Journal of Fish Biology

Atlantic salmon Salmo salar in the chalk streams of England are genetically unique

--Manuscript Draft--

Manuscript Number:	SP 17-009R2
Full Title:	Atlantic salmon Salmo salar in the chalk streams of England are genetically unique
Short Title:	Chalk stream salmon are genetically unique
Article Type:	FSBI Symposium SI Regular Paper
Keywords:	Atlantic salmon; chalk streams; microsatellite; population structure; Salmo salar
Corresponding Author:	Jamie Stevens, PhD University of Exeter Exeter, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Exeter
Corresponding Author's Secondary Institution:	
First Author:	Charles I Ikediashi, PhD
First Author Secondary Information:	
Order of Authors:	Charles I Ikediashi, PhD
	Josephine R Paris, PhD
	Robert Andrew King, PhD
	William R.C. Beaumont
	Anton Ibbotson, PhD
	Jamie Stevens, PhD
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Recent research has identified genetic groups of Atlantic salmon Salmo salar that show association with geological and environmental boundaries. This study focuses on one particular subgroup of the species inhabiting the chalk streams of southern England. These fish are genetically distinct from other British and European Atlantic salmon populations and have previously demonstrated markedly low admixture with salmon in neighbouring regions. The genetic population structure of salmon occupying five chalk streams was explored using 16 microsatellite loci. The analysis provides evidence of the genetic distinctiveness of chalk stream salmon in southern England, in comparison to salmon from non-chalk regions of Western Europe. Little genetic differentiation exists between the chalk populations, and a pattern isolation-by-distance (IBD) was evidenced. Furthermore, evidence of temporal stability of salmon populations across the five chalk streams was found. This work provides new insights into the temporal stability and lack of genetic population sub-structuring within a unique component of the species' range.

Ethics Questionnaire for JFB

Submitted manuscripts will only be considered if the experimental methods employed are ethically justified. Please answer all questions. If you have answered 'yes' to questions 4 to 7, you should include an *Ethics* paragraph in the Methods section of your manuscript which justifies your methods used. You should complete this questionnaire based on all fishes used in your experiment. For example, if you used live fishes as prey in predation experiments, this is a lethal endpoint for the prey fish (see Questions 5 & 6). Please read the Editorial published in JFB **68**, 1-2, for full information on JFB ethics. PLEASE SUBMIT THE COMPLETED QUESTIONNAIRE WITH YOUR MANUSCRIPT ONLINE THROUGH EDITORIAL MANAGER.

Corresponding author's name:

Question 1: Were fishes collected as part of faunal surveys? YES/NO
If 'Yes', have the fishes, where feasible, been killed rapidly or returned to the wild after being held in aquaria and have procedures complied with local and or national animal welfare laws, guidelines and policies? If Yes, state these and provide suitable evidence (e.g. for the U.K. a Home Office PPL number is sufficient) that protocols have undergone an ethical review process by an institutional animal care and use (or similar) committee, a local ethics committee

Question 2: If you have undertaken experimental work, has the care and use of experimental animals complied with local and or national animal welfare laws, guidelines and policies? YES/NO

If 'Yes', state these and provide suitable evidence (e.g. for the U.K. a Home Office PPL number is sufficient), both here and in the manuscript, that protocols have undergone an ethical review process by an institutional animal care and use (or similar) committee, a local ethics committee, or by appropriately qualified scientific and lay colleagues.

If ‘No’, because these laws do not exist in your country, please state this. Alternatively, if you carried out purely observational work so ethical permission was not considered necessary please state this both here and in the manuscript.

Question 3: Were fishes killed during or at the end of your experiment (e.g. for tissue sampling)? YES/NO

If ‘Yes’, what method was used? Please provide details both here and in the manuscript.

Question 4: Have you performed surgical procedures? YES/NO

If ‘Yes’, please give brief details of the surgery here. Full details should be given in the manuscript. If the procedures caused more than slight pain or distress, did you use appropriate sedation, analgesia and anaesthesia, with post-operative care? Please provide full details and justification both here and within the manuscript including type and concentration of anaesthetic.

Question 5: Did you use experimental conditions that severely distressed any fishes involved in your experiments? YES/NO

If 'Yes', state the conditions and how they can be justified. What humane endpoints were used to minimise the effects? Please provide full justification within the methods section of your manuscript.

Question 6: Did any of the experimental procedures, particularly those that involve lethal endpoints (e.g. predation studies, toxicity testing), cause lasting harm to sentient fishes? YES/NO

If 'Yes', provide details both here and in the methods section of your manuscript. Normally these procedures will be considered unacceptable by JFB unless any harm caused can be justified against the benefits gained.

Question 7: Did any of your procedures involve sentient, un-anaesthetised animals paralysed by chemical agents such as muscle relaxants? YES/NO

If 'Yes', provide details both here and in the methods section of your manuscript. Normally these procedures will be considered unacceptable by *JFB*.

Atlantic salmon *Salmo salar* in the chalk streams of England are genetically unique

Charles Ikediashi[1†], Josephine R. Paris[1†], R. Andrew King[1], William R. C. Beaumont[2], Anton Ibbotson[2] and Jamie R. Stevens[1*]

[1]*Department of Biosciences, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter, Devon, EX4 4QD, UK*

[2]*Salmon & Trout Research Centre, Game & Wildlife Conservation Trust (GWCT), East Stoke, Wareham, Dorset BH20 6BB*

Tel: 01392 723775; Fax: 01392 723700

*Corresponding author: J. R. Stevens, email: j.r.stevens@ex.ac.uk

†Joint first authors

Abstract

Recent research has identified genetic groups of Atlantic salmon *Salmo salar* that show association with geological and environmental boundaries. This study focuses on one particular subgroup of the species inhabiting the chalk streams of southern England. These fish are genetically distinct from other British and European *S. salar* populations and have previously demonstrated markedly low admixture with populations in neighbouring regions. The genetic population structure of *S. salar* occupying five chalk streams was explored using 16 microsatellite loci. The analysis provides evidence of the genetic distinctiveness of chalk stream *S. salar* in southern England, in comparison to populations from non-chalk regions of Western Europe. Little genetic differentiation exists between the chalk stream populations, and a pattern of isolation-by-distance (IBD) was evident. Furthermore, evidence of temporal stability of *S. salar* populations across the five chalk streams was found. This work provides new insights into the temporal stability and lack of genetic population sub-structuring within a unique component of the species' range of *S. salar*.

Key words: Atlantic salmon, chalk streams, microsatellite, population structure, *Salmo salar*

Introduction

Atlantic salmon *Salmo salar* L. 1758 is an anadromous fish species, which returns to its natal river after reaching sexual maturity. As a result, the species shows marked population structuring into broad geographic groups, which is readily detectable using genetic methodologies (Stahl, 1987; Verspoor *et al.*, 2005), particularly through analysis of microsatellite markers (e.g. King *et al.*, 2001; Koljonen *et al.*, 2005; Tonteri *et al.*, 2009; Griffiths *et al.*, 2010). Current research suggests that broad genetic groups are largely defined by a combination of geological substrate (Grandjean *et al.*, 2009; Perrier *et al.*, 2011), phylogeography (Finnegan *et al.*, 2013) and environmental factors (Dillane *et al.*, 2007), leading to the suggestion that *S. salar* populations may be locally adapted to their in-river environments (Garcia de Leaniz *et al.*, 2007; Fraser *et al.*, 2011; Perrier *et al.*, 2011).

One sub-group of this species, which resides within the chalk streams of southern England, has been shown to form a genetically distinct unit when compared with groups of geographical neighbour populations in non-chalk rivers (Griffiths *et al.*, 2010; Gilbey *et al.*, 2017). Chalk stream *S. salar* populations also appear to display relatively low levels of admixture with populations in neighbouring regions (Ikediashi *et al.*, 2012). Admixture has, for some time, been associated with a reduction in population differentiation. For example, Stahl (1987) deduced that, in order to maintain genetic differences between two or more *S. salar* sub-populations of 2,500 to 10,000 fish, there had to be fewer than one migrant per year between them. More recent studies in Spain (Ayllon *et al.*, 2006a) and in the Baltic Sea (Vasemägi *et al.*, 2005), showed that reductions in the between-river population structuring of *S. salar* has been identified as a result of admixture with farm escapees. Following this line of argument, it is possible that chalk stream *S. salar*, which engage in relatively little admixture with individuals in neighbouring regions (Ikediashi *et al.*, 2012), may also show reduced genetic structuring between them. However, despite several previous studies having included some fish of chalk stream origin (e.g. Child *et al.*, 1976; Jordan *et al.*, 2005; Finnegan *et al.*, 2013) and their apparent genetic distinctiveness (Griffiths *et al.*, 2010), the

degree of population structure within and between chalk stream *S. salar* has not yet been the subject of detailed exploration.

The reason for the distinction between these *S. salar* populations most likely stems from one or more of the unique abiotic factors of chalk stream geology, which are described in detail by Berrie (1992). The calcareous substrate, upon which chalk streams are formed, is porous, and thus chalk streams are aquifer fed. The water is therefore relatively clear, stable in temperature throughout most of the year, and alkaline (ca. pH 8). Due to these unique environments, several chalk streams have been designated SSSIs (Sites of Special Scientific Interest) in the UK. However, of the 161 rivers classified as chalk streams in England (Environment Agency 2004), major *S. salar* populations are present in just five of these. These rivers include the Frome, Piddle, Avon, Test and Itchen, all of which have each been sampled for the purpose of this study (Figure 1). Crucially, although chalk streams are located between the counties of Yorkshire in north-east England and Dorset in southern England, the five rivers with substantial *S. salar* populations span only some 70 km along the southern English coast. With so few chalk stream populations, each of which has markedly decreased in numbers in recent decades (Environment Agency 2004), there is additional incentive to understand the full extent of chalk stream *S. salar* local population genetic structure.

The primary aim of this study was to assess the population structure of *S. salar* populations from the above five major chalk streams of southern England. First, we assessed the distinctiveness of the chalk stream populations by explicitly comparing them to populations from geographically neighbouring populations residing in non-chalk geologies. Secondly, we assessed whether significant genetic variation exists among the chalk stream populations by exploring population structure, genetic diversity and patterns of isolation-by-distance (IBD). Thirdly, by analysing temporal cohorts, we explored the temporal stability of chalk stream genetic variability over time. In summary, this study represents the first assessment explicitly addressing the distinctiveness of chalk stream *S. salar* populations in southern England, and highlights the importance of managing these unique populations as distinct

genetic entities. We anticipate that this information will be useful for the successful management and conservation of this species within these rivers.

Materials and Methods

Sampling

Juvenile *S. salar* (0+ parr) were sampled from the five chalk streams of southern England that still contain significant populations: the Frome, Piddle, Avon, Test and Itchen (Figure 1; Table 1; Supporting Information Table S1). The Avon, Itchen and Test were sampled by the Environment Agency (EA) during routine national surveys and management programmes between 2004 and 2012. Sampling on the Frome and Piddle was carried out by the Game & Wildlife Conservation Trust (GWCT) in September of 2009 and 2011 (Figure 1) during routine juvenile abundance surveys. Fish were collected by electrofishing; adipose fin clips were then taken and preserved in 100% ethanol, according to national agency ethical guidelines. To avoid issues of small sample sizes we aimed to collect 50 parr samples from each site. For assessment of the chalk stream populations in comparison with those from neighbouring non-chalk geographical regions, salmon from rivers in north-west (NW) France, south-west (SW) England and Norway were included for the population structure analyses (Table 2) and were obtained from a database of salmon genotyped for the SALSEA-Merge project (Ellis *et al.*, 2011, Gilbey *et al.*, 2017).

DNA extraction and microsatellite amplification

DNA was extracted from fin clips using the HOTshot method (Truett *et al.*, 2000) and from scales using the Chelex method (Estoup *et al.*, 1996). Sixteen microsatellite loci were genotyped. Fourteen loci were amplified according to the protocol of Ikediashi *et al.* (2012): Ssa14 (McConnell *et al.*, 1995); Ssa202, SSsp3016, Ssa197 (O'Reilly *et al.*, 1996); SsaF43 (Sánchez *et al.*, 1996); SSspG7, SSsp1605, SSsp2210, SSsp2201, and SSsp2216 (Paterson *et al.*, 2004); Ssa171, Ssa289, Ssa157, and SsaD144 (King *et al.*, 2005). Two

additional loci, Ssosl85 and Ssosl417 (Slettan *et al.*, 1995), were amplified in the first multiplex reaction described by Ikediashi *et al.*, (2012). Potential *S. salar* x brown trout *Salmo trutta* L. 1758 hybrids were recognised by the presence of alleles longer than 350bp for locus SSsp1605, or alleles longer than 135bp for Ssa289 (Finnegan & Stevens, 2008). Hybrid fish were removed from the dataset.

PCR reactions were carried out in 10µl reactions containing ~50ng of extracted *S. salar* template DNA, 3µl water, 5µl of Qiagen Taq PCR Mastermix and 1µl of primer mixture (Supporting Information Table S2). PCR conditions were as follows: an initial denaturation step of 5 min at 95°C, followed by a touchdown PCR consisting of eight cycles with a 30 s denaturation step at 95 °C, a 90 s annealing step starting at 62 °C and decreasing the temperature 2°C every two steps until a touchdown temperature of 47°C was reached, with 3 minutes of extension at 72°C, followed by a final 10 minute extension at 72°C. Size of products of fluorescently labelled PCR products were assessed using a Beckman-Coulter CEQ8000 automatic DNA sequencer and the associated fragment analysis software (Beckman Coulter).

Data checking

MICROCHECKER v2.2.3 (Van Oosterhout *et al.*, 2004) was used to check for scoring errors due to stutter peaks, large allele dropout or null alleles. In order to prevent the false detection of population structure due to the presence of family groups (Allendorf & Phelps, 1981), the program COLONY v2.0.4.1 (Jones & Wang, 2010) was used to identify full siblings. The mating system was defined as polygamous for males and females and without inbreeding. Each run was of medium length, using high precision and the full-likelihood method. Allele frequencies were not updated during the run and no prior sib-ship was assumed. An error rate of 0.02 was used for each locus based on the protocol of Ellis *et al.* (2011). The program was run twice independently, with different starting seeds to check consistency of sibship reconstruction.

Full-sib families were reduced to one representative, if supported by an average likelihood of 0.5 or higher between the two runs.

Linkage disequilibrium and deviations from Hardy Weinberg equilibrium were assessed using GENEPOP v4.2 (Raymond & Rousset, 1995). The 95% significance level for corrections of multiples tests for both procedures were adjusted using the False Discovery Rate (FDR) (Benjamini & Hochberg, 1995).

Descriptive statistics

The number of alleles (N_A), expected heterozygosity (H_E) and observed heterozygosity (H_O) were calculated in Genalex v6.5.02 (Peakall & Smouse, 2012) for each year cohort of *S. salar* from each of the five chalk streams. Allelic richness (A_R) and the heterozygote deficit (F_{IS}) were calculated using the program FSTAT v2.9.3 (Goudet, 1995). In order to determine whether there were any significant differences in genetic diversity between the five chalk streams, two sided permutation tests were performed within FSTAT for A_R, H_O, H_E and F_{IS}, using 1000 permutations of the dataset. Where significant differences were identified, further pairwise comparisons were made in order to determine between which groups the significant differences lay.

The effective population size (N_E) for each river and year was assessed in the program NeEstimator v. 2.01 (Do *et al.*, 2014) using the linkage disequilibrium model under a random mating scenario, using 0.01 as the lowest allele frequency as the critical value cut-off.

Assessment of the genetic uniqueness of chalk stream S. salar

In order to place the chalk stream *S. salar* populations in a wider geographical context, we incorporated genotypes from four other geographical regions: NW France, SW England and Norway (Table 2). Genotypes were obtained from the SALSEA-MERGE dataset (Ellis *et al.*, 2011; Gilbey *et al.*, 2017). To facilitate accurate comparisons across these populations and to allow the

incorporation of previously genotyped loci, two markers, Ssosl417 and Ssosl85, were excluded for these population analyses, resulting in the use of a reduced set of 14 microsatellite loci for all population structure analyses. Two complementary methods were used to assess the population structure between chalk and non-chalk populations and also for the assessment of structure within the chalk stream populations.

Firstly, the program STRUCTURE v2.3.4 (Pritchard *et al.*, 2000) was used to identify the number of distinct genetic units (*k*) across the four geographic regions. STRUCTURE was run from $k = 1$ to $k = 10$ with 150,000 Markov Chain Monte Carlo (MCMC) replicates, after a burn-in of 75,000 replicates from ten independent starting points. The Evanno method (Δk: Evanno *et al.*, 2005) was used to determine the optimum number of genetic units (*k*) from the results. A hierarchical STRUCTURE analysis was conducted based on the most likely number of genetic units (see Results) in order to further assess population sub-divisions and the possible existence of sub-structuring within the chalk stream rivers. In hierarchical analyses of population structure, the same analysis parameters were used as outlined above.

Secondly, an assessment of population structure using a Discriminant Analysis of Principal Components (DAPC) was conducted in R using the adegenet package (Jombart, 2008; Jombart *et al.*, 2010). The optimum alpha score (using the optim.a.score function) was used to assess how many principal components should be retained for each analysis and we assessed structure using five discriminant components. DAPC plots of the first two principal components were derived using ggplot2 (Wickham, 2009).

Assessment of the population structure between chalk stream rivers

To assess the differentiation between sampling sites from across each chalk river, pairwise F_{ST} values from each site and year were calculated. Based on the outcome of this analysis (see Results), fish from individual sample sites were grouped together across each river; annual cohorts from each river were then used for all subsequent analyses of population structure and genetic diversity. Global and pairwise F_{ST} values were calculated for each year cohort

from each of the five chalk stream rivers. All values were tested for significance using 10 000 permutations in MSA (Dieringer & Schlötterer, 2003). Multiple testing correction, as incorporated within MSA was used to assess the 95% confidence level.

Population structure assessment of *S. salar* within each river and across time was assessed using the same methods above (STRUCTURE and DAPC analyses).

To test whether the populations from each of the five chalk stream rivers were structured through a pattern of isolation-by-distance, the genetic distance (F_{ST}/1-F_{ST}) (Rousset, 1997) was tested for significant correlations with geographic distance using a Mantel test in Genalex using 9999 permutations. Geographic distances (in km) were determined between river mouths along the coastal line of southern England using arcGIS v10 (ESRI, 2006).

In order to assess temporal stability, we calculated 'isolation-by-time' using a Mantel test for which a matrix of the difference in years between sampling was correlated with genetic distance (F_{ST}/1-F_{ST}). To further assess temporal stability, the genetic differentiation between sampling year and river was apportioned using an Analysis of Molecular Variance (AMOVA) in Arlequin v 3.5.2 (Excoffier & Lischer, 2010), using standard computations based on the number of different alleles (F_{ST}-like). Significance between the variance components (Va, Vb and Vc) and fixation indices (F_{CT}, F_{SC} and F_{ST}) were accepted at $p < 0.05$.

Results

Number of individuals and grouping of sites over years

In total, 1297 juvenile *S. salar* samples were genotyped at 16 microsatellites across 26 sites in the five chalk stream rivers (Supporting Information Table S1). Two potential *S. salar* x *S. trutta* hybrids were detected within the Frome and five were detected within the Avon. After the removal of hybrids and full

siblings, the final dataset was reduced to 772 samples (Table 1), which were used for all downstream analyses.

After applying the false discovery rate (FDR) correction, linkage disequilibrium was detected at seven out of a total of 3000 comparisons (data not shown). These indicated no consistent pattern between sample sites and, therefore, no loci were removed. Across the 26 sample sites, only two cases of loci not confirming to Hardy-Weinberg equilibrium were found, and therefore, no samples were removed.

Pairwise F_{ST} values between annual sample sites across each of the chalk stream populations were very low (average 0.025), ranging from -0.002 (between AVNbug04 and AVNbri12) to 0.063 (between FROcfmr09 and TESTmem10), and were significant in 278 of the 325 comparisons after multiple comparison corrections (Supporting Information Table S3). Moreover, despite significant F_{ST} values between many of the comparisons, a genetic signal of site differentiation could not be determined over the background of temporal variation in sampling. Furthermore, F_{ST} values for point samples such as these, particularly when sample sizes are small (Supporting Information Table S1), do not provide strong evidence for population differentiation. Accordingly, it was decided to group together sampling sites, irrespective of sampling year, for each river.

Genetic diversity of chalk stream S. salar

Between the year cohorts for each river, the number of alleles (N_A) genotyped in the juvenile *S. salar* from the chalk streams ranged from 6.38 (Piddle 2011) to 10.69 (Frome 2009) and the unbiased measure of allelic richness (A_R) ranged from 4.77 (Test 2010) to 5.47 (Test 2004) (Table 3). Expected heterozygosity (H_E) ranged from 0.66 (Test 2010) to 0.71 (Frome 2011), and observed heterozygosity (H_O) ranged from 0.67 (Itchen 2006 and Itchen 2010) to 0.73 (Piddle 2009 and Test 2004).

Statistical comparisons of diversity were non-significant for A_R ($p = 0.64$) and H_E ($p = 0.46$). However, there were significant differences in H_O ($p = 0.01$) and

294 F_{IS} (p = 0.008). Further analysis indicated that the differences in H_O were
295 between *S. salar* in the Piddle and Itchen (p = 0.001), Avon and Itchen (p =
296 0.039) and Test and Itchen (p = 0.035). These differences were reflected in
297 the statistical significance for F_{IS} between the Piddle and Itchen (p = 0.002)
298 and the Test and Itchen (p = 0.008). The significance of these results is due to
299 the relatively low H_O seen in the Itchen (especially in Itchen 2006 and Itchen
300 2010 cohorts), a pattern that is also reflected by higher values of F_{IS} for the
301 Itchen (Table 3). The differences in F_{IS} suggest a greater amount of
302 inbreeding within the Itchen, although this does not correlate with estimates of
303 effective population size.

304 The Test 2010 showed evidence of the smallest effective population size (N_E)
305 of 22 (95% CI: 18-27) and the highest N_E was observed in the Frome 2009 at
306 315 (95% CI: 249-419). Generally, estimates of N_E appeared stable over time,
307 with the Frome showing the highest N_E, followed by the Avon. The Piddle,
308 Test and Itchen showed relatively smaller values of N_E, with the exception of
309 increases in N_E from the Test 2004, relative to the Test 2010, as well as a
310 slight increases in Itchen 2010 N_E compared to Itchen 2005 and Itchen 2006.

311

312 *Genetic uniqueness of chalk stream S. salar*

313 Using the Δ*k* statistic, the most likely number of genetic clusters ascertained
314 from the STRUCTURE analysis was *k=2*, which illustrated the genetic
315 uniqueness of the five chalk steam *S. salar* populations compared to
316 neighbouring non-chalk populations (Figure 2). Of interest was the genetic
317 similarity of individuals from geographically distant regions on non-chalk
318 geology, compared with the striking distinctiveness of the chalk stream
319 salmon populations. Hierarchical analysis of the NW France, SW England and
320 Norway group showed that the most likely number of genetic clusters was
321 *k=2*, which demonstrated a difference between Norway and the two other
322 non-chalk populations residing in NW France and SW England (Supporting
323 Information Figure S1).

324 The optimum number of PCAs for the DAPC analysis was 34. Results of the
325 population differentiation from the DAPC analysis complimented the

STRUCTURE analysis in also showing the genetic uniqueness of the chalk stream *S. salar* in comparison to all other non-chalk salmon included in this study. Due to the ability of DAPC in uncovering finer-scale hierarchical population structure (Jombart *et al.*, 2010), the Norwegian *S. salar* are observed as a separate genetic unit in the DAPC plot, which was also confirmed in the hierarchical analysis using STRUCTURE. One chalk stream individual from the Frome09 sampling cohort was shown to cluster with the NW France / SW England genetic group. This sample had no missing genotype data so this 'outlier' is most likely a real result (see Discussion).

Lack of population structure and temporal stability within the chalk streams

The global F_{ST} calculated across each annual cohort from each chalk stream population was low but significant (F_{ST} = 0.018, p = 0.001). The average pairwise F_{ST} across all comparisons was 0.028, and ranged from 0.002 (between Frome09 and Frome11) and 0.055 (between Piddle11 and Test04) (Table 4). All pairwise F_{ST} comparisons were significant after FDR correction (except between Piddle09 and Piddle11; Itchen06 and Piddle09; Avon04 and Avon12; Itchen05 and Itchen06; Itchen05 and Itchen10; Itchen06 and Itchen10).

Hierarchical analysis of the chalk stream *S. salar* showed that no significant genetic differentiation occurred across or between the five chalk stream rivers. The Δ*k* statistic showed no single reliable estimate for *k*, as the Δ*k* values were both low, and did not show an obvious peak for any value of potential genetic clusters (Supporting Information Figure S2). This therefore suggests that the chalk stream *S. salar* represent one large genetic group that is not distinguished on the basis of river basin or annual sampling (Figure 3). This was further supported by the DAPC, which could not distinguish any patterns of population differentiation (based on the optimum number of 42 PCAs).

The test for isolation-by-distance (IBD) across the chalk stream salmon was strong and statistically significant (R^2 = 0.2978, p = 0.031) (Figure 4A). This pattern of IBD was also noticeable in the STRUCTURE plot and DAPC.

Assessment of temporal stability using 'isolation-by-time' (IBT) showed no statistically significant relationship between annual cohorts within each river and geographical distances (R^2 = 0.0013, p = 0.422) (Figure 4B). Results from the AMOVA proportioned the majority of the variance (98%) within each sampling cohort (Vc = 5.22, F_{ST} = 0.0197, p < 0.05). Only 0.94% of the genetic variance occurred between rivers (Va = 0.05, F_{CT} = 0.00938, p < 0.05) and just 1.04% of the variance was attributed to between years within rivers (Vb = 0.06, F_{SC} = 0.01046, p < 0.05).

Discussion

Overview

Populations of *S. salar* within the chalk streams of southern England have plummeted in recent decades, yet despite this, and their distinction from other European populations, the genetic population structure of *S. salar* within the chalk streams had not previously been investigated. This study explicitly demonstrated the uniqueness of chalk stream populations in the context of *S. salar* from other non-chalk regions. A significant pattern of isolation-by-distance defines the chalk stream populations, and there is little to no genetic sub-structuring across rivers and across years. Furthermore, patterns in population structure and genetic diversity were shown to be temporally stable. Identification of the homogeneity of the chalk stream fish significantly increases our understanding of the contemporary genetic structure within one of the key reporting regions identified by Griffiths *et al.* (2010) for *S. salar* in the southern part of the species' range. These finding have significant implications for conservation and our understanding of population structure.

Uniqueness of chalk stream S. salar populations

The population structure analyses complement previous findings (Griffiths *et al.*, 2010; Ikediashi *et al.*, 2012) confirming that chalk stream *S. salar* are genetically distinct compared to populations from non-chalk geologies. Extensive analysis of *S. salar* populations from across Europe also confirms

the genetic uniqueness of chalk stream populations in southern England (Gilbey *et al.*, 2017). Interestingly, chalk stream populations appear to be genetically distinct even when compared to populations occupying south-west English rivers, between which a sharp gradient in underlying geology, from chalk to non-chalk, occurs. This is emphasised further by the relative genetic homogeneity of salmon from south-west England and north-west France, which are separated across the English Channel (representing a direct distance of >370km). Notably, even fish from considerably more distant non-chalk *S. salar* populations (Norway) are more genetically similar to English non-chalk stream fish than are the chalk stream *S. salar*.

Geology is known to be a fundamental feature affecting the distribution and abundance of salmonid populations. For example, rainbow trout *Oncorhynchus mykiss* (Walbaum 1792) and cutthroat trout *Oncorhynchus clarki* (Richardson 1837) abundances have been shown to be correlated with particular geologies (Hicks & Hall, 2003), and *S. trutta* condition was shown to decrease in limestone geologies correlated with increased catchment afforestation (Lehane *et al.*, 2004). With chemical cues being a particularly important feature of salmonid homing (Stabell *et al.*, 1984; Tierney *et al.*, 2010), distinctive population structure arising from geologies with especially notable water chemistry features is not surprising. Other research directly investigating the role of geology in *S. salar* population structure across Europe suggests a similar role of geology in structuring local and regional populations (Perrier *et al*., 2011). Despite the increasing appreciation of geological factors on the structuring of salmonid populations, genetic distinctiveness related specifically to geology is not common in the literature.

Furthermore, there appears to be little to no genetic admixture occurring between chalk stream *S. salar* populations and fish from neighbouring rivers. The proportion of straying in salmonids is known to be a significant contributor in re-colonisation events (Vasemägi *et al.*, 2001; Perrier *et al.*, 2009; Griffiths *et al.*, 2011; Ikediashi *et al.*, 2012). Moreover, high rates of straying have been shown to result in patterns of admixture between and among local salmonid populations within a region (Filatre *et al.*, 2003; Ayllon *et al.*, 2006b; King *et al.*, 2016). On the other hand, the potential of stocked fish to swamp local *S.*

salar population structure is not frequently observed, with signals of low admixture between foreign and native genotypes (Finnegan *et al.*, 2008; Hansen *et al.*, 2009; Perrier *et al.*, 2013). The one exception to the apparent low rates of admixture in the chalk stream populations is the occurrence of a single chalk stream individual (genotyped from the Frome), which does not identify – based on its genetic profile – as 'chalk'. As the Frome is the most westerly of the chalk stream rivers, this fish could potentially represent a hybrid from a stray from south-west England crossed with a chalk stream individual. An alternative explanation is that the fish has been illicitly moved by human activity, although, if this were the case, in the short-term such activities might be expected to exhibit a more widespread exogenous signature.

Lack of genetic sub-structuring within the chalk stream S. salar populations

The accuracy of natal homing in salmonids is influenced by a plethora of biotic and abiotic factors (see Keefer & Caudill, 2014). In some cases, evidence of fine-scale natal homing appears high, for example in *S. trutta* populations across 3 km (Carlsson *et al.*, 1999) and in Chinook salmon *Oncorhynchus tshawytscha* (Walbaum 1792) across just 1 km (Neville *et al.*, 2006). On the other hand, Stewart *et al.* (2003) found that, despite phenotypic differences in sockeye salmon *Oncorhynchus nerka* (Walbaum 1792) populations homing to physically similar beaches in Alaska, USA, no evidence of restricted gene flow between the sites was detected. Similarly, genetic variation among *O. nerka* populations in the tributaries of a bay in Alaska were shown to be highly similar (Habicht *et al.* 2006), while relatively weak genetic structure was detected among Coho salmon *Oncorhynchus kisutch* (Walbaum 1792) from different river basins in Oregon (Johnson & Banks 2008).

In this study, the lack of genetic differentiation between chalk stream *S. salar* populations suggests that returning individuals may be homing back to a general chalk geological signature, and, consequently, fine-scale between-river population differentiation is not apparent. We anticipate that a propensity to home to chalk stream waters is likely a fundamental trait of these fish.

Collectively, the chalk stream rivers drain a small area (spanning just 70 km along the southern English coast), and it appears probable that homing accuracy of fish originating within the chalk geology is not further stratified by additional river-specific geochemical features. Furthermore, the chalk stream *S. salar* populations were shown to be temporally stable, which importantly, suggests habitat stability over time (see below).

A marked lack of differentiation across *S. salar* populations from proximal rivers has been noted previously in other parts of Britain. For example, populations in the rivers of north-west England and south-west Scotland that drain into the Solway Estuary (Griffiths *et al.*, 2010, Ikediashi *et al.*, 2012), show little if any consistent genetic differentiation, even when using a large panel of SNPs (Gilbey *et al.*, 2016). While geology may play a role in this scenario, this finding appears best explained by the fact that the rivers in this region share the estuary of the Solway Firth and the Irish Sea, through which returning fish must pass.

Despite a distinct lack of population differentiation between chalk stream *S. salar* populations, significant patterns of isolation-by-distance (IBD) were detected. Isolation-by-distance is prevalent in salmonids at both large continent scales (King *et al.*, 2001), regional scales (Taylor *et al.*, 2003) and within rivers (Griffiths *et al.,* 2009; Primmer *et al.*, 2006). Given the proximity of the river mouths and shared estuaries of the Frome/Piddle and Test/Itchen, higher levels of gene flow and migration between these sites might be expected, and it appears that the geographic distance between the mouths of these rivers does play a role in defining genetic distances between populations.

Temporal stability and chalk stream habitat reliability

The assessment of temporal stability is important in order to understand the extent to which populations exist as dynamic metapopulations punctuated by local extinctions and recolonisations, or in stable patches at gene flow-drift equilibrium. In an assessment of *S. salar* sampled across two consecutive

years across a ~100 km river in Quebec, temporal stability was evident in four out of seven sampling sites, with a high proportion of genetic variance attributable to other factors (Garant *et al.*, 2000). In a study assessing temporal stability over a much longer timeframe (50-100 years), across five rivers (ranging from 3 – 60 km) *S. trutta* populations were shown to be remarkably temporally stable (Hansen *et al.*, 2002). Analysis of net samples from two non-chalk rivers in England, showed temporal stability of the genetic profiles of *S. salar* over more than 20 years (Griffiths *et al.*, 2010). This suggests that the ability to detect temporal stability may depend in part on the window from which the samples originate. Moreover, in order to avoid the confounding effects of ascertainment bias, it is important in assessments of temporal stability to sample the same locations over multiple years.

The samples used in this study spanned an intermediate timeframe (2004 – 2012; 8 years) and were stochastic in terms of sampling site and year. Our results for isolation-by-time (IBT) showed no association between annual sampling, and an AMOVA showed that both sampling between rivers and between years within rivers accounted for only a very small proportion of the explained variance. It should be noted, however, that although variance between years within rivers was significant, it was only marginally higher than variation between rivers. Due to challenges in obtaining samples, the sampling regime in this study was far from ideal; to better address genetic change over time, future assessment of the temporal stability of chalk stream *S. salar* should sample the same sites across a set number of years.

Nonetheless, in the current study, measures of genetic variability were mostly stable across years and diversity estimates of each cohort were comparable to other assessments of *S. salar* using microsatellite markers (Tessier & Bernatchez, 1999; Koljonen *et al.*, 2002; Skaala *et al.*, 2004). This is particularly important given that chalk stream populations are known to have decreased in recent decades. There were significant differences in F_{IS} and H_O ($p<0.005$), which were primarily due to low observed heterozygosity and higher levels of F_{IS} observed from fish in the Itchen. This may reflect differing population dynamics within this river, with more inbreeding within it. However, it is worth noting that the F_{IS} values from the Itchen are low compared to other

studies of *S. salar*. For example, F_{IS} values of 0.11 – 0.13 were found in populations in the Rivers Authie, Valmont and Touques in France (Perrier *et al.*, 2011); therefore, these values alone should not to be a cause for concern.

Given that the studied chalk streams are relatively short in length, estimates of N_E are comparable to estimates obtained from salmonids occupying similar river lengths (Lage & Kornfield, 2006; Jensen *et al.*, 2006; Vähä *et al.*, 2008), although it should be recognised that population dynamics and ecological features can substantially alter such estimates (e.g. Palstra *et al.*, 2007; 2009). One noticeable change was a dramatic drop in N_E in *S. salar* from the Test between 2004 and 2010. It is known that in the past there was a major stocking programme on the River Test and that stocking continued up until the year 2000 (L. Talks, Environment Agency, *pers. comm.*). Interestingly, despite stocking efforts which appear to have temporarily inflated estimates of N_E in this system, apparent effects on population structure and diversity (i.e. admixture effects of stocked fish) are not apparent. More recent estimates of N_E for the Test appear low, but relative decreases in genetic variability were not so apparent. The effects of this stocking activity were also observed in the population structure analyses, where the Test samples deviate in the DAPC, and also show higher Q values for cluster 2 (in blue) in the Structure plot. Evidence suggests that even in populations with small sizes and the potential for future declines, *S. salar* can continue to demonstrate relatively high genetic variability, as has been shown in this study, and in populations in Iberia (Consuegra *et al.*, 2005).

Finally, because *S. salar* typically show considerable variation in the age at which they migrate to sea, such patterns are hypothesised to significantly alter genetic variability and effective population size over time. However, the vast majority of chalk stream fish, at least from the Frome (98%), smolt after one year (R. Lauridsen, GWCT, *pers. comm.*). Future work on the populations assessed here could use molecular analysis to determine the number of years that each generation of chalk stream *S. salar* spends between hatching and spawning, which varies considerably over the range of the species (e.g. Klemetsen *et al.*, 2003; Kusche *et al.*, 2017).

548

549 *Further implications for conservation*

550 The five chalk streams studied are currently managed following county
551 borders and Environment Agency regional borders, so that the Frome, Piddle
552 and Avon are managed within the region of Wessex, while the Test and Itchen
553 are managed within the Solent and South Downs region. This management
554 structure does not appear best suited with their natural population structure,
555 as this study reveals a high degree of connectivity between *S. salar* across all
556 five rivers. The demonstration of the distinctiveness of chalk stream *S. salar*,
557 as well as the lack of sub-structuring between the chalk stream populations,
558 reaffirms the need for bespoke management and conservation of these
559 genetically distinctive fish.

560

561 **Acknowledgements**

562 We thank Chris Gardner (Environment Agency, Hampshire) for help with
563 additional sample collection and fieldwork, and Patrick Hamilton (Exeter) and
564 Rasmus Lauridsen (Game & Wildlife Conservation Trust) for comments on the
565 manuscript. We also thank Sarah-Louise Selly and Andrew Griffiths for help
566 with additional genotyping, and Tom Jenkins for assistance with DAPC
567 analysis. This research was funded by the Game & Wildlife Conservation
568 Trust and the Atlantic Salmon Trust, with additional support from the Salmon
569 and Trout Association, the Westcountry Rivers Trust and the University of
570 Exeter.

571

572

References

ArcGIS, E.S.R.I., 2010. Version 10. *Redlands (CA): ESRI*.

Ayllon, F., Martinez, J. & Garciavazquez, E. (2006a). Loss of regional population structure in Atlantic salmon, Salmo salar L., following stocking. *ICES Journal of Marine Science*, **63**, 1269–1273.

Ayllon, F., Moran, P., & Garcia-Vazquez, E. (2006b). Maintenance of a small anadromous subpopulation of brown trout (*Salmo trutta* L.) by straying. *Freshwater Biology*, **51**, 351–358.

Allendorf, F. & Phelps, S. (1981). Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1507–1514.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289-300.

Berrie, A.. (1992). The chalk-stream environment. *Hydrobiologia*, **248**, 3–9.

Carlsson, J., Olsen, K. H., Nilsson, J., Overli, O., & Stabell, O. B. (1999). Microsatellites reveal fine-scale genetic structure in stream-living brown trout. *Journal of Fish Biology*, **55**, 1290–1303.

Child, A., Burnell, A. & Wilkins, N. (1976). The existence of two races of Atlantic salmon (*Salmo salar* L.) in the British Isles. *Journal of Fish Biology*, **8,** 35–43.

Consuegra, S., Verspoor, E., Knox, D., & García de Leániz, C. (2005). Asymmetric gene flow and the evolutionary maintenance of genetic diversity in small, peripheral Atlantic salmon populations. *Conservation Genetics*, **6,** 823–842.

Dillane, E., Cross, M.C., Mcginnity, P., Coughlan, J.P., Galvin, P.T., Wilkins, N.P. & Cross, T.F. (2007). Spatial and temporal patterns in microsatellite DNA

variation of wild Atlantic salmon, *Salmo salar*, in Irish rivers. *Fisheries Management and Ecology*, **14**, 209–219.

Dieringer, D., & Schlötterer, C. (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167-169.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, **14**, 209–214.

Ellis, J.S., Gilbey, J., Armstrong, a, Balstad, T., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coughlan, J., Cross, T.F., Crozier, W., Dillane, E., Ensing, D., García de Leániz, C., García-Vázquez, E., Griffiths, A.M., Hindar, K., Hjorleifsdottir, S., Knox, D., Machado-Schiaffino, G., McGinnity, P., Meldrup, D., Nielsen, E.E., Olafsson, K., Primmer, C.R., Prodohl, P., Stradmeyer, L., Vähä, J.-P., Verspoor, E., Wennevik, V. & Stevens, J.R. (2011). Microsatellite standardization and evaluation of genotyping error in a large multi-partner research programme for conservation of Atlantic salmon (*Salmo salar* L.). *Genetica*, **139**, 353–67.

Environment Agency. (2004). *The State of England's Chalk Rivers*. The Environment Agency, England and Wales, Bristol BS32 4UD.

Estoup, A., Largiader, C.R., Perrot, E. & Chourrout, D. (1996). Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. *Molecular Marine Biology and Biotechnology*, **5**, 295–298.

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–20.

624 Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: A new series of
625 programs to performpopulation genetics analyses under Linux and Windows.
626 *Molecular Ecology Resources*, **10**, 564-567.

627 Fillatre, E. K., Etherton, P., & Heath, D. D. (2003). Bimodal run distribution in a
628 northern population of sockeye salmon (*Oncorhynchus nerka*): life history and
629 genetic analysis on a temporal scale. *Molecular Ecology*, **12**, 1793–1805.

630 Finnegan A.K, Stevens J.R. (2008) Assessing the long-term genetic impact of
631 historical stocking events on contemporary populations of Atlantic salmon
632 (*Salmo salar* L.). *Fisheries Management and Ecology*, **15**, 315–326.

633 Finnegan, A.K., Griffiths, A.M., King, R.A., Machado-Schiaffino, G., Porcher, J.-P.,
634 Garcia-Vazquez, E., Bright, D. & Stevens, J.R. (2013). Use of multiple
635 markers demonstrates a cryptic western refugium and postglacial colonisation
636 routes of Atlantic salmon (*Salmo salar* L.) in Northwest Europe. *Heredity*, **111**,
637 34–43.

638 Fraser, D.J., Weir, L.K., Bernatchez, L., Hansen, M.M., & Taylor, E.B. (2011). Extent
639 and scale of local adaptation in salmonid fishes: review and meta-analysis.
640 *Heredity*, **106**, 404–20.

641 Garant, D., Dodson, J. & Bernatchez, L. (2000). Ecological determinants and
642 temporal stability of the within-river population structure in Atlantic salmon
643 (*Salmo salar* L.). *Molecular Ecology*, **9**, 615–628.

644 Garcia de Leaniz, C., Fleming, I.A., Einum, S., Verspoor, E., Jordan, W.C.,
645 Consuegra, S., Aubin-Horth, N., Lajus, D., Letcher, B.H., Youngson, A.F.,
646 Webb, J.H., Vøllestad, L. A, Villanueva, B., Ferguson, A & Quinn, T.P. (2007).
647 A critical review of adaptive genetic variation in Atlantic salmon: implications
648 for conservation. *Biological reviews of the Cambridge Philosophical Society*,
649 **82**, 173–211.

650 Gilbey, J., Cauwelier, E., Coulson, M. W., Stradmeyer, L., Sampayo, J. N.,
651 Armstrong, A., Middlemas, S. (2016). Accuracy of assignment of atlantic

652 salmon (*Salmo salar* L.) to rivers and regions in Scotland and northeast
653 England Based on Single Nucleotide Polymorphism (SNP) markers. *PLOS*
654 *ONE*, **11**, e0164327.

655 Gilbey J., Coughlan, J., Wennevik, V., Prodöhl, P. Stevens, J. R., Garcia de Leaniz,
656 C., Ensing, D., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coulson, M. W.,
657 Cross, T. F., Crozier, W., Dillane, E., Ellis, J. S., García-Vázquez, E., Griffiths,
658 A. M., Gudjonsson, S., Hindar, K., Karlsson, S., Knox, D., Machado-
659 Schiaffino, G., Meldrup, D., Nielsen, E. E., Ólafsson, K., Primmer, C. R.,
660 Prusov, S., Stradmeyer, L., Vähä, J.-P., Veselov, A. J., Webster, L. M. I.,
661 McGinnity, P. & Verspoor, E. (2017) A microsatellite baseline for genetic
662 stock identification of European Atlantic salmon (*Salmo salar* L.). *ICES journal*
663 *of Marine Science*, fsx184, https://doi.org/10.1093/icesjms/fsx184.

664 Goudet, J. (1995). FSTAT (version 1.2): a computer program to calculate F-statistics.
665 *Journal of Heredity*, **86**, 485-486.

666 Grandjean, F., Verne, S., Cherbonnel, C. & Richard, A. (2009). Fine-scale genetic
667 structure of Atlantic salmon (*Salmo salar*) using microsatellite markers: effects
668 of restocking and natural recolonization. *Freshwater Biology*, **54**, 417–433.

669 Griffiths, A. M., Ellis, J. S., Clifton-Dey, D., Machado-Schiaffino, G., Bright, D.,
670 Garcia-Vazquez, E., & Stevens, J. R. (2011). Restoration versus
671 recolonisation: The origin of Atlantic salmon (*Salmo salar* L.) currently in the
672 River Thames. *Biological Conservation*, **144**, 2733–2738.

673 Griffiths, A. M., Koizumi, I., Bright, D., & Stevens, J. R. (2009). A case of isolation by
674 distance and short-term temporal stability of population structure in brown
675 trout (*Salmo trutta*) within the River Dart, southwest England. *Evolutionary*
676 *Applications*, **2**, 537–554.

677 Griffiths, A.M., Machado-Schiaffino, G., Dillane, E., Coughlan, J., Horreo, J.L.,
678 Bowkett, A.E., Minting, P., Toms, S., Roche, W., Gargan, P., McGinnity, P.,
679 Cross, T., Bright, D., Garcia-Vazquez, E. & Stevens, J.R. (2010). Genetic

680 stock identification of Atlantic salmon (*Salmo salar*) populations in the
681 southern part of the European range. *BMC Genetics*, **11**, 31.

682 Habicht, C., Seeb, L. W., & Seeb, J.E. (2006). Genetic and ecological divergence
683 defines population structure of sockeye salmon populations returning to
684 Bristol Bay, Alaska, and provides a tool for admixture analysis. *Transactions*
685 *of the American Fisheries Society*, **136**, 82–94.

686 Hansen, M.M., & Mensberg, K.-L.D. (2009). Admixture analysis of stocked brown
687 trout populations using mapped microsatellite DNA markers: indigenous trout
688 persist in introgressed populations. *Biology Letters*, **5**, 656–659.

689 Hansen, M.M., Ruzzante, D.E., Nielsen, E.E., Bekkevold, D., & Mensberg, K.-L.D.
690 (2002). Long-term effective population sizes, temporal stability of genetic
691 composition and potential for local adaptation in anadromous brown trout
692 (*Salmo trutta*) populations. *Molecular Ecology*, **11**, 2523–2535.

693 Hicks, B.J., & Hall, J.D. (2003). Rock type and channel gradient structure salmonid
694 populations in the Oregon Coast Range. *Transactions of the American*
695 *Fisheries Society*, **132**, 468–482.

696 Ikediashi, C., Billington, S. & Stevens, J.R. (2012). The origins of Atlantic salmon
697 (*Salmo salar* L.) recolonizing the River Mersey in northwest England. *Ecology*
698 *and Evolution*, **2**, 2537–48.

699 Jensen, L.F., Hansen, M.M., Carlsson, J., Loeschcke, V., & Mensberg, K.-L.D.
700 (2006). Spatial and temporal genetic differentiation and effective population
701 size of brown trout (*Salmo trutta*, L.) in small Danish rivers. *Conservation*
702 *Genetics*, ***6***, 615–621.

703 Johnson, M.A., & Banks, M.A. (2008). Genetic structure, migration, and patterns of
704 allelic richness among coho salmon (*Oncorhynchus kisutch*) populations of
705 the Oregon coast. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**,
706 1274–1285.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, ***11***, 94.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jones, O.R. & Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, **10**, 551–5.

Jordan, W., Cross, T., Crozier, W.W., Ferguson, A., Galvin, P., Hurrel, R., McGinnity, P., Martin, S., Moffet, I., Price, D., Youngson, A. & Verspoor, E. (2005). Allozyme variation in Atlantic salmon from the British Isles: associations with geography and the environment. *Journal of Fish Biology*, **67**, 146–168.

Keefer, M.L., & Caudill, C.C. (2014). Homing and straying by anadromous salmonids: a review of mechanisms and rates. *Reviews in Fish Biology and Fisheries*, **24**, 333–368.

King, T.L., Eackles, M.S. & Letcher, B.H. (2005). Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. *Molecular Ecology Notes*, **5**, 130–132.

King, T.L., Kalinowski, S.T., Schill, W.B., Spidle, A.P. & Lubinski, B.A. (2001). Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology*, **10**, 807–21.

King, R.A., Hillman, R., Elsmere, P., Stockley, B. and Stevens, J.R. (2016), Investigating patterns of straying and mixed stock exploitation of sea trout, *Salmo trutta,* in rivers sharing an estuary in south-west England. *Fisheries Management and Ecology*, **23**, 376–389.

Klemetsen, A., Amundsen, P.A., Dempson, J.B., Jonsson, B., Jonsson, N., O'Connell, M.F., & Mortensen, E. (2003). Atlantic salmon *Salmo salar* L.,

734 brown trout *Salmo trutta* L. and Arctic charr S*alvelinus alpinus* (L.): a review of
735 aspects of their life histories. *Ecology of Freshwater Fish*, ***12***, 1–59.

736 Koljonen, M., Pella, J. & Masuda, M. (2005). Classical individual assignments versus
737 mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo*
738 *salar*) catches from DNA microsatellite data. *Canadian Journal of Fisheries*
739 *and Aquatic Sciences*, **62**, 2143–2158.

740 Koljonen, M.-L., Tähtinen, J., Säisä, M., & Koskiniemi, J. (2002). Maintenance of
741 genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding
742 programmes and the geographic distribution of microsatellite variation.
743 *Aquaculture*, **212**, 69–92.

744 Kusche, H., Côté, G., Hernandez, C., Normandeau, E., Boivin-Delisle, D., &
745 Bernatchez, L. (2017). Characterization of natural variation in North American
746 Atlantic Salmon populations (Salmonidae: *Salmo salar*) at a locus with a
747 major effect on sea age. *Ecology and Evolution*. DOI: 10.1002/ece3.3132.

748 Lage, C., & Kornfield, I. (2006). Reduced genetic diversity and effective population
749 size in an endangered Atlantic salmon (*Salmo salar*) population from Maine,
750 USA. *Conservation Genetics*, **7**, 91–104.

751 Lehane, B. M., Giller, P. S., O'Halloran, J., & Walsh, P. M. (2004). Relative
752 influences of catchment geology, land use and in-stream habitat on brown
753 trout populations in south-western Ireland. *Biology and Environment:*
754 *Proceedings of the Royal Irish Academy*, **104**, 43–54.

755 McConnell, S., Hamilton, L., Morris, D., Cook, D., Paquet, D., Bentzen, P. & Wright,
756 J. (1995). Isolation of salmonid microsatellite loci and their application to the
757 population genetics of Canadian east coast stocks of Atlantic salmon.
758 *Aquaculture*, **137**, 19–30.

759 Neville, H.M., Isaak, D.J., Dunham, J.B., Thurow, R.F., & Rieman, B.E. (2006). Fine-
760 scale natal homing and localized movement as shaped by sex and spawning

761 habitat in Chinook salmon: insights from spatial autocorrelation analysis of
762 individual genotypes. *Molecular Ecology*, **15**, 4589–4602.

763 O'Reilly, P., Hamilton, L., McConnell, S. & Wright, J. (1996). Rapid analysis of
764 genetic variation in Atlantic salmon (*Salmo salar*) by PCR multiplexing of
765 dinucleotide and tetranucleotide microsatellites. *Canadian Journal of Fisheries*
766 *and Aquatic Sciences*, **2298**, 2292–2298.

767 Palstra, F.P., O'Connell, M.F., & Ruzzante, D.E. (2009). Age structure, changing
768 demography and effective population size in Atlantic salmon (*Salmo salar*).
769 *Genetics*, **182,** 1233-1249.

770 Palstra, F.P., O'Connell, M.F., & Ruzzante, D.E. (2007). Population structure and
771 gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and
772 long-term temporal scales: effects of population size and life history.
773 *Molecular Ecology*, **16**, 4504–4522.

774 Paterson, S., Piertney, S.B., Knox, D., Gilbey, J. & Verspoor, E. (2004).
775 Characterization and PCR multiplexing of novel highly variable tetranucleotide
776 Atlantic salmon (*Salmo salar* L.) microsatellites. *Molecular Ecology Notes*, **4**,
777 160–162.

778 Peakall, R. & Smouse, P. (2012). GenAlEx 6.5: genetic analysis in Excel. Population
779 genetic software for teaching and research—an update. *Bioinformatics*, **1**, 6–
780 8.

781 Perrier, C., Guyomard, R., Bagliniere, J.-L. & Evanno, G. (2011). Determinants of
782 hierarchical genetic structure in Atlantic salmon populations: environmental
783 factors vs. anthropogenic influences. *Molecular Ecology*, **20**, 4231–45.

784 Perrier, C., Baglinière, J.-L., & Evanno, G. (2013). Understanding admixture patterns
785 in supplemented populations: a case study combining molecular analyses and
786 temporally explicit simulations in Atlantic salmon. *Evolutionary Applications*, **6**,
787 218–230.

788 Perrier C, Evanno G, Belliard J, Guyomard R, Baglinière J-L. (2009). Natural
789 recolonization of the Seine River by Atlantic salmon (*Salmo salar*) of multiple
790 origins. *Can. J. Fish. Aquat. Sci.*, **67**,1–4.

791 Primmer, C.R., Veselov, A.J., Zubchenko, A., Poututkin, A., Bakhmet, I., & Koskinen,
792 M.T. (2006). Isolation by distance within a river system: genetic population
793 structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River
794 in northwest Russia. *Molecular Ecology*, **15**, 653–666.

795 Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure
796 using multilocus genotype data. *Genetics*, **155**, 945–59.

797 Raymond, M. & Rousset, F. (1995). GENEPOP (version 1.2): population genetics
798 software for exact tests and ecumenicism. *Journal of Heredity*, **86,** 248–249.

799 Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-
800 statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

801 Sánchez, J.A., Clabby, C., Ramos, D., Blanco, G., Flavin, F., Vázquez, E. & Powell,
802 R. (1996). Protein and microsatellite single locus variability in *Salmo salar* L.
803 (Atlantic salmon). *Heredity*, **77**, 423–32.

804 Skaala, Ø., Høyheim, B., Glover, K., & Dahle, G. (2004). Microsatellite analysis in
805 domesticated and wild Atlantic salmon (*Salmo salar* L.): allelic diversity and
806 identification of individuals. *Aquaculture*, **240**, 131–143.

807 Slettan, A., Olsaker, I. & Lie, Ø. (1995). Atlantic salmon, Salmo salar, microsatellites
808 at the SSOSL25, SSOSL85, SSOSL311, SSOSL417 loci. *Animal genetics*,
809 **26**, 281–282.

810 Stabell, O. (1984). Homing and olfaction in salmonids: a critical review with special
811 reference to Atlantic salmon. *Biological Reviews*, **59**, 333–388.

812 Stahl, G. (1987) Genetic population structure of Atlantic salmon. In Ryman, N. and
813 Utter, F., eds. *Population Genetics and Fishery Management*. Seattle, WA:
814 University of Washington Press, pp. 121–140.

815 Stewart, I.J., Quinn, T.P., & Bentzen, P. (2003). Evidence for fine-scale natal homing
816 among island beach spawning sockeye salmon, *Oncorhynchus nerka*.
817 *Environmental Biology of Fishes*, **67**, 77–85.

818 Taylor, E.B., Stamford, M.D., & Baxter, J.S. (2003). Population subdivision in
819 westslope cutthroat trout (*Oncorhynchus clarki lewisi*) at the northern
820 periphery of its range: evolutionary inferences and conservation implications.
821 *Molecular Ecology*, **12**, 2609–2622.

822 Tessier, N., & Bernatchez, L. (1999). Stability of population structure and genetic
823 diversity across generations assessed by microsatellites among sympatric
824 populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular*
825 *Ecology*, **8**, 169–179.

826 Tierney, K.B., Baldwin, D.H., Hara, T.J., Ross, P.S., Scholz, N.L., & Kennedy, C. J.
827 (2010). Olfactory toxicity in fishes. *Aquatic Toxicology*, **96**, 2–26.

828 Tonteri, A., Veselov, A.J., Zubchenko, A. V., Lumme, J. & Primmer, C.R. (2009).
829 Microsatellites reveal clear genetic boundaries among Atlantic salmon (*Salmo*
830 *salar*) populations from the Barents and White seas, northwest Russia.
831 *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 717–735.

832 Truett, G., Heeger, P. & Mynatt, R. (2000). Preparation of PCR-quality mouse
833 genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques*,
834 **29**, 52–54.

835 Vähä, J.-P., Erkinaro, J., Niemelä, E. & Primmer, C.R. (2008). Temporally stable
836 genetic structure and low migration in an Atlantic salmon population complex:
837 implications for conservation and management. *Evolutionary Applications*, **1**,
838 137–154.

839 Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004). Micro-
840 Checker: Software for identifying and correcting genotyping errors in
841 microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

842 Vasemagi, A., Gross, R., Paaver, T., Kangur, M., Nilsson, J., & Eriksson, L.-O.
843 (2001). Idengification of the origin of an Atlantic salmon (*Salmo salar* L.)
844 population in a recently recolonized river in the Baltic Sea. *Molecular Ecology*,
845 **10**, 2877–2882.

846 Vasemägi, A., Gross, R., Paaver, T., Koljonen, M.-L. & Nilsson, J. (2005). Extensive
847 immigration from compensatory hatchery releases into wild Atlantic salmon
848 population in the Baltic sea: spatio-temporal analysis over 18 years. *Heredity*,
849 **95**, 76–83.

850 Verspoor, E., Beardmore, J.A., Consuegra, S., García de Leániz, C., Hindar, K.,
851 Jordan, W.C., Koljonen, M.-L., Mahkrov, A.A., Paaver, T., Sánchez, J.A.,
852 Skaala, Ø., Titov, S. and Cross, T.F. (2005), Population structure in the
853 Atlantic salmon: insights from 40 years of research into genetic protein
854 variation. *Journal of Fish Biology*, **67**, 3–54.

855 Wickham, H. (2009). *ggplot2*: *Elegant Graphics for Data Analysis*. Springer-Verlag
856 New York.

857

Figure Legends

Figure 1. Map depicting the location of the five chalk stream rivers included in this study and sampling sites. Site codes correspond to those presented in Supporting Information Table S1.

Figure 2. STRUCTURE plot and DAPC of chalk stream *Salmo salar* compared to non-chalk *S. salar* from neighbouring regions of north-west (NW) France, south-west (SW) England, and Norway. Sampled rivers for these regions can be found in Tables 1 and 2. The most likely number of genetic units (*k*) is shown for the STRUCTURE plot (*k = 2*), which distinguishes the chalk stream *S. salar* genotypes as unique compared to non-chalk genotypes. DAPC also distinguishes the chalk stream *S. salar*, and also shows the genetic divergence between NW France/SW England and Norway.

Figure 3. STRUCTURE plot and DAPC of the five chalk stream *Salmo salar* rivers across multiple sampling years (Frome09, Frome11, Piddle09, Piddle11, Avon04, Avon10, Avon12, Test04, Test10, Itchen05, Itchen06 and Itchen10). No genetic groups were defined in the DAPC or STRUCTURE (*k = 2*) plot, but the analyses suggest a pattern of isolation-by-distance (IBD).

Figure 4. Evidence of spatial structuring and temporal stability in *Salmo salar* populations from across the five chalk stream rivers: (A) significant isolation-by-distance (IBD); (B) non-significant isolation-by-time (IBT).

Table 1. Details of sampling for each of the five chalk stream *Salmo salar* populations. Details include the initial sample size and the final sample size after full-sibling removal, together with the percentage of samples removed.

River-Year	No. of sampling sites	Initial sample size	Final sample size	Samples removed (%)
Frome 2009	7	302	221	26.8
Frome 2011	6	454	222	51.1
Piddle 2009	1	32	21	34.4
Piddle 2011	1	89	21	76.4
Avon 2004	2	42	39	52.4
Avon 2010	1	44	20	54.5
Avon 2012	3	117	68	41.9
Test 2004	1	89	45	49.4
Test 2010	1	31	29	6.45
Itchen 2005	1	27	26	3.7
Itchen 2006	1	24	23	4.2
Itchen 2010	1	46	37	19.6

Table 2. Additional rivers sampled from neighbouring (non-chalk) *Salmo salar* populations for inclusion in the STRUCTURE analyses. Genotypes of these populations were obtained through the assessment of 14 microsatellite loci used in the SALSEA-MERGE project. The two loci not included are Ssal417 and Ssosl85. SW indicates south-west.

Country	Sampling site / river	Sample size
France	Sée	47
France	Sélune	48
France	Léguer	47
France	Elorn	47
France	Alune	38
SW England	Exe	142
SW England	Teign	44
SW England	Dart	79
SW England	Tamar	95
SW England	Fowey	55
Norway	Daleelva	105
Norway	Laukhelle	87
Norway	Namsen	90
Norway	Vesterelva	93

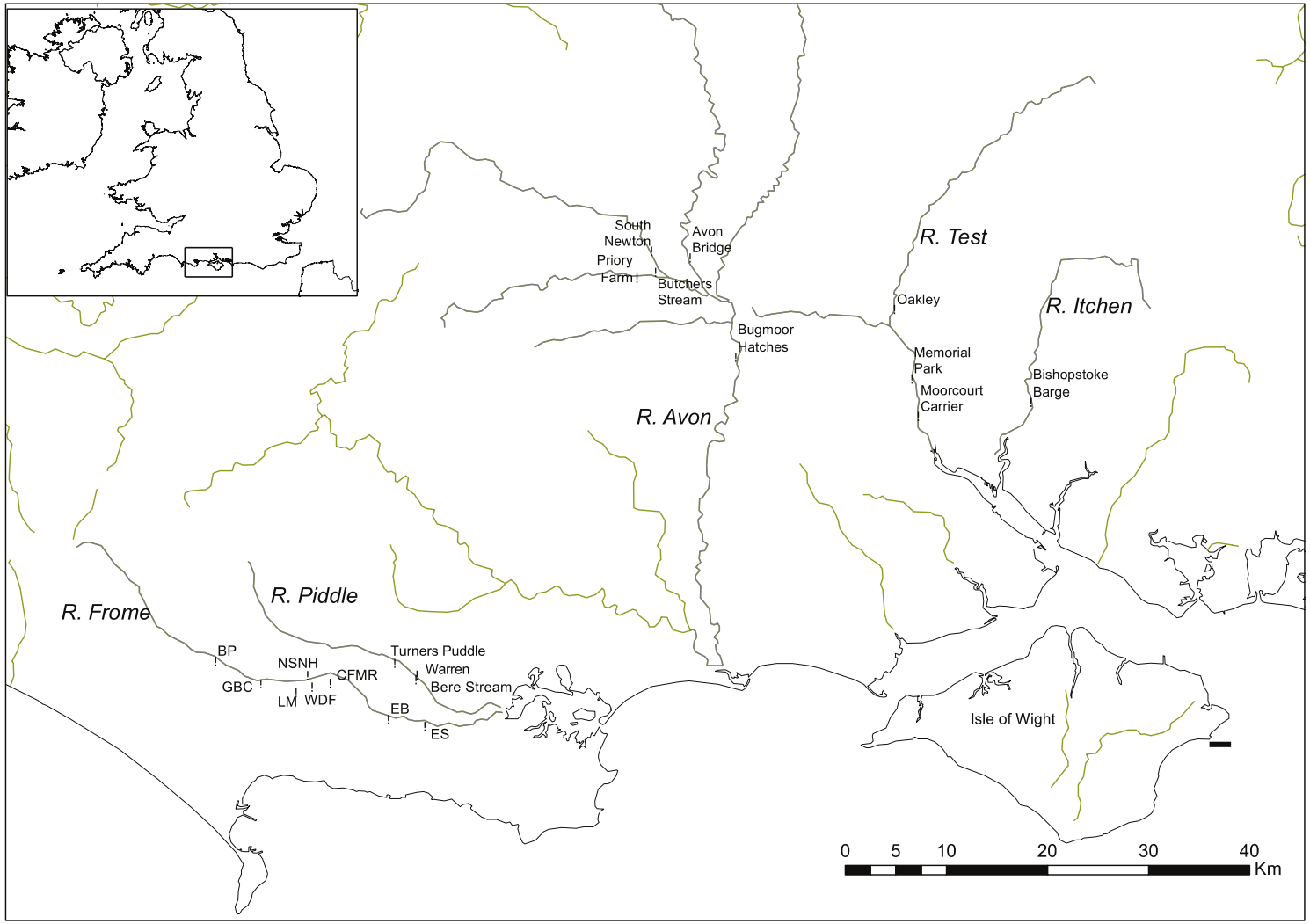
Table 3. Standard population genetics statistics calculated for each of the five chalk stream *Salmo salar* populations. N – sample size; N_A – number of alleles; A_R – allelic richness; H_E – expected heterozygosity; H_O – observed heterozygosity; N_E – effective population size. Numbers in brackets for NE represent 95% CI.

River-Year	N	N_A	A_R	H_E	H_O	F_{IS}	N_E
Frome 2009	221	10.69	5.30	0.7	0.7	0.007	315 (249-419)
Frome 2011	222	10	5.33	0.71	0.69	0.027	228 (194-272)
Piddle 2009	21	6.63	5.03	0.68	0.73	-0.04	39 (27-63)
Piddle 2011	21	6.38	4.95	0.68	0.72	-0.04	53 (37-90)
Avon 2004	39	7.31	5.26	0.7	0.71	0.012	194 (109-682)
Avon 2010	20	6.44	5.02	0.67	0.69	0.002	104 (55-571)
Avon 2012	68	7.69	5.15	0.7	0.71	-0.012	140 (105-204)
Test 2004	45	9	5.47	0.7	0.73	-0.037	132 (89-233)
Test 2010	29	6.5	4.77	0.66	0.69	-0.027	22 (18-27)
Itchen 2005	26	6.75	4.92	0.68	0.7	-0.002	56 (40-90)
Itchen 2006	23	7.19	5.25	0.69	0.67	0.053	99 (60-249)
Itchen 2010	37	7.31	5.03	0.69	0.67	0.036	138 (88-293)

Table 4. Pairwise F_{ST} values calculated for each of the five chalk stream *Salmo salar* populations (river and year indicated). Numbers above the diagonal represent the F_{ST} values and numbers below the diagonal represent the p-value for each comparison (corrected by FDR).

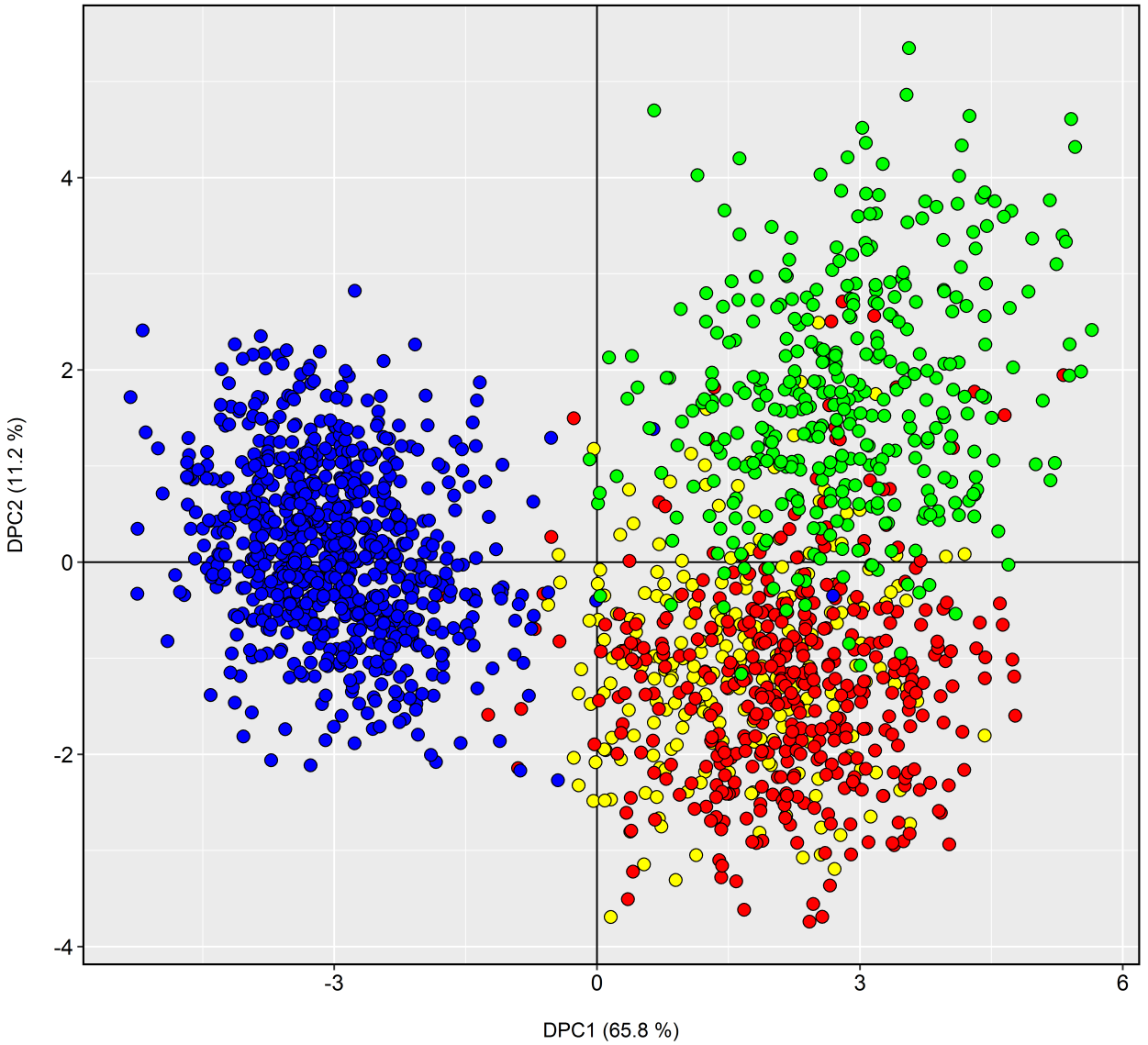
	Frome09	Frome11	Piddle09	Piddle11	Avon04	Avon10	Avon12	Test04	Test10	Itchen05	Itchen06	Itchen10
Frome09		0.002	0.013	0.016	0.013	0.036	0.013	0.032	0.036	0.031	0.016	0.017
Frome11	0.0066		0.010	0.021	0.012	0.033	0.012	0.026	0.032	0.027	0.014	0.014
Piddle09	0.0066	0.033		0.011	0.026	0.042	0.020	0.037	0.038	0.030	0.019	0.025
Piddle11	0.0066	0.0066	n.s.		0.022	0.042	0.025	0.055	0.053	0.051	0.024	0.032
Avon04	0.0066	0.0066	0.0066	0.0066		0.019	0.005	0.030	0.032	0.036	0.018	0.018
Avon10	0.0066	0.0066	0.0066	0.0066	0.0066		0.033	0.049	0.050	0.047	0.040	0.033
Avon12	0.0066	0.0066	0.0066	0.0066	n.s.	0.0066		0.035	0.026	0.038	0.022	0.023
Test04	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066		0.052	0.028	0.018	0.027
Test10	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066		0.043	0.044	0.045
Itchen05	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066		0.007	0.013
Itchen06	0.0066	0.0066	n.s.	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	n.s.		0.008
Itchen10	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066	n.s.	n.s.	

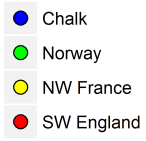
Figure 1 (replacement, as the system kept rejecting my original and I am fast losing the will to live)



Note, this is a pasted in PDF of the file for Figure 1 as your online submission system has randomly stopped accepting my original PDF version of Figure 1 which was submitted fine in the original version.

Figure 2


DPC2 (11.2 %)
DPC1 (65.8 %)
4
2
0
-2
-4
-3
0
3
6


Chalk
Norway
NW France
SW England

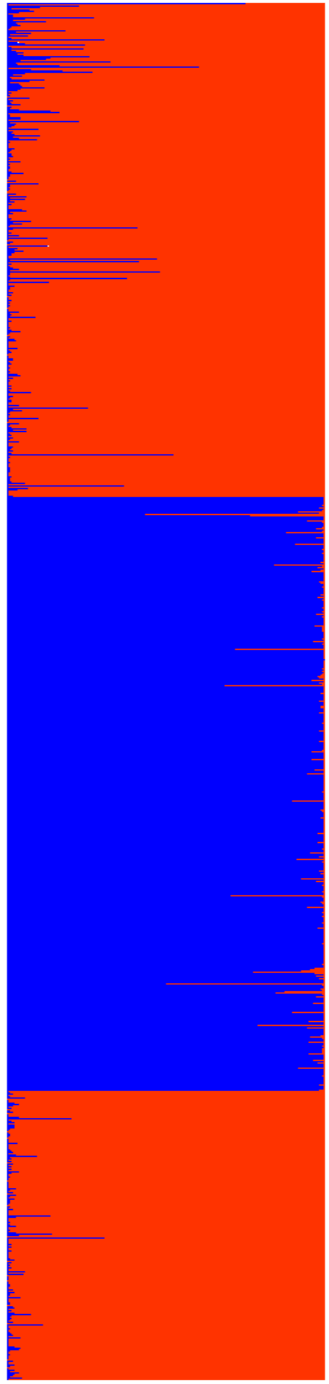
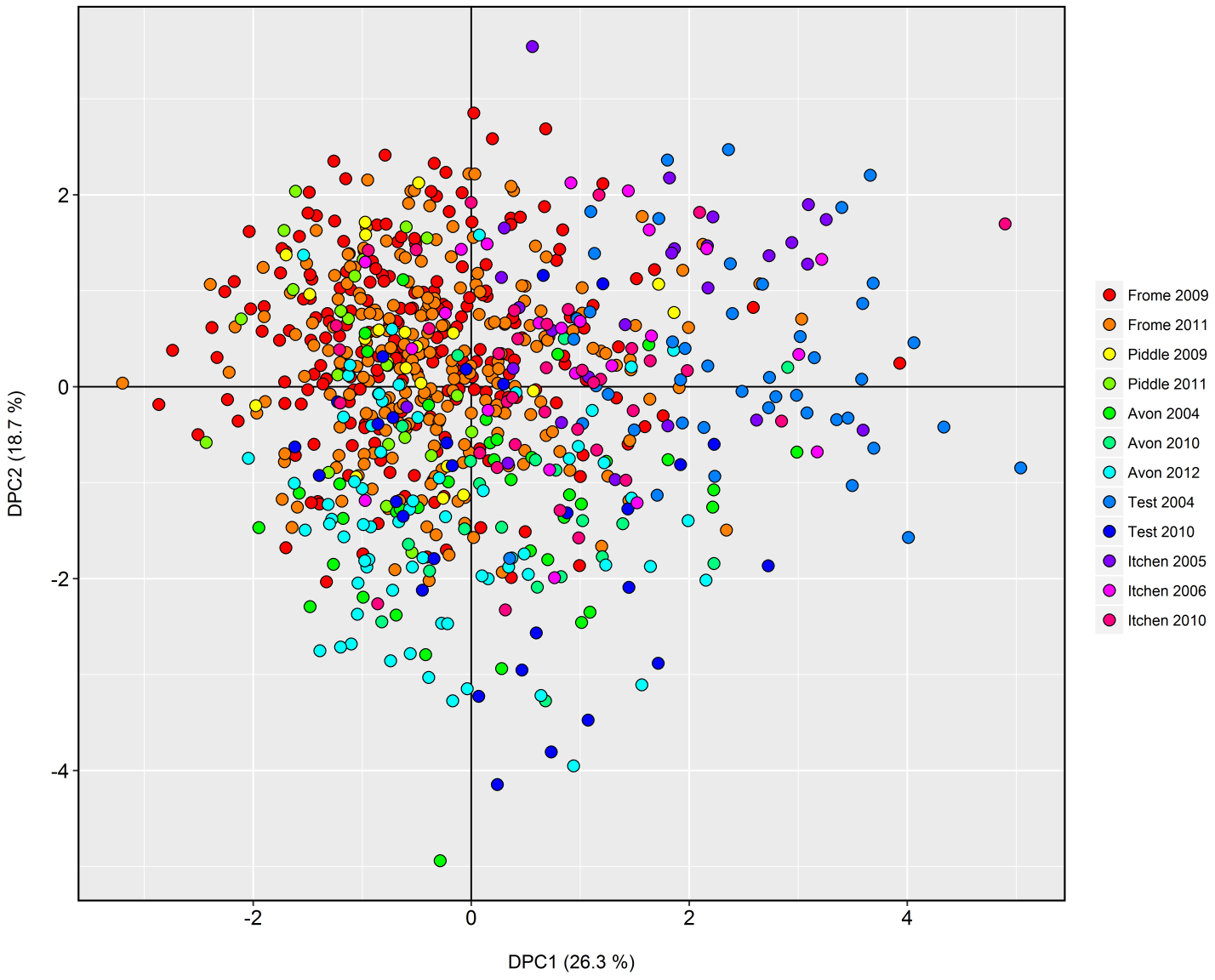

NW France
SW England
Chalk
Norway

Figure 3


DPC2 (18.7 %)
DPC1 (26.3 %)
2
0
-2
-4
-2
0
2
4
Frome 2009
Frome 2011
Piddle 2009
Piddle 2011
Avon 2004
Avon 2010
Avon 2012
Test 2004
Test 2010
Itchen 2005
Itchen 2006
Itchen 2010

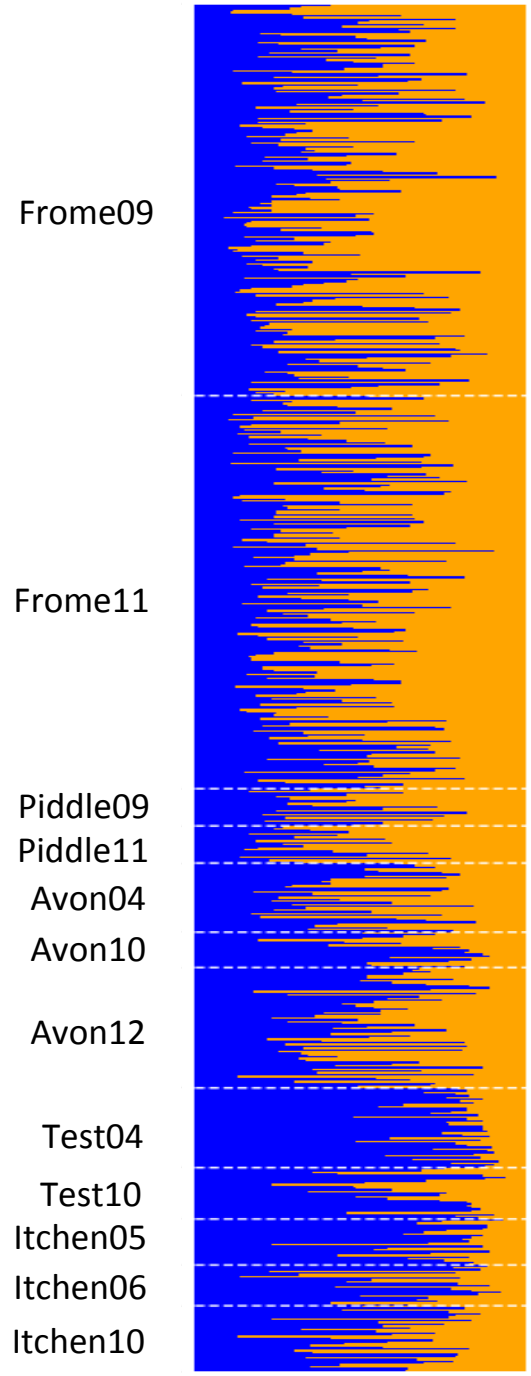
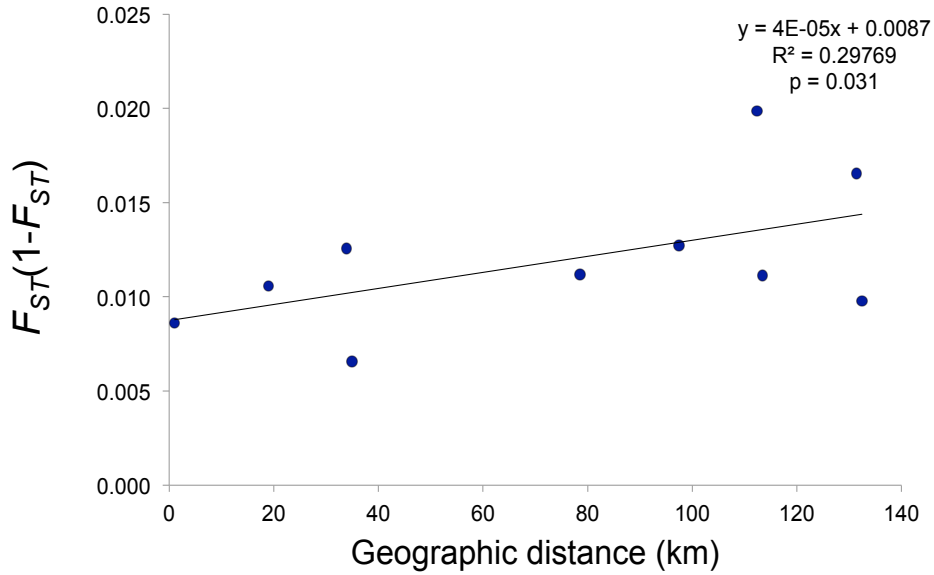

Frome09
Frome11
Piddle09
Piddle11
Avon04
Avon10
Avon12
Test04
Test10
Itchen05
Itchen06
Itchen10

Figure 4


y = 4E-05x + 0.0087
R² = 0.29769
p = 0.031
$F_{ST}(1-F_{ST})$
Geographic distance (km)
0.025
0.020
0.015
0.010
0.005
0.000
0
20
40
60
80
100
120
140

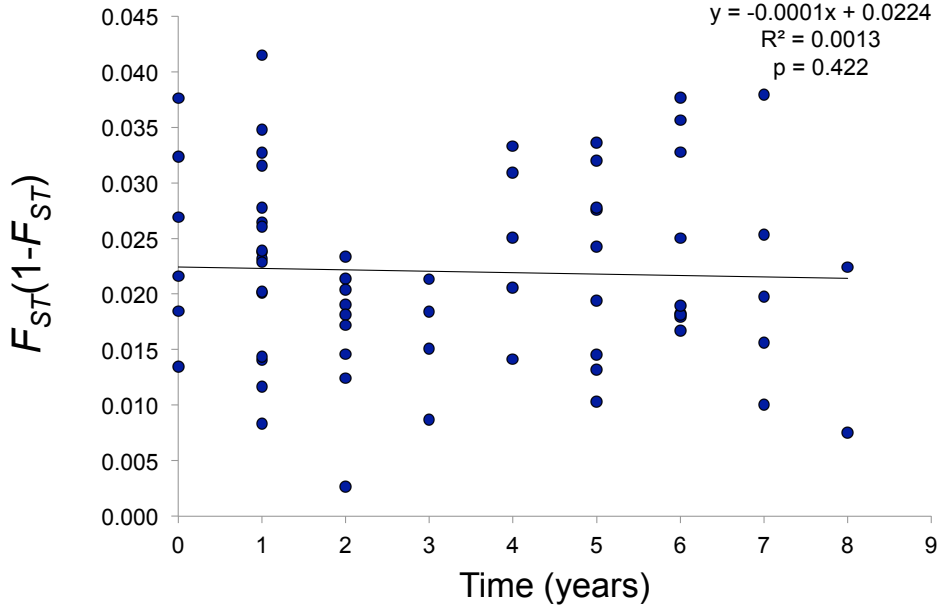

y = -0.0001x + 0.0224
R² = 0.0013
p = 0.422
$F_{ST}(1-F_{ST})$
Time (years)
0.045
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
1
2
3
4
5
6
7
8
9

Table S1. Key for each sample site, including the full name of the sample site, the coordinates, the river and the year sampled, and the original and post sib-ship removal sample sizes.

Population	River	Sampling site	Year of sampling	X coordinate	Y coordinate	Original Sample size	Sample size post sib-ship
FRObp09	Frome	Bradford Peverell	2009	-2.482798	50.736346	42	31
FROgbc09	Frome	Grey Bridge Carrier	2009	-2.419441	50.716733	49	28
FROnsnh09	Frome	North Stream Nine Hatches	2009	-2.359681	50.71588	46	25
FROlm09	Frome	Lewel Mill	2009	-2.369884	50.70913	43	29
FROcfmr09	Frome	Clyffe Farm Main River	2009	-2.322548	50.717482	46	39
FROeb09	Frome	East Burton	2009	-2.240601	50.685788	41	36
FROesg09	Frome	East Stoke	2009	-2.189794	50.679613	37	33
FRObp11	Frome	Bradford Peverell	2011	-2.482798	50.736346	48	27
FROgbc11	Frome	Grey Bridge Carrier	2011	-2.419441	50.716733	49	42
FROnsnh11	Frome	North Stream Nine Hatches	2011	-2.359681	50.71588	95	28
FROcfmr11	Frome	Clyffe Farm Main River	2011	-2.322548	50.717482	93	48
FROeb11	Frome	East Burton	2011	-2.240601	50.685788	47	36
FROesg11	Frome	East Stoke	2011	-2.189794	50.679613	55	41
PIDber09	Piddle	Bere Stream	2009	-2.200775	50.725076	32	21
PIDwar11	Piddle	Warren	2011	-2.202387	50.721071	46	21
AVNbrd04	Avon	Avon Bridge	2004	-1.816891	51.09558	23	20
AVNbrd10	Avon	Avon Bridge	2010	-1.816891	51.09558	44	20
AVNbri12	Avon	Avon Bridge	2012	-1.816891	51.09558	21	21
AVNbut12	Avon	Butchers Stream	2012	-1.866044	51.082822	45	21
AVNprf12	Avon	Priory Farm	2012	-1.892028	51.077579	34	26
TESTmem04	Test	Memorial Park	2004	-1.496397	50.953838	89	45
TESTmem10	Test	Memorial Park	2010	-1.505267	50.987364	31	29
ITCbis05	Itchen	Bishopstoke Barge	2005	-1.337858	50.965754	27	29
ITCbis06	Itchen	Bishopstoke Barge	3006	-1.337858	50.965754	24	23
ITCbis10	Itchen	Bishopstoke Barge	2010	-1.337858	50.965754	46	37

Table S2. Primer quantities and multiplexes.

Primers/36 ul aliquot	
Ssosl417	22.5
Ssa202	9
Ssa14	8
SSsp3016	3.6
SSspG7	14.4
Ssosl85	7.2
SSsp2216	9
SsaF43	24
SSsp2210	16.4
Ssa197	9
SSsp1605	9
SsaD144	7.2
Ssa157	4.5
Ssa171	12
SSsp2201	12
Ssa289	3.2

Multiplex A (FRAG-3)				**Multiplex B (FRAG-3)**		**Multiplex C (FRAG 3-40)**			
Multiplex A1		**Multiplex A2**				**Multiplex C1**		**Multiplex C2**	
Ssosl417	1.6	85	5	SSsp2216	4	SsaD144	5	Ssa157	8
Ssa202	4	Water	90	SsaF43	1.5	Water	90	Ssa171	3
Ssa14	4.5			SSsp2210	2.2			SSsp2201	3
SSsp3016	10			Ssa197	4			Ssa289	11
SSspG7	2.5			SSsp1605	4			Water	50
Water	54.8			Water	68.6				

Table S3. Pairwise F_{ST} between *Salmo salar* sampling sites across the five chalk stream rivers. Sample site codes match those presented in Table S1.

	FROb09	FROgbc09	FROnsnh09	FROl09	FROcfmr09	FROeb09	FROesg09	FROb11	FROgbc11	FROnsnh11	FROcfmr11	FROeb11	FROesg11	PIDber09	PIDwar11	AVNbrd04	AVNbug04	AVNbrd10	AVNbri12	AVNbut12	AVNprf12	TESTme m04	TESTme m10	ITCbis 05	ITCbis 06	ITCbis 10
FROb09		0.0 22	0.02 8	0.0 19	0.02 7	0.0 21	0.0 31	0.0 35	0.0 16	0.03 6	0.02	0.0 38	0.0 29	0.0 27	0.0 36	0.0 33	0.03 5	0.0 52	0.0 41	0.0 33	0.0 34	0.04 6	0.05 6	0.0 43	0.0 4	0.0 32
FROgbc09	0.0 35		0.01 19	0.0 1	0.02 08	0.0 23	0.0 35	0.0 1	0.0 8	0.02 4	0.01	0.0 15	0.0 16	0.0 27	0.0 19	0.0 13	0.00 9	0.0 27	0.0 16	0.0 16	0.0 16	0.03 2	0.04 2	0.0 35	0.0 25	0.0 24
FROnsnh09	0.0 35	0.0 n.s.		0.0 15	0.02	0.0 2	0.0 15	0.0 33	0.0 11	0.02 3	0.01 2	0.0 16	0.0 12	0.0 23	0.0 2	0.0 26	0.01 7	0.0 37	0.0 26	0.0 27	0.0 25	0.03 4	0.05 2	0.0 37	0.0 18	0.0 31
FROlm09	0.0 35	0.0 35	0.03 5		0.02 8	0.0 22	0.0 21	0.0 38	0.0 19	0.02 6	0.01 5	0.0 23	0.0 23	0.0 17	0.0 24	0.0 28	0.02 3	0.0 44	0.0 34	0.0 32	0.0 24	0.03 4	0.05 1	0.0 31	0.0 27	0.0 34
FROcfmr09	0.0 35	0.0 35	0.03 5	0.0 35		0.0 15	0.0 27	0.0 25	0.0 15	0.02 6	0.01	0.0 18	0.0 21	0.0 23	0.0 2	0.0 37	0.02 5	0.0 41	0.0 31	0.0 22	0.0 34	0.04 6	0.06 3	0.0 53	0.0 36	0.0 32
FROeb09	0.0 35	0.0 n.s.	0.03 5	0.0 35	0.03 5		0.0 11	0.0 34	0.0 16	0.02 4	0.00 9	0.0 17	0.0 11	0.0 21	0.0 19	0.0 19	0.02 1	0.0 42	0.0 22	0.0 13	0.0 23	0.04 4	0.04 4	0.0 36	0.0 22	0.0 22
FROesg09	0.0 35	0.0 35	0.03 5	0.0 35	0.03 5	0.0 n.s.		0.0 38	0.0 2	0.02 6	0.00 7	0.0 13	0.0 14	0.0 18	0.0 23	0.0 18	0.02 2	0.0 39	0.0 29	0.0 17	0.0 18	0.03 6	0.03 8	0.0 26	0.0 09	0.0 24
FROb11	0.0 35	0.0 35	0.03 5	0.0 35	0.03 5	0.0 35	0.0 35		0.0 17	0.02 0.04	0.02 7	0.0 32	0.0 42	0.0 33	0.0 38	0.0 44	0.02 6	0.0 43	0.0 38	0.0 32	0.0 43	0.04 3	0.04 5	0.0 51	0.0 48	0.0 36
FROgbc11	0.0 35		n.s.	0.0 35	0.03 5	0.0 35	0.0 35	0.0 35		0.02 1	0.00 9	0.0 16	0.0 14	0.0 18	0.0 16	0.0 15	0.00 5	0.0 25	0.0 15	0.0 17	0.0 2	0.02 3	0.04 1	0.0 26	0.0 18	0.0 13
FROnsnh11	0.0 35	0.0 35	0.03 5	0.0 35	0.03 5	0.0 35	0.0 35	0.0 35	0.0 35		0.01 7	0.0 23	0.0 25	0.0 24	0.0 47	0.0 42	0.02 8	0.0 49	0.0 34	0.0 29	0.0 28	0.03 7	0.05 3	0.0 32	0.0 28	0.0 37
FROcfmr11	0.0 35	0.0 35	0.03 5	0.0 35	0.03 5	n.s.	n.s.	0.0 35	n.s.	0.03 5		0.0 06	0.0 11	0.0 15	0.0 23	0.0 2	0.02	0.0 39	0.0 22	0.0 11	0.0 17	0.02 9	0.03 8	0.0 26	0.0 16	0.0 19
FROeb11	0.0 35	0.0 35	0.03 5	0.0 35	0.03 5	0.0 35	n.s.	0.0 35	0.0 35	0.03 5	n.s.		0.0 16	0.0 23	0.0 39	0.0 21	0.01 8	0.0 4	0.0 25	0.0 23	0.0 27	0.03 1	0.04 4	0.0 35	0.0 2	0.0 22
FROesg11	0.0 35	0.0 35	n.s.	0.0 35	0.03 5	n.s.	0.0 35	0.0 35	0.0 35	0.03 5	0.03 5	0.0 35		0.0 1	0.0 23	0.0 18	0.02 8	0.0 41	0.0 18	0.0 27	0.0 2	0.03 2	0.03 9	0.0 3	0.0 09	0.0 16
PIDber09	0.0 35	0.0 35	0.03 5	n.s.	0.03 5	0.0 35	0.0 35	0.0 35	0.0 35	0.03 5	0.03 5	0.0 35	n.s.		0.0 16	0.0 31	0.03 1	0.0 42	0.0 27	0.0 29	0.0 19	0.03 5	0.04 2	0.0 27	0.0 18	0.0 25

PID																										
war1	0.0	0.0		0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03	0.0	0.0			0.0	0.02	0.0	0.0	0.0	0.0	0.05	0.05	0.0	0.0	0.0
1	35	35	n.s.	35	5	35	35	35	35	5	5	35	35	n.s.		21	1	34	24	28	27	3	7	48	29	32
AVN																										
brd0	0.0		0.03	0.0	0.03	0.0	0.0	0.0		0.03	0.03	0.0	0.0	0.0	0.0		0.00	0.0	0.0	0.0	0.0	0.03	0.03	0.0	0.0	0.0
4	35	n.s.	5	35	5	35	35	35	n.s.	5	5	35	35	35	35		8	24	09	23	1	5	2	38	18	2
AVN																			-							
bug0	0.0			0.0	0.03	0.0	0.0	0.0		0.03	0.03	0.0	0.0	0.0				0.0	0.0	0.0	0.0	0.02		0.0	0.0	0.0
4	35	n.s.	n.s.	35	5	35	35	35	n.s.	5	5	35	35	35	n.s.	n.s.		14	02	12	14	9	0.04	34	24	22
AVN																										
brd1	0.0	0.0	0.03	0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03	0.0	0.0	0.0	0.0				0.0	0.0	0.0	0.04		0.0	0.0	0.0
0	35	35	5	35	5	35	35	35	35	5	5	35	35	35	35	n.s.	n.s.		24	35	39	4	0.05	42	39	31
AVN																										
bri1	0.0		0.03	0.0	0.03	0.0	0.0	0.0		0.03	0.03	0.0	0.0	0.0	0.0			0.0		0.0	0.0	0.03	0.02	0.0	0.0	0.0
2	35	n.s.	5	35	5	35	35	35	n.s.	5	5	35	35	35	35	n.s.	n.s.	35		13	11	3	7	37	25	22
AVN																										
but1	0.0	0.0	0.03	0.0	0.03		0.0	0.0	0.0	0.03		0.0	0.0	0.0	0.0	0.0		0.0			0.0	0.04	0.04	0.0	0.0	0.0
2	35	35	5	35	5	n.s.	35	35	35	5	n.s.	35	35	35	35	35	n.s.	35	n.s.		14	2	1	47	31	26
AVN																										
prf1	0.0	0.0	0.03	0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03	0.0	0.0		0.0			0.0				0.03	0.02	0.0	0.0	0.0
2	35	35	5	35	5	35	35	35	35	5	5	35	35	n.s.	35	n.s.	n.s.	35	n.s.	n.s.		9	8	36	18	3
TEST																										
me	0.0	0.0	0.03	0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03	0.0	0.0	0.0	0.0	0.0	0.03	0.0	0.0	0.0	0.0		0.05	0.0	0.0	0.0
m04	35	35	5	35	5	35	35	35	35	5	5	35	35	35	35	35	5	35	35	35	35		3	24	19	24
TEST																										
me	0.0	0.0	0.03	0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03	0.0	0.0	0.0	0.0	0.0	0.03	0.0	0.0	0.0	0.0	0.03		0.0	0.0	0.0
m10	35	35	5	35	5	35	35	35	35	5	5	35	35	35	35	35	5	35	35	35	35	5		43	39	45
ITCbi	0.0	0.0	0.03	0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03	0.0	0.0	0.0	0.0	0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03		0.0	0.0
s05	35	35	5	35	5	35	35	35	35	5	5	35	35	35	35	35	5	35	35	35	35	5	5	06		11
ITCbi	0.0	0.0		0.0	0.03	0.0		0.0	0.0	0.03	0.03	0.0			0.0		0.03	0.0	0.0	0.0		0.03	0.03	n.s		0.0
s06	35	35	n.s.	35	5	35	n.s.	35	35	5	5	35	n.s.	n.s.	35	n.s.	5	35	35	35	n.s.	5	5	.		09
ITCbi	0.0	0.0	0.03	0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03	0.0	0.0	0.0	0.0	0.0	0.03	0.0	0.0	0.0	0.0	0.03	0.03	n.s	n.s	
s10	35	35	5	35	5	35	35	35	35	5	5	35	35	35	35	35	5	35	35	35	35	5	5	.	.	

Figure S1. Hierarchical STRUCTURE analysis of NW (north-west) France, SW (south-west) England and Norway. *K = 2.*

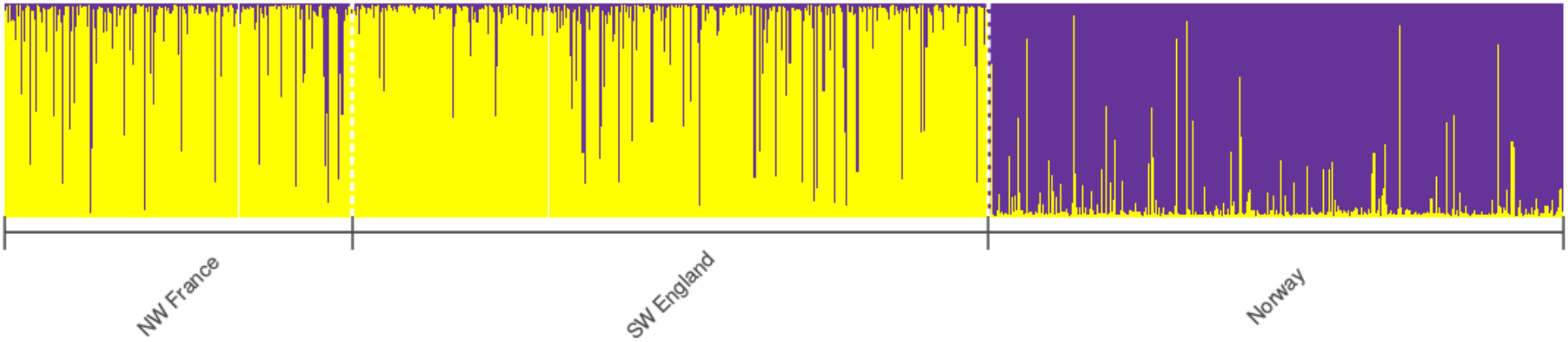
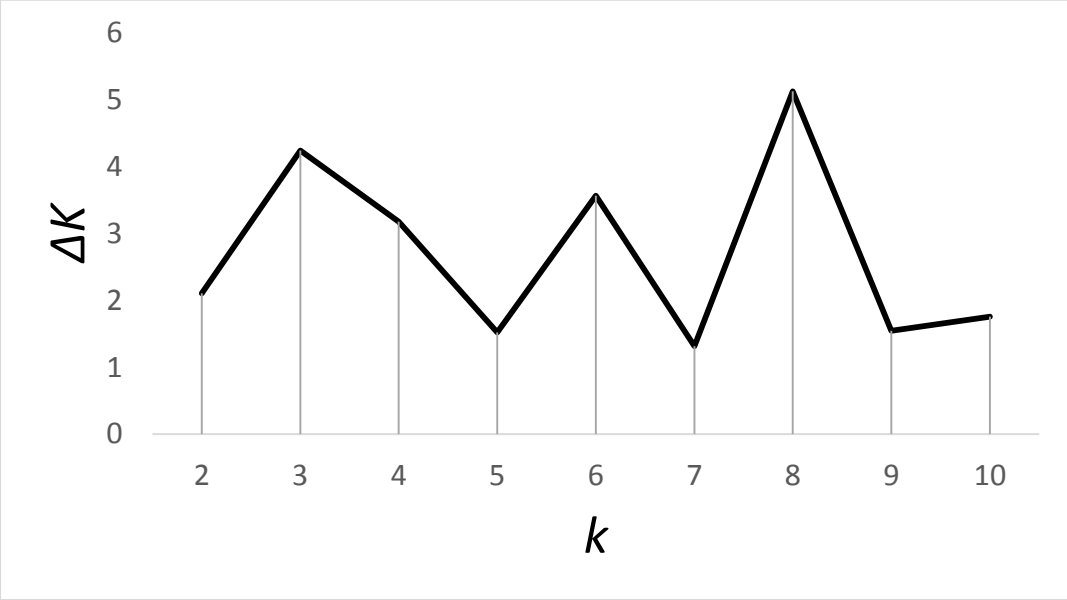


Figure S2. Δ*k* statistic plotted against *k* for the hierarchical STRUCTURE analysis of population differentiation in the chalk stream rivers. Δ*k* values calculated as per Evanno *et al.* (2005) are plotted along the y-axis and potential genetic clusters (*k*) are shown on the x-axis. The low Δ*k* values (max. 5.1) and observation of no clear peak in the statistic show no support for any genetic cluster.



Atlantic salmon *Salmo salar* in the chalk streams of England are genetically unique

Charles Ikediashi[1†], Josephine R. Paris[1†], R. Andrew King[1], William R. C. Beaumont[2], Anton Ibbotson[2] and Jamie R. Stevens[1*]

[1]*Department of Biosciences, University of Exeter, Geoffrey Pope Building, Stocker Road, Exeter, Devon, EX4 4QD, UK*

[2]*Salmon & Trout Research Centre, Game & Wildlife Conservation Trust (GWCT), East Stoke, Wareham, Dorset BH20 6BB*

Tel: 01392 723775; Fax: 01392 723700

*Corresponding author: J. R. Stevens, email: j.r.stevens@ex.ac.uk

†Joint first authors

Abstract

Recent research has identified genetic groups of Atlantic salmon *Salmo salar* that show association with geological and environmental boundaries. This study focuses on one particular subgroup of the species inhabiting the chalk streams of southern England. These fish are genetically distinct from other British and European *S. salar* populations and have previously demonstrated markedly low admixture with populations in neighbouring regions. The genetic population structure of *S. salar* occupying five chalk streams was explored using 16 microsatellite loci. The analysis provides evidence of the genetic distinctiveness of chalk stream *S. salar* in southern England, in comparison to populations from non-chalk regions of Western Europe. Little genetic differentiation exists between the chalk stream populations, and a pattern of isolation-by-distance (IBD) was evident. Furthermore, evidence of temporal stability of *S. salar* populations across the five chalk streams was found. This work provides new insights into the temporal stability and lack of genetic population sub-structuring within a unique component of the species' range of *S. salar*.

Key words: Atlantic salmon, chalk streams, microsatellite, population structure, *Salmo salar*

Introduction

Atlantic salmon *Salmo salar* L. 1758 is an anadromous fish species, which returns to its natal river after reaching sexual maturity. As a result, the species shows marked population structuring into broad geographic groups, which is readily detectable using genetic methodologies (Stahl, 1987; Verspoor *et al.*, 2005), particularly through analysis of microsatellite markers (e.g. King *et al.*, 2001; Koljonen *et al.*, 2005; Tonteri *et al.*, 2009; Griffiths *et al.*, 2010). Current research suggests that broad genetic groups are largely defined by a combination of geological substrate (Grandjean *et al.*, 2009; Perrier *et al.*, 2011), phylogeography (Finnegan *et al.*, 2013) and environmental factors (Dillane *et al.*, 2007), leading to the suggestion that *S. salar* populations may be locally adapted to their in-river environments (Garcia de Leaniz *et al.*, 2007; Fraser *et al.*, 2011; Perrier *et al.*, 2011).

One sub-group of this species, which resides within the chalk streams of southern England, has been shown to form a genetically distinct unit when compared with groups of geographical neighbour populations in non-chalk rivers (Griffiths *et al.*, 2010; Gilbey *et al.*, 2017). Chalk stream *S. salar* populations also appear to display relatively low levels of admixture with populations in neighbouring regions (Ikediashi *et al.*, 2012). Admixture has, for some time, been associated with a reduction in population differentiation. For example, Stahl (1987) deduced that, in order to maintain genetic differences between two or more *S. salar* sub-populations of 2,500 to 10,000 fish, there had to be fewer than one migrant per year between them. More recent studies in Spain (Ayllon *et al.*, 2006a) and in the Baltic Sea (Vasemägi *et al.*, 2005), showed that reductions in the between-river population structuring of *S. salar* has been identified as a result of admixture with farm escapees. Following this line of argument, it is possible that chalk stream *S. salar*, which engage in relatively little admixture with individuals in neighbouring regions (Ikediashi *et al.*, 2012), may also show reduced genetic structuring between them. However, despite several previous studies having included some fish of chalk stream origin (e.g. Child *et al.*, 1976; Jordan *et al.*, 2005; Finnegan *et al.*, 2013) and their apparent genetic distinctiveness (Griffiths *et al.*, 2010), the

degree of population structure within and between chalk stream *S. salar* has not yet been the subject of detailed exploration.

The reason for the distinction between these *S. salar* populations most likely stems from one or more of the unique abiotic factors of chalk stream geology, which are described in detail by Berrie (1992). The calcareous substrate, upon which chalk streams are formed, is porous, and thus chalk streams are aquifer fed. The water is therefore relatively clear, stable in temperature throughout most of the year, and alkaline (ca. pH 8). Due to these unique environments, several chalk streams have been designated SSSIs (Sites of Special Scientific Interest) in the UK. However, of the 161 rivers classified as chalk streams in England (Environment Agency 2004), major *S. salar* populations are present in just five of these. These rivers include the Frome, Piddle, Avon, Test and Itchen, all of which have each been sampled for the purpose of this study (Figure 1)~~, and which are henceforth referred to by their specific names only~~. Crucially, although chalk streams are located between the counties of Yorkshire in north-east England and Dorset in southern England, the five rivers with substantial *S. salar* populations span only some 70 km along the southern English coast. With so few chalk stream populations, each of which has markedly decreased in numbers in recent decades (Environment Agency 2004), there is additional incentive to understand the full extent of chalk stream *S. salar* local population genetic structure.

The primary aim of this study was to assess the population structure of *S. salar* populations from the above five major chalk streams of southern England. First, we assessed the distinctiveness of the chalk stream populations by explicitly comparing them to populations from geographically neighbouring populations residing in non-chalk geologies. Secondly, we assessed whether significant genetic variation exists among the chalk stream populations by exploring population structure, genetic diversity and patterns of isolation-by-distance (IBD). Thirdly, by analysing temporal cohorts, we explored the temporal stability of chalk stream genetic variability over time. In summary, this study represents the first assessment explicitly addressing the distinctiveness of chalk stream *S. salar* populations in southern England, and

highlights the importance of managing these unique populations as distinct genetic entities. We anticipate that this information will be useful for the successful management and conservation of this species within these rivers.

Materials and Methods

Sampling

Juvenile *S. salar* (0+ parr) were sampled from the five chalk streams of southern England that still contain significant populations: the Frome, Piddle, Avon, Test and Itchen (Figure 1; Table 1; Supporting Information Table S1). The Avon, Itchen and Test were sampled by the Environment Agency (EA) during routine national surveys and management programmes between 2004 and 2012. Sampling on the Frome and Piddle was carried out by the Game & Wildlife Conservation Trust (GWCT) in September of 2009 and 2011 (Figure 1) during routine juvenile abundance surveys. Fish were collected by electrofishing; adipose fin clips were then taken and preserved in 100% ethanol, according to national agency ethical guidelines. To avoid issues of small sample sizes we aimed to collect 50 parr samples from each site. For assessment of the chalk stream populations in comparison with those from neighbouring non-chalk geographical regions, salmon from rivers in north-west (NW) France, south-west (SW) England and Norway were included for the population structure analyses (Table 2) and were obtained from a database of salmon genotyped for the SALSEA-Merge project (Ellis *et al.*, 2011, Gilbey *et al.*, 2017).

DNA extraction and microsatellite amplification

DNA was extracted from fin clips using the HOTshot method (Truett *et al.*, 2000) and from scales using the Chelex method (Estoup *et al.*, 1996). Sixteen microsatellite loci were genotyped. Fourteen loci were amplified according to the protocol of Ikediashi *et al.* (2012): Ssa14 (McConnell *et al.*, 1995); Ssa202, SSsp3016, Ssa197 (O'Reilly *et al.*, 1996); SsaF43 (Sánchez *et al.*, 1996); SSspG7, SSsp1605, SSsp2210, SSsp2201, and SSsp2216 (Paterson

et al., 2004); Ssa171, Ssa289, Ssa157, and SsaD144 (King *et al.*, 2005). Two additional loci, Ssosl85 and Ssosl417 (Slettan *et al.*, 1995), were amplified in the first multiplex reaction described by Ikediashi *et al.*, (2012). Potential *S. salar* x brown trout *Salmo trutta* L. 1758 hybrids were recognised by the presence of alleles longer than 350bp for locus SSsp1605, or alleles longer than 135bp for Ssa289 (Finnegan & Stevens, 2008). Hybrid fish were removed from the dataset.

PCR reactions were carried out in 10µl reactions containing ~50ng of extracted *S. salar* template DNA, 3µl water, 5µl of Qiagen Taq PCR Mastermix and 1µl of primer mixture (Supporting Information Table S2). PCR conditions were as follows: an initial denaturation step of 5 min at 95°C, followed by a touchdown PCR consisting of eight cycles with a 30 s denaturation step at 95 °C, a 90 s annealing step starting at 62 °C and decreasing the temperature 2°C every two steps until a touchdown temperature of 47°C was reached, with 3 minutes of extension at 72°C, followed by a final 10 minute extension at 72°C. Size of products of fluorescently labelled PCR products were assessed using a Beckman-Coulter CEQ8000 automatic DNA sequencer and the associated fragment analysis software (Beckman Coulter).

Data checking

MICROCHECKER v2.2.3 (Van Oosterhout *et al.*, 2004) was used to check for scoring errors due to stutter peaks, large allele dropout or null alleles. In order to prevent the false detection of population structure due to the presence of family groups (Allendorf & Phelps, 1981), the program COLONY v2.0.4.1 (Jones & Wang, 2010) was used to identify full siblings. The mating system was defined as polygamous for males and females and without inbreeding. Each run was of medium length, using high precision and the full-likelihood method. Allele frequencies were not updated during the run and no prior sib-ship was assumed. An error rate of 0.02 was used for each locus based on the protocol of Ellis *et al.* (2011). The program was run twice independently, with different starting seeds to check consistency of sibship reconstruction.

Full-sib families were reduced to one representative, if supported by an average likelihood of 0.5 or higher between the two runs.

Linkage disequilibrium and deviations from Hardy Weinberg equilibrium were assessed using GENEPOP v4.2 (Raymond & Rousset, 1995). The 95% significance level for corrections of multiples tests for both procedures were adjusted using the False Discovery Rate (FDR) (Benjamini & Hochberg, 1995).

Descriptive statistics

The number of alleles (N_A), expected heterozygosity (H_E) and observed heterozygosity (H_O) were calculated in Genalex v6.5.02 (Peakall & Smouse, 2012) for each year cohort of *S. salar* from each of the five chalk streams. Allelic richness (A_R) and the heterozygote deficit (F_{IS}) were calculated using the program FSTAT v2.9.3 (Goudet, 1995). In order to determine whether there were any significant differences in genetic diversity between the five chalk streams, two sided permutation tests were performed within FSTAT for A_R, H_O, H_E and F_{IS}, using 1000 permutations of the dataset. Where significant differences were identified, further pairwise comparisons were made in order to determine between which groups the significant differences lay.

The effective population size (N_E) for each river and year was assessed in the program NeEstimator v. 2.01 (Do *et al.*, 2014) using the linkage disequilibrium model under a random mating scenario, using 0.01 as the lowest allele frequency as the critical value cut-off.

Assessment of the genetic uniqueness of chalk stream S. salar

In order to place the chalk stream *S. salar* populations in a wider geographical context, we incorporated genotypes from four other geographical regions: NW France, SW England and Norway (Table 2). Genotypes were obtained from the SALSEA-MERGE dataset (Ellis *et al.*, 2011; Gilbey *et al.*, 2017). To facilitate accurate comparisons across these populations and to allow the

incorporation of previously genotyped loci, two markers, Ssosl417 and Ssosl85, were excluded for these population analyses, resulting in the use of a reduced set of 14 microsatellite loci for all population structure analyses. Two complementary methods were used to assess the population structure between chalk and non-chalk populations and also for the assessment of structure within the chalk stream populations.

Firstly, the program STRUCTURE v2.3.4 (Pritchard *et al.*, 2000) was used to identify the number of distinct genetic units (*k*) across the four geographic regions. STRUCTURE was run from $k = 1$ to $k = 10$ with 150,000 Markov Chain Monte Carlo (MCMC) replicates, after a burn-in of 75,000 replicates from ten independent starting points. The Evanno method (Δk: Evanno *et al.*, 2005) was used to determine the optimum number of genetic units (*k*) from the results. A hierarchical STRUCTURE analysis was conducted based on the most likely number of genetic units (see Results) in order to further assess population sub-divisions and the possible existence of sub-structuring within the chalk stream rivers. In hierarchical analyses of population structure, the same analysis parameters were used as outlined above.

Secondly, an assessment of population structure using a Discriminant Analysis of Principal Components (DAPC) was conducted in R using the adegenet package (Jombart, 2008; Jombart *et al.*, 2010). The optimum alpha score (using the optim.a.score function) was used to assess how many principal components should be retained for each analysis and we assessed structure using five discriminant components. DAPC plots of the first two principal components were derived using ggplot2 (Wickham, 2009).

Assessment of the population structure between chalk stream rivers

To assess the differentiation between sampling sites from across each chalk river, pairwise F_{ST} values from each site and year were calculated. Based on the outcome of this analysis (see Results), fish from individual sample sites were grouped together across each river; annual cohorts from each river were then used for all subsequent analyses of population structure and genetic diversity. Global and pairwise F_{ST} values were calculated for each year cohort

from each of the five chalk stream rivers. All values were tested for significance using 10 000 permutations in MSA (Dieringer & Schlötterer, 2003). Multiple testing correction, as incorporated within MSA was used to assess the 95% confidence level.

Population structure assessment of *S. salar* within each river and across time was assessed using the same methods above (STRUCTURE and DAPC analyses).

To test whether the populations from each of the five chalk stream rivers were structured through a pattern of isolation-by-distance, the genetic distance (F_{ST}/1-F_{ST}) (Rousset, 1997) was tested for significant correlations with geographic distance using a Mantel test in Genalex using 9999 permutations. Geographic distances (in km) were determined between river mouths along the coastal line of southern England using arcGIS v10 (ESRI, 2006).

In order to assess temporal stability, we calculated 'isolation-by-time' using a Mantel test for which a matrix of the difference in years between sampling was correlated with genetic distance (F_{ST}/1-F_{ST}). To further assess temporal stability, the genetic differentiation between sampling year and river was apportioned using an Analysis of Molecular Variance (AMOVA) in Arlequin v 3.5.2 (Excoffier & Lischer, 2010), using standard computations based on the number of different alleles (F_{ST}-like). Significance between the variance components (Va, Vb and Vc) and fixation indices (F_{CT}, F_{SC} and F_{ST}) were accepted at $p < 0.05$.

Results

Number of individuals and grouping of sites over years

In total, 1297 juvenile *S. salar* samples were genotyped at 16 microsatellites across 26 sites in the five chalk stream rivers (Supporting Information Table S1). Two potential *S. salar* x *S. trutta* hybrids were detected within the Frome and five were detected within the Avon. After the removal of hybrids and full

siblings, the final dataset was reduced to 772 samples (Table 1), which were used for all downstream analyses.

After applying the false discovery rate (FDR) correction, linkage disequilibrium was detected at seven out of a total of 3000 comparisons (data not shown). These indicated no consistent pattern between sample sites and, therefore, no loci were removed. Across the 26 sample sites, only two cases of loci not confirming to Hardy-Weinberg equilibrium were found, and therefore, no samples were removed.

Pairwise F_{ST} values between annual sample sites across each of the chalk stream populations were very low (average 0.025), ranging from -0.002 (between AVNbug04 and AVNbri12) to 0.063 (between FROcfmr09 and TESTmem10), and were significant in 278 of the 325 comparisons after multiple comparison corrections (Supporting Information Table S3). Moreover, despite significant F_{ST} values between many of the comparisons, a genetic signal of site differentiation could not be determined over the background of temporal variation in sampling. Furthermore, F_{ST} values for point samples such as these, particularly when sample sizes are small (Supporting Information Table S1), do not provide strong evidence for population differentiation. Accordingly, it was decided to group together sampling sites, irrespective of sampling year, for each river.

Genetic diversity of chalk stream S. salar

Between the year cohorts for each river, the number of alleles (N_A) genotyped in the juvenile *S. salar* from the chalk streams ranged from 6.38 (Piddle 2011) to 10.69 (Frome 2009) and the unbiased measure of allelic richness (A_R) ranged from 4.77 (Test 2010) to 5.47 (Test 2004) (Table 3). Expected heterozygosity (H_E) ranged from 0.66 (Test 2010) to 0.71 (Frome 2011), and observed heterozygosity (H_O) ranged from 0.67 (Itchen 2006 and Itchen 2010) to 0.73 (Piddle 2009 and Test 2004).

Statistical comparisons of diversity were non-significant for A_R ($p = 0.64$) and H_E ($p = 0.46$). However, there were significant differences in H_O ($p = 0.01$) and

F_{IS} (p = 0.008). Further analysis indicated that the differences in H_O were between *S. salar* in the Piddle and Itchen (p = 0.001), Avon and Itchen (p = 0.039) and Test and Itchen (p = 0.035). These differences were reflected in the statistical significance for F_{IS} between the Piddle and Itchen (p = 0.002) and the Test and Itchen (p = 0.008). The significance of these results is due to the relatively low H_O seen in the Itchen (especially in Itchen 2006 and Itchen 2010 cohorts), a pattern that is also reflected by higher values of F_{IS} for the Itchen (Table 3). The differences in F_{IS} suggest a greater amount of inbreeding within the Itchen, although this does not correlate with estimates of effective population size.

The Test 2010 showed evidence of the smallest effective population size (N_E) of 22 (95% CI: 18-27) and the highest N_E was observed in the Frome 2009 at 315 (95% CI: 249-419). Generally, estimates of N_E appeared stable over time, with the Frome showing the highest N_E, followed by the Avon. The Piddle, Test and Itchen showed relatively smaller values of N_E, with the exception of increases in N_E from the Test 2004, relative to the Test 2010, as well as a slight increases in Itchen 2010 N_E compared to Itchen 2005 and Itchen 2006.

Genetic uniqueness of chalk stream S. salar

Using the Δ*k* statistic, the most likely number of genetic clusters ascertained from the STRUCTURE analysis was *k=2*, which illustrated the genetic uniqueness of the five chalk steam *S. salar* populations compared to neighbouring non-chalk populations (Figure 2). Of interest was the genetic similarity of individuals from geographically distant regions on non-chalk geology, compared with the striking distinctiveness of the chalk stream salmon populations. Hierarchical analysis of the NW France, SW England and Norway group showed that the most likely number of genetic clusters was *k=2*, which demonstrated a difference between Norway and the two other non-chalk populations residing in NW France and SW England (Supporting Information Figure S1).

The optimum number of PCAs for the DAPC analysis was 34. Results of the population differentiation from the DAPC analysis complimented the

STRUCTURE analysis in also showing the genetic uniqueness of the chalk stream *S. salar* in comparison to all other non-chalk salmon included in this study. Due to the ability of DAPC in uncovering finer-scale hierarchical population structure (Jombart *et al.*, 2010), the Norwegian *S. salar* are observed as a separate genetic unit in the DAPC plot, which was also confirmed in the hierarchical analysis using STRUCTURE. One chalk stream individual from the Frome09 sampling cohort was shown to cluster with the NW France / SW England genetic group. This sample had no missing genotype data so this 'outlier' is most likely a real result (see Discussion).

Lack of population structure and temporal stability within the chalk streams

The global F_{ST} calculated across each annual cohort from each chalk stream population was low but significant (F_{ST} = 0.018, p = 0.001). The average pairwise F_{ST} across all comparisons was 0.028, and ranged from 0.002 (between Frome09 and Frome11) and 0.055 (between Piddle11 and Test04) (Table 4). All pairwise F_{ST} comparisons were significant after FDR correction (except between Piddle09 and Piddle11; Itchen06 and Piddle09; Avon04 and Avon12; Itchen05 and Itchen06; Itchen05 and Itchen10; Itchen06 and Itchen10).

Hierarchical analysis of the chalk stream *S. salar* showed that no significant genetic differentiation occurred across or between the five chalk stream rivers. The Δ*k* statistic showed no single reliable estimate for *k*, as the Δ*k* values were both low, and did not show an obvious peak for any value of potential genetic clusters (Supporting Information Figure S2). This therefore suggests that the chalk stream *S. salar* represent one large genetic group that is not distinguished on the basis of river basin or annual sampling (Figure 3). This was further supported by the DAPC, which could not distinguish any patterns of population differentiation (based on the optimum number of 42 PCAs).

The test for isolation-by-distance (IBD) across the chalk stream salmon was strong and statistically significant (R^2 = 0.2978, p = 0.031) (Figure 4A). This pattern of IBD was also noticeable in the STRUCTURE plot and DAPC.

Assessment of temporal stability using 'isolation-by-time' (IBT) showed no statistically significant relationship between annual cohorts within each river and geographical distances (R^2 = 0.0013, p = 0.422) (Figure 4B). Results from the AMOVA proportioned the majority of the variance (98%) within each sampling cohort (Vc = 5.22, F_{ST} = 0.0197, p < 0.05). Only 0.94% of the genetic variance occurred between rivers (Va = 0.05, F_{CT} = 0.00938, p < 0.05) and just 1.04% of the variance was attributed to between years within rivers (Vb = 0.06, F_{SC} = 0.01046, p < 0.05).

Discussion

Overview

Populations of *S. salar* within the chalk streams of southern England have plummeted in recent decades, yet despite this, and their distinction from other European populations, the genetic population structure of *S. salar* within the chalk streams had not previously been investigated. This study explicitly demonstrated the uniqueness of chalk stream populations in the context of *S. salar* from other non-chalk regions. A significant pattern of isolation-by-distance defines the chalk stream populations, and there is little to no genetic sub-structuring across rivers and across years. Furthermore, patterns in population structure and genetic diversity were shown to be temporally stable. Identification of the homogeneity of the chalk stream fish significantly increases our understanding of the contemporary genetic structure within one of the key reporting regions identified by Griffiths *et al.* (2010) for *S. salar* in the southern part of the species' range. These finding have significant implications for conservation and our understanding of population structure.

Uniqueness of chalk stream S. salar populations

The population structure analyses complement previous findings (Griffiths *et al.*, 2010; Ikediashi *et al.*, 2012) confirming that chalk stream *S. salar* are genetically distinct compared to populations from non-chalk geologies. Extensive analysis of *S. salar* populations from across Europe also confirms

the genetic uniqueness of chalk stream populations in southern England (Gilbey *et al.*, 2017). Interestingly, chalk stream populations appear to be genetically distinct even when compared to populations occupying south-west English rivers, between which a sharp gradient in underlying geology, from chalk to non-chalk, occurs. This is emphasised further by the relative genetic homogeneity of salmon from south-west England and north-west France, which are separated across the English Channel (representing a direct distance of >370km). Notably, even fish from considerably more distant non-chalk *S. salar* populations (Norway) are more genetically similar to English non-chalk stream fish than are the chalk stream *S. salar*.

Geology is known to be a fundamental feature affecting the distribution and abundance of salmonid populations. For example, rainbow trout *Oncorhynchus mykiss* (Walbaum 1792) and cutthroat trout *Oncorhynchus clarki* (Richardson 1837) abundances have been shown to be correlated with particular geologies (Hicks & Hall, 2003), and *S. trutta* condition was shown to decrease in limestone geologies correlated with increased catchment afforestation (Lehane *et al.*, 2004). With chemical cues being a particularly important feature of salmonid homing (Stabell *et al.*, 1984; Tierney *et al.*, 2010), distinctive population structure arising from geologies with especially notable water chemistry features is not surprising. Other research directly investigating the role of geology in *S. salar* population structure across Europe suggests a similar role of geology in structuring local and regional populations (Perrier *et al*., 2011). Despite the increasing appreciation of geological factors on the structuring of salmonid populations, genetic distinctiveness related specifically to geology is not common in the literature.

Furthermore, there appears to be little to no genetic admixture occurring between chalk stream *S. salar* populations and fish from neighbouring rivers. The proportion of straying in salmonids is known to be a significant contributor in re-colonisation events (Vasemägi *et al.*, 2001; Perrier *et al.*, 2009; Griffiths *et al.*, 2011; Ikediashi *et al.*, 2012). Moreover, high rates of straying have been shown to result in patterns of admixture between and among local salmonid populations within a region (Filatre *et al.*, 2003; Ayllon *et al.*, 2006b; King *et al.*, 2016). On the other hand, the potential of stocked fish ~~in~~ to swamp~~ing~~

local *S. salar* population structure is not frequently observed, with signals of low admixture between foreign and native genotypes (Finnegan *et al.*, 2008; Hansen *et al.*, 2009; Perrier *et al.*, 2013). The one exception to the apparent low rates of admixture in the chalk stream populations is the occurrence of a single chalk stream individual (genotyped from the Frome), which does not identify – based on its genetic profile – as 'chalk'. As the Frome is the most westerly of the chalk stream rivers, this fish could potentially represent a hybrid from a stray from south-west England crossed with a chalk stream individual. An alternative explanation is that the fish has been illicitly moved by human activity, although, if this were the case, in the short-term such activities might be expected to exhibit a more widespread exogenous signature.

Lack of genetic sub-structuring within the chalk stream S. salar populations

The accuracy of natal homing in salmonids is influenced by a plethora of biotic and abiotic factors (see Keefer & Caudill, 2014). In some cases, evidence of fine-scale natal homing appears high, for example in *S. trutta* populations across 3 km (Carlsson *et al.*, 1999) and in Chinook salmon *Oncorhynchus tshawytscha* (Walbaum 1792) across just 1 km (Neville *et al.*, 2006). On the other hand, Stewart *et al.* (2003) found that, despite phenotypic differences in sockeye salmon *Oncorhynchus nerka* (Walbaum 1792) populations homing to physically similar beaches in Alaska, USA, no evidence of restricted gene flow between the sites was detected. Similarly, genetic variation among *O. nerka* populations in the tributaries of a bay in Alaska were shown to be highly similar (Habicht *et al.* 2006), while relatively weak genetic structure was detected among Coho salmon *Oncorhynchus kisutch* (Walbaum 1792) from different river basins in Oregon (Johnson & Banks 2008).

In this study, the lack of genetic differentiation between chalk stream *S. salar* populations suggests that returning individuals may be homing back to a general chalk geological signature, and, consequently, fine-scale between-river population differentiation is not apparent. We anticipate that a propensity to home to chalk stream waters is likely a fundamental trait of these fish.

Collectively, the chalk stream rivers drain a small area (spanning just 70 km along the southern English coast), and it appears probable that homing accuracy of fish originating within the chalk geology is not further stratified by additional river-specific geochemical features. Furthermore, the chalk stream *S. salar* populations were shown to be temporally stable, which importantly, suggests habitat stability over time (see below).

A marked lack of differentiation across *S. salar* populations from proximal rivers has been noted previously in other parts of Britain. For example, populations in the rivers of north-west England and south-west Scotland that drain into the Solway Estuary (Griffiths *et al.*, 2010, Ikediashi *et al.*, 2012), show little if any consistent genetic differentiation, even when using a large panel of SNPs (Gilbey *et al.*, 2016). While geology may play a role in this scenario, this finding appears best explained by the fact that the rivers in this region share the estuary of the Solway Firth and the Irish Sea, through which returning fish must pass.

Despite a distinct lack of population differentiation between chalk stream *S. salar* populations, significant patterns of isolation-by-distance (IBD) were detected. Isolation-by-distance is prevalent in salmonids at both large continent scales (King *et al.*, 2001), regional scales (Taylor *et al.*, 2003) and within rivers (Griffiths *et al.,* 2009; Primmer *et al.*, 2006). Given the proximity of the river mouths and shared estuaries of the Frome/Piddle and Test/Itchen, higher levels of gene flow and migration between these sites might be expected, and it appears that the geographic distance between the mouths of these rivers does play a role in defining genetic distances between populations.

Temporal stability and chalk stream habitat reliability

The assessment of temporal stability is important in order to understand the extent to which populations exist as dynamic metapopulations punctuated by local extinctions and recolonisations, or in stable patches at gene flow-drift equilibrium. In an assessment of *S. salar* sampled across two consecutive

484 years across a ~100 km river in Quebec, temporal stability was evident in four
485 out of seven sampling sites, with a high proportion of genetic variance
486 attributable to other factors (Garant *et al.*, 2000). In a study assessing
487 temporal stability over a much longer timeframe (50-100 years), across five
488 rivers (ranging from 3 – 60 km) *S. trutta* populations were shown to be
489 remarkably temporally stable (Hansen *et al.*, 2002). Analysis of net samples
490 from two non-chalk rivers in England, showed temporal stability of the genetic
491 profiles of *S. salar* over more than 20 years (Griffiths *et al.*, 2010). This
492 suggests that the ability to detect temporal stability may depend in part on the
493 window from which the samples originate. Moreover, in order to avoid the
494 confounding effects of ascertainment bias, it is important in assessments of
495 temporal stability to sample the same locations over multiple years.

496 The samples used in this study spanned an intermediate timeframe (2004 –
497 2012; 8 years) and were stochastic in terms of sampling site and year. Our
498 results for isolation-by-time (IBT) showed no association between annual
499 sampling, and an AMOVA showed that both sampling between rivers and
500 between years within rivers accounted for only a very small proportion of the
501 explained variance. It should be noted, however, that although variance
502 between years within rivers was significant, it was only marginally higher than
503 variation between rivers. Due to challenges in obtaining samples, the
504 sampling regime in this study was far from ideal; to better address genetic
505 change over time, future assessment of the temporal stability of chalk stream
506 *S. salar* should sample the same sites across a set number of years.

507 Nonetheless, in the current study, measures of genetic variability were mostly
508 stable across years and diversity estimates of each cohort were comparable
509 to other assessments of *S. salar* using microsatellite markers (Tessier &
510 Bernatchez, 1999; Koljonen *et al.*, 2002; Skaala *et al.*, 2004). This is
511 particularly important given that chalk stream populations are known to have
512 decreased in recent decades. There were significant differences in F_{IS} and H_O
513 ($p<0.005$), which were primarily due to low observed heterozygosity and
514 higher levels of F_{IS} observed from fish in the Itchen. This may reflect differing
515 population dynamics within this river, with more inbreeding within it. However,
516 it is worth noting that the F_{IS} values from the Itchen are low compared to other

studies of *S. salar*. For example, F_{IS} values of 0.11 – 0.13 were found in populations in the Rivers Authie, Valmont and Touques in France (Perrier *et al.*, 2011); therefore, these values alone should not to be a cause for concern.

Given that the studied chalk streams are relatively short in length, estimates of N_E are comparable to estimates obtained from salmonids occupying similar river lengths (Lage & Kornfield, 2006; Jensen *et al.*, 2006; Vähä *et al.*, 2008), although it should be recognised that population dynamics and ecological features can substantially alter such estimates (e.g. Palstra *et al.*, 2007; 2009). One noticeable change was a dramatic drop in N_E in *S. salar* from the Test between 2004 and 2010. It is known that in the past there was a major stocking programme on the River Test and that stocking continued up until the year 2000 (L. Talks, Environment Agency, *pers. comm.*). Interestingly, despite stocking efforts which appear to have temporarily inflated estimates of N_E in this system, apparent effects on population structure and diversity (i.e. admixture effects of stocked fish) are not apparent. More recent estimates of N_E for the Test appear low, but relative decreases in genetic variability were not so apparent. The effects of this stocking activity were also observed in the population structure analyses, where the Test samples deviate in the DAPC, and also show higher Q values for cluster 2 (in blue) in the Structure plot. Evidence suggests that even in populations with small sizes and the potential for future declines, *S. salar* can continue to demonstrate relatively high genetic variability, as has been shown in this study, and in populations in Iberia (Consuegra *et al.*, 2005).

Finally, because *S. salar* typically show considerable variation in the age at which they migrate to sea, such patterns are hypothesised to significantly alter genetic variability and effective population size over time. However, the vast majority of chalk stream fish, at least from the Frome (98%), smolt after one year (R. Lauridsen, GWCT, *pers. comm.*). Future work on the populations assessed here could use ~~microsatellite~~ molecular analysis to determine the number of years that each generation of chalk stream *S. salar* spends between hatching and spawning, which varies considerably over the range of the species (e.g. Klemetsen *et al.*, 2003; Kusche *et al.*, 2017).

Further implications for conservation

The five chalk streams studied are currently managed following county borders and Environment Agency regional borders, so that the Frome, Piddle and Avon are managed within the region of Wessex, while the Test and Itchen are managed within the Solent and South Downs region. This management structure does not appear best suited with their natural population structure, as this study reveals a high degree of connectivity between *S. salar* across all five rivers. The demonstration of the distinctiveness of chalk stream *S. salar*, as well as the lack of sub-structuring between the chalk stream populations, reaffirms the need for bespoke management and conservation of these genetically distinctive fish.

Acknowledgements

We thank Chris Gardner (Environment Agency, Hampshire) for help with additional sample collection and fieldwork, and Patrick Hamilton (Exeter) and Rasmus Lauridsen (Game & Wildlife Conservation Trust) for comments on the manuscript. We also thank Sarah-Louise Selly and Andrew Griffiths for help with additional genotyping, and Tom Jenkins for assistance with DAPC analysis. This research was funded by the Game & Wildlife Conservation Trust and the Atlantic Salmon Trust, with additional support from the Salmon and Trout Association, the Westcountry Rivers Trust and the University of Exeter.

References

ArcGIS, E.S.R.I., 2010. Version 10. *Redlands (CA): ESRI*.

Ayllon, F., Martinez, J. & Garciavazquez, E. (2006a). Loss of regional population structure in Atlantic salmon, Salmo salar L., following stocking. *ICES Journal of Marine Science*, **63**, 1269–1273.

Ayllon, F., Moran, P., & Garcia-Vazquez, E. (2006b). Maintenance of a small anadromous subpopulation of brown trout (*Salmo trutta* L.) by straying. *Freshwater Biology*, **51**, 351–358.

Allendorf, F. & Phelps, S. (1981). Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1507–1514.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289-300.

Berrie, A.. (1992). The chalk-stream environment. *Hydrobiologia*, **248**, 3–9.

Carlsson, J., Olsen, K. H., Nilsson, J., Overli, O., & Stabell, O. B. (1999). Microsatellites reveal fine-scale genetic structure in stream-living brown trout. *Journal of Fish Biology*, **55**, 1290–1303.

Child, A., Burnell, A. & Wilkins, N. (1976). The existence of two races of Atlantic salmon (*Salmo salar* L.) in the British Isles. *Journal of Fish Biology*, **8,** 35–43.

Consuegra, S., Verspoor, E., Knox, D., & García de Leániz, C. (2005). Asymmetric gene flow and the evolutionary maintenance of genetic diversity in small, peripheral Atlantic salmon populations. *Conservation Genetics*, **6,** 823–842.

Dillane, E., Cross, M.C., Mcginnity, P., Coughlan, J.P., Galvin, P.T., Wilkins, N.P. & Cross, T.F. (2007). Spatial and temporal patterns in microsatellite DNA

599 variation of wild Atlantic salmon, *Salmo salar*, in Irish rivers. *Fisheries*
600 *Management and Ecology*, **14**, 209–219.

601 Dieringer, D., & Schlötterer, C. (2003) Microsatellite analyser (MSA): a platform
602 independent analysis tool for large microsatellite data sets. *Molecular Ecology*
603 *Notes*, **3**, 167-169.

604 Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R.
605 (2014). NeEstimator v2: re-implementation of software for the estimation of
606 contemporary effective population size (N_e) from genetic data. *Molecular*
607 *Ecology Resources*, **14**, 209–214.

608 Ellis, J.S., Gilbey, J., Armstrong, a, Balstad, T., Cauwelier, E., Cherbonnel, C.,
609 Consuegra, S., Coughlan, J., Cross, T.F., Crozier, W., Dillane, E., Ensing, D.,
610 García de Leániz, C., García-Vázquez, E., Griffiths, A.M., Hindar, K.,
611 Hjorleifsdottir, S., Knox, D., Machado-Schiaffino, G., McGinnity, P., Meldrup,
612 D., Nielsen, E.E., Olafsson, K., Primmer, C.R., Prodohl, P., Stradmeyer, L.,
613 Vähä, J.-P., Verspoor, E., Wennevik, V. & Stevens, J.R. (2011). Microsatellite
614 standardization and evaluation of genotyping error in a large multi-partner
615 research programme for conservation of Atlantic salmon (*Salmo salar* L.).
616 *Genetica*, **139**, 353–67.

617 Environment Agency. (2004). *The State of England's Chalk Rivers*. The Environment
618 Agency, England and Wales, Bristol BS32 4UD.

619 Estoup, A., Largiader, C.R., Perrot, E. & Chourrout, D. (1996). Rapid one-tube DNA
620 extraction for reliable PCR detection of fish polymorphic markers and
621 transgenes. *Molecular Marine Biology and Biotechnology*, **5**, 295–298.

622 Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of
623 individuals using the software STRUCTURE: a simulation study. *Molecular*
624 *Ecology*, **14**, 2611–20.

625 Excoffier, L. & Lischer, H.E.L. (2010) Arlequin suite ver 3.5: A new series of
626 programs to performpopulation genetics analyses under Linux and Windows.
627 *Molecular Ecology Resources*, **10**, 564-567.

628 Fillatre, E. K., Etherton, P., & Heath, D. D. (2003). Bimodal run distribution in a
629 northern population of sockeye salmon (*Oncorhynchus nerka*): life history and
630 genetic analysis on a temporal scale. *Molecular Ecology*, **12**, 1793–1805.

631 Finnegan A.K, Stevens J.R. (2008) Assessing the long-term genetic impact of
632 historical stocking events on contemporary populations of Atlantic salmon
633 (*Salmo salar* L.). *Fisheries Management and Ecology*, **15**, 315–326.

634 Finnegan, A.K., Griffiths, A.M., King, R.A., Machado-Schiaffino, G., Porcher, J.-P.,
635 Garcia-Vazquez, E., Bright, D. & Stevens, J.R. (2013). Use of multiple
636 markers demonstrates a cryptic western refugium and postglacial colonisation
637 routes of Atlantic salmon (*Salmo salar* L.) in Northwest Europe. *Heredity*, **111**,
638 34–43.

639 Fraser, D.J., Weir, L.K., Bernatchez, L., Hansen, M.M., & Taylor, E.B. (2011). Extent
640 and scale of local adaptation in salmonid fishes: review and meta-analysis.
641 *Heredity*, **106**, 404–20.

642 Garant, D., Dodson, J. & Bernatchez, L. (2000). Ecological determinants and
643 temporal stability of the within-river population structure in Atlantic salmon
644 (*Salmo salar* L.). *Molecular Ecology*, **9**, 615–628.

645 Garcia de Leaniz, C., Fleming, I.A., Einum, S., Verspoor, E., Jordan, W.C.,
646 Consuegra, S., Aubin-Horth, N., Lajus, D., Letcher, B.H., Youngson, A.F.,
647 Webb, J.H., Vøllestad, L. A, Villanueva, B., Ferguson, A & Quinn, T.P. (2007).
648 A critical review of adaptive genetic variation in Atlantic salmon: implications
649 for conservation. *Biological reviews of the Cambridge Philosophical Society*,
650 **82**, 173–211.

651 Gilbey, J., Cauwelier, E., Coulson, M. W., Stradmeyer, L., Sampayo, J. N.,
652 Armstrong, A., Middlemas, S. (2016). Accuracy of assignment of atlantic

653 salmon (*Salmo salar* L.) to rivers and regions in Scotland and northeast
654 England Based on Single Nucleotide Polymorphism (SNP) markers. *PLOS*
655 *ONE*, **11**, e0164327.

656 Gilbey J., Coughlan, J., Wennevik, V., Prodöhl, P. Stevens, J. R., Garcia de Leaniz,
657 C., Ensing, D., Cauwelier, E., Cherbonnel, C., Consuegra, S., Coulson, M. W.,
658 Cross, T. F., Crozier, W., Dillane, E., Ellis, J. S., García-Vázquez, E., Griffiths,
659 A. M., Gudjonsson, S., Hindar, K., Karlsson, S., Knox, D., Machado-
660 Schiaffino, G., Meldrup, D., Nielsen, E. E., Ólafsson, K., Primmer, C. R.,
661 Prusov, S., Stradmeyer, L., Vähä, J.-P., Veselov, A. J., Webster, L. M. I.,
662 McGinnity, P. & Verspoor, E. (2017) A microsatellite baseline for genetic
663 stock identification of European Atlantic salmon (*Salmo salar* L.). *ICES journal*
664 *of Marine Science*, fsx184, https://doi.org/10.1093/icesjms/fsx184.

665 Goudet, J. (1995). FSTAT (version 1.2): a computer program to calculate F-statistics.
666 *Journal of Heredity*, **86**, 485-486.

667 Grandjean, F., Verne, S., Cherbonnel, C. & Richard, A. (2009). Fine-scale genetic
668 structure of Atlantic salmon (*Salmo salar*) using microsatellite markers: effects
669 of restocking and natural recolonization. *Freshwater Biology*, **54**, 417–433.

670 Griffiths, A. M., Ellis, J. S., Clifton-Dey, D., Machado-Schiaffino, G., Bright, D.,
671 Garcia-Vazquez, E., & Stevens, J. R. (2011). Restoration versus
672 recolonisation: The origin of Atlantic salmon (*Salmo salar* L.) currently in the
673 River Thames. *Biological Conservation*, **144**, 2733–2738.

674 Griffiths, A. M., Koizumi, I., Bright, D., & Stevens, J. R. (2009). A case of isolation by
675 distance and short-term temporal stability of population structure in brown
676 trout (*Salmo trutta*) within the River Dart, southwest England. *Evolutionary*
677 *Applications*, **2**, 537–554.

678 Griffiths, A.M., Machado-Schiaffino, G., Dillane, E., Coughlan, J., Horreo, J.L.,
679 Bowkett, A.E., Minting, P., Toms, S., Roche, W., Gargan, P., McGinnity, P.,
680 Cross, T., Bright, D., Garcia-Vazquez, E. & Stevens, J.R. (2010). Genetic

681 stock identification of Atlantic salmon (*Salmo salar*) populations in the
682 southern part of the European range. *BMC Genetics*, **11**, 31.

683 Habicht, C., Seeb, L. W., & Seeb, J.E. (2006). Genetic and ecological divergence
684 defines population structure of sockeye salmon populations returning to
685 Bristol Bay, Alaska, and provides a tool for admixture analysis. *Transactions*
686 *of the American Fisheries Society*, **136**, 82–94.

687 Hansen, M.M., & Mensberg, K.-L.D. (2009). Admixture analysis of stocked brown
688 trout populations using mapped microsatellite DNA markers: indigenous trout
689 persist in introgressed populations. *Biology Letters*, **5**, 656–659.

690 Hansen, M.M., Ruzzante, D.E., Nielsen, E.E., Bekkevold, D., & Mensberg, K.-L.D.
691 (2002). Long-term effective population sizes, temporal stability of genetic
692 composition and potential for local adaptation in anadromous brown trout
693 (*Salmo trutta*) populations. *Molecular Ecology*, **11**, 2523–2535.

694 Hicks, B.J., & Hall, J.D. (2003). Rock type and channel gradient structure salmonid
695 populations in the Oregon Coast Range. *Transactions of the American*
696 *Fisheries Society*, **132**, 468–482.

697 Ikediashi, C., Billington, S. & Stevens, J.R. (2012). The origins of Atlantic salmon
698 (*Salmo salar* L.) recolonizing the River Mersey in northwest England. *Ecology*
699 *and Evolution*, **2**, 2537–48.

700 Jensen, L.F., Hansen, M.M., Carlsson, J., Loeschcke, V., & Mensberg, K.-L.D.
701 (2006). Spatial and temporal genetic differentiation and effective population
702 size of brown trout (*Salmo trutta*, L.) in small Danish rivers. *Conservation*
703 *Genetics*, ***6***, 615–621.

704 Johnson, M.A., & Banks, M.A. (2008). Genetic structure, migration, and patterns of
705 allelic richness among coho salmon (*Oncorhynchus kisutch*) populations of
706 the Oregon coast. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**,
707 1274–1285.

708 Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal
709 components: a new method for the analysis of genetically structured
710 populations. *BMC Genetics*, ***11***, 94.

711 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic
712 markers. *Bioinformatics*, **24**, 1403–1405.

713 Jones, O.R. & Wang, J. (2010). COLONY: a program for parentage and sibship
714 inference from multilocus genotype data. *Molecular Ecology Resources*, **10**,
715 551–5.

716 Jordan, W., Cross, T., Crozier, W.W., Ferguson, A., Galvin, P., Hurrel, R., McGinnity,
717 P., Martin, S., Moffet, I., Price, D., Youngson, A. & Verspoor, E. (2005).
718 Allozyme variation in Atlantic salmon from the British Isles: associations with
719 geography and the environment. *Journal of Fish Biology*, **67**, 146–168.

720 Keefer, M.L., & Caudill, C.C. (2014). Homing and straying by anadromous
721 salmonids: a review of mechanisms and rates. *Reviews in Fish Biology and*
722 *Fisheries*, **24**, 333–368.

723 King, T.L., Eackles, M.S. & Letcher, B.H. (2005). Microsatellite DNA markers for the
724 study of Atlantic salmon (*Salmo salar*) kinship, population structure, and
725 mixed-fishery analyses. *Molecular Ecology Notes*, **5**, 130–132.

726 King, T.L., Kalinowski, S.T., Schill, W.B., Spidle, A.P. & Lubinski, B.A. (2001).
727 Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide
728 perspective from microsatellite DNA variation. *Molecular Ecology*, **10**, 807–21.

729 King, R.A., Hillman, R., Elsmere, P., Stockley, B. and Stevens, J.R. (2016),
730 Investigating patterns of straying and mixed stock exploitation of sea trout,
731 *Salmo trutta,* in rivers sharing an estuary in south-west England. *Fisheries*
732 *Management and Ecology*, **23**, 376–389.

733 Klemetsen, A., Amundsen, P.A., Dempson, J.B., Jonsson, B., Jonsson, N.,
734 O'Connell, M.F., & Mortensen, E. (2003). Atlantic salmon *Salmo salar* L.,

735 brown trout *Salmo trutta* L. and Arctic charr S*alvelinus alpinus* (L.): a review of
736 aspects of their life histories. *Ecology of Freshwater Fish*, ***12***, 1–59.

737 Koljonen, M., Pella, J. & Masuda, M. (2005). Classical individual assignments versus
738 mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo*
739 *salar*) catches from DNA microsatellite data. *Canadian Journal of Fisheries*
740 *and Aquatic Sciences*, **62**, 2143–2158.

741 Koljonen, M.-L., Tähtinen, J., Säisä, M., & Koskiniemi, J. (2002). Maintenance of
742 genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding
743 programmes and the geographic distribution of microsatellite variation.
744 *Aquaculture*, **212**, 69–92.

745 Kusche, H., Côté, G., Hernandez, C., Normandeau, E., Boivin-Delisle, D., &
746 Bernatchez, L. (2017). Characterization of natural variation in North American
747 Atlantic Salmon populations (Salmonidae: *Salmo salar*) at a locus with a
748 major effect on sea age. *Ecology and Evolution*. DOI: 10.1002/ece3.3132.

749 Lage, C., & Kornfield, I. (2006). Reduced genetic diversity and effective population
750 size in an endangered Atlantic salmon (*Salmo salar*) population from Maine,
751 USA. *Conservation Genetics*, **7**, 91–104.

752 Lehane, B. M., Giller, P. S., O'Halloran, J., & Walsh, P. M. (2004). Relative
753 influences of catchment geology, land use and in-stream habitat on brown
754 trout populations in south-western Ireland. *Biology and Environment:*
755 *Proceedings of the Royal Irish Academy*, **104**, 43–54.

756 McConnell, S., Hamilton, L., Morris, D., Cook, D., Paquet, D., Bentzen, P. & Wright,
757 J. (1995). Isolation of salmonid microsatellite loci and their application to the
758 population genetics of Canadian east coast stocks of Atlantic salmon.
759 *Aquaculture*, **137**, 19–30.

760 Neville, H.M., Isaak, D.J., Dunham, J.B., Thurow, R.F., & Rieman, B.E. (2006). Fine-
761 scale natal homing and localized movement as shaped by sex and spawning

habitat in Chinook salmon: insights from spatial autocorrelation analysis of individual genotypes. *Molecular Ecology*, **15**, 4589–4602.

O'Reilly, P., Hamilton, L., McConnell, S. & Wright, J. (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*, **2298**, 2292–2298.

Palstra, F.P., O'Connell, M.F., & Ruzzante, D.E. (2009). Age structure, changing demography and effective population size in Atlantic salmon (*Salmo salar*). *Genetics*, **182,** 1233-1249.

Palstra, F.P., O'Connell, M.F., & Ruzzante, D.E. (2007). Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. *Molecular Ecology*, **16**, 4504–4522.

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

Peakall, R. & Smouse, P. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **1**, 6–8.

Perrier, C., Guyomard, R., Bagliniere, J.-L. & Evanno, G. (2011). Determinants of hierarchical genetic structure in Atlantic salmon populations: environmental factors vs. anthropogenic influences. *Molecular Ecology*, **20**, 4231–45.

Perrier, C., Baglinière, J.-L., & Evanno, G. (2013). Understanding admixture patterns in supplemented populations: a case study combining molecular analyses and temporally explicit simulations in Atlantic salmon. *Evolutionary Applications*, **6**, 218–230.

789 Perrier C, Evanno G, Belliard J, Guyomard R, Baglinière J-L. (2009). Natural
790 recolonization of the Seine River by Atlantic salmon (*Salmo salar*) of multiple
791 origins. *Can. J. Fish. Aquat. Sci.*, **67**,1–4.

792 Primmer, C.R., Veselov, A.J., Zubchenko, A., Poututkin, A., Bakhmet, I., & Koskinen,
793 M.T. (2006). Isolation by distance within a river system: genetic population
794 structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River
795 in northwest Russia. *Molecular Ecology*, **15**, 653–666.

796 Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure
797 using multilocus genotype data. *Genetics*, **155**, 945–59.

798 Raymond, M. & Rousset, F. (1995). GENEPOP (version 1.2): population genetics
799 software for exact tests and ecumenicism. *Journal of Heredity*, **86,** 248–249.

800 Rousset, F. (1997). Genetic differentiation and estimation of gene flow from F-
801 statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

802 Sánchez, J.A., Clabby, C., Ramos, D., Blanco, G., Flavin, F., Vázquez, E. & Powell,
803 R. (1996). Protein and microsatellite single locus variability in *Salmo salar* L.
804 (Atlantic salmon). *Heredity*, **77**, 423–32.

805 Skaala, Ø., Høyheim, B., Glover, K., & Dahle, G. (2004). Microsatellite analysis in
806 domesticated and wild Atlantic salmon (*Salmo salar* L.): allelic diversity and
807 identification of individuals. *Aquaculture*, **240**, 131–143.

808 Slettan, A., Olsaker, I. & Lie, Ø. (1995). Atlantic salmon, Salmo salar, microsatellites
809 at the SSOSL25, SSOSL85, SSOSL311, SSOSL417 loci. *Animal genetics*,
810 **26**, 281–282.

811 Stabell, O. (1984). Homing and olfaction in salmonids: a critical review with special
812 reference to Atlantic salmon. *Biological Reviews*, **59**, 333–388.

Stahl, G. (1987) Genetic population structure of Atlantic salmon. In Ryman, N. and Utter, F., eds. *Population Genetics and Fishery Management*. Seattle, WA: University of Washington Press, pp. 121–140.

Stewart, I.J., Quinn, T.P., & Bentzen, P. (2003). Evidence for fine-scale natal homing among island beach spawning sockeye salmon, *Oncorhynchus nerka*. *Environmental Biology of Fishes*, **67**, 77–85.

Taylor, E.B., Stamford, M.D., & Baxter, J.S. (2003). Population subdivision in westslope cutthroat trout (*Oncorhynchus clarki lewisi*) at the northern periphery of its range: evolutionary inferences and conservation implications. *Molecular Ecology*, **12**, 2609–2622.

Tessier, N., & Bernatchez, L. (1999). Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **8**, 169–179.

Tierney, K.B., Baldwin, D.H., Hara, T.J., Ross, P.S., Scholz, N.L., & Kennedy, C. J. (2010). Olfactory toxicity in fishes. *Aquatic Toxicology*, **96**, 2–26.

Tonteri, A., Veselov, A.J., Zubchenko, A. V., Lumme, J. & Primmer, C.R. (2009). Microsatellites reveal clear genetic boundaries among Atlantic salmon (*Salmo salar*) populations from the Barents and White seas, northwest Russia. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 717–735.

Truett, G., Heeger, P. & Mynatt, R. (2000). Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). *Biotechniques*, **29**, 52–54.

Vähä, J.-P., Erkinaro, J., Niemelä, E. & Primmer, C.R. (2008). Temporally stable genetic structure and low migration in an Atlantic salmon population complex: implications for conservation and management. *Evolutionary Applications*, **1**, 137–154.

840 Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004). Micro-
841 Checker: Software for identifying and correcting genotyping errors in
842 microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

843 Vasemagi, A., Gross, R., Paaver, T., Kangur, M., Nilsson, J., & Eriksson, L.-O.
844 (2001). Idengification of the origin of an Atlantic salmon (*Salmo salar* L.)
845 population in a recently recolonized river in the Baltic Sea. *Molecular Ecology*,
846 **10**, 2877–2882.

847 Vasemägi, A., Gross, R., Paaver, T., Koljonen, M.-L. & Nilsson, J. (2005). Extensive
848 immigration from compensatory hatchery releases into wild Atlantic salmon
849 population in the Baltic sea: spatio-temporal analysis over 18 years. *Heredity*,
850 **95**, 76–83.

851 Verspoor, E., Beardmore, J.A., Consuegra, S., García de Leániz, C., Hindar, K.,
852 Jordan, W.C., Koljonen, M.-L., Mahkrov, A.A., Paaver, T., Sánchez, J.A.,
853 Skaala, Ø., Titov, S. and Cross, T.F. (2005), Population structure in the
854 Atlantic salmon: insights from 40 years of research into genetic protein
855 variation. *Journal of Fish Biology*, **67**, 3–54.

856 Wickham, H. (2009). *ggplot2*: *Elegant Graphics for Data Analysis*. Springer-Verlag
857 New York.

858

Figure Legends

Figure 1. Map depicting the location of the five chalk stream rivers included in this study and sampling sites. Site codes correspond to those presented in Supporting Information Table S1.

Figure 2. STRUCTURE plot and DAPC of chalk stream *Salmo salar* compared to non-chalk *S. salar* from neighbouring regions of north-west (NW) France, south-west (SW) England, and Norway. Sampled rivers for these regions can be found in Tables 1 and 2. The most likely number of genetic units (*k*) is shown for the STRUCTURE plot (*k = 2*), which distinguishes the chalk stream *S. salar* genotypes as unique compared to non-chalk genotypes. DAPC also distinguishes the chalk stream *S. salar*, and also shows the genetic divergence between NW France/SW England and Norway.

Figure 3. STRUCTURE plot and DAPC of the five chalk stream *Salmo salar* rivers across multiple sampling years (Frome09, Frome11, Piddle09, Piddle11, Avon04, Avon10, Avon12, Test04, Test10, Itchen05, Itchen06 and Itchen10). No genetic groups were defined in the DAPC or STRUCTURE (*k = 2*) plot, but the analyses suggest a pattern of isolation-by-distance (IBD).

Figure 4. Evidence of spatial structuring and temporal stability in *Salmo salar* populations from across the five chalk stream rivers: (A) significant isolation-by-distance (IBD); (B) non-significant isolation-by-time (IBT).

Ethics Questionnaire (Completed)

Figure captions