Landscapes of Ireland*.,
642 illustrated edition edition. Edinburgh: Dunedin Academic Press .

643 Miller, H. C., & D. M. Lambert. (2004). Genetic drift outweighs balancing selection in
644 shaping post-bottleneck major histocompatibility complex variation in New Zealand robins
645 (Petroicidae). *Molecular Ecology* **13,** 3709-3721

646 Muirhead, C. A. (2001). Consequences of population structure on genes under balancing
647 selection. *Evolution* **55,** 1532-1541

648 Narum, S. R., M. Banks, T. D. Beacham, M. R. Bellinger, M. R. Campbell, J. Dekoning, A.
649 Elz, C. M. Guthrie, III, C. Kozfkay, K. M. Miller, P. Moran, R. Phillips, L. W. Seeb, C. T.
650 Smith, K. Warheit, S. F. Young, & J. C. Garza. (2008). Differentiating salmon populations at
651 broad and fine geographical scales with microsatellites and single nucleotide
652 polymorphisms. *Molecular Ecology* **17,** 3464-3477

653 Neff, B., S. Garner, J. Heath, & D. Heath. (2008). The MHC and non-random mating in a
654 captive population of Chinook salmon. *Heredity* **101,** 175-185

655 Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*

39, 197-218

O'Brien, S. J., & J. F. Evermann. (1988). Interactive influence of infectious disease and genetic diversity in natural populations. *Trends in Ecology & Evolution* **3,** 254-259

O'Farrell, B., J. A. H. Benzie, P. McGinnity, J. Carlsson, E. d. Eyto, E. Dillane, C. Graham, J. Coughlan, & T. Cross. (2012). MHC-mediated spatial distribution in brown trout (*Salmo trutta*) fry. *Heredity* **108,** 403-409

O'Reilly, P. T., L. C. Hamilton, S. K. McConnell, & J. M. Wright. (1996). Rapid analysis of genetic variation in Atlantic salmon *(Salmo salar)* by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. *Canadian Journal of Fisheries and Aquatic Sciences* **53,** 2292-2298

Oliver, M. K., X. Lambin, T. Cornulier, & S. B. Piertney. (2009). Spatio-temporal variation in the strength and mode of selection acting on major histocompatibility complex diversity in water vole (*Arvicola terrestris*) metapopulations. *Molecular Ecology* **18,** 80-92

Paterson, S., S. B. Piertney, D. Knox, J. Gilbey, & E. Verspoor. (2004). Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. *Molecular Ecology Notes* **4,** 160-162

Pitcher, T. E., & B. D. Neff. (2006). MHC class IIB alleles contribute to both additive and nonadditive genetic effects on survival in Chinook salmon. *Molecular Ecology* **15,** 2357-2365

Pogson, G. H., K. A. Mesa, & R. G. Boutilier. (1995). Genetic population structure and gene flow in the Atlantic cod *Gadus morhua*: a comparison of allozyme and nuclear RFLP loci. *Genetics* **139,** 375-385

Presa, P., & R. Guyomard. (1996). Conservation of microsatellites in three species of

679 salmonids. *Journal of Fish Biology* **49,** 1326-1329

680 Pritchard, J. K., M. Stephens, & P. Donnelly. (2000). Inference of population structure using
681 multilocus genotype data. *Genetics* **155,** 945-959

682 Prugnolle, F., A. Manica, M. Charpentier, J. Guégan, V. Guernier, & F. Balloux. (2005).
683 Pathogen-Driven Selection and Worldwide HLA Class I Diversity. *Current Biology* **15,**
684 1022-1027

685 Raymond, M., & F. Rousset. (1995). Genepop (Version-1.2) - Population-Genetics
686 Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86,** 248-249

687 Richardson, D. S., & H. Westerdahl. (2003). MHC diversity in two Acrocephalus species:
688 the outbred Great reed warbler and the inbred Seychelles warbler. *Molecular Ecology* **12,**
689 3523-3529

690 Richman, A. (2000). Evolution of balanced genetic polymorphism. *Molecular Ecology* **9,**
691 1953-1963

692 Richman, A. D., L. G. Herrera, & D. Nash. (2003). Evolution of MHC class II E beta
693 diversity within the genus *Peromyscus*. *Genetics* **164,** 289-297

694 Rosenberg, N. A. (2004). Distruct: a program for the graphical display of population
695 structure. *Molecular Ecology Notes* **4,** 137-138

696 Salamon, H., W. Klitz, & S. Easteal. (1999). Evolution of HLA class II molecules: allelic and
697 amino acid site variability across populations. *Genetics* **152,** 393-400

698 Sanchez, J. A., C. Clabby, D. Ramos, G. Blanco, F. Flavin, E. Vazquez, & R. Powell.
699 (1996). Protein and microsatellite single locus variability in *Salmo salar* L. (Atlantic
700 salmon). *Heredity* **77,** 423-432

Sette, A., J. Sidney, B. Livingston, J. Dzuris, C. Crimi, C. Walker, S. Southwood, E. Collins, & A. Hughes. (2003). Class I molecules with similar peptide-binding specificities are the result of both common ancestry and convergent evolution. *Immunogenetics* **54,** 830-841

Shum, B. P., L. Guethlein, L. R. Flodin, M. A. Adkison, R. P. Hedrick, R. B. Nehring, R. J. M. Stet, C. Secombes, & P. Parham. (2001). Modes of salmonid MHC class I and II evolution differ from the primate paradigm. *Journal of Immunology* **166,** 3297-3308

Slade, R. W., & H. I. McCallum. (1992). Overdominant vs. frequency-dependent selection at MHC loci. *Genetics* **132,** 861-864

Slettan, A., I. Olsaker, & O. Lie. (1995). Atlantic salmon, *Salmo salar*, microsatellites at the *Ssosl25*, *Ssosl85*, *Ssosl311*, *Ssosl417* loci. *Animal Genetics* **26,** 281-282

Spurgin, L. G., & D. S. Richardson. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society of London Series B-Biological Sciences* **277,** 979-988

Stet, R. J. M., B. de Vries, K. Mudde, T. Hermsen, J. van Heerwaarden, B. P. Shum, & U. Grimholt. (2002). Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. *Immunogenetics* **54,** 320-331

Sutton, J. T., S. Nakagawa, B. C. Robertson, & I. G. Jamieson. (2011). Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Mol Ecol.* **20,** 4408-4420

Takahata, N., & M. Nei. (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124,** 967-978

Teacher, A. G., T. W. Garner, & R. A. Nichols. (2009). Evidence for directional selection at a novel major histocompatibility class I marker in wild common frogs (*Rana temporaria*) exposed to a viral pathogen (*Ranavirus*). *PLoS One* **4,** e4616

Trivers, R. (1972). Parental investment and sexual selection. Pages 136-179 In B. Campbell editor. Chicago: Aldine.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, & P. Shipley. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4,** 535-538

van Oosterhout, C., D. A. Joyce, S. M. Cummings, J. Blais, N. J. Barson, I. W. Ramnarine, R. S. Mohammed, N. Persad, & J. Cable. (2006). Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*). *Evolution* **60,** 2562-2574

Vasemagi, A., J. Nilsson, & C. R. Primmer. (2005). Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution* **22,** 1067-1076

Vitalis, R., & D. Couvet. (2001). Estimation of effective population size and migration rate from one- and two-locus identity measures. *Genetics* **157,** 911-925

Weber, D. S., B. S. Stewart, J. Schienman, & N. Lehman. (2004). Major histocompatibility complex variation at three class II loci in the northern elephant seal. *Molecular Ecology* **13,** 711-718

Weir, B. S., & C. Cockerham. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38,** 1358-1370

747

748

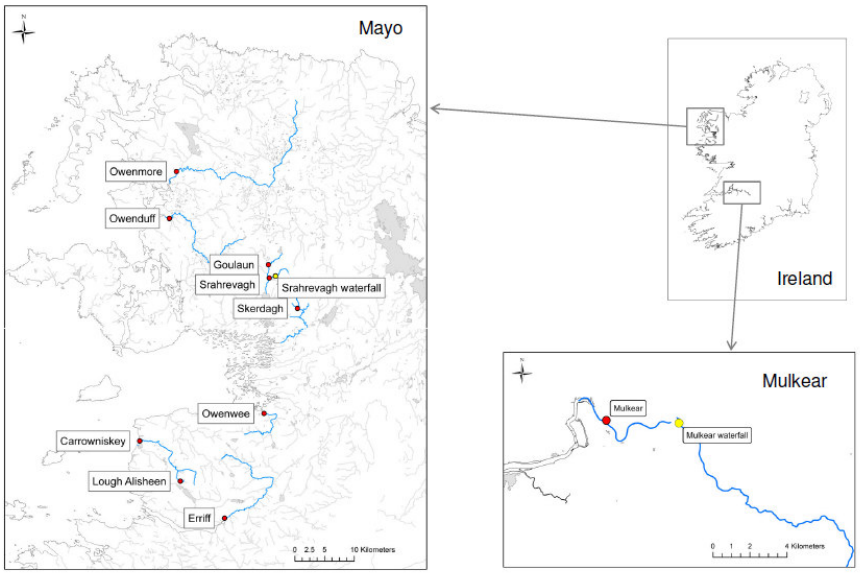
749 **Electronic References**

750

751

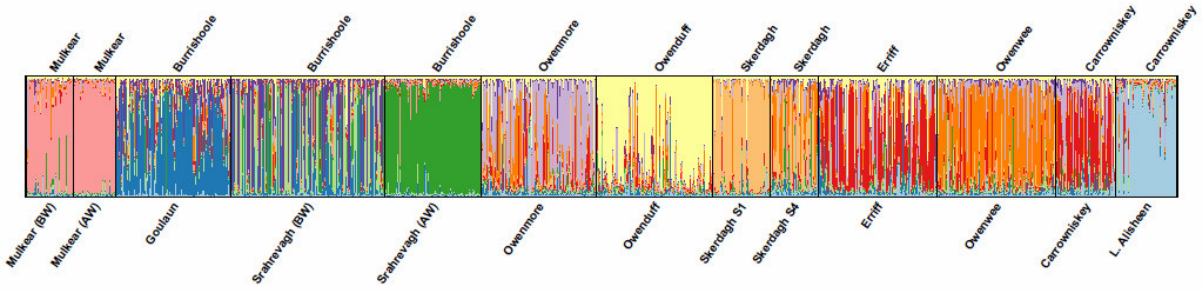
752 Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (2004) Genetix 4.04.
753 http://www.genetix.univ-montp2.fr/genetix/intro.htm#abstract. (last accessed 27 May
754 2012).

755



756

Figure 1



757

Figure 2

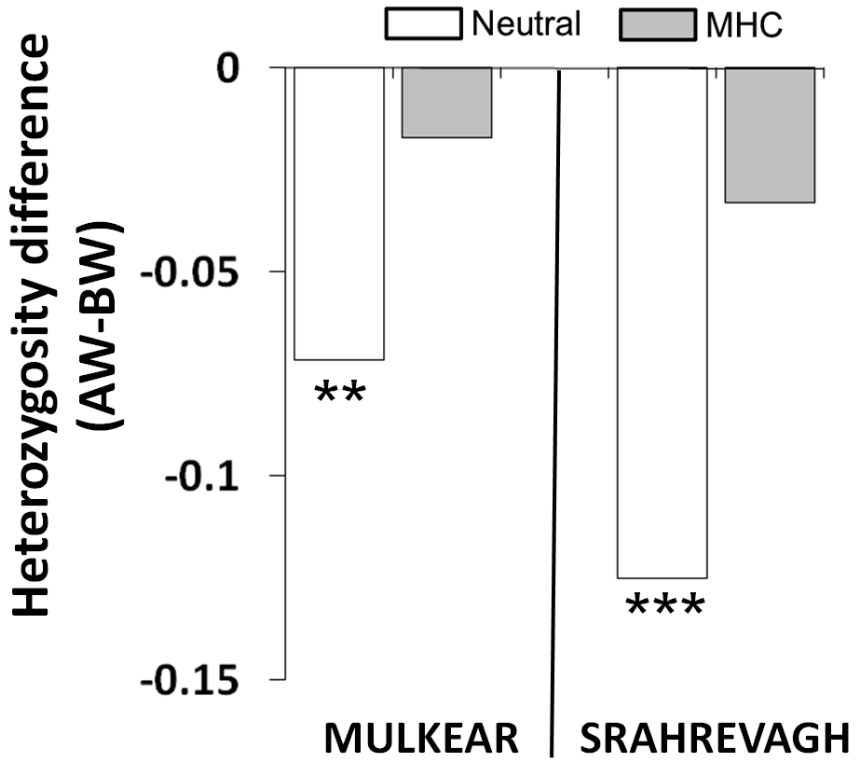


Figure 3

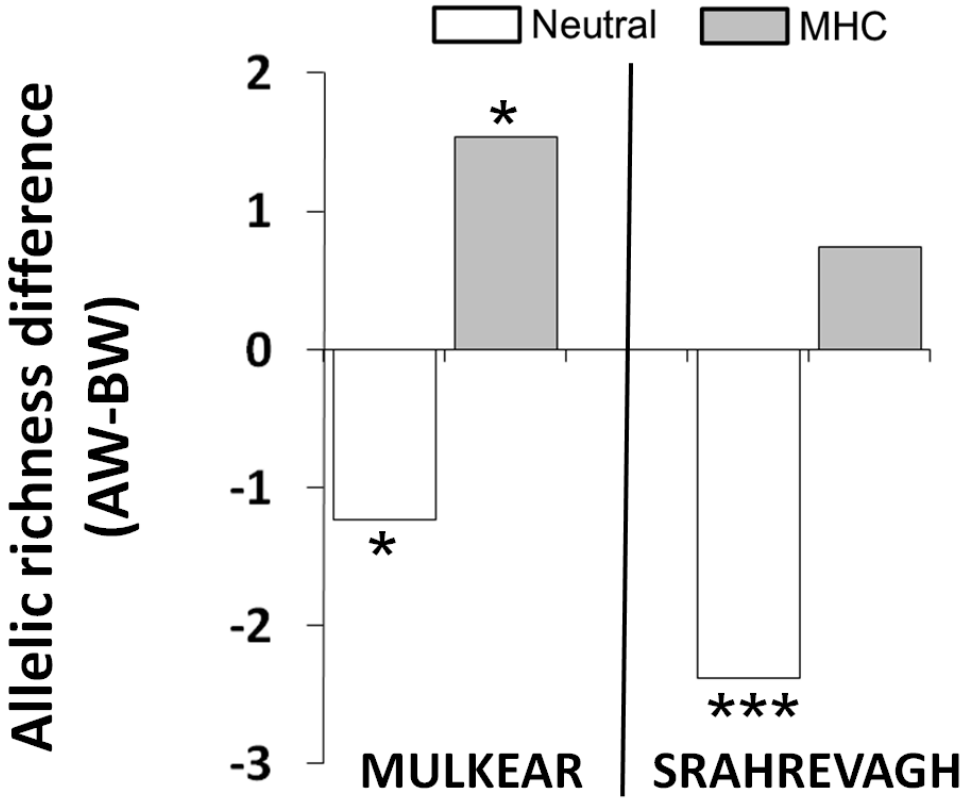


Figure 4

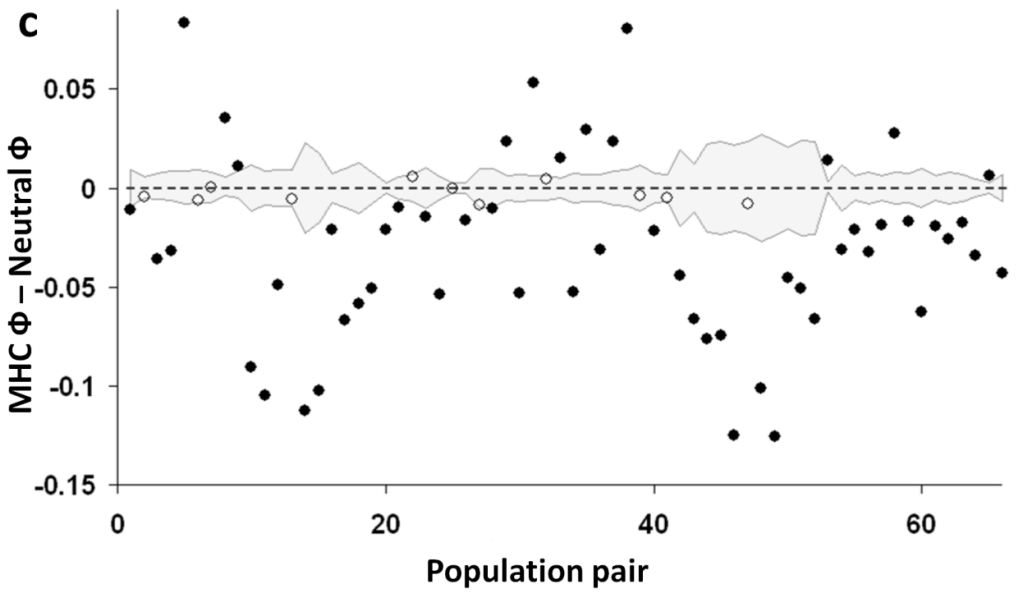


Figure 5

Captions

Figure 1

Map of Ireland showing the areas containing the eight river systems from which brown trout samples in this paper were taken. The Burrishoole system contains the Srahrevagh and Goulaun tributaries shown while the Newport system contains the Skerdagh tributary shown.

Figure 2

STRUCTURE analysis (*K* = 11) of brown trout sampled. The number of clusters was arrived at following the approach of Evanno *et al.* (2005) but then followed by longer runs (Burn-in 100,000 with subsequent 1,000,000 MCMC iterations) over the range K = 9 to K = 14. Trout population names are included at the bottom of the figure with the river

774 catchment from which they are from on the top of the figure. The Mulkear BW and Mulkear
775 AW are presented separately despite lack of significant population differentiation. The
776 populations isolated above waterfalls (Mulkear AW and Srahrevagh AW) demonstrate
777 good assignment of all individuals to a particular cluster. The cryptic population in the
778 Carrowniskey system, now referred to as the Lough Alisheen population, has strong
779 assignment to a unique cluster. This is interesting given there are no physical barriers
780 between this cluster and the downstream population which we term Carrowniskey. Indeed,
781 there is more evidence of admixture between the Carrowniskey and the Erriff population to
782 the South (see Figure 1) than between the Carrowniskey and Lough Alisheen.
783
784
785 **Figure 3** The difference in heterozygosity above waterfalls (AW) and below waterfalls
786 (BW) is presented for neutral loci (median values across individuals) and the MHC-linked
787 locus (proportion of heterozygotes at the locus). Significance levels ($P < 0.001$,***; $P <$
788 0.01,**; $P < 0.05$,*) are also indicated. Heterozygosity is significantly lower at neutral loci
789 above waterfalls than below waterfalls (BW) but not at MHC in both the Srahrevagh and
790 the Mulkear.
791
792 **Figure 4** The difference in allelic richness above waterfalls (AW) and below waterfalls
793 (BW) is presented for neutral loci and the MHC-linked locus. Significance levels ($P <$
794 0.001,***; $P < 0.01$,**; $P < 0.05$,*) are also indicated. Allelic richness above waterfalls (AW)
795 is significantly lower at neutral loci than below waterfalls (BW) but not at MHC in the
796 Srahrevagh and the Mulkear. Allelic richness is actually somewhat higher above waterfalls
797 in both case studies and significantly so in the case of the Mulkear.
798 **Figure 5** Difference between MHC and neutral loci for all population pairs (•=significant,
799 o=not-significant): those which show a significantly lower level of differentiation (φ) at MHC

than at neutral loci are those below the zero line and outside the 99.9% confidence intervals for the neutral loci (grey), those showing higher levels of differentiation at MHC are above the line.

Table I Sample sizes & descriptive data. Note the Mulkear (BW) and Mulkear (AW) are presented separately as the values are relevant to the waterfall case studies and the above waterfall population is implicitly reproductively isolated. It's interesting to note that although neutral allelic richness is lower in the Srahrevagh (AW) sample than all open populations bar the Mulkear (BW) sample, MHC allelic richness is higher than that found in both open populations in the Burrishoole system. The Skerdagh S1 population also shows some signs of reduced allelic richness at both neutral and MHC loci whereas, curiously, the Skerdagh S4 has reduced neutral allelic richness but much higher allelic richness at MHC.

Population	Mulkear BW	Mulkear AW	Goulaun	Srahrevagh BW	Srahrevagh AW	Owenmore	Owenduff	Skerdagh S1	Skerdagh S4	Erriff	Owenwee	Carrowniskey	Lough Alisheen
Sample size	39	45	97	130	74	99	98	48	41	100	100	51	51
River System	Mulkear	Mulkear	Burrishoole	Burrishoole	Burrishoole	Owenmore	Owenduff	Newport	Newport	Erriff	Owenwee	Carrowniskey	Carrowniskey
STRUCTURE Assignment (K=11)	5 (78.9%)	5 (90.5%)	2 (56.0%)	3/10 (29.2%/29.5%)	4 (86.1%)	9 (45.5%)	11 (70.5%)	7 (77.2%)	7 (33.3%)	6 (44.0%)	8 (56.3%)	6 (44.7%)	1 (77.0%)
Allelic Richness Neutral allelic	4.699	3.471	7.513	7.589	5.207	7.746	6.822	5.118	6.073	7.138	6.801	8.0606	6.656

richness (bootstrap)													
MHC allelic richness (bootstrap)	5.000	6.538	7.867	7.841	8.584	11.5551	11.259	5.906	8.799	8.076	9.302	8.988	8.32
Neutral heterozygosity													
Neutral individual hete	0.571	0.500	0.750	0.750	0.625	0.750	0.750	0.625	0.625	0.750	0.625	0.75	0.625

rozygosity (median) MHC (proportion of heterozygotes)	0.795	0.778	0.794	0.885	0.852	0.707	0.776	0.625	0.707	0.750	0.910	0.725	0.902

Table I Sample sizes & descriptive data. Note the Mulkear (BW) and Mulkear (AW) are presented separately as the values are relevant to the waterfall case studies and the above waterfall population is implicitly reproductively isolated. It's interesting to note that although neutral allelic richness is lower in the Srahrevagh (AW) sample than all open populations bar the Mulkear (BW) sample, MHC allelic richness is higher than that found in both open populations in the Burrishoole system. The Skerdagh S1 population also shows some signs of reduced allelic richness at both neutral and MHC loci whereas, curiously, the Skerdagh S4 has reduced neutral allelic richness but much higher allelic richness at MHC.

Population	Mulkear BW	Mulkear AW	Goulaun	Srahrevagh BW	Srahrevagh AW	Owenmore	Owenduff	Skerdagh S1	Skerdagh S4	Erriff	Owenwee	Carrowniskey	Lough Alisheen
Sample size	39	45	97	130	74	99	98	48	41	100	100	51	51
River System	Mulkear	Mulkear	Burrishoole	Burrishoole	Burrishoole	Owenmore	Owenduff	Newport	Newport	Erriff	Owenwee	Carrowniskey	Carrowniskey
STRUCTURE Assignment (K=11)	5 (78.9%)	5 (90.5%)	2 (56.0%)	3/10 (29.2%/29.5%)	4 (86.1%)	9 (45.5%)	11 (70.5%)	7 (77.2%)	7 (33.3%)	6 (44.0%)	8 (56.3%)	6 (44.7%)	1 (77.0%)
Allelic Richness Neutral allelic	4.699	3.471	7.513	7.589	5.207	7.746	6.822	5.118	6.073	7.138	6.801	8.0606	6.656

richness (bootstrap)													
MHC allelic richness (bootstrap)	5.000	6.538	7.867	7.841	8.584	11.5551	11.259	5.906	8.799	8.076	9.302	8.988	8.32
Neutral heterozygosity													
Neutral individual hete	0.571	0.500	0.750	0.750	0.625	0.750	0.750	0.625	0.625	0.750	0.625	0.75	0.625

rozygosity (median) MHC (proportion of heterozygotes)	0.795	0.778	0.794	0.885	0.852	0.707	0.776	0.625	0.707	0.750	0.910	0.725	0.902

825

826

827