Canadian Journal of Fisheries and Aquatic Sciences
Journal canadien des sciences halieutiques et aquatiques

Genetic stock identification of Atlantic salmon and its evaluation in a large population complex

Journal:	*Canadian Journal of Fisheries and Aquatic Sciences*
Manuscript ID	cjfas-2015-0606.R2
Manuscript Type:	Article
Date Submitted by the Author:	27-Jun-2016
Complete List of Authors:	Vähä, Juha-Pekka; University of Turku Erkinaro, Jaakko; Natural Resources Institute Finland Fålkegard, Morten; Norwegian Institute of Nature Research Orell, Panu; Natural Resources Institute Finland (Luke) Niemelä, Eero; Natural Research Institute Finland (Luke)
Keyword:	Fisheries management, Atlantic salmon, Genetic stock identification, MIxed stock fishery

Genetic stock identification of Atlantic salmon and its evaluation in a large population complex

Juha-Pekka Vähä, Jaakko Erkinaro, Morten Falkegård, Panu Orell and Eero Niemelä

Juha-Pekka Vähä*, Kevo Subarctic Research Institute, University of Turku, Finland juha-pekka.vaha@utu.fi

Jaakko Erkinaro, Natural Resources Institute Finland (Luke), Oulu, Finland jaakko.erkinaro@luke.fi

Morten Falkegård, Norwegian Institute for Nature Research, Tromsø, Norway morten.falkegard@nina.no

Panu Orell, Natural Resources Institute Finland (Luke), Oulu, Finland panu.orell@luke.fi

Eero Niemelä, Natural Resources Institute Finland (Luke), Oulu, Finland eero.niemela@luke.fi

Corresponding author: Jaakko Erkinaro, Natural Resources Institute Finland Luke, POB 413, FI-90014 Oulu, Finland, tel. +358 40 5435929, e-mail: jaakko.erkinaro@luke.fi.

**Present address*: Juha-Pekka Vähä, Association for Water and Environment of Western Uusimaa, POB 51, FI-08101 Lohja, Finland, juha-pekka.vaha@luvy.fi

Abstract

Addressing biocomplexity in fisheries management is a challenge requiring an ability to differentiate between distinct populations contributing to fisheries. We produced extensive genetic baseline data involving 36 sampling locations and 33 microsatellite markers, which allowed characterization of the genetic structure and diversity in a large Atlantic salmon population complex of the River Teno system, northernmost Europe. Altogether, we identified 28 hierarchically structured and genetically distinct population segments (Global F_{ST} = 0.065) corresponding exceptionally well with their geographical locations. An assessment of factors affecting the stock identification accuracy indicated that the identification success is largely defined by the interaction of genetic divergence and the baseline sample sizes. The choice between the two statistical methods tested for performance in genetic stock identification, ONCOR and cBAYES, was not critical, albeit the latter demonstrated slightly higher identification accuracy and lower sensitivity to population composition of the mixture sample. The strong genetic structuring among populations together with a powerful marker system allowed for accurate stock identification of individuals and enabled assessment of stock compositions contributing to mixed-stock fisheries.

Introduction

Species diversity support community stability and productivity (e.g. Tilman et al. 1996) as well as ecosystem functioning and services (Gamfeldt et al. 2008; Isbell et al. 2011), while intraspecific phenotypic and genotypic diversity play an important role in population persistence and dynamics (Agashe 2009; Bolnick et al. 2011). Bio-complexity in fish populations - in the form of population diversity, life-history variation and genetic structures - is important for sustainable fisheries (Hilborn et al. 2003; Hutchinson 2008). Population diversity provides resilience to disturbances (e.g. exploitation), which contribute to long-term sustainability, and diverse systems provide more temporally stable ecosystem services; a phenomenon referred to as the 'portfolio effect' (Schindler et al. 2010). Strong or selective harvesting can negatively affect abundance and diversity (e.g. Youngson et al. 2003), and require reductions in harvest to protect less abundant stocks. This can be at odds with socio-economic factors that might point towards continued harvest, to provide food, income and other social goods. Managers balancing this trade-off between harvesting and conservation needs face a particularly challenging quest when the exploitation targets a multitude of species or populations.

In mixed-stock fisheries, stocks at poor status might be targeted to an unknown extent, threatening small and vulnerable populations (Nehlsen et al. 1991; ICES 2015). In the absence of detailed stock-specific knowledge, the precautionary approach would be a substantial reduction or closure of the mixed-stock fishery fishery. Thus, managers need

tools that enable stock-specific estimates of exploitation to enable a stock-specific management of mixed-stock fisheries, tailored towards safeguarding vulnerable populations whilst allowing the continued harvest of healthy stocks (Crozier et al. 2004).

Molecular genetic techniques and the application of genetic markers have not only revolutionized our understanding of population diversity, but genetic aspects have become an integral part of fishery and wildlife management (Mills 2012; Allendorf et al. 2013). Genetic monitoring of the population composition of mixed stock catches provide a way to ensure a stock-specific management of the mixed stock fisheries, including the knowledge needed to establish targeted stock-specific regulatory measures tailored towards safeguarding vulnerable populations. With the advent of powerful genetic markers, reduced costs of analysing large numbers of samples accompanied with the development of tailored statistical methods, genetic stock identification (GSI) is now one of the most successful biological tools available for effective monitoring and stock-specific management actions (e.g. Beacham et al. 2008; Ensing et al. 2013).

Anadromous fish populations undertake feeding migrations between fresh water and open ocean areas and return to spawn in their natal rivers after sexual maturation. Accurate homing to natal rivers (e.g. Quinn 1993) provides the potential for genetic population structuring between and even within river systems (Vähä et al. 2007; Hess et al. 2014). At sea, anadromous fish populations may mix and harvesting during the migratory stage targets multiple populations resulting in a mixed-stock fishery. GSI has been used to manage Pacific salmon (*Oncorhynchus* spp.) fisheries for three decades (Milner et al. 1985).

Ecological applications of GSI on Pacific salmon species cover a variety of spatial and temporal scales from ocean distribution of juveniles and adults (e.g. Beacham et al. 2006; Habicht et al. 2010) to in-river investigations of population of origin among migrating individuals (e.g. Beacham et al. 2008; Hess et al. 2014). GSI is a routine tool for monitoring mixed-stock fisheries of Pacific salmon, providing real-time information on catch stock composition to managers and fishermen (e.g. http://pacificfishtrax.org).

In contrast, GSI has less been used in management of Atlantic salmon (*Salmo salar*) mixed-stock fisheries. There are only some marine examples, e.g. genetic mixed-stock analyses resolving stock group proportions in the Baltic sea (Koljonen 2006), contributions of North American Atlantic salmon stocks to the West-Greenland salmon fishery (Gauthier-Ouellet et al. 2009), and some coastal studies demonstrating the potential for such approach, e.g. from England (Gilbey et al. 2012), Ireland (Ensing et al. 2013), Canada (Bradbury et al. 2015), St. Pierre & Miguelon of France (Bradbury et al. 2016), and the Barents Sea coast in Norway and Russia (Vähä et al. 2014). Freshwater examples include lake-run brown trout (*Salmo trutta*; Swatdipong et al. 2013), lake trout (*Salvelinus namaycush*; Northrup et al. 2010) and broad whitefish (*Coregonus nasus* ; Harris and Taylor 2010). However, genetic mixed-stock analyses of Atlantic salmon fisheries in large rivers have so far received little or no attention (Vähä et al. 2011).

The River Teno (Tana in Norwegian) in the northernmost Europe is one of the few remaining large river systems that still support multiple and abundant wild Atlantic salmon populations (Niemelä et al. 2006; Vähä et al. 2007). Genetic studies have revealed a

103 structured population complex consisting of a number of demographically independent,
104 genetically distinct and temporally stable population segments (Vähä et al. 2007; 2008).
105 Genetic assessment of systematically collected salmon scale samples from the mixed-stock
106 fishery of the Teno main stem can provide a fine-resolution estimation of the origin of the
107 captured salmon (Vähä et al. 2011). In the present study we set out to 1) complement the
108 baseline data on salmon population structure of the Teno complex, 2) assess factors
109 affecting the GSI success, and 3) investigate the feasibility of using GSI to monitor the
110 harvest of multiple salmon populations in the mixed-stock fishery of the River Teno main
111 stem.

112

113 Material and methods

114

115 Baseline data

116 The examined genetic baseline consisted of 1) previously described baseline data from 12
117 tributaries (n=1076; Vähä et al. 2007; 2008; 2011), which all were supplemented by 2) a new
118 collection of samples (n=866), 3) samples from 15 previously unsampled tributaries
119 (n=770), and 4) samples from eight different parts of the Teno main stem (n=611; TMS TB,
120 AK, PI, SI, GJ, KO, YK, OUT) and from the Inarijoki (n=71) (Figure 1). The old baseline
121 data were collected by sampling adult salmon catches (Vähä et al. 2007) while the new
122 data consisted of juvenile samples.

123 Juvenile salmon were electrofished and sampled for adipose fin tissue. Only one
124 individual of both major juvenile salmon age groups (age-0, age-1) were sampled at one
125 site, and the next pair of fish were sampled 50-100 m apart, depending on the area, in
126 order to minimize the probability of sampling siblings. License for sample collection was
127 issued by the County Governor of Finnmark, Norway, and the Center for Economic
128 Development, Transport and the Environment in Finland. Individual tissue samples were
129 stored in 96% ethanol. In total, the updated baseline data consisted of 3323 salmon from 36
130 sites within the Teno river system (Figure 1, Table S1[1]). Due to the temporal stability of the
131 genetic structure within the Teno system (Vähä et al. 2008), temporally replicated samples
132 were pooled for each site.

133

134 [1]Supplementary data are available through the journal Web site at xxx

135

136 ## Test samples

137 In order to test the performance of the new baseline, we used test samples of known
138 origin. Two sets of test samples were used: 1) late season samples of adult salmon from
139 five areas of the main stem (TS1-5, Figure 1), and 2) juvenile samples collected from
140 spawning grounds at the lowermost part of the River Utsjoki, a major tributary of the Teno
141 (TS6, Figure 1). The first set comprised samples from the old baseline data (Vähä et al.
142 2007) which were not included in the current baseline. While these individuals were
143 sampled in late August after migration period and assumed to originate from that area,

144 the samples may include transient fish from nearby populations (Vähä et al. 2007; Erkinaro
145 et al. 2010).

146

147 ## Microsatellite analysis

148 Total genomic DNA was extracted from scale or fin tissue as previously described (Vähä
149 et al. 2007). In the present study, we used genotypes of 33 microsatellite markers, of which
150 32 were described in detail in Vähä et al. (2007; 2011). In addition, locus Sssp3016
151 (Paterson et al. 2004) was added to marker panel. Primer sequences for amplifying
152 alternative MHCI locus amplicons (Grimholt et al. 2002) were F:
153 GAAGGTGCTGAAGAGGAACGTC and R: GTTTCAATTACCACAAGCCCGCTC.

154 All microsatellite loci were amplified by multiplex PCR slightly modified from Vähä et al.
155 (2011) and details are available from the authors upon request. Varying volumes of the
156 PCR amplified products were pooled and electrophoresis was then performed on ABI
157 3130xl (Applied Biosystems). Electropherograms were inspected and allele scoring
158 performed with GeneMapper V4.0 (Applied Biosystems) followed by manual corrections.
159 Electropherograms and allele scores were reviewed by two persons independently.

160

161 ## Microsatellite variability

162 All 33 microsatellite loci examined were polymorphic in all 36 baseline samples displaying
163 527 alleles in total (4 - 33 / locus). Tests of conformance to Hardy–Weinberg equilibrium
164 expectations over all samples indicated that loci MHC I ($p<0.001$) and Sssp2201 ($p<0.01$)

departed significantly from expectation. In locus MHC I, the departure from HWE was caused by the presence of a null allele: when a subset of individuals (n = 2064) from all baseline samples was amplified with an alternative set of primers, null allele was detected with a frequency varying from 0.01 to 0.12 in 14 of the samples. An excess of homozygotes in locus Sssp2201 may arise from large allele dropouts since the locus is highly variable (33 alleles) with long amplicons (243bp-387bp).

Over all loci, excluding MHC I and SSsp2201, only the Anárjohka sample deviated significantly from conformity to HW proportions expressing an excess of homozygotes (Bonferroni correction applied; $\alpha = 0.05$). Deficiency of heterozygotes was most likely due to allelic dropouts stemming from low quality DNA extracts (inferred from peak intensity of electropherograms) from archived juvenile scales collected in 1996.

Regardless, all loci and samples were retained in the baseline as several previous studies have indicated that inclusion of additional marker data despite aberrations add to the accuracy of genetic stock identification (Beacham et al. 2006; Carlsson 2008; Griffiths et al. 2010).

Statistical analysis

Genetic differentiation between samples and deviations from Hardy–Weinberg equilibrium (HWE) within and across loci for each locality and globally were tested using *F*-statistics of Weir and Cockerham (1984) with significance levels calculated with a randomization procedure (3300 permutations) as implemented in FSTAT v2.9.3 (Goudet

2001). FreeNA (Chapuis and Estoup 2007) was used to assess the presence of null alleles and their effect on the global and pairwise F_{ST} estimates as well as on the tree topology.

Genetic relationships among samples were estimated as Cavalli-Sforza and Edward's Dc (1967) genetic distance using PowerMarker v3.25 (Liu and Muse 2005) and the genetic relationships were visualized by the neighbor-joining method (Saitou and Nei 1987) in SplitsTree4 v4.11.3 (Huson and Bryant 2006). Robustness of clustering was evaluated by calculating split support values by bootstrapping 10,000 times over loci using PHYLIP (Felsenstein 2005).

The Bayesian clustering approach of STRUCTURE v2.3.2. (Pritchard et al. 2000) with the correlated allele frequency model (Falush et al. 2003) was used to create plots of ancestry (i.e., admixture coefficients (Q)) for evaluating genetic clustering of individuals (Figure S1[1]). Pairwise and global F_{ST} values as well as the variance components among groups of populations were calculated using ARLEQUIN v3.5.1 (Excoffier et al. 2005). Population-specific F_{ST} values were calculated using GESTE v2.0 (Foll and Gaggiotti 2006).

For testing isolation-by-distance patterns, Mantel tests were performed using PASSaGE 2 (Rosenberg and Anderson 2011). Geographical distances (km) between sampling localities were plotted against the estimates of $F_{ST} / (1 - F_{ST})$ (Rousset 1997) and the significance was tested through a randomization with 10,000 permutations of the data.

The effects of genetic divergence and sample size of the baseline population on GSI success were analyzed using generalized linear mixed models (GLMM). Identification events per trials from analyses of known-origin mixtures constructed by resampling

individuals (see below) were used as the response variable, whereas population-specific F_{ST} and baseline sample size were set as fixed effects and treated as continuous variables. Analyses were performed using the GLIMMIX procedure of SAS v9.3 (SAS Institute Inc.) with a logit link function and a binomial error term.

[1]Supplementary data are available through the journal Web site at xxx

GSI methods

GSI performance was tested using ONCOR (Kalinowski et al. 2007) and cBAYES (Neaves et al. 2005). ONCOR implements the method of Rannala and Mountain (1997) which uses an equal probability Dirichlet density as the prior for the allele frequencies at a locus assigning a frequency of 1/(n+1) to an allele not found in a population. The prior densities updated with the observed baseline data give the posterior probability densities of allele frequencies.

cBAYES (Neaves et al. 2005) implements the Bayesian method of Pella and Masuda (2001). The prior distribution of alleles at a locus follows the mean of the allele frequencies over all stocks and posterior distributions of the baseline allele frequencies are the product of priors and the observed allele frequencies. Shrinking the observed values toward central values takes advantage of the genetic similarity of populations and is thought to minimize estimation error in allele frequencies. Further, the allele frequencies of mixture individuals

assigned to a baseline population, at each MCMC step, are used to update the baseline allele frequencies.

All analyses with cBAYES were performed using eight independent chains of 100K iterations starting from four random stocks. The number of replications was increased if diagnostics indicated convergence issues (shrink factor by population >1.2).

Power analysis

The discriminatory power of the baseline for stock identification was first evaluated through simulation procedures implemented in the program ONCOR, which has been shown to produce essentially unbiased estimates of GSI accuracy (Anderson et al. 2008).

A second power test analyzed known-origin mixtures constructed by resampling individuals without replacement from each of the baseline samples. Our purpose was to compare the performance of the GSI methods when the data included missing genotypes and potential genotyping errors, typical to microsatellite data. Each mixture sample (12 in total) was composed of ~10% randomly chosen individuals from each baseline sample (n=3816) and was analyzed against remaining baseline samples.

Both methods above use the estimated stock composition of the mixture sample, which might cause bias. Resampled baseline test individuals were therefore analyzed jointly with varying combinations of test samples (as described above) to assess the sensitivity of the power estimates to mixture sample composition. In order to evaluate the effect of mixture

sample composition while retaining full baseline data, mixture samples with varying combinations of test samples were constructed, analyzed and their effect on individual assignment patterns assessed.

Throughout the paper, the proportion of correctly identified individuals is referred to as the correct identification rate, applicable only for samples with known origin. The proportion of correctly identified individuals of all individuals assigned to a specific baseline population is referred to as the correct assignment rate. Efficiency was defined as the proportion of individuals in a group that were correctly identified and accuracy as the proportion of an identified group that truly belongs to that category (see Vähä and Primmer 2006 for details).

Results

Genetic structure of Teno salmon

Overall, genetic structuring within the Teno salmon population complex was strong and highly significant (Global F_{ST} = 0.065, p<0.001). All pairwise comparisons of genetic differentiation were statistically highly significant (p<0.001) apart from those among five Teno main stem samples (AK, PI, SI, GJ and KO) and between Teno main stem YK and GJ (Figure 1; F_{ST} =0.002, p=0.024; Bonferroni adjusted p-value for α = 0.05 was 0.022 across 630 correlated tests, Pearson's r =0.87). Estimates of genetic differentiation were affected by null alleles (mean d = 0.0007, paired t-test t = 21.9, d.f. 629, p<0.001), but this did not have a significant effect on the overall pattern (Pearson's r=0.999, p<0.001) or the overall level of

differentiation (mean pairwise F_{ST} 0.058 vs. 0.057 for null allele corrected F_{ST}; t=0.32, p=0.75).

Neighbor joining analysis based on genetic distance with and without null-allele correction (Cavalli-Sforza and Edwards 1967) provided the same tree topology (Figure 2). In general, patterns of genetic relationships among samples corresponded well with their geographical locations. There was a clear distinction between the Teno main stem and the headwater river systems, as well as between groups of tributaries draining to lower and upper parts of the main stem. Analysis of molecular variance (AMOVA), made in accordance with these geographical groupings apportioned 1.2% of the total genetic variation among groups and 5.6% between populations within groups. Variation among populations was highest in the 'lower tributaries' -group (10%) and lowest in the 'main stem' -group (0.64%, Table 1).

Closer inspection of the genetic structuring revealed less distinct patterns within the Inarijoki headwater group and Teno river main stem. In addition to very low levels of genetic differentiation between Iškorasjohka and Goššjohka (F_{ST} 0.008) as well as between Vuomajoki and Inarijoki (F_{ST} 0.008) within the Inarijoki basin, genetic clustering of individuals with the program STRUCTURE suggested no significant divergence between these samples (results not shown in detail). Similarly, in the Teno main stem, model-based clustering of individuals without sample location information did not support explicit population boundaries. However, despite low levels of genetic structure, a signal of isolation by distance was observed from Tana Bru to Outakoski (Mantel's r_{xy} = 0.638,

289 p=0.010) as well as from Teno main stem AK to YK (Mantel's r_{xy} = 0.58, p=0.026, see Figure
290 1 for sample locations).

291 The observed strong clustering of geographically close populations and the configuration
292 of regional groups of populations allowed their proper use as reporting groups in the
293 mixed-stock analyses. For subsequent power analyses, baseline samples from Teno main
294 stem AK, PI, SI, GJ and KO were pooled and referred to as the Teno main stem lower. This
295 pooled area alongside with the Teno main stem at TB, YK and OUT, and all the tributary
296 samples were tested separately for the power of resolution.

297

298 ## Power analyses

299 An analysis of simulated single-stock mixture samples provided the first reference point
300 for evaluating the resolving power of the baseline data. Across 32 baseline samples,
301 average identification rate was 91% (±12%) with 18 sites showing >95% identification
302 accuracy, implying a high resolution GSI power within the Teno river system (Table 1).

303 Analyzing simulated mixtures with equal proportions from all 32 baseline samples
304 provided a pattern similar to 100% simulations (Pearson's r=0.975 p<0.001, Figure 3).
305 Despite a lower identification rate (82% ± 19%) compared to single stock simulations, 15
306 baseline samples showed >95% correct identification rates. The correct assignment rate
307 was slightly higher (85% ± 19%) and higher than 95% for 16 sites.

ONCOR and cBAYES performed equally well in identifying the stock of origin for individuals in mixtures constructed by resampling (83.9% and 84.8%, respectively). The average site-specific identification rates (ONCOR: 77% ± 20% and cBAYES: 79% ± 21%) and the correct assignment rates (ONCOR: 84% ± 18% and cBAYES: 83% ± 18%) were slightly lower than results from simulated mixtures (Figure 3).

An assessment of factors affecting stock identification accuracy of the two methods indicated that the success rate is defined by the interaction of genetic divergence and baseline sample size (Figure 4; Table S2[1]). The effect of baseline sample size is large when the population divergence is low (F_{ST} < 0.05), while increasing sample size beyond 150 benefits only very little with highly diverged populations (F_{ST} > 0.07). The models predict cBAYES to require smaller sample size than ONCOR at any population divergence level for the same identification success.

There was a clear tendency for misassignments to occur between nearby tributaries or locations; as a corollary, higher stock identification success to regions or groups of nearby rivers was achieved (Table 1). For example, in equal proportion simulations 74% of the incorrectly identified TMS YK individuals were assigned to TMS lower, 13% to TMS OUT and 3% to TMS TB. Thus, despite low (36%) identification success to exact site of origin, 95% of TMS YK individuals were successfully identified to originate from the Teno main stem.

While estimated contributions to mixtures were relatively accurate (±10%) for the majority of sites, the tendency of some stocks to receive largest proportions of the misidentified

329 samples implied potential for biased estimates (Table 1). For example, in the equal
330 proportion mixtures, the contribution of TMS lower was overestimated by a factor of 2.1
331 and Inarijoki main stem by 1.4. Again, since misassignments occurred largely within
332 regions, significantly more accurate estimates were obtained at the regional level (1.12 and
333 0.95 for the Teno main stem and the Inarijoki region, respectively).

334 [1]Supplementary data are available through the journal Web site at xxx

335

336 ## Analysis of test samples

337 The estimated origins of test samples were to a large extent according to expectations for
338 both ONCOR and cBAYES (Table 2). Both methods correctly estimated all but one
339 (127/128) individual salmon (test sample 6) to originate from the Teno main stem lower
340 population. On the other hand, test sample 3 from TMS OUT fishery appeared to include
341 high proportion of transient individuals originating from the Inarijoki region. The most
342 significant difference in assignment patterns between the methods was observed for test
343 sample 4, which was presumed to include individuals originating from populations
344 within the Inarijoki region, including the Karigasjoki tributary. While cBAYES assigned
345 97% of the test individuals to expected local populations, the same was true only for 88%
346 of the samples by using ONCOR.

347 Analyses of mixture samples containing varying proportions of individuals from test
348 samples 4 (Inarijoki region), 5 (Kárášjohka-Iešjohka region) and from a pool of test
349 samples 1, 2, and 6 (Teno main stem) indicated that cBAYES performed better than

ONCOR in estimating the region of origin (Table 3). cBAYES identified more than 90% of the individuals from the headwater tributaries in all mixture assemblies, while the identification success with ONCOR varied from 78% to 92%. ONCOR systematically assigned lower than expected proportions of individuals to the Inarijoki region and higher than expected to Teno main stem. The effect of this bias is best illustrated when no Teno main stem samples were expected present in a mixture assembly; main stem contribution was then estimated at 7% with ONCOR and 1% with cBAYES.

A comparison of individual stock estimates indicated that results for both ONCOR and cBAYES were affected by mixture sample composition. Joint analyses of resampled mixture samples and different combinations of real test samples obtained the same stock estimate for 97.7% (±0.5%) of individuals with ONCOR and for 96% (±1.7%) of individuals with cBAYES. For the test samples the pattern was the opposite and more significant. For example, when test samples from the headwater regions were analyzed separately and in combination with varying proportions of the Teno main stem samples, ONCOR provided the same stock estimate only for 91% of individuals while 96% of the estimates were unaffected with cBAYES.

GSI of mixed-stock fishery samples

The present study was based on an improved, complemented baseline that allows for fine-scale assignment of individual salmon to 28 genetic stocks, whilst the original baseline, used in a study assessing run-timing of one-sea-winter Atlantic salmon (Vähä et al. 2011),

371 contained only 14 baseline stocks. A reanalysis of the previous study showed that regional
372 genetic stock estimates remained unchanged for 83% of the individuals. The largest
373 changes were observed for the Teno main stem (20%) and Inarijoki regions (31%) which
374 were expected given the now more thorough sampling and better understanding of the
375 genetic structure in these areas. The largest relative change in the estimated contribution
376 to the mixed-stock fishery was observed for the upper main stem tributaries (from 3.9% to
377 8.1%). The observed tendency of ONCOR to assign lower than expected proportion of
378 individuals to Inarijoki region and higher than expected to Teno main stem was evident
379 also in the apportionment of the mixed-stock fishery samples (Figure 5).

380

381 ## Discussion

382 In this study, we established a comprehensive genetic baseline reflecting the Atlantic
383 salmon population structure of the River Teno, with an assessment of factors affecting the
384 performance of genetic stock identification (GSI) within a population complex displaying
385 large genetic variation. In general, the large genetic variation among populations within
386 the river system coupled with the powerful marker system allowed individual salmon
387 from mixed-stock samples to be accurately identified to the population of origin.
388 Furthermore, the strong coherence of geographical and genetic relationships of
389 populations allowed defining rational units for monitoring and management of mixed-

390 stock fisheries. Below we discuss the main points relevant to the accuracy of the genetic
391 stock estimates and the applicability of GSI in fisheries management.

392

393 ## Genetic structure

394 Populations within the Teno river system display large variation in their degree of
395 differentiation and diversity, and the improved baseline revealed a significantly improved
396 geographical pattern from the earlier structure (Vähä et al. 2007). This was facilitated by
397 including several previously unsampled tributary populations, but also by replacing adult
398 salmon samples from the Teno main stem and the Inarijoki by juvenile samples in the
399 baseline data. In addition, the more dense sampling coverage allowed clarification of the
400 genetic structure among different areas of the main stem and discrimination between the
401 Teno main stem and the Inarijoki.

402 A robust genetic clustering of populations was highly coherent with geographical
403 locations as illustrated by the neighbor-joining analyses. Broadly, local populations
404 constituted five distinct groups of populations, wherein more local and strong clusters
405 were evident. This illustrates a strong hierarchical sub-structuring of Teno salmon and
406 necessitates adjusting management strategies accordingly. Altogether, after excluding
407 three small tributaries that likely do not have self-sustaining populations (Vuomajoki,
408 Iškorasjohka and Luovttejohka) and treating the middle part of the main stem (between
409 the large rapid sections Alaköngäs (TMS AK) and Yläköngäs (TMS YK)) as one, we

identified 28 genetically distinct demes within the Teno river system that is clearly more than the 16 demes identified using the earlier baseline (Vähä et al. 2007).

The level of genetic structuring observed within the Teno river system (mean pairwise F_{ST} 0.058) exceeds those generally reported for Atlantic salmon elsewhere, e.g. 0.02 in the River Moy (Ireland; Dillane et al. 2008) and 0.014 in the River Varzuga (Russian Kola Peninsula; Primmer et al. 2006). The genetic structure of Teno salmon is also high compared to salmon populations within larger geographical areas, e.g. an F_{ST} value of 0.02 in for salmon populations in Ireland (Ensing et al. 2013), 0.03 in Scotland (Gilbey et al. 2012) and 0.04 across southern Europe (Griffiths et al. 2010). The only geographic region with higher reported genetic divergence is the Baltic Sea area with F_{ST} at 0.12 (Koljonen 2006). As noted by Dionne et al. (2009), the overall level of genetic structuring and the distribution of genetic variation between local populations observed within the Teno river system is special in contrast to many other large river systems (cf. Primmer et al. 2006). The underlying reasons for this are unknown, but are likely linked with post-glacial colonization history of the Teno salmon (cf. Verspoor et al. 2007) and catchment-specific characteristics (Vähä et al. 2007; Ozerov et al. 2012). The observed high level of genetic differentiation among populations provides an excellent basis for deploying GSI as a tool enabling stock-specific management of the Teno mixed-stock fishery.

Components of genetic stock identification success

Conforming to the expectations, the overall accuracy of GSI, estimated on simulated single sample mixtures, was high (91% ± 12%) and 18 of the 32 baseline stocks demonstrated >95% accuracy. Illustrating the importance of genetic distinctiveness to successful stock estimation, 16 of the 17 populations with lowest pairwise genetic distance above 0.015 had assignment success higher than 95%. Unfortunately, the general scarcity of within-river studies and the lack of a GSI study of similar resolution do not allow for proper comparison, but the levels of accuracy in wider geographical surveys of Atlantic salmon stocks are commonly 20% or lower (e.g. Griffiths et al. 2010; Gilbey et al. 2012). In general, 90% accuracy levels are obtained only after defining larger regional groups (Ensing et al. 2013; Gauthier-Ouellet et al. 2009). Pacific salmon generally display higher genetic differentiation than Atlantic salmon (F_{ST} 0.06-0.10, Beacham et al. 2011 and references therein). Accordingly, similar or even higher levels of GSI accuracy have been reported for Pacific salmon populations both within river systems (e.g. Beacham et al. 2004) and over wider geographical areas (e.g. Beacham et al. 2006). These accuracy levels are similar to those reported in the present study.

In our study, GSI accuracy of a population depended on both the level of genetic differentiation and the baseline sample size (Figure 4). Genetically less distinct populations required larger baseline sample sizes than more diverged populations; a result in accordance with the findings by Beacham et al. (2011). Our mixed model for assignment success implied an optimal baseline sample size of >100 individuals for highly

450 diverged populations (population specific F_{ST} >0.07), and >200 for less diverged
451 populations. Mean sample size per population in the current Teno baseline was 103 (range
452 23-318; highest in combined Teno main stem population; Table S1) and the model
453 estimates imply that stock estimation accuracy could markedly benefit from increasing
454 sample size. Increasing the baseline sample size to ~200 individuals would nearly double
455 the identification rate in some populations. A population such as the Akujoki (n=53, F_{ST} =
456 0.021, identification rate ~50%) may present accuracy as low as 60% of the theoretical
457 maximum determined by the employed set of markers (see Beacham et al. 2011). On the
458 other hand, the Teno main stem lower population with high baseline sample size (n=318)
459 showed higher assignment success than expected by the level of genetic differentiation
460 alone (F_{ST} = 0.01, identification rate ~80%).

461 Stock estimation accuracy is critically dependent on the information content of the
462 employed genetic markers (Banks and Jakobson 2004). While some of the markers we
463 employed showed tendency for large allele dropout with low quality DNA extracts or
464 presence of null-alleles, inclusion of loci deviating from the Hardy-Weinberg proportions
465 in the baseline data have been proven not to hamper GSI (Carlsson 2008; Griffiths et al.
466 2010). GSI accuracy generally increases with number of loci and number of alleles per
467 locus (Kalinowski 2004; Beacham et al. 2006). In this respect, our baseline data, which
468 included genotypes from 33 microsatellite markers displaying more than 500 alleles in
469 total, is large. The number of markers applied in the present study is roughly a double
470 compared to common GSI studies, albeit depending on application and aimed resolution,

471 even very low number of markers may be sufficient (e.g. seven employed by Ensing et al.
472 2013).

473 Due to many practical reasons (e.g. variability, accessibility, and availability of statistical
474 methods) microsatellites have been the marker of choice for GSI studies for a long time.
475 Recent developments in SNP (Single-nucleotide polymorphism) techniques are however
476 making SNP markers more attractive also for studies where genetic information from
477 large numbers of samples is required. Recent studies have shown SNP markers to
478 significantly improve stock identification success in GSI studies, especially when the
479 discrimination power with microsatellites is initially low (Ozerov et al. 2013; Moore et al.
480 2014). However, despite the promising results, the costs for analyzing adequate number of
481 SNP markers is still high for GSI studies when thousands or tens of thousands of samples
482 are to be analyzed (see Moore et al. 2014). Microsatellite markers remain a cost effective
483 choice especially in systems where genetic structure and the suite of applied loci provide
484 adequate and sufficient resolution for the requirements.

485

486 ## Resolution power of GSI within Teno river system

487 In the absence of samples with known origin that are not included in the baseline data, the
488 accuracy of GSI is commonly estimated through simulations on equal proportion
489 mixtures. This approach provided lower level of stock identification accuracy (82% ± 19%)
490 than simulated single-sample mixtures (91% ± 12%). In accordance with the large variation
491 in genetic differentiation and sample sizes of the baseline populations, there was a large

492 variation among populations in the accuracy estimates, ranging from 36% for Teno main
493 stem (TMS YK) individuals to 100% for seven different tributaries. Overall, the lowest GSI
494 accuracies were observed for the Teno main stem and Inarijoki populations as well as for
495 the Nilijoki and Akujoki tributaries located in the upper main stem area. It is, however,
496 important to note that in all cases misidentified individuals were assigned to neighboring
497 populations within a region.

In the Teno main stem, high proportions of misassignments were observed between Teno main stem lower and TMS YK, and TMS OUT and TMS YK. However, treating the main stem as a single reporting unit resulted in high identification accuracy (94.4%). Similarly, there were substantial proportions of misassignments among some of the tributaries within the Inarijoki headwater system, but the identification rate of Inarijoki salmon at the group level was 92% with only 2.7% of individuals being misassigned from other regions. A third region, where a group level identification might more appropriate was the upper tributary group where the identification rate was 90%. Evidently, due to high proportions of misassignments among some of the tributaries (Table 1), interpretation of stock- or tributary-specific estimates within these regions deserves caution. As discussed above, increasing the baseline sample size to ~200 for individual tributaries will likely allow better population-specific identification within the upper tributaries and the Inarijoki system. Nevertheless, with the current baseline samples and the applied marker system, Teno salmon can be assigned to at least 17 geographically and genetically distinct groups with more than 90% identification rate.

Comparison of the methods

The selected statistical method may affect the accuracy of stock estimates and define the power of resolution. In accordance with Araujo et al. (2014), but in contrast to Griffiths et al. (2010), we found only small inconclusive differences between statistical methods in our simulation and resampling analyses. However, even though stock estimates were expected to be affected by the population composition of the mixture sample (since both

ONCOR and cBAYES utilize this information), the most salient difference between the two methods was detected in their sensitivity to population composition of the mixture sample. At the stock level, cBAYES appeared to be slightly more sensitive than ONCOR when all stocks were present in the mixture sample. However, the changes appeared largely among populations within region and ONCOR appeared significantly more sensitive at the region level. ONCOR significantly overestimated the contribution of Teno main stem stocks as illustrated with mixtures including only headwater samples in which case the Teno main stem contribution was estimated to 8% by ONCOR. This outcome, in addition to slightly higher overall accuracies, indicates a preference for using cBAYES over ONCOR when analyzing mixed-stock fishery samples; an observation in agreement with Griffiths et al. (2010) but in contrast with Araujo et al. (2014). Notwithstanding, the sensitivity of stock estimates to the population composition of mixture sample warrants dividing large mixture samples in subsets composed, as far as possible, of single stocks. For example, large mixture samples can be divided in subsets based on location, time or life-history characteristics of individuals (cf. Vähä et al. 2011).

Management implications and future prospects

The present study demonstrates that a comprehensive marker panel in combination with a baseline representation of populations enables reliable genetic identification of individual populations from mixed-stock samples. The resulting resolving power is a necessity for a sustainable stock-specific management of mixed-stock harvesting (e.g. Begg et al. 1999),

reducing the risk of overexploitation of components of the population complex and biodiversity loss (Crozier et al. 2004).

Use of the salmon resource in the Teno river system is important for the local communities, their economy and the indigenous Sámi culture. However, stock status assessments based on biological reference points (spawning targets; Falkegård et al. 2014), following the estimation procedure described by Forseth et al. (2013), reveal poor target attainment for several populations within the Teno (Anon. 2015). The resulting tradeoff, between conservation factors pointing towards a need for reduced harvesting and socio-economic factors pointing towards continued harvesting, can only be resolved through stock-specific knowledge.

The GSI procedure presented in this study is the primary solution to this. The GSI allows estimates of stock proportions in the Teno mixed-stock fishery, which can then be used to estimate stock-specific exploitation rates which are necessary to regulate fishing activity to sustainable levels (Thorley et al. 2007). Furthermore, the GSI reveals stock-specific information on run timing (Vähä et al. 2011), necessary for implementing temporal stock-specific fisheries restrictions. Combined with population-specific biological reference points (Falkegård et al. 2014), the GSI of the Teno mixed-stock fishery catches is an imperative requirement for future adaptive management of this diverse salmon population complex. This approach follows the guidelines of the North Atlantic Salmon Conservation Organization (NASCO) where both abundance and diversity criteria must be considered in a precautionary approach to salmon management (NASCO 2009). In

addition, estimation of population-specific spawning target attainment and cumulative fishing mortality will also enable the estimation of yearly abundance of salmon at sea prior to any fisheries (pre-fishery abundance; e.g. Potter et al. 2004). Protecting individual populations from overharvesting is required to maintain the diversity that stabilizes variance in salmon returns (cf. Schindler et al. 2010). Without conserving the roles of individual populations, the resilience that population diversity provides to fisheries will deteriorate well before the Teno salmon is extirpated.

Acknowledgements

We thank Mikhail Ozerov and two anonymous referees for valuable comments on earlier versions of the manuscript, Jari Haantie and Matti Kylmäaho for scale archive mining and scale pattern analyses and Kristiina Haapanen for laboratory assistance. Funding: the Academy of Finland (Project no: 133565) and the Norwegian Environment Agency.

References

Agashe, D. 2009. The stabilizing effect of intraspecific genetic variation on population dynamics in novel and ancestral habitats. Am. Nat. **174**: 255-267

Allendorf, F.W., Luikart, G., and Aitken, S.N. 2013. Conservation and the Genetics of Populations. Wiley-Blackwell.

Anderson, E.C., Waples, R.S., and Kalinowski, S.T. 2008. An improved method for estimating the accuracy of genetic stock identification. Can. J. Fish. Aquat. Sci. **65**: 1475-1486.

584 Anon. 2015. Status of the River Tana salmon populations 2015. Report of the Working Group on Salmon
585 Monitoring and Research in the Tana River System. Available at:
586 http://mmm.fi/documents/1410837/1801204/2015_Tenon_lohikantojen_tila_Status-of-the-
587 river-Tana-salmon-populations.pdf/aced31a7-af81-40aa-9c82-9e6147e58134

588 Araujo, H.A., Candy, J.R., Beacham, T.D., White, B., and Wallace, C. 2014. Advantages and challenges of
589 Genetic Stock Identification in fish stocks with low genetic resolution. Trans. Am. Fish. Soc. **143**:
590 479-488.

591 Banks, M.A., and Jacobson, D.P. 2004. Which genetic markers and GSI methods are more appropriate for
592 defining marine distribution and migration of salmon? North Pacific Anadromous Fish
593 Commission Technical Note **5**: 39–42.

594 Beacham T. D., Lapointe, M., Candy, J. R., McIntosh, B., MacConnachie, C., Tabata, A., Kaukinen, K., Deng,
595 L., Miller, K. M., and Withle,r R. E. 2004. Stock identification of Fraser River sockeye salmon
596 (*Oncorhynchus nerka*) using microsatellites and major histocompatibility complex variation.
597 Trans. Am. Fish. Soc. **133**: 1117–1137.

598 Beacham, T.D., Candy, J.R., Jonsen, K.L., Supernault, J., Wetklo, M., Deng, L., Miller, K. M., Withler, R.E., and
599 Varnavskaya, N. 2006. Estimation of stock composition and individual identification of Chinook
600 salmon across the Pacific Rim by use of microsatellite variation. Trans. Am. Fish. Soc. **135**: 861–
601 888.

602 Beacham, T.D., Winther, I., Jonsen, K.L., Wetklo, M., Deng, L., and Candy, J. R. 2008. The application of rapid
603 microsatellite-based stock identification to management of a Chinook salmon troll fishery off the
604 Queen Charlotte Islands, British Columbia. N. Am. J. Fish. Manage. **28**: 849–855.

605 Beacham, T.D., McIntosh B., Wallace C.G. 2011. A comparison of polymorphism of genetic markers and
606 population sample sizes required for mixed-stock analysis of sockeye salmon (*Oncorhynchus*
607 *nerka*) in British Columbia. Can. J. Fish. Aquat. Sci. **68**: 550–562.

608 Begg, G.A., Friedland, K.D., and Pearce, J.B. 1999. Stock identification and its role in stock assessment and
609 fisheries management: an overview. Fish. Res. **43**: 1-8.

610 Bolnick, D.I., Amarasekare P., Araujo M.S., Burger R., Levine J.M., Novak M., Rudolf V.H.W., Schreiber S.J.,
611 Urban M.C., and Vasseur D.A. 2011. Why intraspecific trait variation matters in community
612 ecology. TREE **26**: 183-192.

613 Bradbury, I.R., Hamilton, L.C., Rafferty, S., Meerburg, D., Poole, R., Dempson, J.B., Robertson, M.J., Redding,
614 D.G., Bourret, V., Dionne, M., Chaput, G., Sheehan, T.F., King, T.L., Candy, J.R., and Bernatchez, L.
615 2015. Genetic evidence of local exploitation of Atlantic salmon in a coastal subsistence fishery in
616 the Northwest Atlantic. Can. J. Fish. Aquat. Sci. **72**: 83-95.

617 Bradbury, I.R., Hamilton, L.C., Chaput, G., Robertson, M.J., Goraguer, H., Walsh, A., Morris, V., Reddin, D.,
618 Dempson, J.B., Sheehan, T.F., King, T., and Bernatchez, L. 2016. Genetic mixed stock analysis of
619 an interceptor Atlantic salmon fishery in the Northwest Atlantic. Fish. Res. **174**: 234-244.

620 Carlsson, J. 2008. Effects of microsatellite null alleles on assignment testing. J. Heredity **99**: 616-623.

621 Cavalli-Sforza, L.L., and Edwards A.W.F. 1967. Phylogenetic analysis: models and estimation procedures.
622 Evolution **21**: 550—570.

623 Chapuis, M.P., and Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation.
624 Mol. Biol. Evol. **24**: 621-631.

625 Crozier, W.W., Schön, P.-J., Chaput, G., Potter, E.C.E., Ó Maoiléidigh, N., and MacLean, J.C. 2004. Managing
626 Atlantic salmon in the mixed stock environment: challenges and considerations. ICES J. Mar. Sci.
627 **61**: 1344-1358.

628 Dillane, E., McGinnity, P., Coughlan, J.P., Cross, M.C., De Eyto, E., Kenchington, E., Prodöhl, P., and Cross,
629 T.F. 2008. Demographics and landscape features determine intrariver population structure in
630 Atlantic salmon (*Salmo salar* L.): the case of the River Moy in Ireland. Mol. Ecol. **17**: 4786–4800

631 Dionne, M., Caron, F., Dodson, J.J., and Bernatchez, L. 2009. Comparative survey of within-river genetic
632 structure in Atlantic salmon; relevance for management and conservation. Cons. Gen. **10**: 869-
633 879

634 Ensing, D., Crozier, W.W., Boylan, P., O'Maoiléidigh, N., and McGinnity, P. 2013. An analysis of genetic stock
635 identification on a small geographical scale using microsatellite markers, and its application in
636 the management of a mixed–stock fishery for Atlantic salmon (*Salmo salar*) in Ireland. J. Fish
637 Biol. **82**: 2080–2094.

638 Excoffier, L., Laval, G., and Schneider, S. 2005. Arlequin ver. 3.0: An integrated software package for
639 population genetics data analysis. Evolutionary Bioinformatics Online **1**: 47-50.

640 Erkinaro, J., Niemelä, E., Vähä, J.-P. , Primmer, C.R. , Brørs, S., and Hassinen, E. 2010. Distribution and
641 biological characteristics of escaped farmed salmon in a major subarctic salmon river.
642 Implication for monitoring. Can. J. Fish. Aquat. Sci. **67**: 130–142.

643 Falkegård, M., Foldvik, A., Fiske, P., Erkinaro, J., Orell, P., Niemelä, E., Kuusela, J., Finstad, A.G., and Hindar,
644 K. 2014. Revised first generation spawning targets for the Tana/Teno river system. NINA Report
645 1087. Available from http://www.nina.no/archive/nina/PppBasePdf/rapport/2014/1087.pdf

646 Falush, D., Stephens, M., and Pritchard, J.K. 2003. Inference of population structure using multilocus
647 genotype data: Linked loci and correlated allele frequencies. Genetics **164**: 1567–1587.

648 Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author.
649 Department of Genome Sciences, University of Washington, Seattle.

650 Foll, M., and Gaggiotti O.E., 2006. Identifying the environmental factors that determine the genetic
651 structure of Populations. Genetics **174**: 875-891.

652 Forseth, T., Fiske, P., Barlaup, B., Gjøsæter, H., Hindar, K., and Diserud, O.H. 2013. Reference point based
653 management of Norwegian Atlantic salmon populations. Env. Cons. **40**: 356-366.

654 Gamfeldt, L., Hillebrand, H., and Jonsson, P.R. 2008. Multiple functions increase the importance of
655 biodiversity for overall ecosystem functioning. Ecology **89**: 1223–1231.

656 Gauthier-Ouellet, M., Dionne, M., Caron, F., King, T.L., and Bernatchez, L. 2009. Spatiotemporal dynamics of
657 the Atlantic salmon (*Salmo salar*) Greenland fishery inferred from mixed-stock analysis. Can. J.
658 Fish. Aquat. Sci. **66**: 2040-2051.

659 Gilbey, J., Stradmeyer, L., Cauwelier, E., and Middlemas, S. 2012. Genetic investigation of the North East
660 English net fisheries. Marine Scotland Science Report 04/12.

661 Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3).
662 Available from http://www2.unil.ch/popgen/softwares/fstat.htm. Updated from Goudet (1995)

663 Griffiths, A.M., Machado-Schiaffino, G., Dillane, E., Coughlan, J., Horreo, J. L., Bowkett, A.E., Minting, P.,
664 Toms, S., Roche, W., Gargan, P., McGinnity, P., Cross, T., Bright, D., Garcia-Vazquez, E., and
665 Stevens, J.R. 2010. Genetic stock identification of Atlantic salmon (*Salmo salar*) populations in
666 the southern part of the European range. BMC Genetics **11**: 31.

667 Grimholt, U., Drablos, F., Jorgensen, S.M., Høyheim, B., and Stet, R.J.M. 2002. The major histocompatability
668 class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein
669 modelling. Immunogenetics **54**: 570–581

670 Habicht, C., Seeb, L.W., Myers, K.W., Farley, E.V., and Seeb, J.E. 2010. Summer–fall distribution of stocks of
671 immature sockeye salmon in the Bering Sea as revealed by single-nucleotide polymorphisms
672 (SNPs). Trans. Am. Fish. Soc. **139**: 1171–1191.

673 Harris, L.N., and Taylor, E.B. 2010. Genetic population structure of broad whitefish, *Coregonus nasus*, from
674 the Mackenzie River, Northwest Territories: implications for subsistence fishery management.
675 Can. J. Fish. Aquat. Sci. **67**: 905-918.

676 Hess, J.E., Whiteaker, J.M., Fryer, J.K., and Narum, S.R. 2014. Monitoring stock-specific abundance, run
677 timing, and straying of Chinook salmon in the Columbia River using genetic stock identification
678 (GSI). N. Am. J. Fish. Manage. **34**: 184–201.

679 Hilborn, R., Quinn, T. P., Schindler, D. E., and Rogers, D. E. 2003. Biocomplexity and fisheries sustainability.
680 Proc. Nat. Acad. Sci. **100**: 6564–6568

681 Huson, D.H., and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol.
682 Evol. **23**: 254–267.

683 Hutchinson, W.F. 2008. The dangers of ignoring stock complexity in fishery management: the case of the
684 North Sea cod. Biol. Lett. **4**: 693–695.

685 ICES. 2015. Report of the Working Group on North Atlantic Salmon (WGNAS) ICES CM 2015, ACOM:09.

686 Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W.S., Reich, P.B., Scherer-Lorenzen, M., Schmid, B.,
687 Tilman, D., Van Ruijven, J., Weigelt, A., Wilsey, B. J., Zavaleta, E.S., and Loreau, M. 2011. High
688 plant diversity is needed to maintain ecosystem services. Nature **477**: 199-202.

689 Kalinowski ,S.T., Manlove, K.R., and Tape,r M.L. 2007. ONCOR A computer program for Genetic Stock
690 Identification. Department of Ecology, Montana State University, Bozeman MT 59717. Available
691 for download from http://www.montana.edu/kalinowski

692 Kalinowski, S.T. 2004. Genetic polymorphism and mixed stock fisheries analysis. Can. J. Fish. Aquat. Sci. **61**:
693 1075–1082.

694 Koljonen, M.-L. 2006. Annual changes in the proportions of wild and hatchery Atlantic salmon (*Salmo salar*)
695 caught in the Baltic Sea. ICES J. Mar. Sci. **63**: 1274-1285.

696 Liu, K., and Muse, S.V. 2005. PowerMarker: Integrated analysis environment for genetic marker data.
697 Bioinformatics **21**: 2128-2129.

698 Mills, L.S. 2012. Conservation of Wildlife Populations. Demography, Genetics and Management. Wiley-
699 Blackwell.

700 Milner, G. B., Teel, D. J., Utter, F. M., and Winans, G. A. 1985. A genetic method of stock identification in
701 mixed populations of Pacific salmon, *Oncorhynchus* spp. Mar. Fish. Rev. **47**: 1–8.

702 Moore J.-S., Bourret V., Dionne M., Bradbury I., O'Reilly P., Kent M., Chaput G., and Bernatchez L. 2014.
703 Conservation genomics of anadromous Atlantic salmon across its North American range: outlier
704 loci identify the same patterns of population structure as neutral loci. Mol. Ecol. **23**: 5680-5697.

705 NASCO. 2009. Guidelines for the Management of Salmon Fisheries. North Atlantic Salmon Conservation
706 Organization (NASCO), Edinburgh, Scotland, UK. NASCO Council Document CNL 43.

707 Neaves, P.I., Wallace, J.R., Candy, J.R., and Beacham, T.D. 2005. CBayes: computer program for mixed stock
708 analysis of allelic data, version v4.02. Available: pac.dfo-mpo.gc.ca/sci/mlg/Cbayes_e.htm.

709 Nehlsen, W., Williams, J.E., and Lichatowich, J.A. 1991. Pacific Salmon at the Crossroads: Stocks at Risk from
710 California, Oregon, Idaho, and Washington. Fisheries **16**: 4-21.

711 Niemelä, E., Erkinaro, J., Julkunen, M., Hassinen, E., Länsman, M. and Brørs, S. 2006. Temporal variation in
712 abundance, return rate and life histories of previously spawned Atlantic salmon in a large
713 subarctic river. J. Fish Biol. **68**: 1222–1240.

714 Northrup, S., Connor, M., and Taylor, E.B. 2010. Population structure of lake trout (*Salvelinus namaycush*) in
715 a large glacial-fed lake inferred from microsatellite DNA and morphological analysis. Can. J. Fish.
716 Aquat. Sci. **67**: 1171-1186.

717 Ozerov, M., Veselov, A.E., Lumme, J. and Primmer, C.R. 2012. "Riverscape" genetics: river characteristics
718 influence the genetic structure and diversity of anadromous and freshwater Atlantic salmon
719 (*Salmo salar*) populations in northwest Russia. Can. J. Fish. Aquat. Sci. **69**: 1947–1958.

720 Ozerov, M., Vasemägi, A., Wennevik, V., Diaz-Fernandez, R. Niemelä, E., Prusov, S.,Kent, M.P., and Vähä, J.-
721 P. 2013. Finding markers that make a difference: DNA pooling and SNP-arrays identify
722 population informative markers for genetic stock identification. PLOS One **8**: e82434.
723 doi:10.1371/journal.pone.0082434

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

727 Pella, J., and Masuda, M. 2001. Bayesian methods for analysis of stock mixtures from genetic characters.
728 Fish. Bull. **99**: 151–167.

729 Potter, E.C.E., Crozier, W.W., Jan-Schön, P.-J., Nicholson, M.D., Maxwell, D.L., Prevost, E. Erkinaro, J.,
730 Gudbergsson, G., Karlsson, L., Hansen, L P, Maclean, J.C., O'Maoileidigh, N., and Prusov, S. 2004.
731 Estimating and forecasting pre-fishery abundance of Atlantic salmon in the Northeast Atlantic
732 for the management of mixed stock fisheries. ICES J. Mar. Sci. **61**: 1359-1369.

733 Primmer, C.R., Veselov, A.E., Zubchenko, A., Poututkin, A., Bakhmet, I., and Koskinen, M.T. 2006. Isolation
734 by distance within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*,
735 in tributaries of the Varzuga River in northwest Russia. Mol. Ecol. **15**: 653–666.

736 Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus
737 genotype data. Genetics **155**: 945–959.

738 Quinn, T. P. 1993. A review of homing and straying of wild and hatchery-produced salmon. Fish. Res. **18**: 29-
739 44.

740 Rannala, B., and Mountain, J.L. 1997. Detecting immigration by using multilocus genotypes. Proc. Nat. Acad.
741 Sci. USA **94**: 9197–9201.

742 Rosenberg, M.S., and Anderson, C.D. 2011. PASSaGE: Pattern Analysis, Spatial Statistics and Geographic
743 Exegesis. Version 2. Methods in Ecology and Evolution **2**: 229-232.

744 Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by
745 distance. Genetics **145**: 1219–1228.

746 Saitou, N., Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic
747 trees. Mol. Biol. Evol. **4**: 406–425.

748 Schindler, D.E., Hilborn, R., Chasco, B., Boatright, C.P., Quinn, T.P., Rogers, L.A., and Webster, M.S. 2010.
749 Population diversity and the portfolio effect in an exploited species. Nature **465**: 609–612.

750 Swatdipong, A., Vasemägi, A., Niva, T., Koljonen, M.-L., and Primmer, C.R. 2013. Genetic mixed-stock
751 analysis of lake-run brown trout *Salmo trutta* fishery catches in the Inari Basin, northern Finland:
752 implications for conservation and management. J. Fish Biol. **83**: 598-617.

753 Thorley, J.L., Youngson, A.F., and Laughton, R. 2007. Seasonal variation in rod recapture rates indicates
754 differential exploitation of Atlantic salmon, *Salmo salar*, stock components. Fish. Manage. Ecol.
755 **14**: 191-198.

756 Tilman, D., Wedln, D., and Knops, J. 1996. Productivity and sustainability influenced by biodiversity in
757 grassland ecosystems. Nature **379**: 718-720.

758 Vähä, J.-P., and Primmer, C.R.. 2006. Efficiency of model–based Bayesian methods for detecting hybrid
759 individuals under different hybridization scenarios and with different numbers of loci. Mol. Ecol.
760 **15**:63–72.

761 Vähä, J.-P., Erkinaro, J., Niemelä, E., and Primmer, C. 2007. Life-history and habitat features influence the
762 within-river genetic structure of Atlantic salmon. Mol. Ecol. **16**: 2638–2654.

763 Vähä, J.-P., Erkinaro, J., Niemelä, E., and Primmer, C. 2008. Temporally stable genetic structure and low
764 migration in an Atlantic salmon population complex: implications for conservation and
765 management. Evol. Appl. **1**: 137–154.

766 Vähä, J.-P., Erkinaro, J., Niemelä, E., Saloniemi, I., Primmer, C.R., Johansen, M., Svenning, M., and Brørs, S.
767 2011. Temporally stable population-specific differences in run timing of one-sea-winter Atlantic
768 salmon returning to a large river system. Evol. Appl. **4**: 39–53.

769 Vähä, J.-P., Wennevik, V., Ozerov, M., Diaz-Fernandez, R., Unneland, L., Haapanen, K., Niemelä, E.,
770 Svenning, M., Falkegård, M., Prusov, S., Lyzhov, I., Rysakova, K., Kalske, T., Christiansen, B., and
771 Ustyuzhinsky, G. 2014. Genetic structure of Atlantic salmon in the Barents region and genetic
772 stock identification of coastal fishery catches from Norway and Russia. Kolarctic ENPI CBC –
773 Kolarctic salmon project (KO197) report. Available at www.fylkesmannen.no/kolarcticsalmon.

774 Weir, B. S., and Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure.
775 Evolution **38**: 1358–1370.

776 Verspoor, E., Stradmeyer, L., and Nielsen, J.L. (eds) 2007. The Atlantic salmon: genetics, conservation and
777 management. Blackwell.

778 Youngson, A. F., Jordan, W. C., Verspoor, E., McGinnity, P., Cross, T., and Ferguson, A. 2003. Management
779 of salmonid fisheries in the British Isles: towards a practical approach based on population
780 genetics. Fish. Res. **62**: 193-209.

781

782 **Table 1** Genetic differentiation of baseline samples and results from single and multi-sample simulations in ONCOR. Column abbreviations: FST, population
783 specific estimate of genetic differentiation (Foll and Gaggiotti 2006); FST within, percentage of variation among populations within group; 100% simulation,
784 discriminatory power of the baseline for stock identification as predicted from single stock simulations in ONCOR (Kalinowski et al. 2007); Equal prop. id.
785 rate, proportion of correctly identified individuals in equal-proportion multi-sample simulations in ONCOR; Largest misassignment to, baseline stock with
786 largest proportion of incorrectly identified individuals; stock contribution, estimated contribution of stock to equal-proportion multi-sample simulation
787 mixtures; id.rate / stock contributions, proportion of correctly identified individuals and estimated stock contributions to a regional group.

baseline sample	F_{ST}	F_{ST} within	100% simulation	equal prop. id. rate	largest misassignment to		stock contributions	id. rate / stock contributions
Iešjohka headwaters								
Iešjohka Upper	0.035		92 %	79 %	Iešjohka Lower	12 %	82 %	90% / 97%
Iešjohka Lower	0.020		98 %	87 %	Kárášjohka	3 %	112 %	
Kárášjohka headwaters		3 %						
Kárášjohka	0.027		100 %	95 %	Iešjohka Lower	2 %	114 %	97% /105%
Bavttájohka	0.067		99 %	96 %	Kárášjohka	2 %	96 %	
Geáimmejohka	0.101		100 %	100 %			100 %	
Inarijoki headwaters								
Kietsimäjoki	0.072		99 %	96 %	Inarijoki MS	3 %	98 %	92% / 95%
Anárjohka	0.078		100 %	98 %	Inarijoki MS	1 %	99 %	
Inarijoki MS[1]	0.022		92 %	79 %	Inarijoki MS	6 %	144 %	
Cáščemjohka	0.106	3.6 %	97 %	92 %	Inarijoki MS	7 %	92 %	
Vuomajoki	0.032		67 %	47 %	Inarijoki MS	23 %	53 %	
Goššjohka	0.034		83 %	64 %	Iškorasjohka	12 %	80 %	
Iškorasjohka	0.029		79 %	61 %	Inarijoki MS	17 %	95 %	
Upper tributaries								
Váljohka	0.075		100 %	99 %	Kárášjohka	0.3 %	100 %	93% / 96%
Karigasjoki	0.032		96 %	85 %	Akujoki	4 %	104 %	
Akujoki	0.021		73 %	49 %	Nilijoki	19 %	83 %	
Báišjohka	0.058	4.23 %	87 %	72 %	Nilijoki	19 %	83 %	
Nilijoki	0.033		78 %	63 %	Akujoki	19 %	114 %	
Levajohka	0.038		93 %	83 %	Nilijoki	5 %	93 %	
Kuoppilasjoki	0.067		99 %	97 %	Nilijoki	2 %	102 %	
lower tributaries								
Tsarsjoki	0.221		100 %	100 %			100 %	100% / 100%
Kevojoki	0.083		100 %	100 %			100 %	
Utsjoki	0.094		100 %	100 %			100 %	
Vetsijoki	0.038	10 %	100 %	98 %	TMS lower	1 %	101 %	
Lakšjohka	0.169		100 %	100 %			100 %	
Ylä-Pulmankijoki	0.136		100 %	100 %			100 %	

	Galddasjoki	0.187		100 %	100 %			100 %	
	Máskejohka	0.042		100 %	98 %	TMS lower	0.3 %	100 %	
Teno main stem									
	TMS OUT	0.012	0.64 %	75 %	56 %	TMS YK	20 %	93 %	96% / 112%
	TMS YK	0.008		54 %	36 %	TMS lower	48 %	108 %	
	TMS Lower[2]	0.010		93 %	79 %	TMS YK	15 %	214 %	
	TMS TB	0.024		84 %	63 %	TMS lower	23 %	79 %	
	Luovttejohka	0.024		77 %	58 %	TMS lower	23 %	65 %	

788 [1]Inarijoki main stem

789 [2]Baseline samples from Teno main stem AK, PI, SI, GJ and KO were pooled and referred to as the Teno main stem lower, see text.

790

791

792 **Table 2** Estimated origins for the adult (TS 1-5) and juvenile (TS 6) salmon test individuals collected in different parts of the Teno river system (see Figure 1
793 for locations). Numbers within a grey shaded area are according to a priori expectations.

	Test sample 1		Test sample 2		Test sample 3		Test sample 4		Test sample 5		Test sample 6	
	ONCOR	cBayes	ONCOR	cBayes	ONCOR	cBayes	ONCOR	cBayes	ONCOR	cBayes	ONCOR	cBayes
01_Iešjohka Upper									2	3		
02_Iešjohka Lower	2	2			3	2	5	4	78	76		
03_Kárášjohka	1	1	1	1	3	2			76	80		
04_Bavttájohka									4	4		
05_Geáimmejohka					1	1			13	13		
06_Kietsimäjoki					3	3	7	19				
07_Anarjohka							5	5				
08_Inari MS	7	10	3	7	20	18	67	54				
09_Cáščemjohka							3					
10_Vuomajoki			1		4	7	1	4				
11_Goššjohka	3	1	2	2	15	26	73	125				
12_Iskurasjoki	2	2	1		8	4	45	15				
13_Karigasjoki					2		12	12				
14_Váljohka					1	1						
15_Akujoki					3	5						
16_Báišjohka												
17_Nilijoki												
18_Levajohka			1									
19_Kuoppilasjoki												
20_Tsarsjoki											1	1
21_Kevojoki	1	1										
22_Utsjoki	1											
23_Vetsijoki							3	1				
24_Lakšjohka	1	1										
25_Ylä-Pulmankijoki												
26_Galddasjoki												
27_Máskejohka									2	1		
28_TMS Outakoski	4	4	1	1	10	5	11					
29_TMS YK	6	4	4	6	26	25	7		6	4		
30_TMS lower	69	71	54	52			1	2			127	127
31_TMS TB	1	1	5	4								
32_Luovttejohka							1					

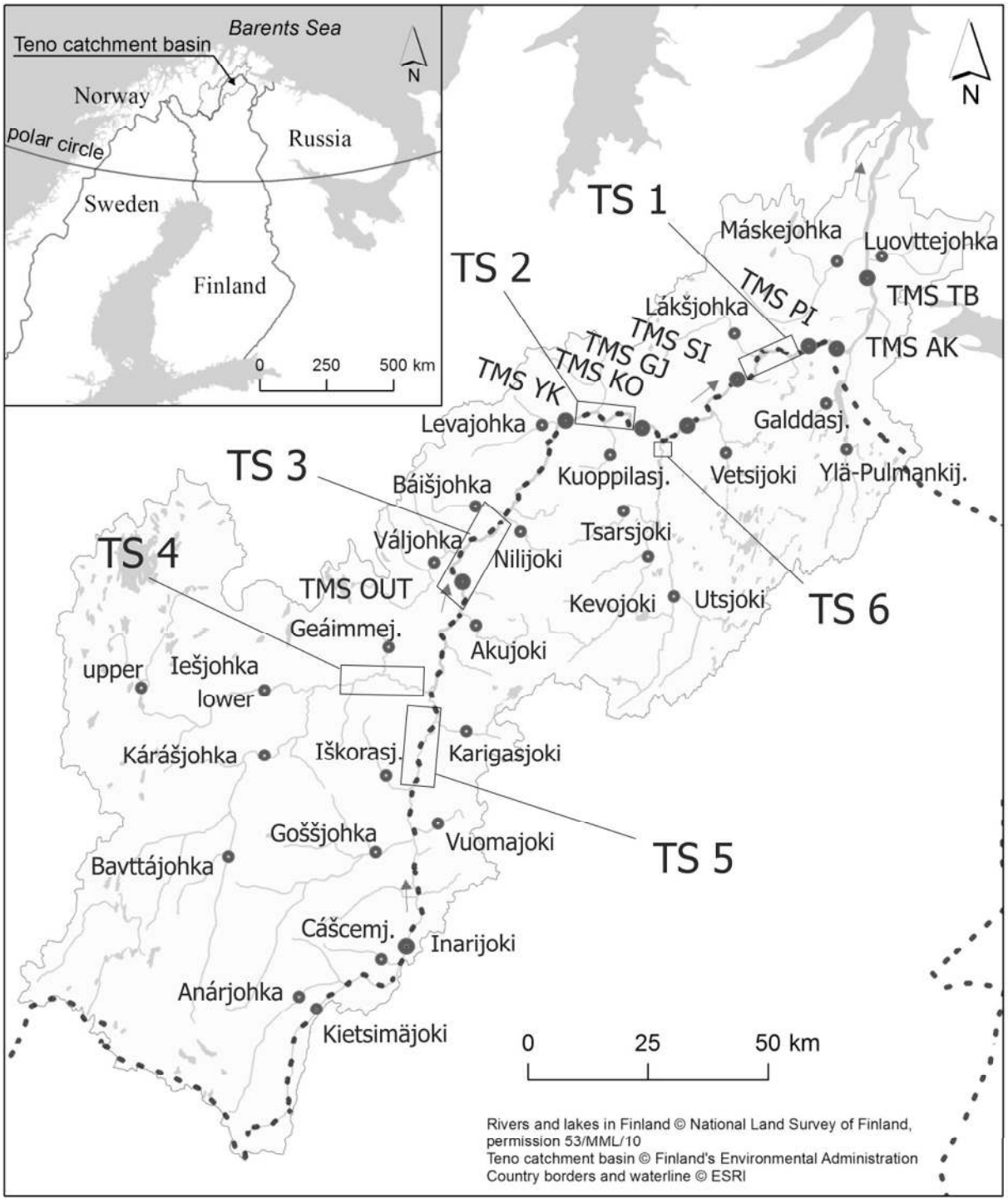
794

795 **Table 3** Effect of mixture sample stock composition to estimated stock proportions in ONCOR and cBAYES. 'Assembly of mixture sample' denotes the stock
796 composition of the mixture sample. 'Identified' denotes the proportion of correctly identified samples from each region. 'Assigned' denotes the proportion
797 of correctly identified individuals of all assigned to a region. Regions: INA=Inarijoki, KAR-IES= Kárášjohka- Iešjohka, TMS=Teno main stem.

798

assembly of mixture sample	**INA / KAR-IES / TMS 57% / 43% / 0%**		**INA / KAR-IES / TMS 40% / 40% / 20%**		**INA / KAR-IES / TMS 33% / 33% / 33%**		**INA / KAR-IES / TMS 20% / 20% / 60%**	
	ONCOR	cBAYES	ONCOR	cBAYES	ONCOR	cBAYES	ONCOR	cBAYES
INARI								
identified	84 %	95 %	81 %	94 %	81 %	94 %	78 %	90 %
assigned	98 %	99 %	93 %	93 %	88 %	88 %	77 %	77 %
est.contr	49 %	55 %	35 %	41 %	30 %	36 %	20 %	24 %
KAR-IES								
identified	92 %	97 %	91 %	94 %	89 %	92 %	88 %	91 %
assigned	92 %	95 %	92 %	95 %	91 %	95 %	90 %	94 %
est.contr	43 %	44 %	40 %	40 %	33 %	32 %	20 %	20 %
TMS								
identified	n.a.	n.a.	81 %	79 %	84 %	84 %	89 %	89 %
assigned	n.a.	n.a.	68 %	86 %	79 %	90 %	91 %	94 %
est.contr	7 %	1 %	24 %	19 %	35 %	31 %	58 %	56 %

799



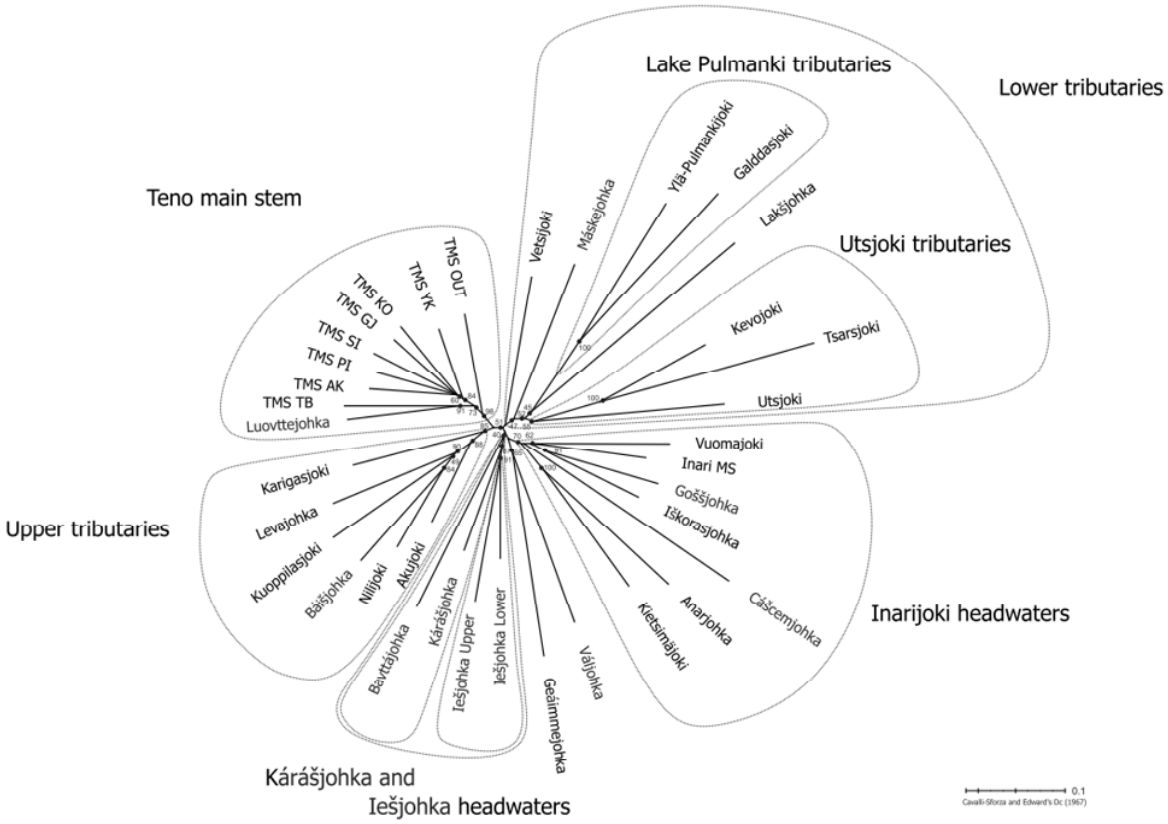
800

801 **Figure 1** Map of the Teno River system and its location in northernmost Europe. Locations of the sampled
802 baseline populations are indicated with circles (small circles – tributary samples, large circles – main stem
803 samples). Rectangles indicate sites where test samples (TS1-6) were collected.

804

805

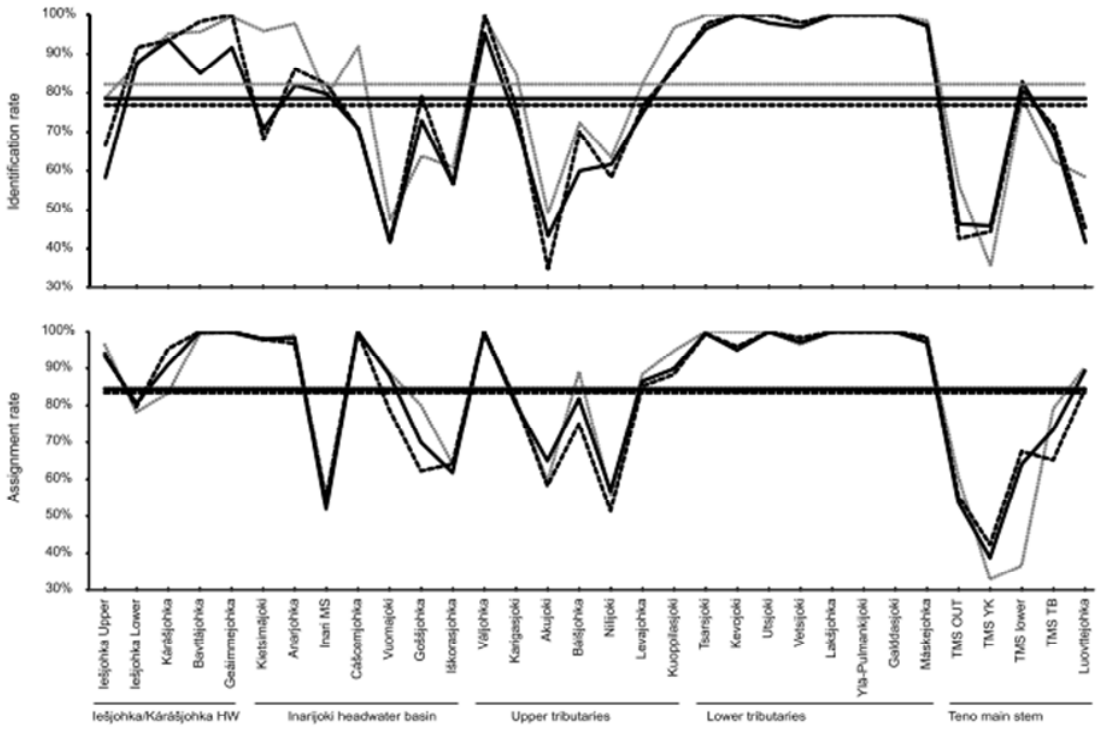
806



807

808 **Figure 2** Unrooted Neighbor-joining phylogram based on Cavalli-Sforza and Edwards' genetic distances
809 among samples collected from 36 sites within the Teno River system. Bootstrap values shown are in
810 percentage of 10 000 replicates.

811

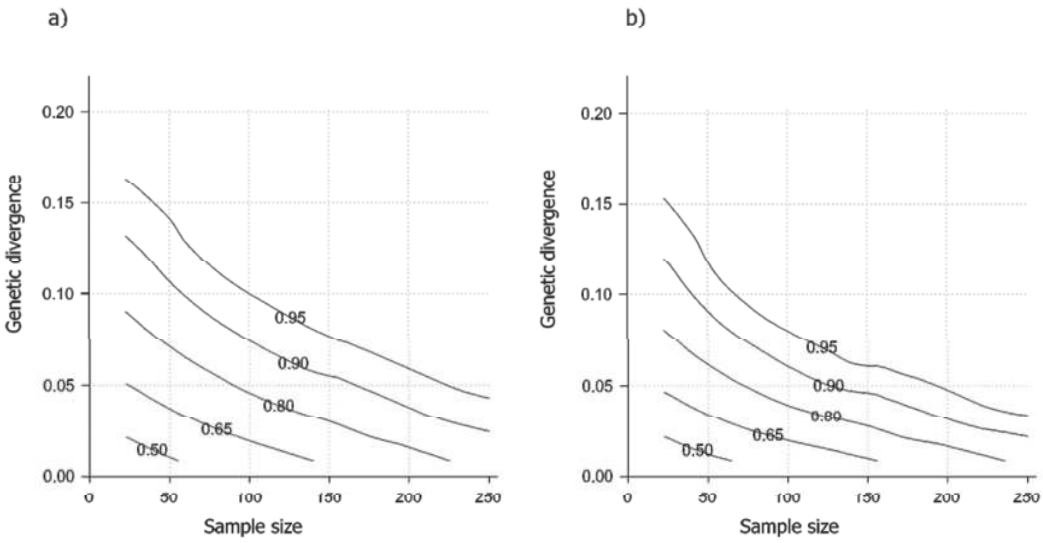


812

813 **Figure 3** Stock specific identification and assignment rate estimates from simulated equal proportion
814 mixtures with ONCOR (grey line) and from resampled baseline mixtures with ONCOR (solid black line) and
815 cBAYES (dashed black line). Horizontal lines show mean rates over all stock estimates. HW=large headwater
816 tributaries.

817

818

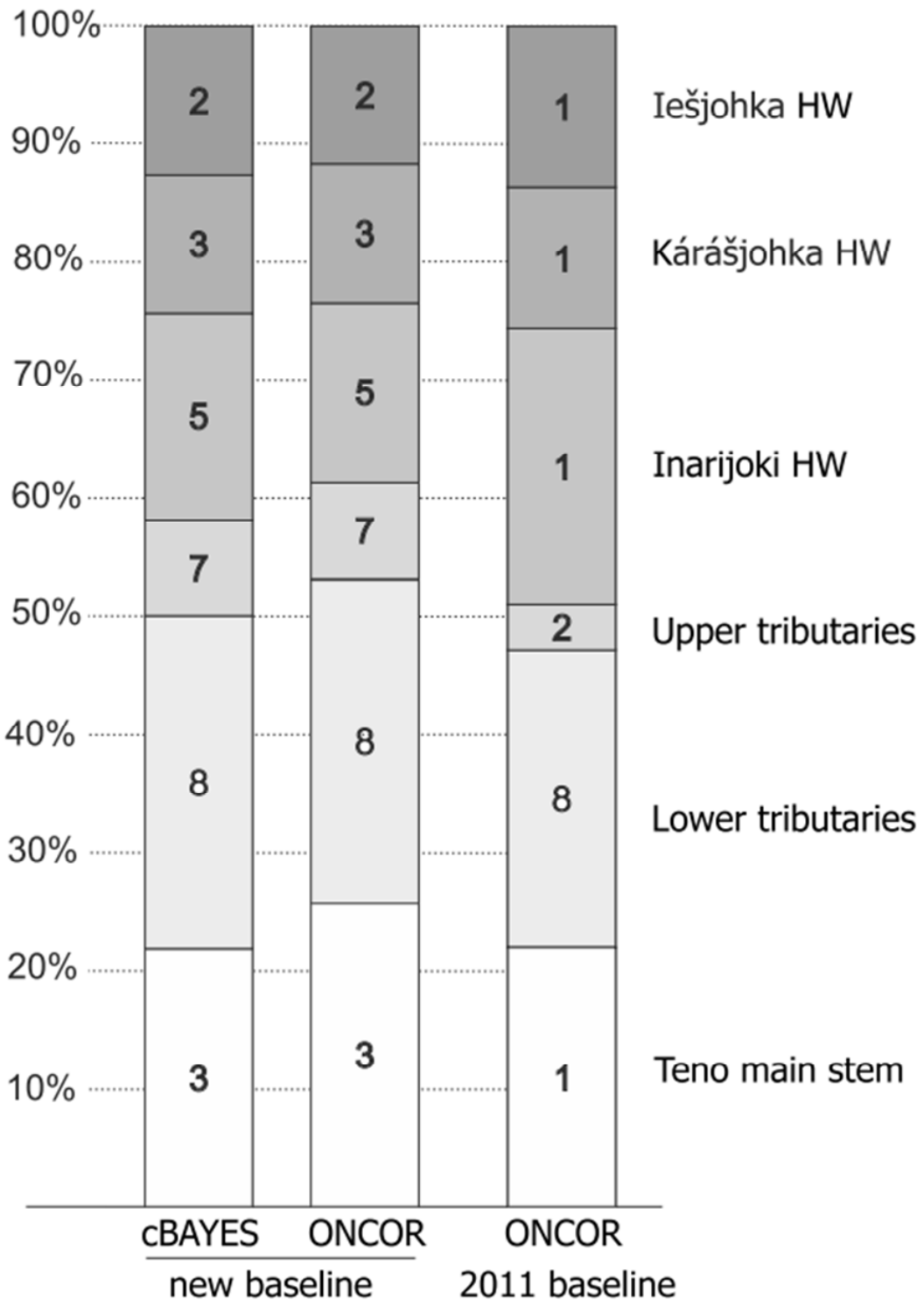


819

820 **Figure 4** Contour plot views of GLMM model predictions for identification success using a) ONCOR or b)
821 cBAYES versus genetic differentiation and baseline sample size.

822

823



824

825 **Figure 5** Estimated mixture proportions of mixed-stock fishery samples used by Vähä et al. (2011) as
826 inferred with cBAYES and ONCOR applying the new, improved baseline in comparison with the original
827 2011 baseline. Numbers in boxes refer to number of baseline populations within a region. Size of the box
828 refers to the estimated contribution of a regional group to the total sample. HW=large headwater tributary.

Vähä et al cjfas-2015-0606

Supplementary Data

Table S1 Sample details and genetic variation indices as estimated using 33 microsatellite loci. Column abbreviations: n, sample size; sample arrangement, reference to type of update from previously published data (see methods); H_E, expected heterozygosity; H_O, observed heterozygosity, A_R , allelic richness in 30 genes; PA_R , private allelic richness in 30 genes; F_{ST} within, percentage of variation among populations within 5 groups (see Figure 2); F_{ST} , GESTE (Foll and Gaggiotti 2006) estimates of genetic differentiation for each population; mean pw F_{ST}, genetic differentiation measured by average of the pairwise F_{ST} values; 100% simulations identification rate, discriminatory power of the baseline for stock identification as predicted from single stock simulations in ONCOR (Kalinowski et al. 2007).

Location (see Figure 1)	n	sample arrangement	# temporal samples	He	Hobs	AR (30 gene	pAR	F_{ST} within groups	F_{ST}	ave pw fst	100% simulations Identification rate
01_Iešjohka Upper	49	new	1	0.70	0.68	6.8	2.28		0.035	0.050	92 %
02_Iešjohka Lower	152	substituted	3	0.69	0.69	7.1	1.58		0.020	0.043	98 %
03_Kárášjohka	270	substituted	4	0.69	0.69	6.9	2.41	3 %	0.027	0.046	100 %
04_Bavttájohka	59	new	1	0.66	0.67	6.0	2.79		0.067	0.059	99 %
05_Geáimmejohka	52	new	1	0.65	0.65	5.4	0.57		0.101	0.082	100 %
06_Kietsimäjoki	61	new	1	0.66	0.64	5.8	0.24		0.072	0.066	99 %
07_Anarjohka	60	new	1	0.65	0.60	5.8	2.90		0.078	0.074	100 %
08_Inari MS	71	substituted	1	0.69	0.71	6.9	0.53		0.022	0.041	92 %
09_Cáščemjohka	23	new	1	0.65	0.69	5.5	1.81	3.6 %	0.106	0.084	97 %
10_Vuomajoki	31	new	2	0.68	0.69	6.6	0.88		0.032	0.050	67 %
11_Goššjohka	45	new	1	0.68	0.69	6.5	0.17		0.034	0.052	83 %
12_Iškorasjohka	52	new	1	0.68	0.68	6.6	0.40		0.029	0.046	79 %
13_Váljohka	79	supplemented	3	0.67	0.65	5.9	1.23		0.075	0.070	100 %
14_Karigasjoki	75	new	2	0.69	0.69	6.6	1.05		0.032	0.048	96 %
15_Akujoki	53	new	1	0.68	0.66	6.7	1.50		0.021	0.041	73 %
16_Báišjohka	51	new	1	0.66	0.65	6.0	0.86	4.2 %	0.058	0.061	87 %
17_Nilijoki	55	new	1	0.68	0.68	6.5	0.37		0.033	0.049	78 %
18_Levajohka	51	new	1	0.68	0.66	6.5	2.36		0.038	0.058	93 %
19_Kuoppilasjoki	112	supplemented	3	0.66	0.66	5.8	1.25		0.067	0.072	99 %
20_Tsarsjoki	196	supplemented	3	0.61	0.60	4.5	0.51		0.221	0.139	100 %
21_Kevojoki	165	supplemented	3	0.68	0.66	6.0	1.05		0.083	0.070	100 %
22_Utsjoki	129	supplemented	4	0.67	0.66	5.8	0.70		0.094	0.078	100 %
23_Vetsijoki	212	supplemented	5	0.69	0.69	6.7	3.18	10.0 %	0.038	0.049	100 %
24_Lakšjohka	71	supplemented	2	0.58	0.56	4.7	1.21		0.169	0.122	100 %
25_Ylä-Pulmankijoki	279	supplemented	6	0.65	0.63	5.5	3.48		0.136	0.097	100 %
26_Galddasjoki	85	supplemented	2	0.61	0.60	4.7	1.15		0.187	0.126	100 %

27_Máskejohka	121	supplemented	5	0.70	0.67	6.7	3.43		0.042	0.053	100 %
28_TMS Outakoski	91	substituted	1	0.70	0.70	7.2	1.44		0.012	0.041	75 %
29_TMS YK	128	substituted	1	0.70	0.70	7.3	0.70		0.008	0.039	54 %
30_TMS lower KO	73	substituted	2	0.71	0.70	7.3	1.17				
31_TMS lower GJ	58	substituted	1	0.71	0.71	7.3	1.93				
32_TMS lower SI	56	substituted	1	0.72	0.72	7.3	1.01	0.64 %	0.010	0.038	93 %
33_TMS lower PI	71	substituted	2	0.71	0.71	7.3	1.32				
34_TMS lower AK	60	substituted	1	0.71	0.71	7.2	1.12				
35_TMS TB	74	substituted	1	0.71	0.72	7.0	1.91		0.024	0.048	84 %
36_Luovttejohka	53	new	1	0.71	0.73	6.9	1.98		0.024	0.044	77 %

Table S2 The results of the GLMM models showing the variables influencing genetic stock identification success of a) ONCOR and b) cBAYES.

a)

Fixed effects	Estimate	±SE	Lower CL	Upper CL	DF	t value	p
Intercept	-0.548	0.148	-0.851	-0.246	28	-3.71	0.001
genetic differentiation	17.151	3.758	9.454	24.847	28	4.56	<0.001
baseline sample size	0.006	0.001	0.004	0.008	28	6.64	<0.001
gen. diff. x sample size	0.113	0.034	0.042	0.183	28	3.28	0.0028

b)

Fixed effects	Estimate	±SE	Lower CL	Upper CL	DF	t value	p
Intercept	-0.555	0.154	-0.870	-0.240	28	-3.61	0.001
genetic differentiation	17.758	4.279	8.994	26.523	28	4.15	<0.001
baseline sample size	0.005	0.001	0.003	0.007	28	4.67	<0.001
gen. diff. x sample size	0.203	0.041	0.118	0.287	28	4.93	<0.001

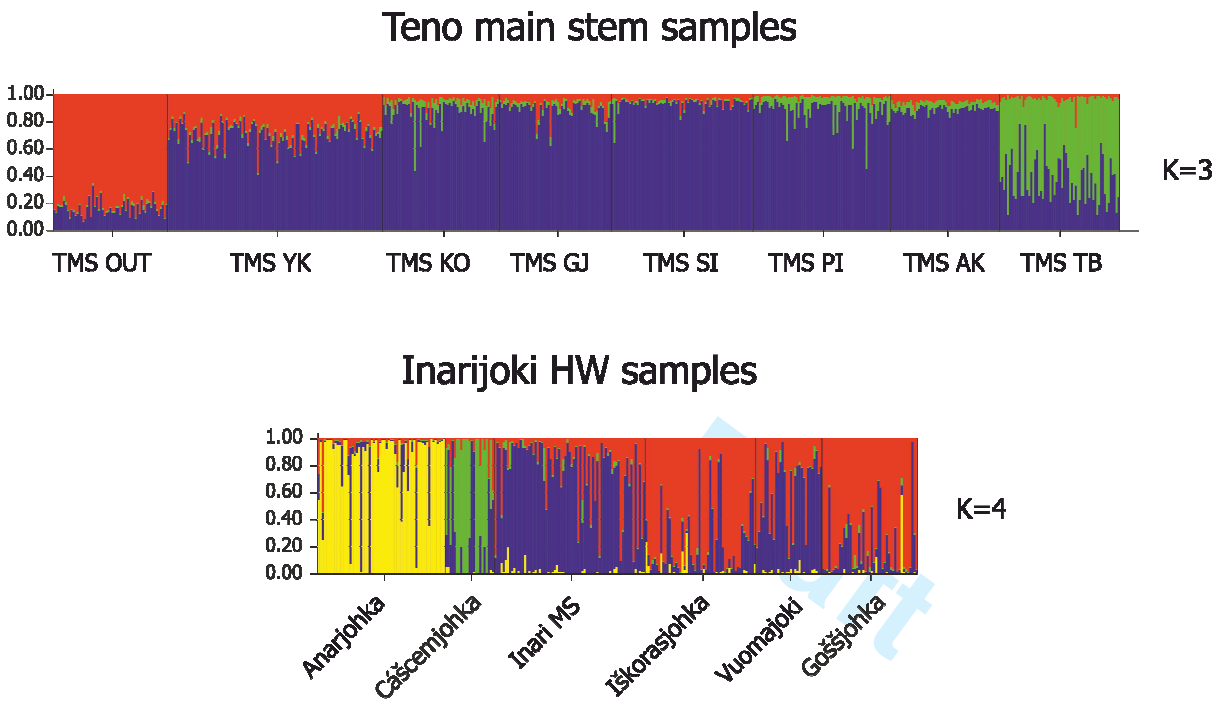


Figure S1. Population structure within Teno main stem and Inarijoki headwaters as inferred by STRUCTURE analysis. Each individual is represented by a vertical bar, which is partitioned into K-colored segment representing individual's estimated membership fractions in K clusters. Black lines separate individuals from different sampling sites. Given the number of K, the model of STRUCTURE pursues clustering solutions that are, as far as possible, in Hardy–Weinberg and linkage equilibrium. Used value for the K paratemer is shown next to clustering solution.