Low but significant genetic differentiation underlies biologically meaningful phenotypic divergence in a large Atlantic salmon population

Tutku Aykanat[1], Susan E. Johnston [1,2], Panu Orell [3], Eero Niemelä [3], Jaakko Erkinaro [3], Craig R. Primmer [1].

[1] Division of Genetics and Physiology, Department of Biology, University of Turku, Itäinen Pitkäkatu 4, Turku FI-20520, Finland.

[2] Institute of Evolutionary Biology, University of Edinburgh, Charlotte Auerbach Road, Edinburgh, EH9 3FL, United Kingdom.

[3] Natural Resources Institute Finland, Utsjoki FI-99980, Finland

Keywords: Cryptic population structure, life-history variation, age at maturity, growth, SNPs, population genetics.

Corresponding author: Craig Primmer, Division of Genetics and Physiology, University of Turku, Turku FIN-20014, Finland. Tel. +358 2 333 5571; craig.primmer@utu.fi

Running Title: Cryptic genetic structuring in Atlantic salmon

Abstract

Despite decades of research assessing the genetic structure of natural populations, the biological meaning of low yet significant genetic divergence often remains unclear due to a lack of associated phenotypic and ecological information. At the same time, structured populations with low genetic divergence and overlapping boundaries can potentially provide excellent models to study the eco-evolutionary dynamics in cases where high resolution genetic markers and relevant phenotypic and life history information are available. Here, we combined SNP-based population inference with extensive phenotypic and life history data to identify potential biological mechanisms driving fine scale sub-population differentiation in Atlantic salmon (*Salmo salar*) from the Teno River, a major salmon river in Europe. Two sympatrically occurring sub-populations had low but significant genetic differentiation (F_{ST} = 0.018) and displayed marked differences in the distribution of life history strategies, including variation in juvenile growth rate, age at maturity and size within age classes. Large, late-maturing individuals were virtually absent from one of the two sub-populations and there were significant differences in juvenile growth rates and size-at-age after oceanic migration between individuals in the respective sub-populations. Our findings suggest that different eco-evolutionary processes affect each sub-population and that hybridization and subsequent selection may maintain low genetic differentiation without hindering adaptive divergence.

Introduction

Defining populations based on genetic markers has a long history in evolutionary biology (reviewed by Waples & Gaggiotti 2006). The emergence of each new type of molecular marker has seen new discoveries in the extent and scale at which genetic divergence is detected (reviewed by Avise 1994; Wright & Bentzen 1994; Morin *et al.* 2004; Schlotterer 2004). Most recently, studies using single-nucleotide polymorphisms (SNPs) have identified low but statistically significant genetic differentiation in a number of cases where populations were previously thought to be panmictic (O'Reilly *et al.* 2004; Ackerman *et al.* 2011; Zarraonaindia *et al.* 2012; Catchen *et al.* 2013; Garroway *et al.* 2013; Milano *et al.* 2014). Such information is frequently used as the basis for designing management and conservation plans, and in many cases may represent the only information available on population differences. However, the ecological meaning of low but significant genetic differentiation often remains unexplored (Waples & Gaggiotti 2006; Knutsen *et al.* 2011) and relative roles of adaptation, gene flow and the effects of the environment in shaping the genetic structure is not well understood. Likewise, genetically similar populations with dissimilar life histories and morphology may provide insights to the onset of ecological speciation and reproductive isolation (Hendry 2009). Such issues are particularly relevant when considering species or populations of conservation concern and/or harvested species as their interpretation can affect management strategies (Allendorf & Luikart 2007). Integrative approaches, where demographic and phenotypic information are simultaneously assessed alongside genetic analyses, are pivotal for establishing well founded basis for testing ecological-evolutionary hypotheses. However, such breadth of data is often lacking in non-model, wild systems.

Atlantic salmon (*Salmo salar*) is a species of both commercial importance and conservation concern (Verspoor *et al.* 2007). As a result, considerable population genetics research has been conducted on this species, with a variety of molecular markers at various geographic scales (King 2000; King *et al.* 2001; Nilsson *et al.* 2001; Consuegra *et al.* 2002; Verspoor *et al.* 2005; Tonteri *et al.* 2009; Perrier *et al.* 2011; Bourret *et al.* 2013a; Moore *et al.* 2014). Genetic diversity is generally partitioned hierarchically, starting at the continental, followed by basin and then river levels (King *et al.* 2007; Bourret *et al.* 2013a). However, genetic divergence within rivers has also been reported on a number of occasions, where population subdivision at tributary levels are likely to be maintained due to strong homing behaviour (i.e. restricted gene flow) of returning adults and sometimes also local adaptation to different

73 demes (Garant *et al.* 2000; Primmer *et al.* 2006; Dillane *et al.* 2007, 2008; Dionne *et al.* 2008;
74 Olafsson *et al.* 2014).

75 One of the clearest cases of genetic sub-structuring in wild Atlantic salmon within a river
76 basin has been reported in the Teno River, a large river system in northern Finland and
77 Norway. Microsatellite analyses have revealed surprisingly high levels of genetic divergence
78 across scales of tens of kilometres among tributaries, with average F_{ST} being around 0.1
79 (ranging from 0.015 to 0.201; Vähä *et al.* 2007). This divergence was shown to be temporally
80 stable and genetic diversity in the sub-populations was associated with life history variation
81 (Vähä *et al.* 2008). These findings support the notion that sub-populations may be locally
82 adapted. A more recent study using a medium density SNP chip (≈4,300 SNPs) identified
83 several sympatric subpopulation clusters within the river mainstem, with F_{ST} values at the
84 lower end of those earlier reported (F_{ST} < 0.0121, Johnston *et al.* 2014). Differences in the
85 distribution of age at maturity ("sea-age"- see below) between sub-population clusters were
86 detected, however, the study focussed on the sea-age phenotype only, and did not include
87 detailed analyses of sub-population structuring within the mainstem and thus the biological
88 significance of the cryptic population structuring remained unclear (Johnston *et al.* 2014).

89 Sea age at maturity and growth are heritable complex life-history traits closely linked to
90 fitness in salmonid fishes (Garant *et al.* 2003; Schaffer 2003; Hutchings 2011; Jonsson &
91 Jonsson 2011). The variation in these traits maintained within and among Atlantic salmon
92 populations are excellent targets for studying evolutionary trade-offs. For example, later
93 maturation at sea is associated with larger size, and therefore higher fecundity in females and
94 higher reproduction success in males, but comes with a cost of higher risk of mortality prior
95 to reproduction (Schaffer 2003). In addition, smaller tributaries with lower water levels are
96 more hospitable to smaller sized, earlier maturing fish, thus providing fitness advantages to
97 younger sea age fish in such tributaries (Garant *et al.* 2003; Niemelä et al. 2006). Likewise,
98 growth, which is inherently linked to several fitness metrics including maturation, survival,
99 and egg size, is likely to be under adaptive constraints associated with intraspecific
100 competition and predator avoidance during juvenile life-history phases (Reid & Peichel 2010;
101 Jonsson & Jonsson 2011), and genetic variation is maintained by context dependent
102 performance in different environments (Gillespie & Turelli 1989; Mackay *et al.* 2009, Reid *et*
103 *al.* 2012). On the other hand, the underlying genetic and environmental factors shaping
104 reproductive isolation and sea age variation between populations are not well understood

Thus, low genetic differentiation combined with substantial life-history variation within the Teno mainstem populations provides an excellent system for a detailed assessment of whether low but significant genetic differentiation is associated with biologically meaningful phenotypic divergence.

In this study, we build on the Teno River Atlantic salmon data-set reported in (Johnston *et al.* 2014) to identify the potential biological mechanisms associated with sympatric population divergence. First, we adopted a model-based Bayesian method to refine population structure inference, and subsequently elucidated the distribution of the inferred sub-populations throughout the river. Second, using a wealth of phenotypic and demographic information obtained from fishing records and scale measurements, we provide a detailed account of individual growth rates during different life history stages and demographic properties of each sub-population. Our results suggest that despite only subtle genetic divergence, the sub-populations harbour substantial, potentially adaptive, phenotypic divergence including differences in growth rates and size within age classes.

Materials and Methods

Study site and sample collection

The Teno River, located in far-north Europe (68–70°N, 25–27°E) runs between Finland and Norway, drains north into the Tana Fjord at the Barents Sea (Figure 1). It supports one of the world's largest wild Atlantic salmon populations, with up to 50000 individuals being harvested by local fishers and recreational fisheries annually (Johansen *et al.* 2008), accounting for up to 20% of the riverine Atlantic salmon catches in Europe (ICES 2013). A notable feature of the population is the extensive life-history variation observed: age at smoltification (i.e. age of outward migration to sea) varies between two and eight years while the time spent in the marine environment prior to maturation, also called sea-age, varies from one to five years with a proportion of individuals also returning to spawn a second or third time (Niemelä *et al.* 2006). This high diversity of age structure contributes to generally high temporal genetic stability in the system (Vähä *et al.* 2007). Scale samples of returning anadromous adult Atlantic salmon are routinely collected and fish length and weight are recorded by co-operating, trained, fishers within the system. Scales were consistently sampled from below the adipose fin and just above the lateral line (using standard guidelines provided by ICES 2011) and were dried and archived in paper envelopes by Natural

Resources Institute Finland (formerly known as the Finnish Game and Fisheries Research Institute). We used the scale sample set reported in Johnston *et al.* (2014) which consisted of fish that return to spawn following one or three to five consecutive winters spent at sea (sea winters, hereafter 1SW (N = 253) and 3SW (N = 283), respectively), and added samples with intermediate maturity time i.e. two sea winter fish (hereafter 2SW, N = 189). A small number of four sea winter (4SW, N= 18) and one five sea winter fish were grouped with the 3SW group (i.e. multi sea winter, MSW); these fish were excluded from growth trait analyses (see below). All fish had been captured along a ~130km stretch of the mainstem Teno River, reaching c. 190km from the sea (Figure 1) between 2001 and 2003. Sampling targeted fish captured during the last 4 weeks of the fishing season in August, which is 2-4 weeks after most individuals have entered the river (Erkinaro *et al*. 2010). As within-river migration to spawning grounds and exploratory movement beyond home spawning areas is limited during the sampling period (Økland *et al*. 2001; Karppinen *et al.* 2004), it is therefore likely that the sampling location is reflective of spawning region in the vast majority of cases. The genetic sex of each fish was determined using the protocol outlined in Yano *et al*. (2013).

Quantifying morphological and life history traits

We assessed a number of morphological and life history traits extrapolated from scale-derived measurements to determine the biological significance of fine-scale genetic structuring. Scale measures were conducted by trained technicians at the Natural Resources Institute Finland and age and growth rate were determined using the internationally agreed guidelines for Atlantic salmon scale reading (ICES 2011). Seasonal growth variation is reflected in the scale ring patterns, which are used to infer the age of fish (e.g. Friedland & Haas 1996). Likewise, inter-annuli distance (the scale growth between two adjacent annulus rings) is highly correlated to fish growth in the same period (e.g. r=0.96 for juvenile and ocean caught coho salmon (*Oncorhynchus kisutch*), Fisher & Pearcy 1990) and has long been used as proxy for growth rates (e.g. Pierce *et al.* 1996; Erkinaro *et al.* 1997). In the current data, the correlation between scale growth and adult size was high (Pearson's r = 0.92), and a similarly high correlation is observed for scale growth and juvenile size in a sample from the same river sytem (Pearson's r = 0.96 , Supp. figure 1). This high correlation between the scale growth and the phenotypes indicates that measurement error should not have a major effect on variance component analysis.

Growth indices were recorded for both the juvenile period (i.e. from the phase in fresh water prior to sea migration) and marine period (feeding phase at sea). In addition to age at smoltification (number of years spent in the fresh water prior to migration to the sea; *FW Age*), several juvenile growth indices were analysed: growth until the end of year one (*Growth*$_{FW1,}$ the radius of the scale from the focus to first year annulus), freshwater growth between year one and year two (*Growth*$_{FW2,}$ the radius of the scale from the first year annulus to second year annulus), freshwater growth between year two and year three (*Growth*$_{FW3,}$ the radius of the scale from the second year annulus to third year annulus), and total freshwater growth (*Growth*$_{FWtot}$, scale growth from focus until the end of freshwater growth zone, the point when fish migrates to the sea). In our dataset, all but one individual for which freshwater age data were available spent at least three years in fresh water, therefore *Growth*$_{FW1}$, *Growth*$_{FW2}$ and *Growth*$_{FW3}$ were common metrics for all but one sample. Marine phase indices were: sea age at first maturity (*SW Age*, number of winters spent at sea prior to first migration back to fresh water), first year growth at sea (*Growth*$_{SW1}$, the radius of the scale from the end of the freshwater growth to the first year summer annulus). *Growth*$_{SW1}$ was the only marine growth parameter that was common to all fish in the data-set. Two terminal traits recorded by the fishers were also included in the analysis; total length at capture (*Length*, i.e. length of the fish from the tip of the snout to the end of the tail) and weight at capture (*Weight*). We also measured body robustness by Fulton's condition factor at capture (CF = 100 x Weight x Length^{-3}; Ricker 1975). Phenotypic measurements were available for >90% of samples in all cases except for the yearly freshwater growth parameters (*Growth*$_{FW1}$, *Growth*$_{FW2,}$ and *Growth*$_{FW3}$), which were available for 77% of samples. This was because of the difficultly in confidently assigning annual rings (i.e. annulus) in the freshwater period, which are more prone to scale damage and regeneration of scales.

DNA extraction, sex determination and genotyping

DNA extraction, sex determination and SNP genotyping for all samples was carried out on individual archived scale samples using the same protocols described in Johnston *et al.* (2014). All 744 samples were genotyped at 5568 SNP loci using a custom-designed Illumina® iSelect SNP-array, the majority of which have been mapped to 29 linkage groups (Lien *et al.* 2011; Bourret *et al.* 2013b). Individual genotypes were scored using the clustering algorithm implemented in the Illumina® GenomeStudio Genotyping Analysis Module v2011.1. Samples with a call rate less than 0.98 were discarded from the analysis. A

SNP locus was filtered out if the call rate was less than 0.95, the minor allele frequency (MAF) was less than 0.05 and/or if the heterozygote excess/deficit was significant following false discovery rate adjustment (FDR=0.1), after which 684 individuals remained in the dataset. SNPs in high linkage disequilibrium (LD) were pruned using PLINK's pruning routine (command --*indep*), using window size=50, sliding window= 5, and variance inflation factor (VIF) = 1.11, the latter corresponding to multiple correlation coefficient of r^2=0.1 (Purcell *et al.* 2007). After the pruning step, 2874 SNPs and 685 individuals remained in the dataset. SNPs that were out of Hardy-Weinberg equilibrium were retained, since any population structure may result in HW disequilibrium.

Migrants from distant populations or undetected farmed aquaculture escapees (i.e. among individuals with missing scale growth parameters) were detected from the dataset by calculating pairwise allele sharing between samples using the *ibs* function of the GENABEL package v1.8.0 (Aulchenko *et al.* 2007) implemented in R v 3.1.0 (R Core Development Team 2012). Individuals with average allele sharing distances > 3.09 standard deviations from the median of the distribution (type I error rate probability = 0.001 assuming a normal distribution) were marked as outliers and removed from the analysis. Twenty three (3%) individuals were filtered out at this stage (Supp. figure 2) and a total of 662 individuals remained in the dataset (Supp. table 1).

Analysis of population structure

Population structure was inferred based on the 2874 SNP markers described above using STRUCTURE Unix version 2.3.3 (Pritchard *et al.* 2000), with 110000 MCMC runs and a burn-in length of 10000, using the correlated allele frequency method (Falush *et al.* 2003) and without defining prior population structure or location. Population structure was inferred by estimating the optimum number of clusters (K) as suggested by Pritchard & Wen (2004) and Evanno *et al.* (2005), in which the smallest K capturing the most structure is concluded as the optimum number of populations explaining the genetic data. K values ranged from one to seven, and each run with a particular K value was replicated 12 times. We then identified each individual's membership to inferred clusters using a cut off value of q=0.80 (probability of an individual belonging to a group), where q values were averaged over 12 replicated runs. The q=0.80 threshold is conservative for assigning individuals to populations (see Vähä & Primmer 2006), and also allows the distinction of some hybrid classes from pure-breds (e.g.

backcross hybrids are expected to have a q-value around 0.75; see below). Individuals not assigned to any population cluster (q-value < 0.80) were defined as "admixed".

Following population inference, Weir's and Cockerham's pairwise F_{ST} (Weir & Cockerham 1984) and within sub-population genetic diversity indices (i.e. observed and expected heterozygosity) were estimated within and among the inferred sub-populations using the HIERFSTAT package v0.04-10 (Goudet 2005) in R v3.1.0. Diversity indices of inferred sub-populations were compared with Kruskal-Wallis test.

Demographic and phenotypic properties of sub- populations

To evaluate genetic isolation by distance in the data-set, associations between individual-level genetic distances (i.e. allele sharing) and geographic distances (i.e. approximate river position) were assessed using a Mantel test, and significance was evaluated by permuting the data 10,000 times using the VEGAN package v2.0-10 in R v 3.1.0 (Oksanen *et al.* 2013). In addition to isolation by distance, we also tested for a possible isolation by region signature along the lower and the upper section of the mainstem, which are separated by a 40 km stretch of sandy river habitat that is generally unsuitable for salmon reproduction and nursery (Niemelä *et al.* 1999, Figure 1). Because of this, we also included a test of genetic isolation by region where genetic similarity of fish from the lower (< 140km) and the upper (> 180 km) stretches of the river were compared. A small number of fish sampled within this sandy region (3% of the final dataset) were excluded from this Mantel test. We constructed the distance matrix as follows: any two fish that were sampled in the same region were scored as "0" in the distance matrix (i.e. no distance between them), whereas fish that were not sampled in the same region were scored "1". Finally, we quantified the relative contribution of distance (km) vs sub-region (upper vs lower) effect in explaining the pairwise genetic distance between individuals. The two matrices (distance matrix vs sub-region matrix) are inherently confounded, thus we used a partial Mantel test to identify the relative contribution of each one, in which the correlation between the genetic distance matrix and either of the spatial matrices are conditioned on the other spatial matrices (using *mantel.partial* function in the VEGAN package v2.0-10). Significance was assessed at alpha value of 0.00625, after Bonferroni correction for multiple testing.

The among sub-population variation in continuous growth traits was evaluated using a linear mixed effect model, where parameters were estimated with maximum likelihood using the

LME4 package v1.1-7 in R v 3.1.0 (Bates 2010). The model included the sub-population of origin (as inferred by structure analysis at q = 0.8), *SW age*, *FW age*, and the genetically assigned sex as fixed effects, and year of sampling as a random effect. These covariates were chosen because they are either inherently or likely to be associated with the traits of interest. For example, *SW age* and sex are both strong predictors of sea growth, while *FW age* is a good predictor of freshwater growth and total size in the fresh water. The model was parametrically bootstrapped 10000 times using the *bootMer* function in LME4, from which the sampling median and 95% confidence interval of the parameters were calculated. Finally, the null hypothesis, that the parameter has no effect on the response variable, was evaluated at two alpha values, 0.05 and 0.001, which denote the proportion of (bootstrapped) parameter estimates with an opposite sign to the null. All phenotypic measurements other than *CF* were log scaled to achieve normality. In addition to the continuous traits, the two categorical traits *FW age* and *SW age* were tested for association with sub-population of origin, using a generalized linear model (Poisson error function and log link), where *SW age* was modelled as number of years that maturation was delayed beyond *SW age* = 1, otherwise with the same procedure as above. We then extended the phenotype analysis to assess a potential isolation barrier between the upper and lower sections of the river that are separated by a sandy stretch of river that is mostly unsuitable for spawning and juvenile rearing. Therefore, we re-formulated the above linear mixed effect by replacing the "sub-population" term with "sub-population and region" effect, where each sub-population and region combination was accounted as a categorical fixed effect in the model. Similar to the previous model, the parameter confidence intervals were estimated by parametric bootstrapping with 10000 permutations.

Genome wide association with phenotypes

Genome wide association studies (GWAS) were performed on all 10 phenotypic traits outlined above. Eight continuous traits were modelled using general linear models with a Gaussian error structure, fitting SNP genotypes and all covariates significantly associated with the response variable as fixed effects; two traits (FW Age and Sea Age) were modelled using Poisson function as the link, where *SW age* was modeled as number of years maturation was delayed beyond *SW age* = 1. A GWAS of 1SW vs 3-4SW individuals was conducted earlier (Johnston *et al.* 2014), however here a larger data-set including 2SW individuals and additional phenotypic traits was investigated. Population stratification was accounted for

293 either by including the significant principal components to the model as fixed effects, or
294 using genomic control whereby the test statistic was divided by the genomic inflation factor
295 (i.e. λ, Price *et al. 2010*). Principal components were added sequentially until the inflation
296 factor (lambda) was less than 1.1. The significance threshold for genome-wide association
297 after multiple testing at α= 0.05 was calculated using the Bonferroni method.

298 *Adaptive divergence among populations*

299 We evaluated the role of adaptive divergent selection among populations using a P_{ST}–F_{ST}
300 comparison (Brommer 2011). This is an extension of the Q_{ST}–F_{ST} framework, in which the
301 proportion of additive genetic contribution to population divergence is estimated within a
302 range of values to infer the robustness of the selection signal. This was determined using the
303 following equation:

304 $$P_{ST} = (c/h^2)*\sigma_{GB}^2/(c/h^2)*\sigma_{GB}^2 + 2\sigma_{GW}^2,$$

305

306 where σ_{GB}^2 and σ_{GW}^2 are the variances between and within each population, respectively (i.e.
307 residuals of the model); h^2 is heritability; and c is the proportion of the total variance that is
308 presumed to be due to additive genetic effects across populations (Leinonen *et al.* 2006;
309 Brommer 2011). We estimated the among population variation using a mixed model
310 approach, where significant covariates (as evaluated in the linear model above) were included
311 as fixed terms and population provenance as a random term using a restricted maximum
312 likelihood approach (REML) as implemented in the LME4 package v1.1-7 (Bates and
313 Maechler 2009) in R 3.0.2 (R Core Team). *FW Age* and *SW Age* were fitted using a
314 generalized model with a Poisson link, where *SW age* was modeled as number of years
315 maturation was delayed beyond *SW age* = 1. In this analysis, we included only individuals
316 that were confidently assigned to a population (q>0.8). Finally, models were bootstrapped
317 10000 times using the *bootMer* function in LME4 (with *use.u=T* option), from which the
318 confidence interval of the parameters were calculated. We calculated F_{ST} distribution by
319 performing an F_{ST} -outlier analysis in ARLEQUIN 3.5 (Excoffier *et al.* 2005, Beaumont &
320 Nichols 1996). The highest non-significant F_{ST} value at α = 0.05 was taken as the upper
321 threshold for the neutral expectations. In natural populations, the empirical values of c and h^2
322 are often unknown; therefore, we tested the robustness of P_{ST}–F_{ST} comparisons within a
323 specified range of c/h^2 ratios (0 to 2) as recommended by (Brommer 2011).

Estimating admixture between the inferred sub-populations

In order to gain further insight into the patterns of gene flow among sub-populations (e.g. Taylor 2003), we estimated the composition of different hybrid classes within the admixed individuals. To do this, we used the q-value of an individual as a proxy for its hybrid index (Vähä & Primmer 2006). First, we assessed the expected q-value distribution of different hybrid classes by simulating individuals using the empirical frequency distribution of inferred sub-populations. We simulated three different hybrid classes, assuming no linkage: 1) F_1 hybrids; 2) F_2 hybrids (i.e. F_1 x F_1); and 3) Backcross hybrids (F_1 x pure-bred sub-population 1 or 2). A baseline of pure type individuals (N = 400 for each population) was generated by sampling the observed allele frequency distributions (using genotypes inferred in the population structure analysis), and the population of origin for this group were marked *a priori* in the STRUCTURE analysis (using POPFLAG = 1). Next, 200 individuals from each hybrid class were simulated and q-value distributions were retrieved using STRUCTURE software using the same parameters as above. The q-value distributions of simulated hybrid classes were visually compared to the distribution of empirical q-values in order to infer the possible hybrid structure within the empirical data.

Results

Analysis of population structure

The STRUCTURE analysis showed a rapid increase in the log likelihood value from K=1 to K=2, followed by a plateau (Figure 2a), suggesting K=2 as the optimal number of sub-populations identified within the genetic data. This conclusion was also supported by the ΔK method of (Evanno *et al.* 2005), where ΔK was highest at K=2 (Supp. figure 3). Using a conservative q-value threshold of 0.80 (see Materials and Methods), 52% (N = 347) and 26% (N = 171) of individuals were assigned to the two main clusters, whereas 22% (*N* = 144) were assigned as admixed (Figure 2b, Supp. figure 4). Therefore, we refer to these two distinct sub-populations as “Sub-population 1”, and “Sub-population 2” hereafter, while the remaining samples are referred to as “admixed”. Individuals assigned to clusters in the STRUCTURE analysis also grouped together in the principle component analysis (PCA), where

353 the first two principle component (PC) explained 6.7 and 5.6% of the genetic variation
354 respectively (Figure 2c).

355 Expected and observed heterozygosity was marginally but significantly larger in Sub-
356 population 1 compared to Sub-population 2 (Kruskal-Wallis test, Table 1). Genetic
357 differentiation between the two sub-populations was F_{ST} = 0.018 (95% CI= 0.017 -0.019,
358 Weir and Cockerham's F_{ST}).

359 *Demographic and phenotypic properties of within and among the inferred sub-populations*

360 Fish from distinct sub-populations were not distributed evenly, nor grouped completely
361 separately along the sampled stretch of the mainstem. A higher proportion of Sub-population
362 1 fish was present in the lower Teno, and Sub-population 2 fish were more common in the
363 upper Teno (Figure 3). There was a marked change in the proportions of sub-populations
364 around the river stretch that is unsuitable for spawning after c. 130 km (Figure 3). There were
365 no significant differences in sampling time between populations, sea age or their interaction,
366 suggesting both populations, and the different sea age groups within them, are likely to have
367 similar spawning periods (Supp. table 2).

368 Individual-level isolation by distance (IBD) within Sub-population 1 revealed a marginal but
369 non-significant signal after the multiple test correction (Mantel's r= 0.063, p = 0.007; Table
370 2). Sub-population 2 showed slightly weaker IBD patterns (Mantel's r = 0.032, p = 0.020;
371 Table 2). The isolation by region analysis testing for genetic isolation between upper and
372 lower Teno mainstem samples was significant for Sub-population 1 (Mantel's r= 0.093, p =
373 0.002), but not for Sub-population 2 (Mantel's r = 0.036, p = 0.018). Partial Mantel tests, by
374 which confounded effects of linear distance and region on genetic distance were partitioned,
375 suggested that the genetic divergence in Sub-population 1 was driven primarily by restricted
376 gene flow between regions (Mantel's r= 0.075, p = 0.001, Table 2), but this was not the case
377 in Sub-population 2 (Mantel's r= 0.006, p = 0.455, Table 2), suggesting lack of divergence
378 between upper and lower Teno fish from Sub-population 2.

379 There were striking differences in the proportion of sea age classes assigned to each sub-
380 population and the sex-ratios within each population (Figure 4). Most 3SW fish were
381 assigned to Sub-population 1 (88% of 264 3SW fish) while only 11 (4%) were assigned to
382 Sub-population 2. Almost all 1SW fish assigned to Sub-population 1 were male (78 of 82

383 fish, 95%). In contrast, there were no apparent differences in the distribution of 2SW fish
384 between sub-populations (Figure 4). No difference in the freshwater age distribution was
385 observed between sub-populations, nor was there any association between sea age and
386 freshwater age (Table 3).

387 Continuous growth traits were also significantly different between sub-populations. Out of
388 nine growth/size traits measured, six showed significant differences between sub-populations
389 (Table 3, Figure 5). In general, freshwater growth rate was faster for Sub-population 2,
390 however, following the marine period, this was reversed and at the time of sampling, fish
391 from Sub-population 1 were significantly larger in length and weight, and had higher
392 condition factors than Sub-population 2 individuals (Figure 5, Table 3). For example, the
393 average weight differences between individuals of the same sea age classes from Sub-
394 populations 1 and 2 were 0.21 kg (11%) and 3 kg (34%) for 1SW and 3SW fish, respectively
395 (see Table 3 for parameters and log scale CIs). Sex was a significant determinant for growth
396 at sea traits, such that males grew more in the first year at sea ($Growth_{SW1}$) and were longer
397 and heavier at return (Table 3). Males had also grown more by the end of the freshwater
398 period ($Growth_{FWtot}$, Table 3). Finally, higher sea age at maturity (*SW Age*) was significantly
399 associated with slower freshwater growth ($Growth_{FW2}$ and $Growth_{FWtot}$; Table 3).

400 When sampling location was taken into account, we observed significant differences in the
401 freshwater growth trajectories between sub-population 1 individuals from the upper and
402 lower main-stem regions with higher growth in the upper region (i.e. $Growth_{FW2}$ and
403 $Growth_{FWtot}$ in Supp. figure 5). However, this fast growth appears to slow down in the first
404 year at sea (i.e. $Growth_{SW1}$) and both upper Teno and lower Teno Sub-population 1 attain
405 similar size at return (Supp. figure 5). Unlike Sub-population 1, Sub-population 2 fish
406 sampled in the upper and lower Teno exhibited similar growth both in the fresh water and in
407 the sea (Supp. figure 5).

408 *Genome wide association studies.*

409 None of the 2874 SNP loci showed a genome-wide significant association with any trait, after
410 correction for population stratification using the principal component method (Price *et al.*
411 2006). A number of SNPs were significant at a non-conservative alpha value of 0.01, but
412 allelic substitution effects of these SNPs did not explain phenotypic variation within sub-
413 populations, more than by chance alone (Supp. figure 6), further indicating that these loci are

likely false positives. The only exception was the condition factor, where 3.9% and 6.6% of phenotypic variation were explained by the top 28 significant SNPs in each sub-populations respectively ($p<0.01$), suggesting a small polygenic effect on condition factor can be explained by these SNPs (Supp. figure 7. See figure legend for details**)**. When using the genomic control method alone to account for population stratification, a significant association between several genome regions and *SW Age* was observed; this is consistent with the significant genome regions identified Johnston et al. (2014) when comparing 1SW and 3SW fish using the genomic control method, but not with correction using principle components or when modelling identity-by-state between individuals. However, as acknowledged in the previous study, effective population sizes within the Teno mainstem are high, whilst genome-wide levels of linkage disequilibrium are low. Therefore, we cannot rule out that absence of associations are due to low heritability and/or a polygenic basis of these traits, or if marker density and sample size are insufficient to capture variation at markers in strong linkage disequilibrium with causal variants (see discussion in Johnston et al. 2014).

Adaptive divergence among populations

The phenotypic differences between sub-populations in terminal traits including *Length*, *CF*, and *Sea Age*, were consistent with selection contributing to the divergence, whereby P_{ST} estimates and 95% CI of these traits were larger than neutral range, which was also robust across a wide range of c/h^2 values (Figure 6). P_{ST} estimates for these traits remained above the neutral range at c/h values as low as 0.5, suggesting that trait variances may be subjected to divergent selection even when the proportion of additive genetic affect among populations is half of the within population value (see Brommer 2011). Population variance between freshwater traits was not significantly different from neutral expectations, although median P_{ST} estimates for juvenile growth during later years in the river (i.e. $Growth_{FW3}$, and $Growth_{FWTot}$) was larger than the neutral range at higher c/h^2 values, weakly suggesting divergent selection may potentially influence these traits.

Population admixture between the inferred sub-populations

A substantial proportion of sampled fish (21.8%; Figure 2b and Figure 3) had intermediate q-values, suggesting that admixture in the system was common. The empirical q-value distribution of admixed fish was skewed towards Sub-population 1 which suggests genomes from admixed individuals contain a higher proportion of alleles from Sub-population 1

(Supp. figure 8). The relatively "flat" distribution of q-values suggests that the admixed individuals also include higher order hybrids (Supp. figure 8). On the other hand, the admixed group had high F_{IS}, which cannot be explained by inbreeding (i.e. overall high *Ho* of the group, see Table 1). However, a heterogeneous origin of populations within a group would elevate the F_{IS} signal, suggesting some fish in the admixed group may have origins other than the two sub-populations in the study, perhaps other sub-populations from other tributaries in the Teno river system.

Discussion:

We combined SNP-based sub-population inference with extensive phenotypic and life history data to obtain a detailed account of fine-scale population differentiation in Atlantic salmon from the mainstem of the Teno River, a major salmon river in Europe. Our results suggest that despite only subtle genetic divergence (F_{ST} = 0.018), the two sub-populations do indeed harbour substantial phenotypic divergence, including differences in life-history strategy structure, growth rates and size within age classes. Although both sub-populations inhabited overlapping sections of the river, Sub-population 2 appeared to have a broader range extending towards the upper Teno mainstem. This suggests that different eco-evolutionary processes may maintain divergence between these two genetically similar, overlapping sub-populations. Furthermore, strong signatures of adaptive divergence at sea, coupled with seemingly similar spawning timing and location leave open the possibility of a link between reproductive isolation and divergence at sea. In this discussion, we consider the potential processes that may be driving this population structuring, as well as the broader significance of the findings from both evolutionary and conservation management perspectives.

Partial reproductive isolation in sympatry: possible mechanisms

Detailed spatial analyses indicated that members of each sub-population were distributed throughout the main-stem of the river, suggesting that the two sub-populations occur in sympatry. Sympatric reproductive isolation has been reported for a number of salmonid fish populations. However, in the vast majority of these cases, reproductive isolation between populations is mediated by extensive dichotomy in life history variation: examples include anadromous vs resident strategies in Atlantic salmon (Verspoor & Cole 1989; Vuorinen & Berg 1989) and steelhead/rainbow trout, *O. mykiss* (Docker & Heath 2003; Narum *et al.* 2004; Pearse *et al.* 2009; Hecht *et al.* 2013); run timing variation in pink salmon, *O.*

gorbuscha (Gharrett *et al.* 2013); and freshwater (kokanee) and marine (sockeye) migrating populations of *O. nerka*, (Taylor 1999). Likewise, species pairs with diverged ecotypes, which may have overlapping breeding ranges, show discontinuous adaptive variation and strong genetic differentiation as a result of established post- and pre- zygotic reproductive isolation (Gislason *et al.* 1999; Taylor 1999; Saint-Laurent *et al.* 2003; Østbye *et al.* 2005; Landry *et al.* 2007; Hendry 2009; Power *et al.* 2009; Kapralova *et al.* 2011; May-McNally *et al.* 2015).

In comparison, the results reported here provide a novel case of phenotypic divergence between populations with very subtle genetic divergence, where gene flow between populations is restricted despite an overlapping breeding range, similar basic life histories (e.g. both sub-populations are anadromous) and similar spawning periods. The potential mechanisms maintaining the population structure are therefore less clear than in some earlier cases. Below, we consider potential pre- and post-zygotic isolation mechanisms that could potentially lead to the observed genetic and phenotypic divergence.

A potential pre-zygotic reproductive isolation mechanism is micro-geographic separation of spawning areas throughout the mainstem Teno River. It is known that breeding site preference in Atlantic salmon is partly driven by gravel size (Louhi *et al.* 2008), whereby areas with faster flowing water and larger gravel size are only accessible to larger females (Fleming & Einum 2010). Given that Sub-population 1 is essentially devoid of small, 1SW females, whereas Sub-population 2 almost completely lacks large 3SW females, size-assortative breeding site selection could provide the means for at least partial reproductive isolation on a micro-geographic scale. On the other hand, this argument does not explain the genetic divergence satisfactorily, since size-assortative breeding sites of females may not restrict gene flow via males. Moreover, gravel size is not known to be different in the upper and lower section of the mainstem (J. Erkinaro, *unpubl. data*).

Inference of possible post-zygotic reproductive mechanisms assumes that there is a fitness disadvantage for hybrid individuals (Turelli *et al.* 2001, Servedio & Noor 2003), which in turn requires the assumption that the two sub-populations are locally adapted. Although the relatively flat distribution of admixed q-values observed here suggests that the admixed fish can survive and reproduce for more than few generations, there is some circumstantial evidence that could provide a basis for post-zygotic isolation if the sub-populations are indeed locally adapted. Firstly, size at return from the marine migration is significantly

different between sub-populations, and consistent with adaptive divergence. For example, 3SW female fish from Sub-population 1 are ~9.9 kg in weight (N=108) compared to ~7.6 kg for the few 3SW fish from Sub-population 2 (N=9, see also Table 3 for parameters and log scale CI). Likewise, 2SW and 1SW fish from Sub-population 1 are about 2.0 kg (N = 65 and 63 for subpopulations) and 0.25 kg heavier (N = 69 and 97 for subpopulations), respectively (after adjusting for sex). In addition to size, condition factor is also significantly different between sub-populations, with fish from Sub-population 1 having a higher condition factor on return from the sea (Figure 5). This dramatic difference in size and condition of fish following the marine feeding phase could be explained either by the sub-populations exploiting different marine feeding grounds, or by differences in their efficiency to exploit the same feeding grounds. Very little is known about the marine feeding phases of most salmon populations (Haugland *et al.* 2006; Chaput 2012; MacKenzie *et al.* 2012), and thus this issue requires further research. Nevertheless, the pronounced size difference in returning adults provide a plausible post-zygotic isolation mechanism if the marine feeding strategy/behaviour of hybrids was sub-optimal, and therefore hybrids had lower survival compared to the pure-breds of either sub-population. The high P_{ST} values in these traits is also consistent with divergent selection in the marine environment (Figure 6) thus further supporting the significance of the marine habitat for population structure. .

Faster freshwater growth – earlier sea age at maturity:

Our results suggest Sub-population 1 was mostly confined to the lower Teno mainstem, while Sub-population 2, which seemingly performed poorer at sea, was inhabiting the entire sampling range of the mainstem. Intriguingly, even in the lower mainstem where individuals of the two sub-populations occur sympatrically, individuals of Sub-population 2 had higher growth in fresh water, suggesting the growth differences are not due to spatial geographical variation (Supp. figure 5). This variation in early growth may be explained through differing growth efficiency due to differential metabolic activity (Reid *et. al* 2012) or through behavioural differences between populations e.g. in feeding aggressiveness (Armstrong et al. 2003; Amundsen & Gabler 2008) or by within-river migration to nursery brooks for better growth opportunities (e.g. Erkinaro & Niemelä 1995). Temporal or microspatial variation in the environment, food availability and predation may maintain growth variation among populations (Amundsen & Gabler 2008,Ward *et al.* 2011, Reid *et. al* 2012, Jonsson & Jonsson 2011). On the other hand, P_{ST}-F_{ST} analysis indicated that, in general, divergence

between freshwater traits (other than third year fresh water growth ($Growth_{FW3}$; Figure 6) generally did not deviate from neutral expectations and therefore variation between the sub-populations may be explained by neutral processes alone.

It is also of interest to determine if freshwater growth properties may be mechanistically linked to sea age at maturity variation between sub-populations. Larger juvenile size in salmonids is associated with lower mortality (e.g. O'Connell & Ash 1993, Hutchings & Jones 1998, Grover 2005, Jonsson & Jonsson 2011). Therefore, higher freshwater growth of Sub-population 2 individuals may imply lower mortality both in fresh water and during the early marine phase, which predicts a younger age at maturity in Sub-population 2 compared to Sub-population 1 (e.g. Hutchings & Jones 1998, Schaffer 2003). A genetic basis for freshwater growth variation may result in differential optimum age structures in these sub-populations (e.g. Garant et al. 2003), and differences in migratory behaviour may further re-inforce post-zygotic isolation between them and help to maintain diversity and population structure within the mainstem. Neither genetic by environment interactions, nor the mechanistic basis of sea age variation is clearly understood in salmonids and therefore resolving this issue awaits further research.

Implications for conservation

Age at maturity is one of the key traits for the management of Atlantic salmon, as larger multi-sea winter fish are favoured in fisheries. In addition, older age at maturity within a population is correlated with higher genetic diversity and is therefore important for genetic stability of populations and maintaining ecosystem services (Vähä *et al.* 2007; Schindler *et al.* 2010). However, sea age structure is shifting towards younger age classes in many populations (Hansen & Quinn 1998; Niemelä *et al.* 2006; Friedland *et al.* 2009; Chaput 2012; Otero *et al.* 2012). The importance for conservation and management of preserving variation in sea-age within the Teno system has already been recognised (Vähä et al. 2007; Johnston et al. 2014). The results reported here build upon this by providing additional support for targeted preservation programmes, as well as the details necessary for their implementation. Although sea-age has been an obvious target, our assessment of additional phenotypic traits indicated that the phenotypic divergence between the two sub-populations extends beyond sea-age composition, with several growth parameters, including both freshwater and marine growth, differing significantly between sub-populations (Figure 5). Therefore, actions to preserve sea-age variation and/or both sub-populations will serve to preserve diversity in life-

history variation expressed during the marine and freshwater phases of the Atlantic salmon life cycle. Detailed population genetic analyses provide further information enabling the targeting of preservation of both sub-populations; for example, even though the two sub-populations occur sympatrically throughout the mainstem, sub-population 2 is more common in the upper reaches. Assessment of historical phenotypic proportions of the sub-populations, which is feasible via the long-term scale archive (Niemelä *et al.* 2006), may be warranted to determine if anthropogenic factors may have altered their life-history make-up and/or sub-population distribution over recent decades and if so, which potential solutions should be proposed.

More generally, our results further indicate that low but significant differentiation revealed by molecular markers can indeed be biologically meaningful, and such subtle, fine scale population differentiation may be overlooked without an integrated analysis of demographic, phenotypic and genetic data. As few within-river genetic studies on salmonids have been conducted with as many genetic markers as used here, it remains to be seen whether Teno River Atlantic salmon represent an exception for the occurrence of such fine scale differentiation in sympatry or whether these findings may be generalized to other large salmon river systems or even more broadly. Likewise, the system appears to be an excellent wild model to study the evolution of life history trade-offs and to improve our understanding of the dynamics of life history evolution both at population and meta-population levels.

Acknowledgements

We acknowledge the fishermen and women on the Teno River who contributed scales and phenotypic information to the Natural Resources Institute Finland. Scale analyses were carried out by Jari Haantie and Jorma Kuusela. The samples were prepared for SNP genotyping by Katja Salminen and SNP genotyping was conducted at the Centre for Integrative Genomics with the assistance of Matthew Kent and Sigbjørn Lien. Financial support was provided by the Academy of Finland (grants 272836 and 284941).

References

Ackerman MW, Habicht C, Seeb LW (2011) Single-Nucleotide Polymorphisms (SNPs) under Diversifying Selection Provide Increased Accuracy and Precision in Mixed-Stock Analyses of Sockeye Salmon from the Copper River, Alaska. *Transactions of the American Fisheries Society* **140**, 865-881.

604 Allendorf FW, Luikart G (2007) *Conservation and the genetics of populations* Blackwell
605 Pub., Malden, MA ; Oxford.
606 Amundsen PA, Gabler HM (2008) Food consumption and growth of Atlantic salmonSalmo
607 salarparr in sub-Arctic rivers: empirical support for food limitation and competition.
608 Journal of Fish Biology 73, 250-261.Anon. (2012) Status of the River Tana Salmon
609 Population. , Working group on salmon monitoring and research in the Tana River
610 system. http://www.mmm.fi/.
611 Aulchenko YS, Ripke S, Isaacs A, van Duijn CM (2007) GenABEL: an R library for
612 genome-wide association analysis. *Bioinformatics* **23**, 1294-1296.
613 Avise JC (1994) *Molecular markers, natural history and evolution* Chapman & Hall, New
614 York.
615 Bates DM (2010) lme4: Mixed-effects modeling with R. http://lme4.r-forge.r-
616 project.org/lMMwR/lrgprt.pdf
617 Bourret V, Dionne M, Kent MP, Lien S, Bernatchez L. (2013a). Landscape Genomics in
618 Atlantic Salmon (*Salmo salar*): searching for gene-environment interactions driving
619 local adaptation. *Evolution*. **67**, 3469-3487.
620 Bourret V, Kent MP, Primmer CR, *et al.* (2013) SNP-array reveals genome-wide patterns of
621 geographical and potential adaptive divergence across the natural range of Atlantic
622 salmon (*Salmo salar*). *Molecular Ecology* **22**, 532-551.
623 Catchen J, Bassham S, Wilson T, *et al.* (2013) The population structure and recent
624 colonization history of Oregon threespine stickleback determined using restriction-site
625 associated DNA-sequencing. *Molecular Ecology* **22**, 2864-2883.
626 Chaput G (2012) Overview of the status of Atlantic salmon (*Salmo salar*) in the North
627 Atlantic and trends in marine mortality. *ICES Journal of Marine Science* **69**, 1538-
628 1548.
629 Consuegra S, Garcia de Leaniz C, Serdio A, *et al.* (2002) Mitochondrial DNA variation in
630 Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular*
631 *Ecology* **11**, 2037-2048.
632 Dillane E, Cross MC, McGinnity P, *et al.* (2007) Spatial and temporal patterns in
633 microsatellite DNA variation of wild Atlantic salmon, *Salmo salar*, in Irish rivers.
634 *Fisheries Management and Ecology* **14**, 209-219.
635 Dillane E, McGinnity P, Coughlan JP, *et al.* (2008) Demographics and landscape features
636 determine intrariver population structure in Atlantic salmon (*Salmo salar* L.): the case
637 of the River Moy in Ireland. *Molecular Ecology* **17**, 4786-4800.
638 Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Comparative survey of within-river
639 genetic structure in Atlantic salmon; relevance for management and conservation.
640 *Conservation Genetics* **10**, 869-879.
641 Docker MF, Heath DD (2003) Genetic comparison between sympatric anadromous steelhead
642 and freshwater resident rainbow trout in British Columbia, Canada. *Conservation*
643 *Genetics* **4**, 227-231.
644 Erkinaro J, Dempson JB, Julkunen M, Niemelä E (1997) Importance of ontogenetic habitat
645 shifts to juvenile output and life history of Atlantic salmon in a large subarctic river:
646 an approach based on analysis of scale characteristics. *Journal of Fish Biology* **51**,
647 1174-1185.
648 Erkinaro J, Niemelä E (1995) Growth differences between the Atlantic salmon parr, *Salmo*
649 *salar*, of nursery brooks and natal rivers in the river teno watercourse in Northern
650 Finland. *Environmental Biology of Fishes* **42**, 277-287.
651 Erkinaro J, Niemelä E, Vähä J-P, *et al.* (2010) Distribution and biological characteristics of
652 escaped farmed salmon in a major subarctic wild salmon river: implications for
653 monitoring. *Canadian Journal of Fisheries and Aquatic Sciences* **67**, 130-142.

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

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

Fisher JP, Pearcy WG (1990) Spacing of scale circuli versus growth-rate in young coho salmon. *Fishery Bulletin* **88**, 637-643.

Fleming IA, Einum S (2010) Reproductive Ecology: A Tale of Two Sexes. In: *Atlantic Salmon Ecology*, pp. 33-65. Wiley-Blackwell.

Friedland KD, Haas RE (1996) Marine post-smolt growth and age at maturity of Atlantic salmon. *Journal of Fish Biology* **48**, 1-15.

Friedland KD, MacLean JC, Hansen LP, *et al.* (2009) The recruitment of Atlantic salmon in Europe. *ICES Journal of Marine Science* **66**, 289-304.

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

Garant D, Dodson JJ, Bernatchez L (2003) Differential reproductive success and heritability of alternative reproductive tactics in wild Atlantic salmon (*Salmo salar* L.). *Evolution* **57**, 1133-1141.

Garroway CJ, Radersma R, Sepil I, *et al.* (2013) Fine-scale genetic structure in a wild bird population: The role of limited dispersal and environmentally based selection as causal factors. *Evolution* **67**, 3488-3500.

Gharrett AJ, Joyce J, Smoker WW (2013) Fine-scale temporal adaptation within a salmonid population: mechanism and consequences. *Molecular Ecology* **22**, 4457-4469.

Gillespie JH, Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. Genetics 121, 129-138.

Gislason D, Ferguson M, Skulason S, Snorrason SS (1999). Rapid and coupled phenotypic and genetic divergence in Icelandic Arctic char (*Salvelinus alpinus*). *Canadian Journal of Fisheries and Aquatic Sciences* **56**, 2229-2234.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5**, 184-186.

Grover MC (2005) Changes in size and age at maturity in a population of kokanee *Oncorhynchus nerka* during a period of declining growth conditions. *Journal of Fish Biology* **66**, 122-134.

Hansen LP, Quinn TR (1998) The marine phase of the Atlantic salmon (*Salmo salar*) life cycle, with comparisons to Pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 104-118.

Haugland M, Holst J, Holm M, Hansen L (2006) Feeding of Atlantic salmon (*Salmo salar* L.) post-smolts in the Northeast Atlantic. *ICES Journal of Marine Science* **63**, 1488-1500.

Hecht BC, Campbell NR, Holecek DE, Narum SR (2013) Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology* **22**, 3061-3076.

Hendry AP (2009) Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences* **66**, 1383-1398.

Hutchings JA (2011) Old wine in new bottles: reaction norms in salmonid fishes. *Heredity (Edinb)* **106**, 421-437.

Hutchings JA, Jones MEB (1998) Life history variation and growth rate thresholds for maturity in Atlantic salmon, *Salmo salar*. *Canadian Journal of Fisheries and Aquatic Sciences* **55**, 22-47.

ICES (2011) Report of the workshop on age determination of salmon (WKADS), Galway, Ireland.

ICES (2013) Report of the Working group on North Atlantic salmon (WGNAS), 3-12 April, 2013. In: *ICES CM 2013/ACOM:09, pp376.*, Copenhagen, Denmark.

Johansen M, Erkinaro J, Niemelä E, *et al.* (2008) Atlantic salmon monitoring and research in the Tana river system. http://www.tenojoki.fi/.

Johnston SE, Orell P, Pritchard VL, *et al.* (2014) Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*.

Jonsson N, Hansen LP, Jonsson B (1991) Variation in age, size and repeat spawning of adult Atlantic salmon in relation to river discharge. *Journal of Animal Ecology* **60**, 937-947.

Jonsson B, Jonsson N (2011). *Ecology of Atlantic salmon and brown trout : habitat as a template for life histories*. Dordrecht, Springer.

Kapralova KH, Morrissey MB, Kristjánsson BK *et al.* (2011) Evolution of adaptive diversity and genetic connectivity in Arctic charr (*Salvelinus alpinus*) in Iceland. *Heredity*, **106**, 472–487.

Karppinen P, Erkinaro J, Niemelä E, Moen K, Økland F (2004) Return migration of one-sea-winter Atlantic salmon in the River Tana. *Journal of Fish Biology* **64**, 1179-1192.

King T (2000) Mitochondrial DNA diversity in North American and European Atlantic salmon with emphasis on the Downeast rivers of Maine. *Journal of Fish Biology* **57**, 614-630.

King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA (2001) Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology* **10**, 807-821.

King TL, Verspoor E, Spidle AP, *et al.* (2007) Biodiversity and Population Structure. In: *The Atlantic Salmon: Genetics, Conservation and Management*, pp. 117-166. Blackwell Publishing Ltd.

Knutsen H, Olsen EM, Jorde PE, *et al.* (2011) Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Molecular Ecology* **20**, 768-783.

Landry L, Vincent WF, Bernatchez L (2007) Parallel evolution of lake whitefish dwarf ecotypes in association with limnological features of their adaptive landscape. *Journal of Evolutionary Biology*, **20**, 971–984.

Lien S, Gidskehaug L, Moen T, *et al.* (2011) A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics* **12**, 615.

Lorenzen K, Beveridge MC, Mangel M (2012) Cultured fish: integrative biology and management of domestication and interactions with wild fish. Biol Rev Camb Philos Soc 87, 639-660.

Louhi P, Mäki-Petäys A, Erkinaro J (2008) Spawning habitat of Atlantic salmon and brown trout: general criteria and intragravel factors. River Research and Applications 24, 330-339.

Mackay TF, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. Nat Rev Genet 10, 565-577.

MacKenzie KM, Trueman CN, Palmer MR, *et al.* (2012) Stable isotopes reveal age-dependent trophic level and spatial segregation during adult marine feeding in populations of salmon. *ICES Journal of Marine Science* **69**, 1637-1645.

May-Mcnally SL, Quinn TP, Woods PJ, Taylor EB (2015) Evidence for genetic distinction among sympatric ecotypes of Arctic char (Salvelinus alpinus) in south-western Alaskan lakes. *Ecology of Freshwater Fish* (online early).

753 Milano I, Babbucci M, Cariani A, *et al.* (2014) Outlier SNP markers reveal fine-scale genetic
754 structuring across European hake populations (*Merluccius merluccius*). *Molecular*
755 *Ecology* **23**, 118-135.
756 Moore JS, Bourret V, Dionne M, *et al.* (2014) Conservation genomics of anadromous
757 Atlantic salmon across its North American range: outlier loci identify the same
758 patterns of population structure as neutral loci. *Mol Ecol* **23**, 5680-5697.
759 Morin PA, Luikart G, Wayne RK, *et al.* (2004) SNPs in ecology, evolution and conservation.
760 *Trends in Ecology and Evolution* **19**, 208-216.
761 Narum, Talbot, Powell, Contor (2004) Genetic divergence of sympatric resident and
762 anadromous forms of *Oncorhynchus mykiss* in the Walla Walla River, U.S.A. *Journal*
763 *of Fish Biology* **65**, 471-488.
764 Niemelä E, Erkinaro J, Julkunen M, *et al.* (2006) Temporal variation in abundance, return
765 rate and life histories of previously spawned Atlantic salmon in a large subarctic river.
766 *Journal of Fish Biology* **68**, 1222-1240.
767 Niemelä E, Julkunen M, Erkinaro J (1999) Densities of the juvenile Atlantic salmon (*Salmo*
768 *salar* L) in the subarctic Teno River watercourse, northern Finland. *Boreal*
769 *Environment Research*.
770 Nilsson J, Gross R, Asplund T, *et al.* (2001) Matrilinear phylogeography of Atlantic salmon
771 (*Salmo salar* L.) in Europe and postglacial colonization of the Baltic Sea area.
772 *Molecular Ecology* **10**, 89-102.
773 O'Connell MF, Ash EGM (1993) Smolt size in relation to age at first maturity of Atlantic
774 salmon (*Salmo salar*): the role of lacustrine habitat. *Journal of Fish Biology* **42**, 551-
775 569.
776 O'Reilly PT, Canino MF, Bailey KM, Bentzen P (2004) Inverse relationship between F and
777 microsatellite polymorphism in the marine fish, walleye pollock (*Theragra*
778 *chalcogramma*): implications for resolving weak population structure. *Molecular*
779 *Ecology* **13**, 1799-1814.
780 Økland F, Erkinaro J, Moen K, *et al.* (2001) Return migration of Atlantic Salmon in the River
781 Tana: Phases of migratory behaviour. *Journal of Fish Biology* **59**, 862-874.
782 Olafsson K, Pampoulie C, Hjorleifsdottir S, Gudjonsson S, Hreggvidsson GO (2014) Present-
783 day genetic structure of Atlantic salmon (*Salmo salar*) in Icelandic rivers and ice-cap
784 retreat models. *PLoS One* **9**, e86809.
785 Oksanen J, Blanchet GF, Kindt R, *et al.* (2013). vegan: Community Ecology Package. R
786 package version 2.0-10. http://CRAN.R-project.org/package=vegan
787 Otero J, Jensen AJ, L'Abee-Lund JH, *et al.* (2012) Contemporary ocean warming and
788 freshwater conditions are related to later sea age at maturity in Atlantic salmon
789 spawning in Norwegian rivers. *Ecology and Evolution* **2**, 2192-2203.
790 Østbye K, Naesje TF, Bernatchez L, Sandlund OT, Hindar K (2005). Morphological
791 divergence and origin of sympatric populations of European whitefish (*Coregonus*
792 *lavaretus L.*) in Lake Femund, Norway. *Journal of Evolutionary Biology* **18**, 683-702.
793 Pearse DE, Hayes SA, Bond MH et al. (2009) Over the falls? Rapid evolution of ecotypic
794 differentiation in steelhead/rainbow trout (*Oncorhynchus mykiss*). *Journal of*
795 *Heredity*, **100,** 515–525.
796 Perrier C, Guyomard R, Bagliniere JL, Evanno G (2011) Determinants of hierarchical genetic
797 structure in Atlantic salmon populations: environmental factors vs. anthropogenic
798 influences. *Mol Ecol* **20**, 4231-4245.
799 Pierce CL, Rasmussen JB, Leggett WC (1996) Back-calculation of fish length from scales:
800 Empirical comparison of proportional methods. *Transactions of the American*
801 *Fisheries Society* **125**, 889-898.

802 Power M, Power G, Reist JD, Bajno R (2009). Ecological and genetic differentiation among
803 the Arctic charr of Lake Aigneau, Northern Québec. *Ecology of Freshwater Fish* **18**,
804 445-460.
805 Price AL, Patterson NJ, Plenge RM, et al. (2006) Principal components analysis corrects for
806 stratification in genome-wide association studies. Nat Genet 38, 904-909.
807 Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population
808 stratification in genome-wide association studies. Nat Rev Genet 11, 459-
809 463.Primmer CR, Veselov AJ, Zubchenko A, *et al.* (2006) Isolation by distance
810 within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*,
811 in tributaries of the Varzuga River in northwest Russia. *Molecular Ecology* **15**, 653-
812 666.
813 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
814 multilocus genotype data. *Genetics* **155**, 945-959.
815 Pritchard JK, Wen W (2004) Documentation for structure software: version 2, Department of
816 Human Genetics, University of Chicago, Chicago. *http://pritchardlab.stanford.edu/*
817 Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: a tool set for whole-genome
818 association and population-based linkage analyses. *The American Journal of Human*
819 *Genetics* **81**, 559-575.
820 Reid DT, Peichel CL (2010) Perspectives on the genetic architecture of divergence in body
821 shape in sticklebacks. Integr Comp Biol 50, 1057-1066.
822 Reid D, Armstrong JD, Metcalfe NB (2012) The performance advantage of a high resting
823 metabolic rate in juvenile salmon is habitat dependent. J Anim Ecol 81, 868-
824 875.Ricker WE (1975) Computation and interpretation of biological statistics of fish
825 populations. *Bulletin of the Fisheries Research Board of Canada* **191**, 1-382.
826 Schaffer W (2003) Life histories, evolution, and salmonids. In: *Evolution illuminated:*
827 *Salmon and Their Relatives* (eds. Hendry AP, Stearns SC), pp. 20-51. Oxford
828 University Press.
829 Schindler DE, Hilborn R, Chasco B, *et al.* (2010) Population diversity and the portfolio effect
830 in an exploited species. *Nature* **465**, 609-612.
831 Schlotterer C (2004) The evolution of molecular markers-just a matter of fashion? *Nature*
832 *Reviews Genetics* **5**, 63-69.
833 Saint-Laurent R, Legault M, Bernatchez L (2003). Divergent selection maintains adaptive
834 differentiation despite high gene flow between sympatric rainbow smelt ecotypes
835 (Osmerus mordax Mitchill). *Molecular Ecology* **12**, 315-330.
836 Servedio MR, Noor MAF (2003) The role of reinforcement in speciation: Theory and Data.
837 Annual Review of Ecology, Evolution, and Systematics 34, 339-364.
838 Taylor EB (1999) Species pairs of north temperate freshwater fishes: Evolution, taxonomy,
839 and conservation. *Reviews in Fish Biology and Fisheries* **9**, 299-324.
840 Taylor EB (2003) Evolution in mixed company: evolutionary inferences from studies of
841 natural hybridization in Salmonidae. In: *Evolution illuminated: Salmon and Their*
842 *Relatives* (eds. Hendry AP, Stearns SC). Oxford University Press.
843 Tonteri A, Veselov AJ, Zubchenko AV, Lumme J, Primmer CR (2009) Microsatellites reveal
844 clear genetic boundaries among Atlantic salmon (*Salmo salar*) populations from the
845 Barents and White seas, northwest Russia. *Canadian Journal of Fisheries and*
846 *Aquatic Sciences* **66**, 717-735.
847 Turelli M, Barton NH, Coyne JA (2001) Theory and speciation. *Trends in Ecology and*
848 *Evolution* **16**, 330-343.
849 Vähä JP, Erkinaro J, Niemelä E, Primmer CR (2007) Life-history and habitat features
850 influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology* **16**,
851 2638-2654.

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

Vähä JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology* **15**, 63-72.

Verspoor E, Beardmore JA, Consuegra S, *et al.* (2005) Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. *Journal of Fish Biology* **67**, 3-54.

Verspoor E, Cole LJ (1989) Genetically distinct sympatric populations of resident and anadromous Atlantic salmon, *Salmo salar*. *Canadian Journal of Zoology-Revue Canadienne de Zoologie* **67**, 1453-1461.

Verspoor E, Stradmeyer L, Nielsen JL (2007) *The Atlantic Salmon: Genetics, Conservation and Management* Wiley.

Vincent B, Dionne M, Kent MP, Lien S, Bernatchez L (2013) Landscape genomics in Atlantic salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. *Evolution* **67**, 3469-3487.

Vuorinen J, Berg OK (1989) Genetic divergence of anadromous and nonanadromous Atlantic Salmon (*Salmo salar*) in the River Namsen, Norway. *Canadian Journal of Fisheries and Aquatic Sciences* **46**, 406-409.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

Waples RS, Gustafson RG, Weitkamp LA, *et al.* (2001) Characterizing diversity in salmon from the Pacific Northwest. *Journal of Fish Biology* **59**, 1-41.

Ward DM, Nislow KH, Folt CL (2011) Seasonal shift in the effects of predators on juvenile Atlantic salmon (Salmo salar) energetics. Canadian Journal of Fisheries and Aquatic Sciences 68, 2080-2089.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. *Evolution* **38**, 1358-1370.

Wright J, Bentzen P (1994) Microsatellites: genetic markers for the future. *Reviews in Fish Biology and Fisheries* **4**, 384-388.

Yano A, Nicol B, Jouanno E, *et al.* (2013) The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications* **6**, 486-496.

Zarraonaindia I, Iriondo M, Albaina A, *et al.* (2012) Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in European anchovy (*Engraulis encrasicolus* L.). *PLoS One* **7**, e42201.

Data accessibility

Sampling locations, phenotype data, Structure paramfiles and raw results, and SNP genotypes will be uploaded to Dryad upon acceptance.

Author Contributions

J.E., E.N. and P.O. co-ordinated the collection of samples. C.R.P., T.A., J.E., P.O. and S.E.J. designed the study. T.A. analysed the data. T.A. and C.R.P. wrote the first version of the paper. All authors contributed significantly to revisions.

Table 1: Diversity indices of the main Teno River salmon clusters inferred by STRUCTURE.

	H_O	H_s	F_{IS}
Sub-population 1	0.35882	0.36041	0.0040
Sub-population 2	0.35182	0.35464	0.0076
Admixed	0.35591	0.36381	0.0214

900

901

Table 2: Isolation by distance analyses in the two Teno River salmon sub-populations.

Sub-population	N [a]	distance matrix	Partial Mantel correction matrix[2]	Mantel's r	p value [b]
Sub-population 1	347	distance		0.063	0.007
	347	sub-region		0.093	**0.002**
	347	distance	sub-region	-0.021	0.892
	347	sub-region	distance	0.075	**0.001**
Sub-population 2	171	distance		0.032	0.020
	171	sub-region		0.036	0.018
	171	distance	sub-region	0.011	0.356
	171	sub-region	distance	0.006	0.455

[a] Number of individuals in the analysis.

[b] Bold letters indicate significant α values after multiple test correction; α= 0.00625

902

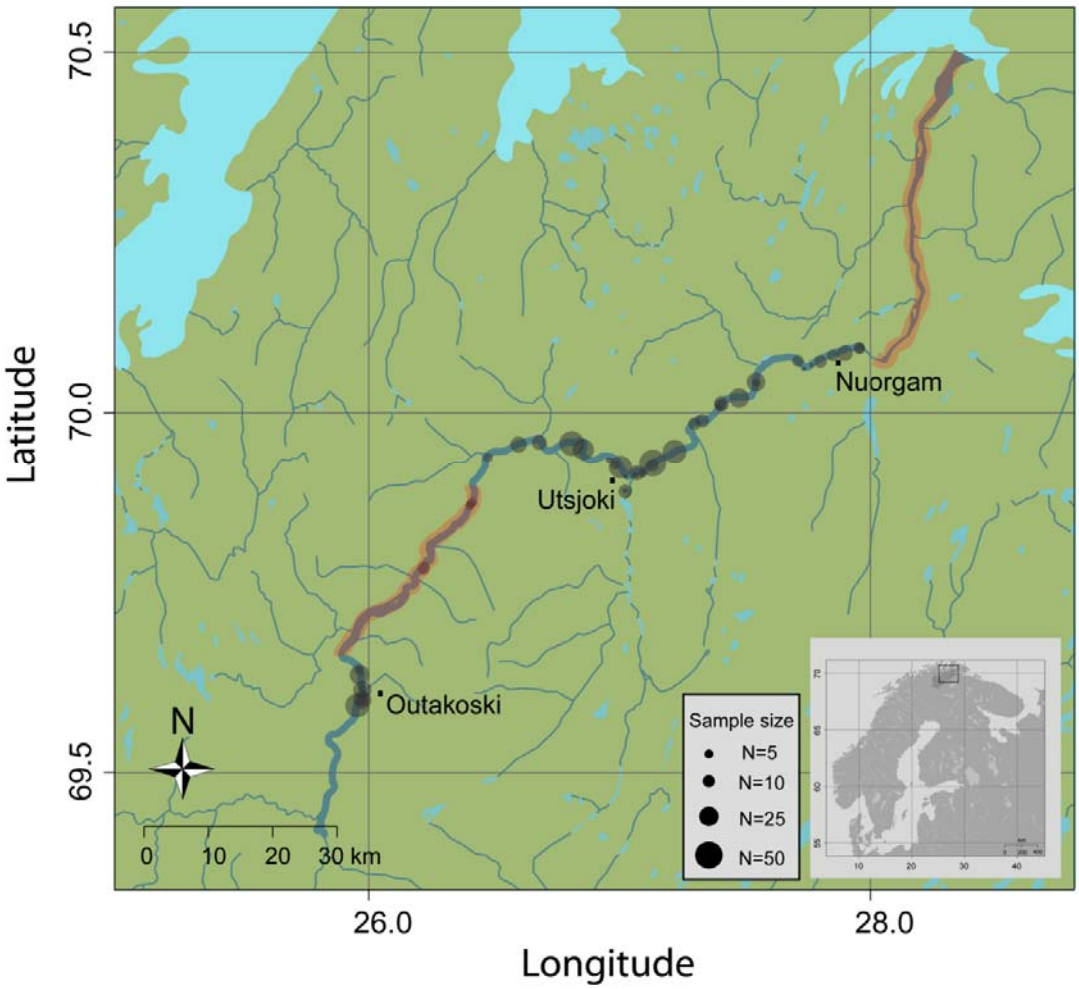
903

Table 3: Estimated fixed effects, random variance components in the mixed model analysis. The 95% confidence intervals, estimated by parametric bootstrapping, are given in parentheses. Asterisks denote effect sizes significantly different from zero[1] (*** = 0.001, *= 0.05).All continuous traits other than condition factor are log scaled

Response variable	μ (mean)	Fixed effect estimates (95% CI)				Random variances	
		pop (pop 2)	FW age	SW age	sex (male)	Residuals (σ^2_R)	σ^2_{Year}
Gaussian error models[1]							
$Growth_{FW1}$	-1.59 (-1.72, -1.47)	0.021 (-0.014, 0.06)	-0.11 *** (-0.13, -0.088)	-0.015 (-0.036, 0.006)	-0.036 * (-0.068, -0.004)	0.0274 (0.0242, 0.0307)	0.0006 (0, 0.0023)
$Growth_{FW2}$	-0.693 (-0.895, -0.494)	0.066 * (0.010, 0.121)	-0.256 *** (-0.290, -0.221)	-0.066 *** (-0.098, -0.035)	0.023 (-0.026, 0.071)	0.0645 (0.0571, 0.0723)	0.0012 (0, 0.0044)
$Growth_{FW3}$	-0.191 (-0.365, -0.017)	0.103 *** (0.055, 0.151)	-0.349 *** (-0.379, -0.318)	-0.013 (-0.040, 0.015)	0.018 (-0.025, 0.062)	0.0505 (0.0448, 0.0566)	$1x10^{-04}$ ($0, 4x10^{-04}$)
$Growth_{FWtot}$	-0.479 (-0.577, -0.384)	0.050 * (0.023, 0.078)	0.048 *** (0.031, 0.065)	-0.015 * (-0.031, 0.001)	0.030 * (0.006, 0.053)	0.0187 (0.0166, 0.0209)	$2 x10^{-04}$ ($0, 7x10^{-04}$)
$Growth_{SW1}$	0.154 (0.057, 0.248)	$1x10^{-04}$ (-0.027, 0.028)	0.012 (-0.005, 0.029)	-0.00321 (-0.0192, 0.0125)	0.027 * (0.002, 0.052)	0.0192 (0.0172, 0.0214)	$2x10^{-05}$ ($0, 1x10^{-04}$)
Weight	6.40 (6.24, 6.56)	-0.208 *** (-0.254, -0.163)	0.039 * (0.011, 0.068)	0.853 *** (0.826, 0.880)	0.148 *** (0.107, 0.189)	0.0503 (0.0447, 0.0562)	$1x10^{-04}$ ($0, 4x10^{-04}$)
Length	3.71 (3.66, 3.76)	-0.04 *** (-0.06, -0.03)	0.008 ($-3x10^{-04}$, 0.017)	0.276 *** (0.268, 0.283)	0.052 *** (0.040, 0.064)	0.0046 (0.0041, 0.0051)	$5x10^{-06}$ ($0, 3x10^{-05}$)
CF	0.876 (0.8, 0.952)	-0.062 *** (-0.083, -0.041)	0.010 (-0.003, 0.023)	0.030 *** (0.017, 0.042)	-0.003 (-0.022, 0.016)	0.0104 (0.0092, 0.0116)	$2x10^{-04}$ ($0, 7x10^{-04}$)
Poisson error models							
SW age	1.28 (0.76, 1.80)	-1.24 *** (-1.47, -1.03)	-0.021 (-0.129, 0.086)	NA	-0.59 *** (-0.74, -0.44)	1	$8x10^{-03}$ (0, 0.033)
FW age	1.45 (1.22, 1.68)	-0.044 (-0.14, 0.06)	NA	-0.01 (-0.06, 0.04)	-0.04 (-0.13, 0.05)	1	$3x10^{-04}$ (0, 0.002)

1- For fixed effects, 99.9 % (p=0.001) and 95 % (p=0.05) of 10000 permutations, the effect size is on the same side of zero. For the random component, the same proportion of permutations should deviate from zero.

905 Figure 1



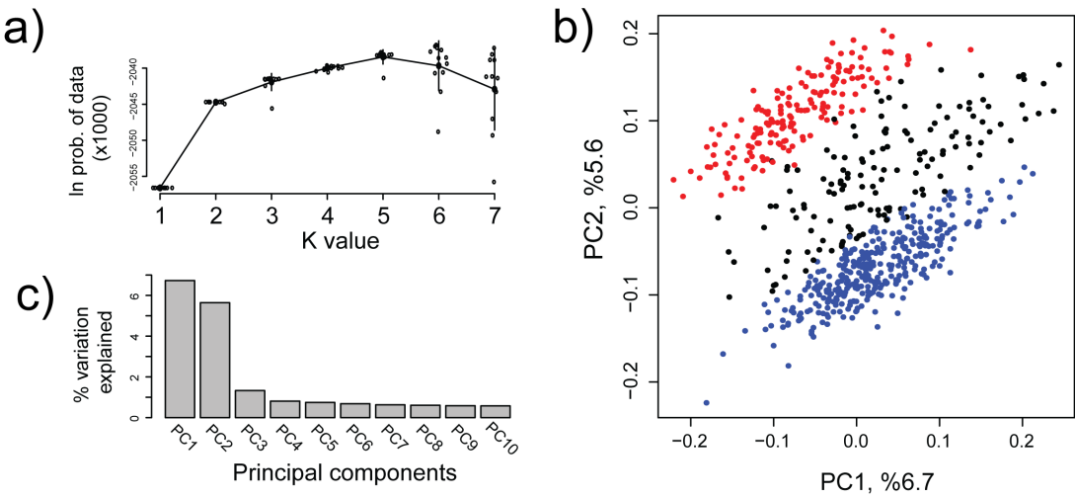
906

907 **Figure 1:** Map of the Teno River and basin with sampling locations along the mainstem
908 (highlighted with a thicker blue line). Stretches of the mainstem not suitable as spawning
909 grounds or juvenile nurseries are highlighted in red.

910

911 Figure 2

912

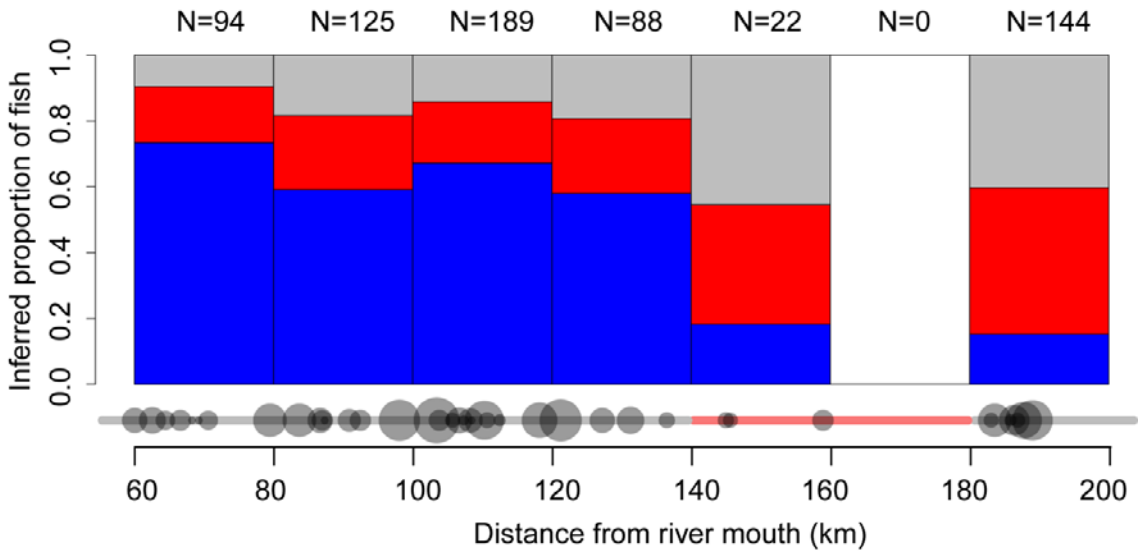


913

914 **Figure 2:** Structure and principal component analyses of Atlantic salmon sampled from the
915 Teno River main-stem. a) Plot of the first two major PC axes, where colors show sub-
916 populations inferred by the STRUCTURE analysis at the optimum *K* value of two. Blue, red
917 and, black colors show Sub-population 1, Sub-population 2, and admixed individuals,
918 respectively. b) The estimated *ln* probability of data given the K value. Error bars are
919 standard deviations after 12 replicate runs. The results for each of the 12 replicate run are
920 given with open circles. (c) Percent variation explained in the first 10 PC axes.

921

922 Figure 3

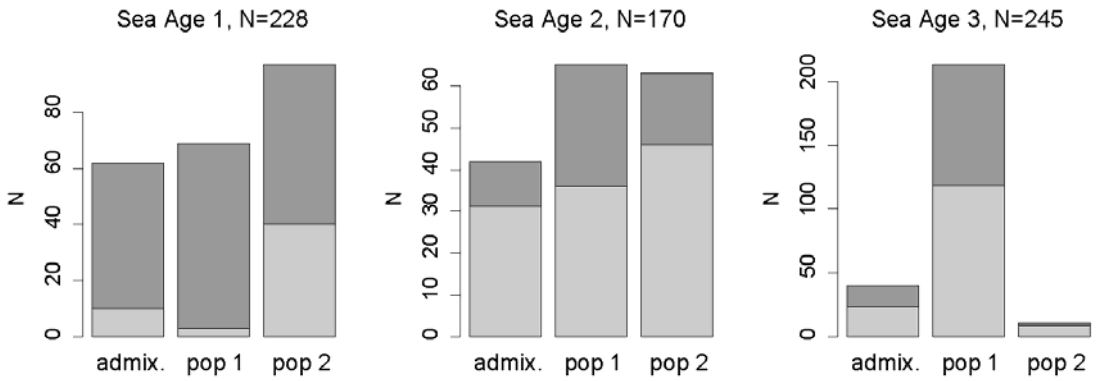


923

924 **Figure 3:** Proportions of the inferred sub-populations over the sampling range along the
925 Teno River mainstem. Blue, red and, grey colors indicate Sub-population 1, Sub-population
926 2, and the admixed group, respectively. Proportional sample sizes for specific locations along
927 the main-stem are indicated by circle diameter and total sample size for each 20km interval is
928 listed above the bars. The sandy stretch of river that is mostly unsuitable for spawning and
929 juvenile rearing is indicated with red on the lower horizontal line.

930

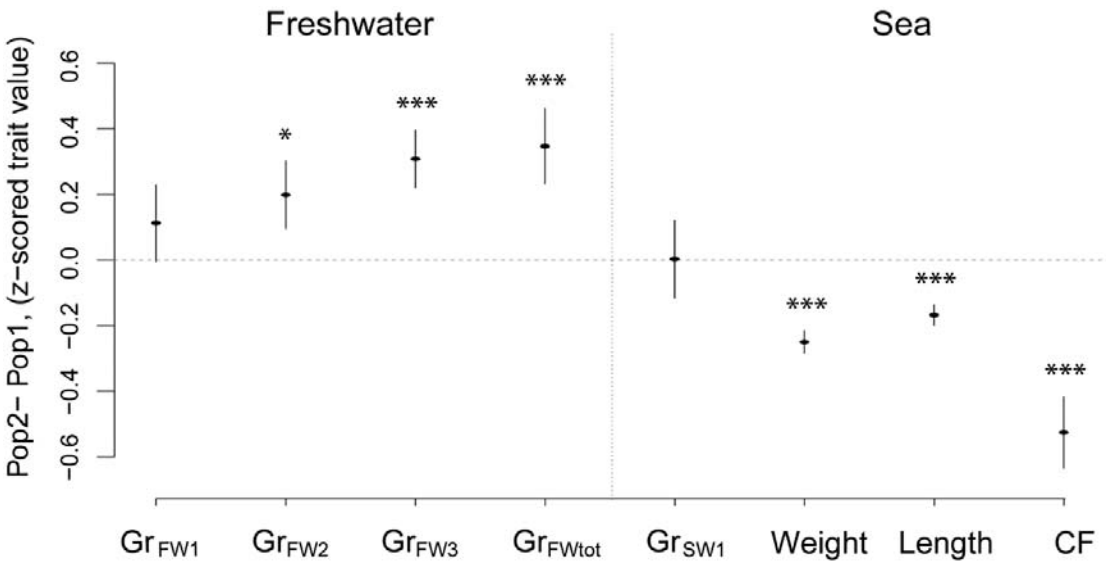
931 Figure 4



932

933 **Figure 4:** Sex distribution (males in dark-grey, females in light-grey) among sub-populations
934 and sea age classes.

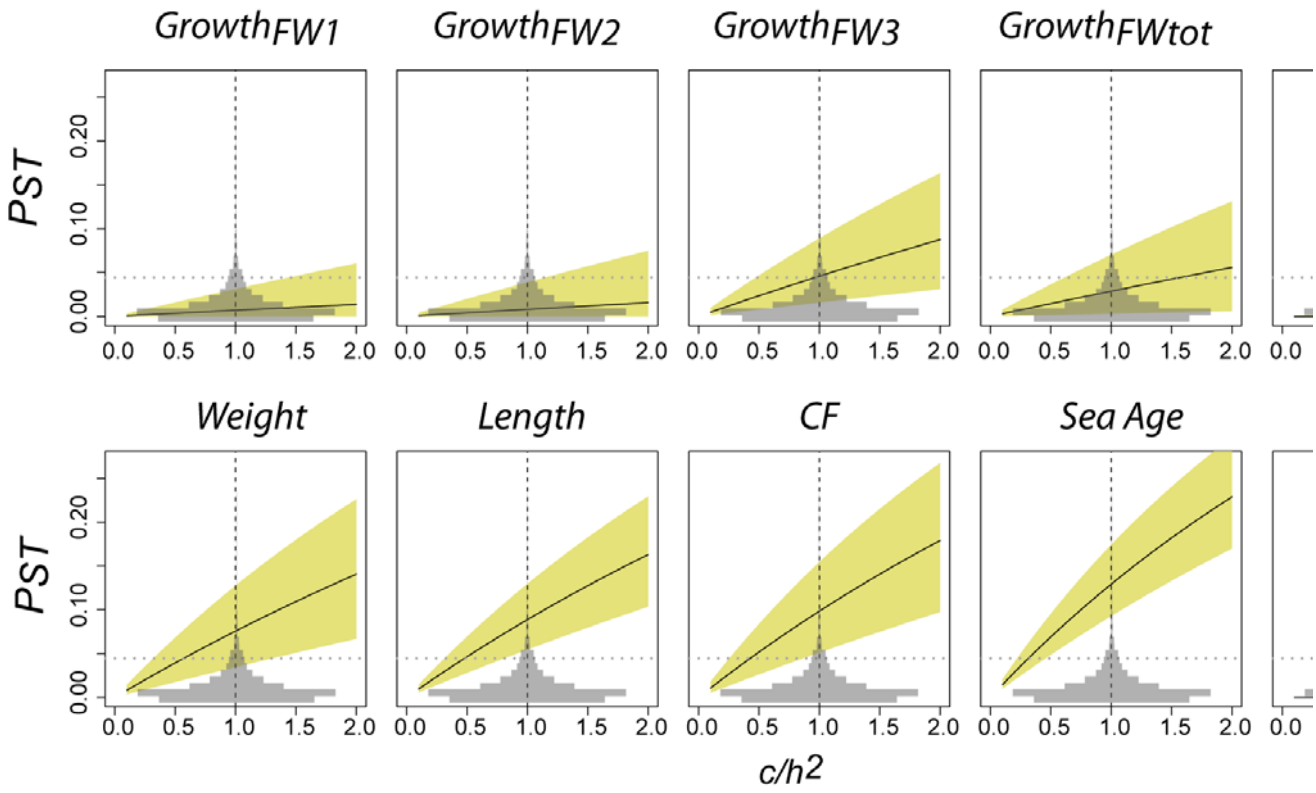
935 Figure 5



936

937 **Figure 5:** Population specific differences between phenotypic trait values of Teno River sub-
938 populations. Bars shows standard deviation of the difference inferred from 10000
939 permutations. Asterisks denote significant differences between populations (*** = 0.001, *=
940 0.05). Here, only population specific effects are accounted for after being inferred by the
941 linear model (See Table 2 for details).

942 Figure 6



943

944 **Figure 6.** The relationship between P_{ST} and F_{ST} between the two Teno mainstem Atlantic
945 salmon populations under different c/h^2 ratio scenarios for the 10 phenotypic traits assessed
946 in this study. The SNP F_{ST} distribution is plotted in light grey and the upper neutral F_{ST}
947 estimate is indicated with a grey horizontal line within each plot. The vertical dashed line in
948 each panel shows the c/h^2 value at 1, where the relative contribution of additive genetic
949 effects to population variation (c) is equal to (h^2). The median P_{ST} estimate is shown with
950 solid black line, and the green shaded area covers 95 % CI of the P_{ST} estimate.

951

952 **Supplementary material:**

953 **Suppl. table 1:** Sea age and sex distribution of Teno River salmon samples used in the study. F and M denote
954 female and male, respectively.

	Distance from the river mouth (km)														
	60-80		80-100		100-120		120-140		140-160		160-180		180-200		
sea age	F	M	F	M	F	M	F	M	F	M	F	M	F	M	total
1	4	30	10	27	10	54	7	25	5	6	0	0	17	33	228
2	11	7	25	13	31	14	10	4	1	7	0	0	35	12	170
3	23	16	26	21	40	31	30	10	1	2	0	0	28	17	245
4	2	0	0	3	2	7	0	2	0	0	0	0	0	2	18
5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
total	40	54	61	64	83	106	47	41	7	15	0	0	80	64	662

955

956 **Supp. table 2:** Sampling day in relation to population provenance and sea age.

	df	*SS*	*MS*	*F*	*p-value*
Populations	1	0	0.033	0.0004	0.984
Sea age	1	80	79.751	0.971	0.3249
Population x sea age	1	109	108.827	1.325	0.2503
Residuals	482	39588	82.133		

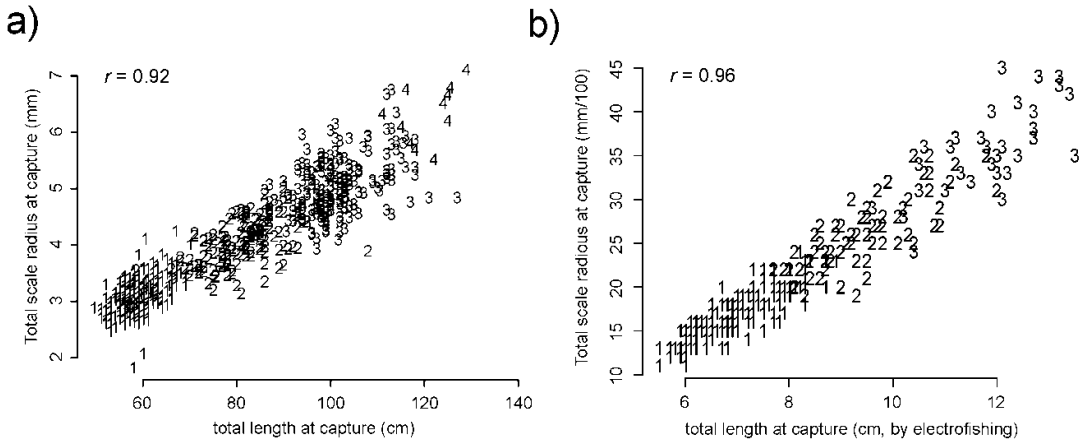
957

958

959

960

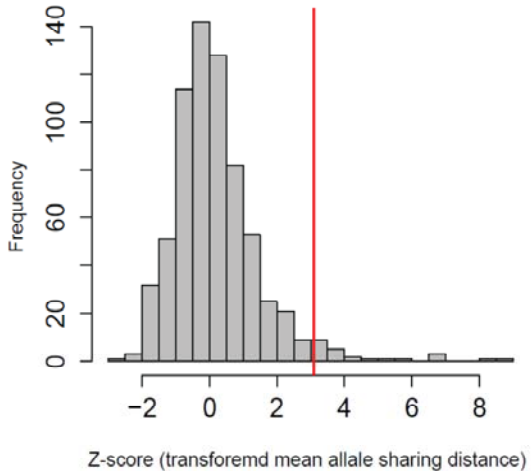
961



962

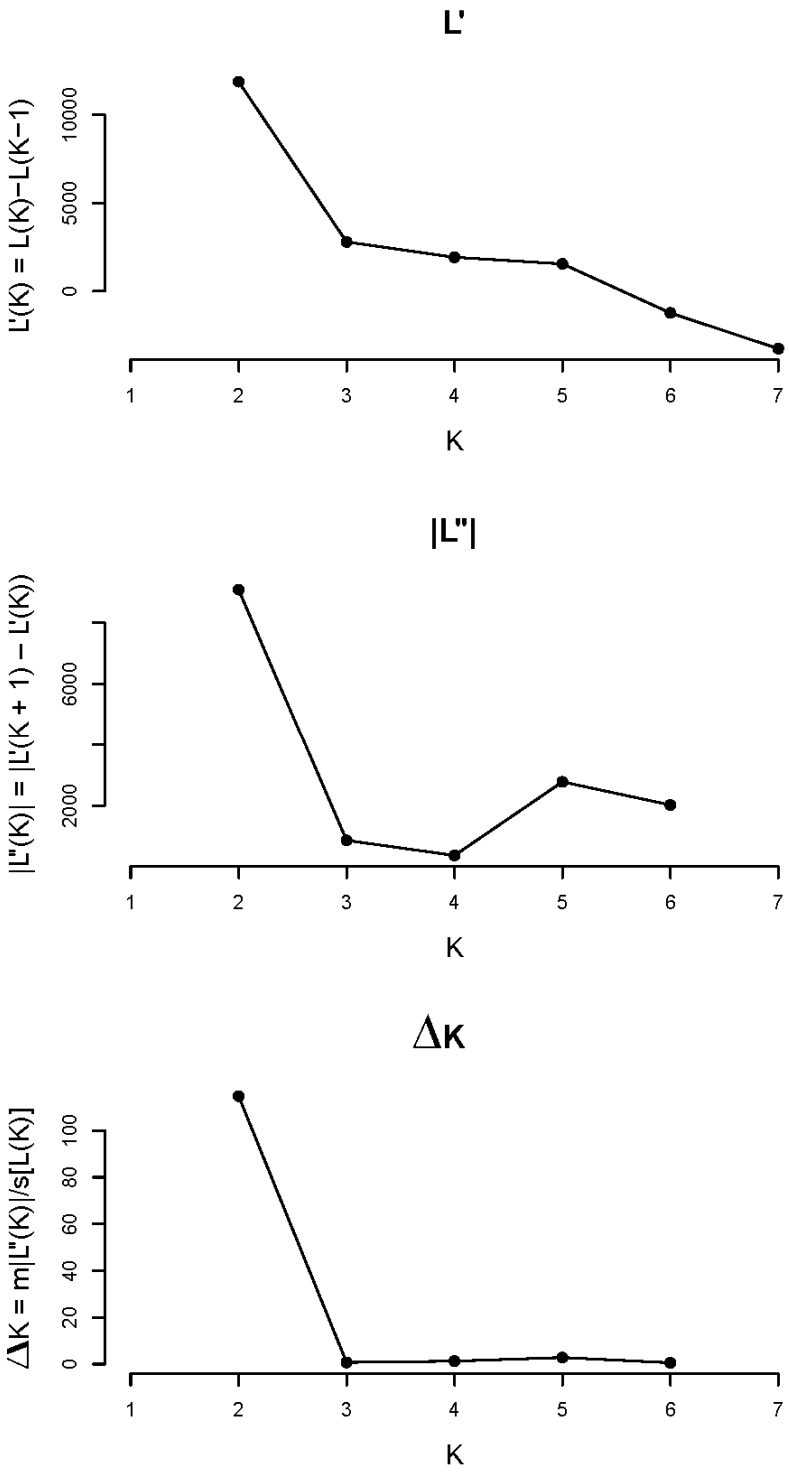
963 **Supp. figure 1:** Correlation between total scale growth (mm) and length in (a) adult fish sampled for this study
964 and. (b) Juveniles electrofished from the mainstem. Data points in a and b are depicted by numbers which
965 indicate Sea age or FW age of fish at the time of capture, respectively.

966



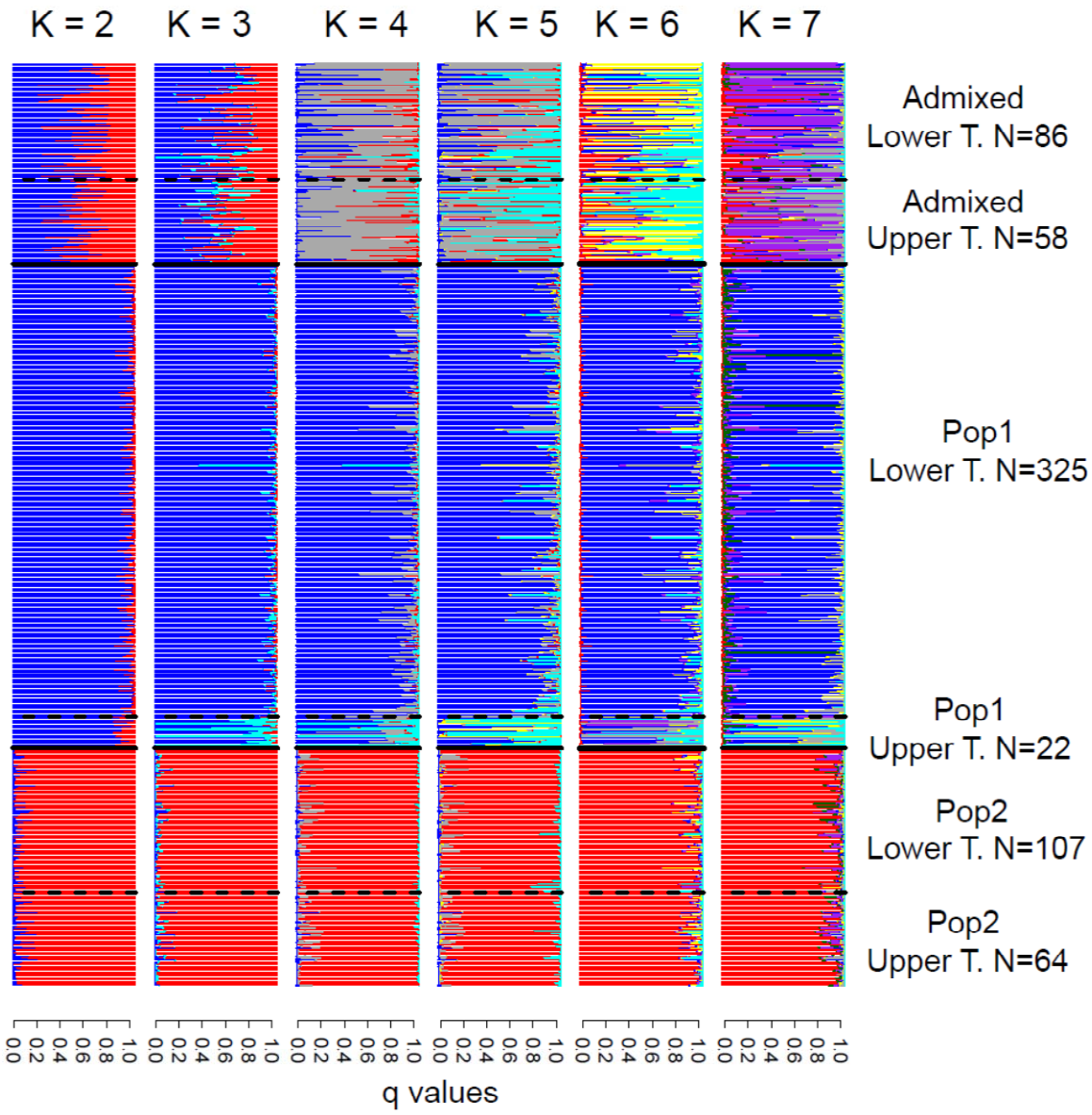
967

968 **Supp. figure 2:** Z-transformed allele sharing distance. Individuals 3.09 standard deviation away from the mean
969 (the left of the red vertical line) were excluded from the study.



970

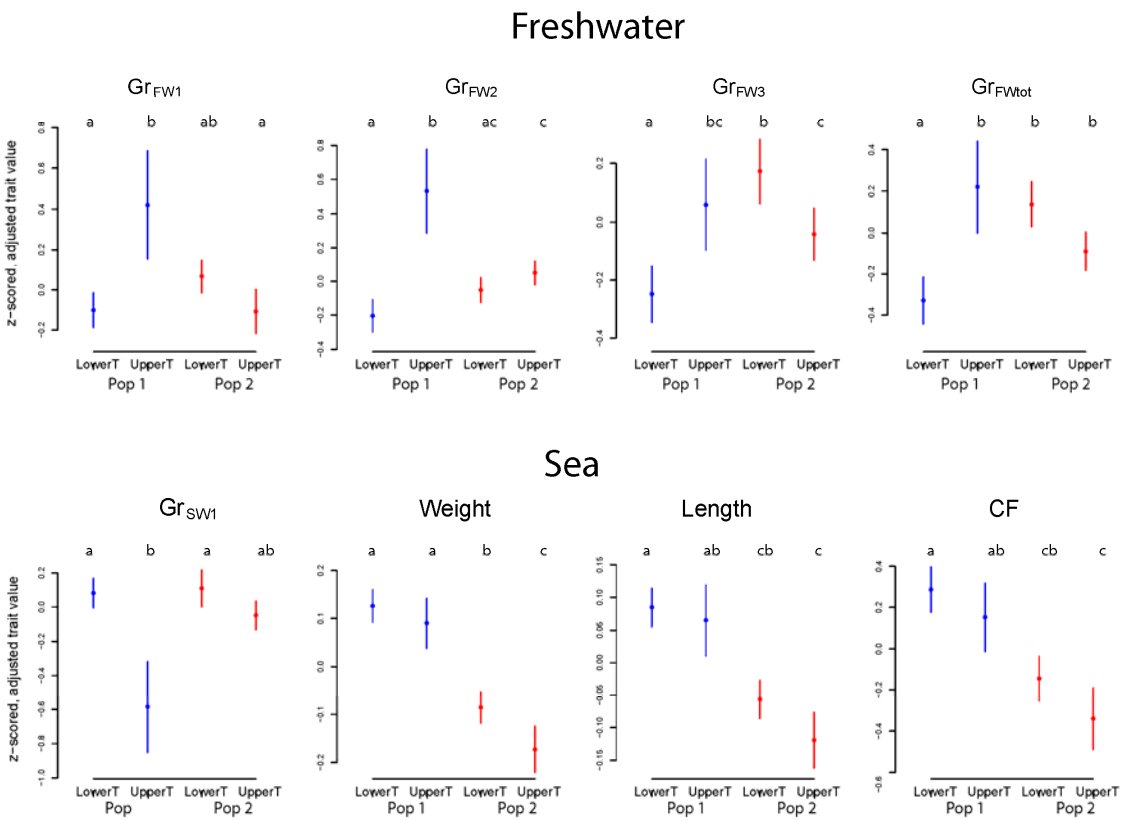
971 **Supp. figure 3:** The parameters for the graphical method in the Evanno et al (2005) for detection of the true
972 number of groups K. Note that the method cannot be employed at K=1.



973

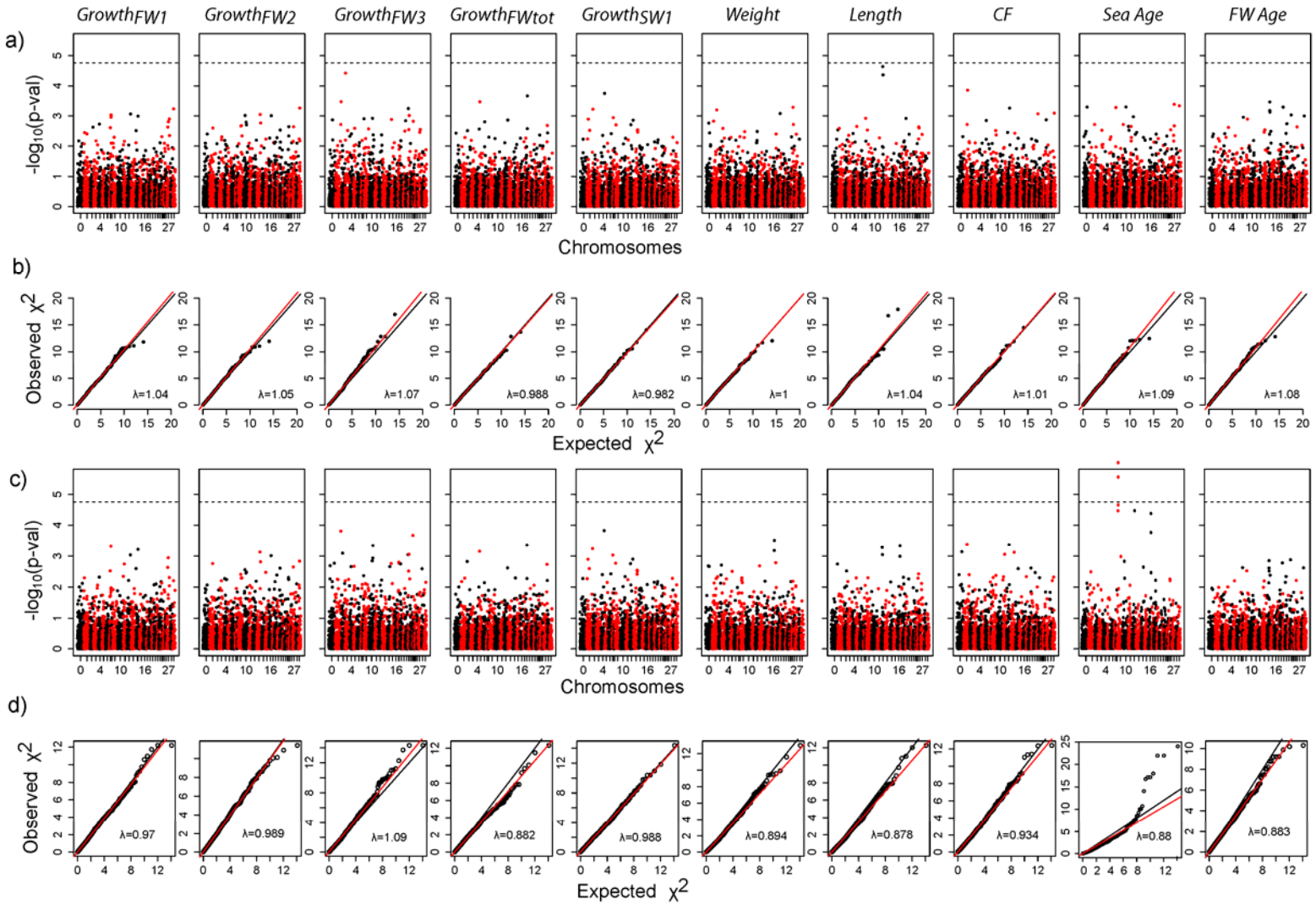
974 **Supp. figure 4:** q value plots of individuals for K values between 2 and 7. Individuals are grouped together
975 based on their assigned clusters inferred at K=2 (Sub-population 1, Sub-population 2 and admixed), and each
976 cluster is grouped by sampling location (i.e. upper Teno vs lower Teno main stem).

977



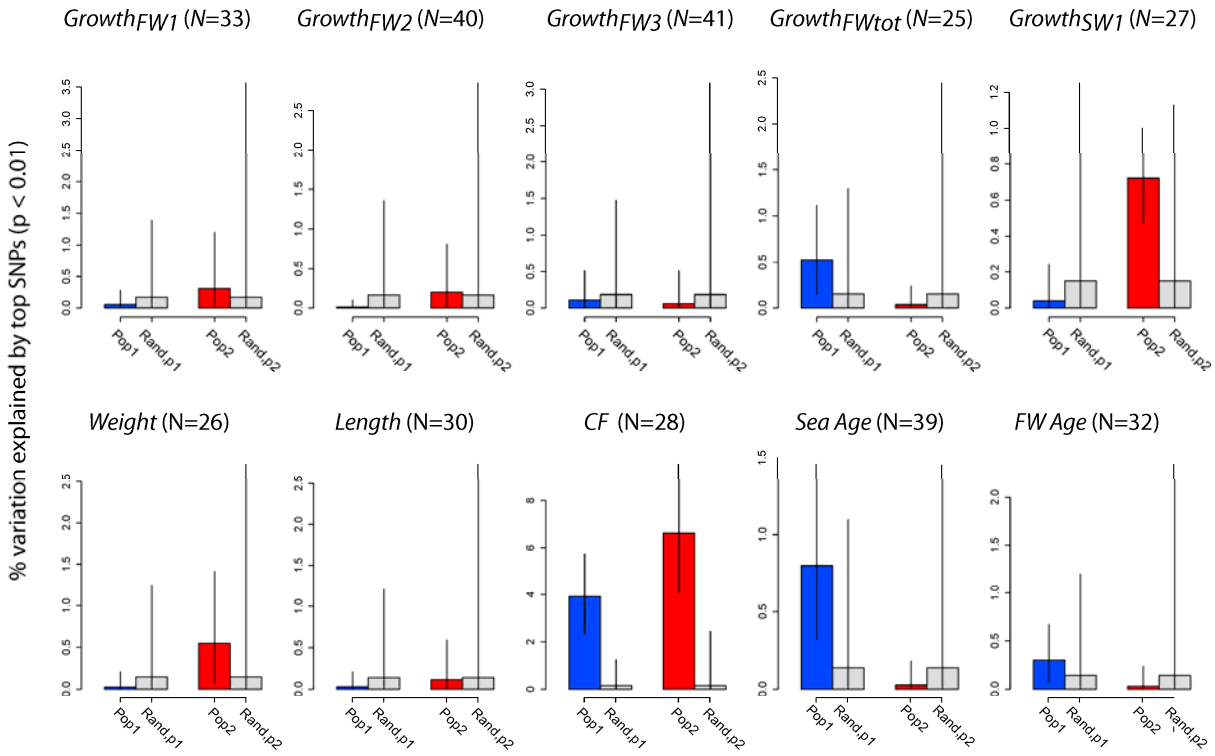
978

979 **Supp. figure 5:** Phenotypic variation among sub-populations subdivided by sampling location along the Teno River main stem (upper and lower Teno main stem). The y-axis
980 denotes the standardised parameter estimates for each sub-population/location class after the liner mixed effect model (see materials and methods). Bars shows standard
981 deviation of the difference inferred from 10000 permutations. Blue and red colors indicate Sub-population 1, Sub-population 2, respectively.Letters denote significant
982 differences between groups after (p-value = 0.01), such that the trait is significantly different between two grouped that do not share a common latter.



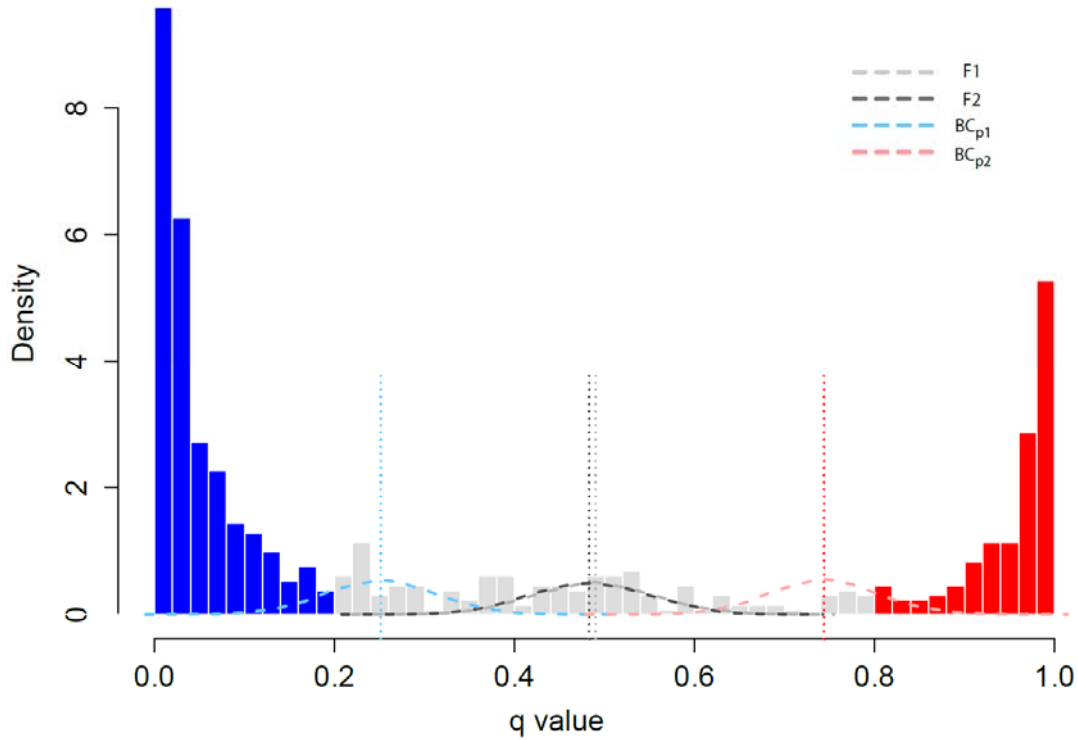
Supp. Figure 6: The genome wide association study across 10 phenotypes. Top two panels (a,b) shows results using principal component method to correct for of population structure, and bottom two panels (c,d) shows correction with genomic control. Manhattan plots (a,c) show association fo SNPs along the 29 linkage groups and a set of unmapped SNPs (i.e. linkage group zero). Horizontal dashed line shows the genome wide significant level after Bonferroni

corrected alpha value 0.05. Distribution of the $\chi 2$ test statistic is given after (b) principal component and (b) genomic control corrections. The solid line indicates the distribution under neutral expectations with no stratification.



Supp. figure 7: Cumulative per cent variation explained in the GWAS analyses (i.e. PC correction methods, Price *et al*. 2006) by the top significant loci ($p<0.01$), within Sub-Population 1 (blue bars) and Sub-Population 2 (red bars). Gray bars adjacent to coloured bars are per cent variation explained by equal number of SNPs drawn randomly from SNP set at each population respectively (N=2874). Error bars of coloured bars represent 95% sampling distribution, as calculated by permuting effect size and its standard errors of each SNP (as inferred from linear models in the GWAS analyses, see Materials and Methods), and calculating the cumulative (additive) effect 10000 times. Error bars of gray bars represent 95% random sampling distribution, as calculated by effect sizes of 10000 random independent SNP sets (drawn independently each time). (Effects size was calculated by first obtaining the residuals of the model as described in

Table 3. Then by performing a linear regression, where residuals of the initial model were modelled as response variable and the cumulative allelic substitution effect of SNPs as independent variable, the adjusted R^2 value were obtained as the proportion of variance expaliend by the polygenic SNP set at $p < 0.01$)



Supp. figure 7: Q value distribution of the dataset with empirical (bars) and simulated hybrid classes (dashed lines) genotypes assuming based on allele frequency distribution of Sub-population 1 (blue) and Sub-population 2 (red) and assuming the intermediates are first generation hybrids. Q-values are given for K=2, in which individuals with $q < 0.2$ are assigned to Sub-population 1 and individuals with $q > 0.8$ are assigned to Sub-population 2.