Journal of FISH BIOLOGY

fsbi An International Society for Fish Biology

Journal of Fish Biology (2016) **89**, 2658–2679
doi:10.1111/jfb.13149, available online at wileyonlinelibrary.com

From population genomics to conservation and management: a workflow for targeted analysis of markers identified using genome-wide approaches in Atlantic salmon *Salmo salar*[a]

T. Aykanat, M. Lindqvist, V. L. Pritchard and C. R. Primmer*

Department of Biology, University of Turku, Turku, 20014, Finland

A genotyping assay for the Ion Torrent Ion PGM platform was developed for fast and cost-effective targeted genotyping of key single nucleotide polymorphisms (SNPs) earlier identified using a genome-wide SNP array in Atlantic salmon *Salmo salar*. The method comprised a simple primer design step for multiplex-polymerase chain reaction (PCR), followed by two rounds of Ion Torrent Ion PGM sequencing to empirically evaluate marker efficiency in large multiplexes and to optimise or exclude them when necessary. Of 282 primer pairs initially tested, 217 were successfully amplified, indicating good amplification success (>75%). These markers included the *sdy* partial gene product to determine genetic sex, as well as three additional modules comprising SNPs for assessing neutral genetic variation ($N_{SNP} = 150$), examining functional genetic variation associated with sea age at maturity ($N_{SNP} = 5$), and for performing genetic subpopulation assignment ($N_{SNP} = 61$). The assay was primarily developed to monitor long-term genetic changes in *S. salar* from the Teno River, but modules are likely suitable for application in a wide range of *S. salar* populations. Furthermore, the fast and versatile assay development pipeline offers a strategy for developing targeted sequencing assays in any species.

Key words: diagnostic SNPs; genotyping by sequencing; multiplex; primer design; sea age at maturity.

INTRODUCTION

With the advent of the genomics era, the opportunity to use genomic tools to inform species conservation and management was greeted with much enthusiasm (Primmer, 2009; Allendorf *et al.*, 2010; Ouborg *et al.*, 2010). It has been noted, however, that such genomic approaches have yet to fulfill their potential in this area (Shafer *et al.*, 2015). Population genomic research on species of conservation concern is undoubtedly increasing (Thompson *et al.*, 2013; Harrisson *et al.*, 2014; Hoffmann *et al.*, 2015). There are relatively few cases, however, where results have been adapted for the practical needs of conservation and management, for example to identify management units (Finger *et al.*, 2011; Larson *et al.*, 2014), or develop diagnostic assays of key molecular

*Author to whom correspondence should be addressed. Tel.: +358 2 333 5571; email: craig.primmer@utu.fi

[a]This paper was presented at the FSBI Symposium, Bangor, in July 2016. Its content may not follow the usual style and format of the *Journal of Fish Biology*.

markers (Campbell *et al.*, 2012; Nussberger *et al.*, 2013; Candy *et al.*, 2015), possibly because of a lack of straightforward approaches for converting genomics discoveries into routine assays that can be applied in conservation and management (Shafer *et al.*, 2015).

Nevertheless, it is clear that analysis of population genomic data can inform many aspects of population management. Amongst other things, population-level genomic datasets enable identification of regions of the genome associated with specific phenotypic traits (for example, *via* genome-wide association analysis) and identification of genomic regions potentially underlying adaptive divergence (*via* F_{ST} outlier analyses and similar approaches). From a management point of view, such analyses enable the identification of relatively small subsets of markers that are associated with particular ecologically important phenotypes or that discriminate locally adapted populations. For example, genomic studies have enabled the development of panels of single nucleotide polymorphism (SNP) markers to identify interspecific hybrids (Campbell *et al.*, 2012; Pujolar *et al.*, 2014), or assign individuals to populations or management units (Russello *et al.*, 2012; Ogden *et al.*, 2013; Ozerov *et al.*, 2013; Candy *et al.*, 2015).

For such marker panels to be practically applied in species management, they need to be assayed rapidly and inexpensively in a large number of individuals (Narum *et al.*, 2013). Although several methods to rapidly genotype tens to hundreds of user-identified SNPs have been developed over the past decade, all these require access to dedicated platforms and have limitations to their flexibility and how economically they can be utilized. For example, microfluidic arrays (Fluidigm; www.fluidigm.com) allow the rapid and simultaneous genotyping of 96 or 384 SNPs on 96 or 384 individuals, but their cost-effectiveness for different combinations of SNP–individual number is limited by this fixed format. Custom SNP genotyping using mass spectrometry (Sequenom iPlex Gold; www.sequenom.com; Gabriel *et al.*, 2009) is more flexible for similar or smaller numbers of SNPs but the initial financial investment required for the platform is much larger. Furthermore, the previous platform of choice for genotyping a few hundred to a few thousand custom SNPs, the Illumina Golden Gate (www.illumina.com; Chao & Lawley, 2015), is no longer in production. Finally, a recently described method called Rapture (Ali *et al.*, 2016) combines restriction site associated DNA (RAD) sequencing and target selection using biotinylated bait sequences, which may provide a cost-efficient genotyping by sequencing alternative. This approach, however, is limited to SNPs previously identified from RAD sequencing and therefore on known RAD tags. It is likely to be difficult to optimize for target SNPs in other genomic locations, such as those previously identified through Sanger sequencing or those linked to candidate genes of interest.

Now that small-scale next generation sequencing machines are available for the laboratory benchtop (*e.g.* IonTorrent Ion PGM, Illumina MiSeq), genotyping of SNPs by targeted sequencing has become an attractive and flexible alternative. Campbell *et al.* (2015) recently presented a targeted sequencing protocol for genotyping SNP panels that had previously been assayed using a polymerase chain reaction (PCR) approach on the Fluidigm platform. When applied to several thousand individuals, this targeted sequencing approach demonstrated much greater cost efficiency than the previous methodology with similar genotyping reliability. The SNP markers described in that paper, however, had already been well-characterized and optimized for PCR-based genotyping. Thus, it is unclear whether the optimization of this genotyping protocol for newly discovered markers will be similarly straightforward.

In this paper, an optimization pathway is presented for developing modules of informative SNP markers identified from genomic studies of Atlantic salmon *Salmo salar* L. 1758, a species of both economic interest and conservation concern. Wild populations of *S. salar* support food fisheries and tourist industries, but have declined over recent decades as a result of over-fishing and habitat loss (Parrish *et al.*, 1998). In addition, the *S. salar* aquaculture industry is a major source of income in several countries. The cultural and economic importance of *S. salar* has resulted in the development of many genomic resources, including dense linkage maps (Moen *et al.*, 2004; Lien *et al.*, 2011), arrays enabling the simultaneous genotyping of many thousands of SNPs (Bourret *et al.*, 2013*a*, *b*; Barson *et al.*, 2015) and an annotated genome (Lien *et al.*, 2016). These resources have been used to address a wide range of evolutionary and ecological questions. For example, a SNP array interrogating seven thousand (7k) genome-wide markers (Illumina 7k array) has been utilized for a variety of purposes, including investigation of phylogenetic structure (Bourret *et al.*, 2013*a*), characterization of neutral and adaptive genetic variation (Bourret *et al.*, 2013*b*; Aykanat *et al.*, 2015), genetic stock identification (Ozerov *et al.*, 2013; Moore *et al.*, 2014) and identification of regions of the genome associated with phenotypic variation (Gutierrez *et al.*, 2012, 2014, 2015; Johnston *et al.*, 2014).

Recently, a single genomic region explaining a substantial amount of the variance in an ecologically important fitness-related trait, sea age at maturity has been pinpointed in European *S. salar* (Barson *et al.*, 2015). Age at maturity is influenced by both genotype and freshwater and marine environmental conditions (Jonsson *et al.*, 1991, 2016; Jonsson & Jonsson, 2007; Otero *et al.*, 2012; Barson *et al.*, 2015). The distribution of sea age at maturity in a population can be shaped by human activity, both indirectly *via* anthropogenic environmental change and directly *via* size selective fishing. There has been a recent global trend of decreasing sea age at maturity in *S. salar*, but the factors contributing to this trend are unknown (Chaput, 2012; ICES, 2013). Thus, a simple, inexpensive diagnostic assay for screening sea age genetic status of large numbers of individuals would be useful for management. Also, in the main stem of the Teno River of northern Finland, two cryptic subpopulations of *S. salar* have recently been identified that exhibit substantial phenotypic differentiation in size and age at maturity (Aykanat *et al.*, 2015). This is in addition to previously identified genetically distinct populations in various Teno River tributaries (Vähä *et al.*, 2007). A four-decade series of scale samples (yielding DNA) and demographic and environmental information is available for the Teno River *S. salar* stock, enabling detailed temporal investigation of how genetic and environmental factors interact to determine variation in age at maturity.

Here, the Teno *S. salar* population is used as a case study to outline a detailed procedure for transforming genome-wide information into diagnostic assays suitable for routine management use. Modules of SNPs were designed for different purposes: for sexing, for routine monitoring of key polymorphisms in several genes associated with sea age at maturity, baseline loci for standard population genetic analyses and highly diverged loci among populations (*i.e.* between the Teno River *S. salar* main-stem subpopulations) to facilitate subpopulation assignment. The method provides a flexible approach for designing primers for multiplex assays to suit the needs of a particular management or conservation question, as well as subsequent cost-efficient targeted sequencing.

MATERIALS AND METHODS

SNP MODULE CONSTRUCTION

A series of modules were designed, whereby each module consisted of a set of markers with differing objectives. These modules could be applied separately for a specific objective, with the advantage of sequencing more individuals (by reducing the number SNPs sequenced), used together in projects with multiple objectives in mind, or combined with new modules (*e.g.* SNPs to investigate hybridization between *S. salar* and escapees from aquaculture; Pritchard *et al.*, 2016). Modules introduced here are described below.

Baseline SNP module

This module consisted of SNPs that were selected in order to estimate the overall genome-wide level of differentiation among populations within the Teno system and quantify genetic drift over time ($N = 196$ baseline SNPs prior to optimisation; see SNP panel name in Table SI, Supporting Information). In order to maximize compatibility with other datasets, all SNPs were selected from those on the 7k SNP array that have been genetically mapped and previously utilized by Bourret *et al.* (2013*a*, *b*); Ozerov *et al.* (2013) and Aykanat *et al.* (2015). Although the 7k SNP array was originally optimized for use with Norwegian aquaculture *S. salar* strains and markers are not all expected to be informative in the Teno system, most selected SNPs were from the final dataset of Aykanat *et al.* (2015) ($N = 2874$) and therefore had been already filtered for high information content (minor allele frequency, $f_{MA} > 0{\cdot}05$). Here, information content was further increased by preferentially selecting SNPs that exhibited heterozygosity (H_O) $> 0{\cdot}2$ in the in the dataset (Aykanat *et al.*, 2015), though shift in H_O distribution in the selected SNPs was not significantly different from zero (Mann–Whitney $U = 118760$, $N_{1,2} = 152{,}2874$, $P > 0{\cdot}05$). The number of SNPs on a chromosome was roughly proportional to the length of the chromosome (Fig. S1, Supporting Information) and SNPs were selected to have linkage disequilibrium (LD) $< 0{\cdot}1$ with physically adjacent SNPs (r^2, Hill & Robertson, 1968). Pairwise LD was estimated from genotype information for the Teno River main-stem subpopulation 1 ($N = 347$; Aykanat *et al.*, 2015), using the r2fast function in GENABEL package 1.8.0 (Aulchenko *et al.*, 2007) implemented in R (www.r-project.org).

Outlier SNP module

This module consisted of SNPs from the 7k array that were highly diverged between the two sympatrically occurring *S. salar* subpopulations in the Teno River main stem. Although these two populations exhibit only low genetic differentiation ($F_{ST} \approx 0{\cdot}018$), they exhibit substantial phenotypic divergence (Aykanat *et al.*, 2015), thus for management it is important to discriminate them. The purpose of this SNP module was to assign individual fish sampled in the Teno main stem to their subpopulation of origin. SNPs were selected that displayed significantly higher differentiation ($P < 0{\cdot}05$) between the two Teno main-stem subpopulations than neutral expectations, *i.e.* outlier loci, corresponding to $F_{ST} \geq 0{\cdot}07$ as in Aykanat *et al.* (2015) (Table SI, Supporting Information), with Weir's and Cockerham's pairwise F_{ST} (Weir & Cockerham, 1984) using the HIERFSTAT package 0.04-10 (Goudet, 2005) in R 3.1.0.

Sea-age module

The sea-age module comprised six SNPs within the genomic regions identified by Barson *et al.* (2015) to be associated with sea age at maturity in *S. salar*. Five of these SNPs were within the associated region on chromosome 25 and included the SNP with the most significant association in the genome-wide association study (GWAS) analysis ($vgll3_{TOP}$ SNP) and two SNPs causing mis-sense mutations in the primary candidate gene for determining age at maturity (*vgll3*). The sixth SNP was within a chromosome 9 region strongly associated with the sea age at maturity phenotype prior to correction for population stratification (Barson *et al.*, 2015). This SNP is 34·5 kb from, and in complete LD with, the SNP named as $SIX6_{TOP}$ in Barson *et al.* (2015).

Sexing assay

The *sexing assay* comprised a region of the *sdy* gene that determines sex in *S. salar* (present = male, absent = female; Yano *et al.*, 2013). Although this gene is not expected to be present in females, contamination (*e.g.* during sampling in the wild) may result in PCR amplification of a small number of *sdy* gene copies in females. Thus, presence or absence of the *sdy* gene was determined by examining read coverage for the assayed region, corrected by total read coverage of all genomic regions sequenced. Clear divergence was expected in normalized *sdy* coverage between sexes, where females are expected to have zero, or close to zero, coverage after normalization.

SNP SELECTION AND PRIMER DESIGN

The SNP selection process differed for modules that targeted specific regions (the sexing and sea-age at maturity modules) and those that included a sub-set of the genome-wide SNPs included on the 7k SNP array (the baseline and outlier SNP modules). In all cases, BatchPrimer3 1.0 (You *et al.*, 2008) was used for primer design. Parameters for primer design were chosen to optimize the success of multiplex PCR and generate fragment sizes suitable for sequencing on the Ion Torrent (Thermo Fisher Scientific; www.thermofisher.com). For the initial design of all primers, parameters more stringent than the default parameters were employed: melting temperatures (T_M) were restricted to a narrow range (59–61° C), product size was limited to 75–120 bp and primers were constrained to have low 3′ self-complementary (≤2), self-complementary match (≤4) and 3′ poly-X (≤3). The choice of product size range was intended to generate fragments that could be optimally sequenced using the Ion Torrent whilst allowing the BatchPrimer3 software some flexibility in primer design around the SNP site. Since there were numerous candidate SNPs on the Illumina 7k SNP array for inclusion in the baseline and outlier SNP modules, only SNPs for which primers could be designed using these stricter parameters were retained. In contrast, the sexing and sea-age at maturity modules included specific loci of importance and thus the BatchPrimer3 primer design parameters described above were relaxed when necessary in order to identify suitable primers.

LIBRARY PREPARATION AND OPTIMIZATION

Library preparation comprises DNA extraction followed by two rounds of PCR and clean ups (Fig. 1). DNA was extracted from eight *S. salar* fin clips using QIAamp DNA mini kit (Qiagen; www.qiagen.com) and from eight scale samples with a protocol slightly modified from Elphinstone *et al.* (2003), where DNA is bound on silica beads in a 96 well filter plate (Vaha *et al.*, 2011). Extracted DNA was quantified with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific) and each sample normalized to 50 ng μl^{-1}. Library preparation was performed in two PCR steps: in the first step (PCR-1), SNP loci were amplified using primers that consisted of a SNP specific sequence and an additional tag sequence at the 5′-end of both the forward (tagF, 5′ ACGACGTTGTAAAA 3′) and reverse (tagR, 5′ CATTAAGTTCCCATTA 3′) primers (Hayden *et al.*, 2008; Bybee *et al.*, 2011). In the second PCR step (PCR-2), the Ion Torrent specific adapter sequences and sample specific barcodes were incorporated by targeting the tag sequences from step one (Fig. 1 and Table SII, Supporting Information). Thus the forward primer sequence (5′–3′) in step two consisted of three parts: Ion A adapter + Ion barcode1-96 + tagF and the reverse primer sequence (5′–3′) consisted of two parts: Ion trP1 + tagF (Table SI, Supporting Information).

Round one library optimization

Amplification for first step (PCR-1) was done in a multiplexed 10 µl reaction consisting of 1× Qiagen multiplex mastermix (QMP), equal amounts of forward and reverse primers at a final concentration of 0·1 μM and 50 ng of genomic DNA. Thermal cycling conditions were as follows: initial denaturation at 95° C for 15 min, seven cycles of 95–58–72° C for 30–60–45 s, respectively, and 15 cycles of 95–62–72° C for 30–60–45 s, respectively. Amplification for the second step (PCR-2) consisted of 1× Kapa Hifi HS ready-mix (Kapa Biosystems; www.kapabiosystems.com), forward and reverse primers at a final concentration of 0·3 μM

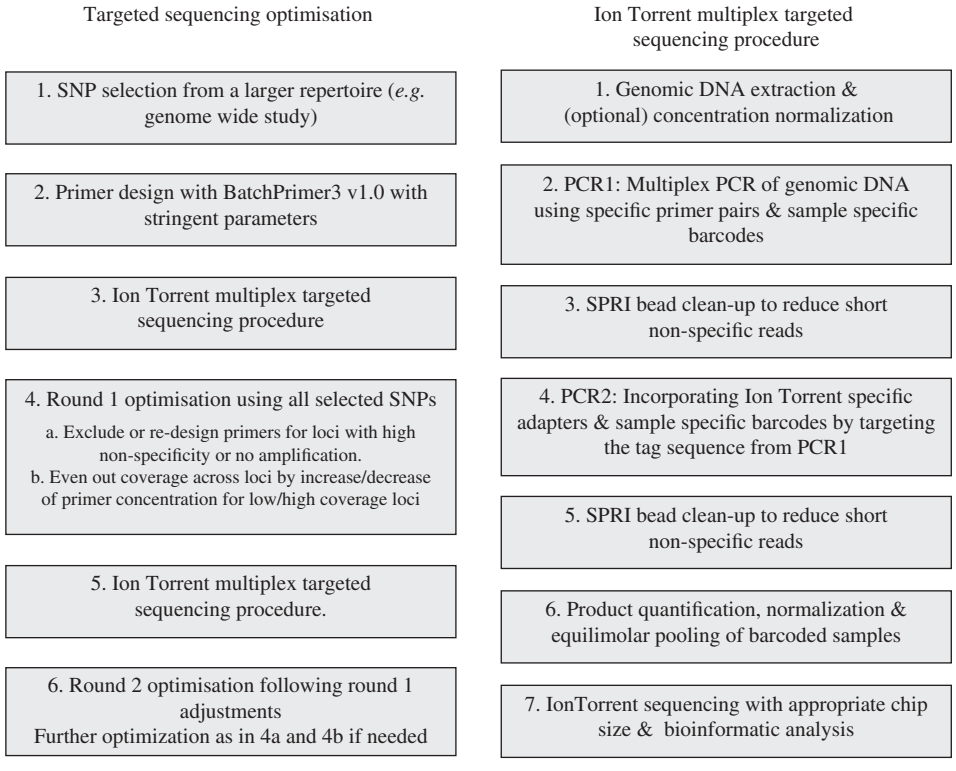

FIG. 1. Flowcharts summarising optimisation, and routine targeted genotyping procedures in this study of *Salmo salar* single nucleotide polymorphisms (SNPs). PCR, polymerase chain reaction; SPRI, solid phase reversible immobilization.

and 1 µl of diluted (1:10) PCR product from PCR-1. The total volume of PCR-2 was 10·8 µl. Thermal cycling conditions were as follows: initial denaturation at 95° C for 4 min, 15 cycles of 98–60–72° C for 20–15–20 s, respectively, followed by a final extension step at 72° C for 3 min. After PCR-2 the products were cleaned with solid phase reversible immobilization (SPRI) beads (Sera Mag, GE Healthcare Life Sciences; www.gelifesciences.com) to remove unincorporated primers and potential non-specific amplification products (Rohland & Reich, 2012; Vesterinen *et al.*, 2016).

Round two library optimization

The same protocols were used for PCR-1 and PCR-2 as in round one, except that primer concentrations were modified in PCR-1 based on the results from the first round optimization (0·05, 0·1 or 0·2 μM). Also, an extra SPRI-bead-cleaning step was employed after PCR-1 to reduce short reads originating from non-specific amplification. In order to evaluate the efficiency of this extra clean-up step, it was applied to 96 individually bar-coded reactions (in PCR-2), but skipped for 24 bar codes (Table SII, Supporting Information). To allow more samples to be multiplexed in the final library, an additional 6 bp bar code was added to the reverse primer used in PCR-2.

Library sequencing

Following the two rounds of PCR, the SPRI purified individual libraries were quantified by Qubit 2.0 fluorometer (Invitrogen; www.invitrogen.com) and pooled into one library in equal

amounts. The pooled library was quantified by running three replicates on a Bioanalyzer HS DNA chip (Agilent; www.agilent.com) and the mean value of replicates was used as the final concentration. The library was then diluted for the template preparation with an Ion PGM Hi-Q OT2 kit according to recommendations for Ion AmpliSeq DNA Library and OT2 for 200 bp reads and enrichment steps (ES) were done by following the manufacturer's instructions (Catalog Nr A27739, Publication Nr MAN0010902, Rev A.0). The Ion PGM Hi-Q sequencing kit was used for sequencing the templated sample on an Ion 316 Chip 2 following guidelines in the respective protocol (Catalog Nr A25592, Publication Nr MAN0009816, Revision D.0).

PCR multiplex optimization

A multi-round approach was adopted to optimize the PCR multiplex reaction used to generate fragments for sequencing (Fig. 1). Initially, all primer pairs were tested in two multiplexes: one for the baseline module and one for the outlier + sex + sea-age modules, and the resulting products were sequenced on the Ion Torrent. Primer pairs were excluded from the subsequent round of multiplex PCR if they failed to adequately amplify the targeted genomic region (mean depth of coverage of targeted region <10×) or strongly amplified non-target regions (mean depth of coverage of non-targeted regions coverage >20×). When more than one primer set was involved in the amplification of non-target regions, the primer set that had lower amplification (coverage) for the targeted region was excluded from the next round optimisation and the other set was retained. Additionally, the relative depth of sequencing coverage for the genomic regions targeted by the primer pairs was examined and adjustments were made to the concentrations of particular primer pairs in the multiplex with the aim of equalizing coverage. More specifically, primer concentrations of markers with either high (>100×) or low (10–20×) coverage were decreased or increased by two fold, respectively.

DATA ANALYSIS

Unless otherwise stated, all statistical analyses were performed using R software 3.2.5. A simple analytical pipeline, coded in R (similar to Angilletta *et al.* 2008) was used to process raw fastq output from the Ion Torrent server, align locus sequences and score SNP genotypes. Forward and reverse bar codes were used as individual identifiers to assign reads to uniquely bar coded samples (*e.g.* individuals). SNP genotypes were then identified by using a two-step procedure that was implemented to improve speed and performance. Reads were first assigned to a locus by matching either the forward or reverse primer sequence for that locus (allowing two mismatches). All reads assigned to a particular locus were then scanned for a perfect match to the 9 bp region surrounding the SNP site (4 bp upstream, 4bp downstream) and any non-matching reads discarded. Any read with phred score < 20 on the SNP base was also discarded at this stage. The remaining reads for each locus are henceforth termed on-target reads. Finally, the proportion of coverage for each allele was evaluated to score the genotype, as in Campbell *et al.* (2015). Briefly, genotypes were assigned based on the ratio of number coverage between allele 1 and allele 2. Allele 1–allele 2 > 10 was assigned as homozygous for allele 1, allele 1–allele 2 < 0·1 was assigned as homozygous for allele 2, allele 1–allele 2 between 0·2 and 5 was assigned as heterozygous and any proportion in between was not assigned a genotype. Likewise, genotypes were not assigned when total read coverage was <10. The genotyping fidelity of Ion Torrent genotyping was compared with the 7k Illumina SNP array with eight individuals that were also genotyped by the latter platform in Aykanat *et al.* (2015).

The proportion of non-specific primer products to the total number of reads was then evaluated and primer sets that contributed highly to non-specific read ratio were excluded. This was calculated by counting the number of non-specific primer–primer products, or primer dimers by searching every possible pairwise primer combination matches across all reads. In this search algorithm, only the first 10 5′ bp of the primer were used in order not to skip primer–primer dimers that may anneal imperfectly in the final product. This procedure was performed between forward and reverse primers only, because the directed PCR (PCR-2) uses forward and reverse primer tails as primers thus amplifying only forward–reverse primer combinations in the second PCR step.

The power of the different modules to successfully assign an individual to subpopulations in the Teno River main stem was assessed using data from Aykanat *et al.* (2015), in which individuals were assigned either to one of the subpopulations in the Teno River main stem or categorised as admixed (*e.g.* unassigned). For that, first the likelihood of a genotype to occur in a population was calculated using a frequency based method (Paetkau *et al.*, 1995), then the likelihood scores were evaluated by a comparison with the scores of 1000 simulated individuals per population (individuals simulated based on the observed genotyped frequencies in the two populations (Rannala & Mountain, 1997) using custom R scripts. Two criteria were used to assign an individual to either of the populations: an individual was assigned to a population if its likelihood (to occur in the population) was not within or below the least 5% of the simulated data and if the difference in log-likelihood scores for the alternative populations was greater than three (that is, an individual is 20·1 (e^3) times more likely to originate from one population relative to other). Otherwise it was marked as admixed. Assignment power was assessed with four different sets of SNPs: using all SNPs in Aykanat *et al.* (2015) ($N = 2874$), only with the SNPs in the outlier module, only with the baseline module, and using SNPs in two modules combined. In this procedure, the individual being evaluated for population assignment was excluded from the population pool when calculating population allele frequencies and missing genotype calls of tested individuals were accounted for by excluding the same SNPs when calculating the likelihood of simulated genotypes. Further, SNP loci with a minor allele frequency lower than 0·05 were adjusted to have a frequency = 0·05, to avoid zero or very small assignment probabilities when relying on a small number of loci.

RESULTS

Two optimization rounds of sequencing generated 2·87 and 2·00 million reads in the final libraries (Table I), with loading ratios of 77 and 47% and 37 and 25% polyclonalities, respectively. Of the total number of reads, 95·2 and 98·5% were successfully assigned to unique bar codes (forward and reverse identifiers were present in the reads).

TABLE I. Sequencing details for two optimization rounds for targeted analysis of markers identified using genome-wide approaches in *Salmo salar*

	Round 1	Round 2
Total number of single nucleotide polymorphisms	282	221
Total individually processed and bar-coded genomic DNA*	96	120
Total reads	$2{\cdot}87 \times 10^6$	$2{\cdot}00 \times 10^6$
Total on-target reads	$0.\ 93 \times 10^6$ (32·4%)	$1{\cdot}29 \times 10^6$ (64·7%)
Non-specific PCR products†	$1{\cdot}13 \times 10^6$ (39·3%)	$0{\cdot}15 \times 10^6$ (7·5%)
Mean number of markers per barcode‡	128·6 ($N = 282$), 101·1.0 ($N = 221$)	105·5 ($N = 221$)
Mean number on-target coverage per marker	69·0 ($N = 282$), 84·0 ($N = 221$)	98·2
% > 10× on-target reads	76·9 ($N = 282$), 92·1% ($N = 221$)	96·9%
% > 20× on-target reads	68·3 ($N = 282$), 82·7% ($N = 221$)	92·6%

PCR, polymerase chain reaction.

*Sixteen individuals were individually barcoded and processed to have 96 and 120 individually processed samples for round 1 and round 2, respectively. See also Table SII, Supporting Information.

†Non-specific PCR products consist of reads with non-specific primer match detected.

‡The number of markers included per barcode varies. See Table SII for details.

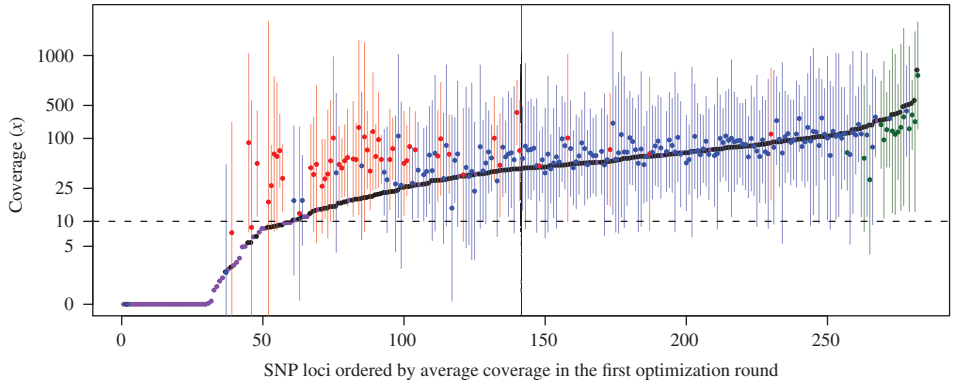

FIG. 2. Coverage per locus averaged across individuals (mean ± 2 S.D.) for each *Salmo salar* single nucleotide polymorphisms (SNP) in the first ($N = 282$; ● and ●), and second ($N = 221$, ●, ● and ●) Ion Torrent sequencing runs. Loci marked with purple were excluded for round 2 due to having little or no coverage or for having high unspecific coverage. —●—, markers for which primer concentrations were increased in round 2; —●—, markers where primer concentrations were decreased; —●—, blue indicates no adjustment in primer concentrations. NB. The *y*-axis is in the logarithmic scale.

The proportion of on-target reads in the first optimisation round was 32·4% (Table I). The mean coverage for on-target reads was 69·0× and included many loci with very low mean coverage (lower than 10×) and a few with very high coverage (Fig. 2). The non-target sequences resulting from non-specific primer pairing included 1·13 million reads, representing 39·3% of the total reads (Table I). This large number of non-specific primer products was driven by a small number of primers (Table SIII, Supporting Information) and non-specific amplification with coverage >20× accounted for 96·7% of the total detected non-specific amplification (Table SIII, Supporting Information). These primer sets were removed and 221 of 282 primer sets (78.4%) retained for the second optimisation round.

The second optimization round had substantially better performance than the first round. After excluding non-amplifying loci ($N = 49$) and loci with high non-specificity ($N = 12$), the proportion of on-target reads was 64·5%, in contrast with 32·4% for the first optimisation round (Table SIII, Supporting Information). Correspondingly, excluding primers with high non-specificity dramatically reduced the proportion of non-specific reads to 7·5% of all reads (Table I). In addition, primer concentration adjustments evened out coverage considerably (Figs 2 and 3 and Table SI, Supporting Information). The coverage was improved to a mean of 98·2 reads per marker compared with 84·0 reads in the first round (Fig. 2), despite the overall lower read number in the second round. Further, the samples with additional clean-up after PCR-1 showed significantly higher mean coverage compared with samples without this step (see Table SII, Supporting Information for comparisons). In the baseline module, mean ± S.D. coverage was 76·1 ± 41·7 and 48·4 ± 31·4 with and without clean-up treatment, respectively (*t*-test, $P < 0{\cdot}01$). Similarly, in the outlier + sea age + sexing modules, mean ± S.D. coverage was 110·0 ± 52·9 and 90·5 ± 52·9 with and without clean-up treatment, respectively (*t*-test, $P < 0{\cdot}05$).

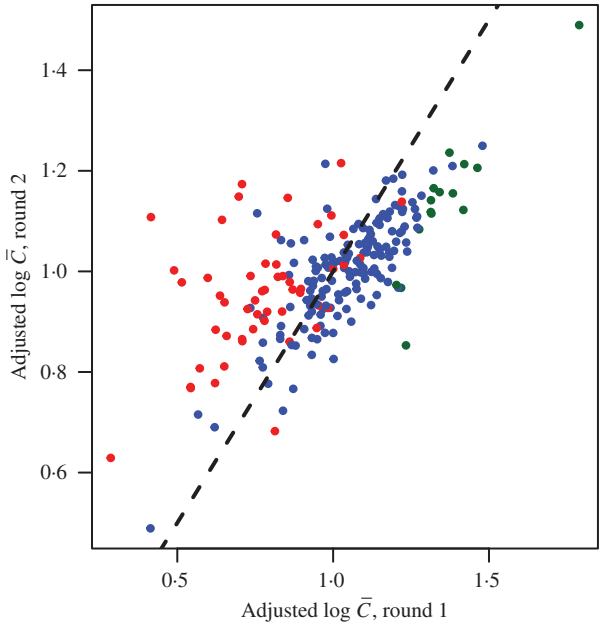

FIG. 3. Effect of primer concentration adjustment on (log) mean coverage ($\bar{C}$) after adjusting for on target coverages in round 1 and round 2 (*i.e.* round 1 and round 2 coverages were adjusted to be equal). ●, markers for which primer concentrations were increased for round 2; ●, markers where primer concentrations were decreased; ●, no adjustment in primer concentrations. Both decreased and increased primer concentrations resulted in significant changes in coverage in the desired direction [$P < 0{\cdot}001$, *t*-test compared adjustments groups with no adjustment, using log difference in coverage (round 1 *v*. round 2) as data points].

Molecular sexing estimated by read count of the *sdy* gene (normalised by mean read count) provided unambiguous results. It was also in perfect concordance with the phenotypically identified sex and *sdy* genotyping results obtained from gel electrophoresis (Fig. 4).

GENOTYPING

The overall genotyping call rate was 98·7% after excluding one individual bar code and one SNP locus that failed to amplify completely (Fig. S2, Supporting Information). Three loci had <85% and 16 (7·3%) had <95% genotyping success. In general, the distinction between heterozygote and homozygote calls was clear, although there were some indications of multisite variants (MSV; Gidskehaug *et al.*, 2011), as evidenced by the allele coverage ratios in some loci falling between the 50:50 and 100:0–0:100 ratios expected for heterozygotes and homozygote, respectively (Fig. 5). Within platform genotyping concordance was as high as 99·6% between replicate samples (*i.e.* DNA from same individuals bar coded and processed separately; see also Table SII, Supporting Information). Genotype concordance with eight individuals previously genotyped with the 7k Illumina iSelect platform (Aykanat *et al.*, 2015), was also high (95·7%). Most non-concordance between the two platforms appeared to be associated with presence of MSVs.

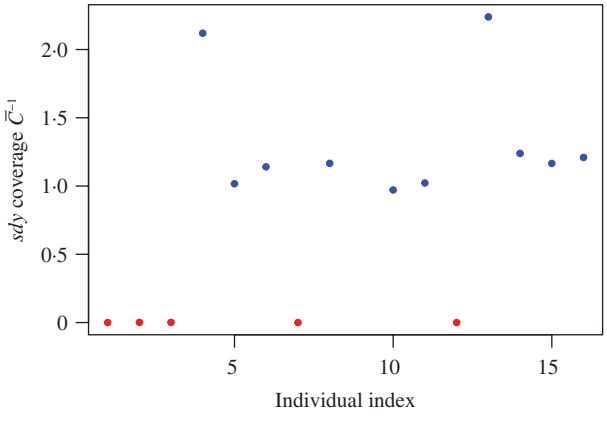

FIG. 4. Coverage of the *sdy* sexing locus after adjusting for average mean coverage ($\overline{C}$) per individual *Salmo salar*. Note that individual 9 failed in genotyping, hence sex was not determined. ●, phenotypically identified female and ●, male individuals.

UTILITY OF THE SELECTED MODULES

Overall there were 221 markers that were evaluated in optimization round 2. Of these, one marker was excluded for having no SNP in the product (N_1819) and four markers for very low genotyping success (*i.e.* <0·80, AKAP11_2, TN_2216, TN_2392). Thus, 61 markers remained in the outlier module, five in the sea age module, one in the sexing module and 150 in the baseline module ($N = 217$ in total). LD between markers was low, with a mean ± S.D. pairwise r^2 of 0·018 ± 0·018 and 0·016 ± 0·018 between the nearest SNPs on the same chromosome in the outlier and baseline modules, respectively. Among all pairwise LD tests ($N = 1830$ and 10878 for outlier and baseline modules, respectively), all but two comparisons had linkage < 0·15 (r^2), suggesting low LD overall, hence good separation of SNPs across the genome (Fig. S4, Supporting Information). Genetic differentiation between the two cryptic Teno River subpopulations estimated by the final set of SNPs in the baseline module ($F_{ST} = 0{\cdot}015$) was not significantly different from that estimated by the full SNP set reported in Aykanat *et al.* (2015) ($F_{ST} = 0{\cdot}018$; Mann-Whitney $U = 189880$, $N_{1,2} = 2874{,}137$, $P > 0{\cdot}05$; Fig. S3 and Table SI, Supporting Information). As expected, the outlier SNP module had much higher F_{ST} with a mean of 0·121.

When all 2874 SNPs from Aykanat *et al.* (2015) were used for population assignment, 94·4% of individuals that originated from subpopulation 1 or subpopulation 2 were correctly assigned to their respective populations (Fig. 6). Using just the 61 SNPs in the outlier module resulted in similarly high assignment power with 94·6% correct assignments. The baseline module correctly assigned only 76·0% of the individuals to their respective populations and combining both, assignment power was 95·0%. On the other hand, mis-assignment of admixed individuals was high in all panels tested, which is possibly a result of several hybrid cross types present within the admixed group combined with the small genetic difference between subpopulation 1 and subpopulation 2 (Aykanat *et al.*, 2015).

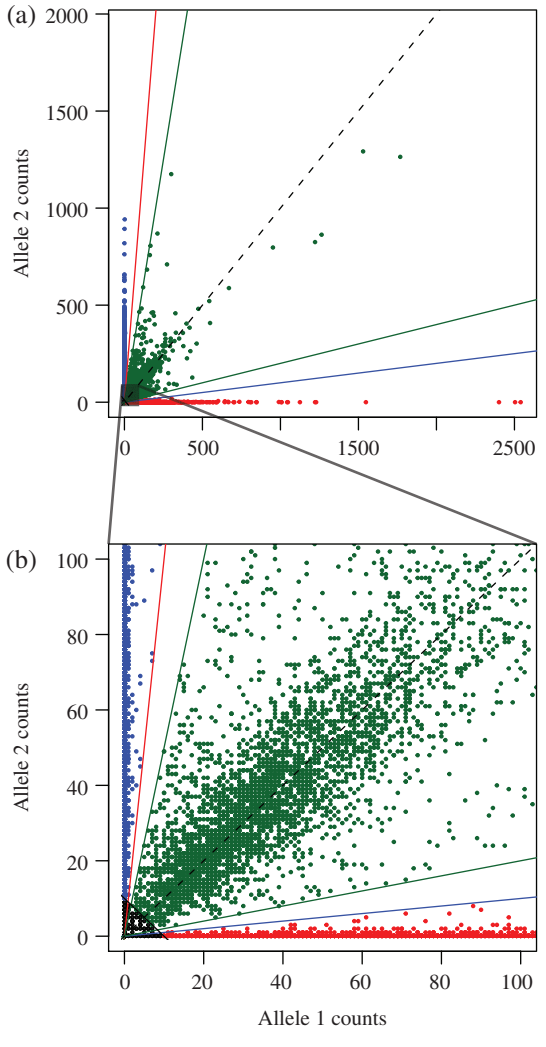

FIG. 5. (a) *Salmo salar* genotype calling based on allelic coverages shown for each data point: ●, homozygote genotype for allele 1; ●, homozygote genotype for allele 2; ●, a heterozygote genotype; ●, a no call genotype; ——, heterozygote thresholds where proportion of allele1 to allele 2 coverage is 0·2 and 5; ——, homozygote threshold where the proportion of allele 1 to allele 2 coverage is 0·1; ——, homozygote threshold where the proportion of allele1 to allele 2 coverage is 10. - - - -, equal coverage between alternative alleles. Samples excluded due to low total coverage are indicated below ——. The figure is drawn after the round 2 optimization and includes 120 individually barcoded samples and 219 loci (two loci with no single nucleotide polymorphisms regions including the sex determining locus and one additional locus were excluded from the analysis). (b) The data-rich section of (a), *i.e.* the area shown as ■.

DISCUSSION

This paper has presents an optimisation framework for developing modules of informative SNP markers derived from larger genomic datasets, in order to enable subsequent targeted resequencing in a cost-effective manner. The efficiency of this

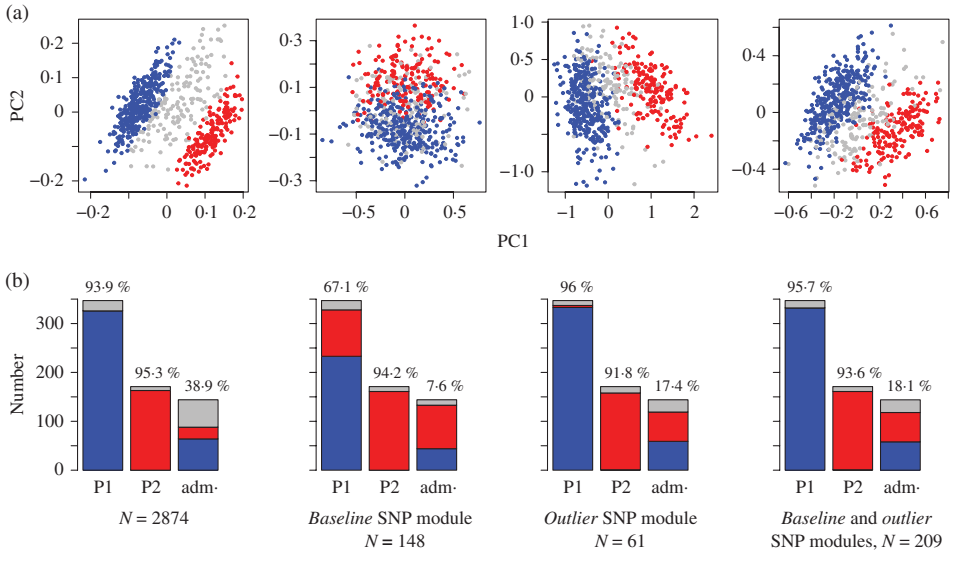

FIG. 6. Divergence and success of assigning individuals to (two subpopulations of *Salmo salar* in the Teno River main stem). (a) Plot of the first two principle components illustrating divergence: ●, individuals originating from subpopulation 1 ($N = 347$); ●, subpopulation 2 ($N = 171$); ●, admixed origin ($N = 144$). (b) Assignment success of individual *S. salar* to: ■, subpopulation 1; ■, subpopulation 2; ■, the admixed group. The percentage of correctly assigned individuals for each group of *S. salar* is shown above each bar plot. Results are given for all single nucleotide polymorphisms (SNP; $N = 2874$), and for the SNPs common between results from Aykanat *et al.* (2015) and the baseline and outlier modules presented here ($N = 148$ and 61, respectively).

framework was demonstrated by developing resequencing assays for >200 SNPs (from original Illumina 7k and Affymetrix 220k SNP arrays) in four modules, each designed to elucidate different aspects of the biology of *S. salar* populations in the Teno River main stem. The primers from the initial design step had a success rate >75% in the outlier and baseline modules (211 out 272 were retained after two rounds of optimisations). This demonstrates that the present simple primer design approach, without *in silico* optimization of primer multiplexing, is both suitable and versatile. In the targeted SNP sets, which included the sexing and sea-age modules, the only crucial locus that failed was the AKAP11 mis-sense SNP. Subsequent reassessments suggested the primer design failed as a result of inadvertent error in preparing the fasta files and a newly designed primer set was subsequently successfully tested and is presented in Table SI, Supporting Information. The approach enables rapid progress from extensive genome-wide information to more affordable targeted resequencing of specific SNPs in modules of particular interest or importance and is thus ideally suited for larger-scale monitoring of population over large spatial and temporal scales.

The main aim of the study was to present a framework enabling the simple optimisation of any set of SNP markers. Although the specific SNP modules presented here were designed based on data from Teno River main-stem subpopulations, the large number of loci described and presence of loci with significant functional properties (*i.e.* sex and sea age at maturity modules) makes it likely that these specific modules will also be of use in other populations. In particular, the genomic regions assayed by the sea-age

module are associated with life-history variation in at least three divergent evolutionary lineages of *S. salar* (Barson *et al.*, 2015). Further, the baseline module (potentially in combination with the outlier module) is expected to be useful for examining population genetic structure in *S. salar* populations beyond the Teno River system. Ozerov *et al.* (2013) suggested that 198 random SNPs (roughly *c.* 28 microsatellite loci with 460 independent alleles) should result in >95% correct assignment of origin across 23 *S. salar* populations. This suggests that the modules described here should be sufficient for successful genetic stock identification across the European range of *S. salar*. Although it should be noted that the discriminatory power of the outlier module alone is probably not as strong in other similarly diverged populations (*i.e.* $F_{ST} \approx 0{\cdot}015$) as in the Teno River main-stem subpopulations.

In the assay development pipeline, the presence of non-specific primer products (*e.g.* heterodimers) was quantitatively evaluated and primers with high non-specificity were excluded in the interest of increasing on-target reads and hence coverage. Multiplexing large numbers of loci reduces PCR and labour costs, but also increases the likelihood of nonspecific heterodimer formation. In this study, an empirical assessment methodology rather than *in silico* prediction was adopted to identify such artefacts, since information from the latter may not be sufficient to refine primer sets in large multiplexes. This was demonstrated by comparing the maximum binding energy (ΔG) values resulting from heterodimer formation of all possible pairwise primer combinations using the MultiPLX 2.1 software (Kaplinski *et al.*, 2005), with the empirical data. It was evident that non-specific product formation was related to higher ΔG values (Table SIV, Supporting Information). Many non-specific primer pairs with zero non-specific products, however, also had high binding energies (Table SIV, Supporting Information) and would have been excluded from the dataset if the assessment was made by theoretical means, suggesting empirical assessment of non-specific product formation is preferable over theoretical predictions. Likewise, an empirical assessment for non-specific product formation should be performed every time a new SNP is added to a multiplex module, which is not needed in hybridization based methods such as in rapture (Ali *et al.*, 2016).

The two-step optimisation protocol greatly improved on-target reads and at the same time balanced out the mean coverage per loci, thus reducing the final cost of sequencing to as low as €6·6 per sample for targeted resequencing of 200 SNPs with 50× coverage in 210 individuals (Table II). This cost could have been reduced even further, to as low as €6·2 per sample, if the SNPs were amplified in a single multiplex (which was not tested). The current cost estimate is slightly higher than the cost estimates of Campbell *et al.* (2015), which had similar number of loci in their SNP panel ($N = 192$). The reduced per-sample cost in the latter study is mostly a result of analysis of very high number of individuals per sequencing (*i.e.* >1000) by the use of a much larger scale next generation sequencing platform (Illumina HiSeq). In contrast, the lower capacity of the Ion Torrent Ion PGM platform results in a higher per sample cost. For example, the number of reads that can be obtained by Ion 316 Chip 2 is given as three million by the manufacturer, which equals to a capacity of genotyping only 300 individuals across 200 loci with an mean 50× coverage with 100% on-target reads (Table II). On the other hand, Ion Torrent is a bench-top platform accessible by many small to medium-scale laboratories and may benefit projects with medium-scale input such as monitoring populations with small census or sampling sizes. In addition, targeted regions in the in the optimised assays are relatively short, ranging from 97 and 156 bp (including 5′ and

TABLE II. A cost assessment overview for different sample inputs and expected sequencing outcomes in a targeted analysis of markers identified using genome-wide approaches in *Salmo salar*. Estimates are based on the anticipated number of individuals to be included in the sequencing run and subsequent adjustment of other variables

Number of loci	Number of individuals	Number of MPs*	Mean coverage	On-target read†	Total reads‡	Sequencing on Ion PGM (€)§	DNA extraction cost (€)¶	Library preparation cost (€)‖	Cost per sample (€)	Cost per genotype (€)
200	105	2	100	$2{\cdot}1\times10^6$	3×10^6	643	53–222	351	10·0–11·6	0·049–0·057
200	105	1	100	$2{\cdot}1\times10^6$	3×10^6	643	53–222	301	6·5–11·1	0·047–0·055
150	140	1	100	$2{\cdot}1\times10^6$	3×10^6	643	70–296	380	7·8–9·4	0·051–0·061
100	210	1	100	$2{\cdot}1\times10^6$	3×10^6	643	105–444	540	6·1–7·7	0·060–0·077
50	420	1	100	$2{\cdot}1\times10^6$	3×10^6	643	210–888	1017	4·5–6·1	0·090–0·121
200	210	2	50	$2{\cdot}1\times10^6$	3×10^6	643	105–444	657	6·6–8·2	0·033–0·041
200	210	1	50	$2{\cdot}1\times10^6$	3×10^6	643	105–444	557	6·2–7·8	0·031–0·039
150	280	1	50	$2{\cdot}1\times10^6$	3×10^6	643	140–592	716	5·0–6·6	0·036–0·046
100	420	1	50	$2{\cdot}1\times10^6$	3×10^6	643	210–888	1034	4·1–5·7	0·045–0·061
50	840	1	50	$2{\cdot}1\times10^6$	3×10^6	643	410–1776	1990	3·2–4·8	0·072–0·104

PCR, polymerase chain reaction.

*Anticipated number of multiplex reactions in the first PCR step.

†On-target reads are estimated to be 70% of the total reads.

‡Total read number is based on manufacturer's estimate.

§Including template preparation and sequencing on Ion PGM Hi-Q kits using one Ion 316 Chip 2 (+€10 for consumables).

¶DNA extraction costs are estimated for a range of methods spanning Chelex-100 (€0·50 per sample) and QIAamp DNA mini kit (€2·11 per sample).

‖Includes PCRs, two clean up steps, concentration measurements for pooling individual libraries (Qubit), concentration measurement of pooled library (Bioanalyzer, 1 HS DNA chip).

3′ tags) with SNP positions ranges between 33 and 112 (*e.g.* SNP position is within 100 bp from both ends in 90% of the loci; Table SI, Supporting Information), making the modules developed here suitable for application with many other next-generation sequencing platforms. Coupling the optimisation pipeline with a high capacity platform is likely to further reduce costs considerably.

The 7k SNP array was optimized for SNPs variable in aquaculture lines of Norwegian *S. salar* and as a result, a number of the SNPs are invariant or have low *maf* (and hence low heterozygosity) in wild *S. salar* populations (Bourret *et al.*, 2013*a*, *b*). The information content of such SNPs for population genetic applications (including, *e.g.* parentage assignment) within the Teno River system is therefore low. SNPs in the baseline and outlier modules have high diversity in the Teno River main-stem samples, as they were selected from SNPs in a dataset that had already been filtered to exclude low minor allele frequency loci (*maf* < 0·05, Aykanat *et al.*, 2015) and was further enriched for higher heterozygosity. This may have implications for application of the baseline module to *S. salar* populations that are genetically diverged from the Teno River population. This ascertainment bias, however, can be accounted for in population genetic inferences, *e.g.* allele frequency spectrum of neutral divergence can be modelled as a function of heterozygosity when estimating demographic histories across other populations (Excoffier *et al.*, 2013) and random genetic expectations (Ozerov *et al.*, 2015).

The assignment power of the *outlier* module to assign individuals of two subpopulations in the Teno River main stem was high despite the small number of SNPs in this module compared with the baseline module. Admixed individuals, however, had higher mis-assignment rate to either of the populations. This is probably due to stringent criteria employed in the original analysis in Aykanat *et al.* (2015) to distinguish backcross hybrids from pure-type fish combined with the low genetic structure between subpopulations. Correct identification of different hybrid classes is instrumental to estimate relative abundance of hybrids in the system and their variance structure in phenotypic traits, which would further help to understand mechanisms of reproductive isolation between two sympatric populations of the Teno River main stem (Aykanat *et al.*, 2015). Therefore, panels should be further evaluated for their power to distinguish different hybrid classes.

Results from the sexing assay were fully concordant with phenotypic sex as determined by fishers collecting the samples (Fig. 4). Recently though, some disagreement has been reported regarding the reliability of the *sdy* assay, *e.g.* in Chinook salmon *Oncorhynchus tshawytscha* (Walbaum 1792), where the *sdy* region amplified in *c.* 10% of females, suggesting the involvement of other factors in the sex determination pathway in this species (Cavileer *et al.*, 2015). The sample size of the present study ($N = 15$ fish; Fig. 4), was not sufficient to contribute to this discussion, but the primary motivation of including this assay is to provide sex information for field-collected samples when phenotypic sex determination is unreliable (*i.e.* sampling juveniles or immature adults). Furthermore, this sexing assay may be also potentially be utilised to identify the presence and basis of disconcordance in phenotypic and genotypic sex, if high samples sizes and additional life-history–physiological information is assured.

Large-scale sequencing, SNP discovery and genotyping efforts have always been a challenge in salmonids due to high rates of MSV SNP sites (Davidson *et al.*, 2010; Gidskehaug *et al.*, 2011; Lien *et al.*, 2011). MSV SNPs in *S. salar* stem from paralogous regions with high homologies, which are abundant in salmonid genomes as the

result of a recent whole genome duplication event (Lien *et al.*, 2016). In PCR applications, this results in primers that can bind to both paralogous copies and inadvertent amplification of the paralogous locus in addition to the target region. Such an artefact will result in alternative alleles deviating from expected ratios. When detected, correcting for such artefacts is relatively trivial by adjustment of the expected baseline of calls (Gidskehaug *et al.*, 2011). In this study, MSV markers could be identified by clear deviation of allelic coverages from expected proportions, which in turn resulted in disconcordant genotype calls compared with the Illumina 7k array. More than half of these loci, however, were not specified as MSVs in the Illumina 7k array (13 out of 25), which suggests re-designing primers for targeted sequencing may result in changes in MSV signatures. It should be relatively straightforward to model and correct for MSV signature once a sufficient number of individuals have been genotyped (Gidskehaug *et al.*, 2011).

In this study, a multi-purpose SNP panel with 217 SNPs and four modules is presented for use in conservation, management and evolutionary genetic studies in *S. salar*. The targeted genotyping approach provides a cost-effective alternative to genome scanning approaches when sample size is prohibitively large, such as in time-series analysis and large scale monitoring of populations, or when dense genome scans are simply not necessary to address the research of interest (Narum *et al.*, 2013). For example, targeted genotyping approaches are likely to be critical for data demanding analyses, such as estimation of the environmental and genetic components of long term changes in the life histories of populations and to predict future population composition (Hendry *et al.*, 2011; Piou & Prevost, 2013; Carroll *et al.*, 2014). *Salmo salar* populations have been intensively monitored over the past 50 years, which provides researchers and managers ample genetic resources by means of residual DNA in archived scale samples. The developed assay is intended to utilize this large genetic resource, particularly with Teno River main-stem populations in mind, but several modules are probably suitable for application in a wide range of populations at larger spatial scales. More generally, the fast and versatile assay development pipeline offers a strategy for developing targeted sequencing assays in any species.

A. Vasemägi is thanked for valuable discussions as is K. Salminen for laboratory assistance. This study was supported by the Maj and Tor Nessling Foundation (to T.A.; project number 201600445) and the Academy of Finland (to C.R.P.; grants 284941 and 137710). Codes for module optimisation and genotype calling are available in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.g0870.

Supporting Information

Supporting Information may be found in the online version of this paper:

Table SI. Details of single nucleotide polymorphisms (SNPs) targeted in this study of *Salmo salar.*

Table SII. Different multiplex conditions tested for different modules in optimization rounds 1 and 2.

Table SIII. Details of decision making to remove primers associated with high unspecific coverage (generally >20× per single nucleotide polymorphism pair).

Table SIV. Enthapies (ΔG) for pairwise primer pairs, ordered by strongest affinity to form a dimer. The top six pairs with highest unspecific coverage are marked in

bold (see also Table Unspecific). Only the top 76 combinations are shown (additional 22689 entries with $\Delta G < 10{\cdot}93$ are not shown). Although the six primer pairs producing exceptionally high non-specific coverage are in the top fraction of ΔG values, there are 92 unique primers with similar ΔG values but limited unspecific products. NA entries indicate at least one of the indicated primers was not included in the second optimization round.

FIG. S1. Location of 209 single nucleotide polymorphism loci used in the outlier and baseline modules across chromosomes, as a function of chromosome length together (a), and separately for (b) the outlier module, (c) the baseline module. Data point numbers indicate chromosome number. The chromosomal position of two loci in the outlier loci was not known and they are thus are not included here.

FIG. S2. Genotyping success rate (a) per individual barcode and (b) per locus.

FIG. S3. (a) Single nucleotide polymorphism (SNP) divergence between two Teno River main stem sub-populations. Background SNPs (pale colouring) are ordered by chromosomal position (data taken from Aykanat *et al.*, 2015). Overlaid red and blue triangles indicate SNPs from the neutral and outlier modules respectively. (b–d) Frequency histogram of SNPs shown in (a), for (b) all SNPs from Aykanat *et al.* (2015), the (c) baseline module and (d) outlier modules.

FIG. S4. Pairwise LD among single nucleotide polymorphisms (SNPs) in the (a) null and (b) outlier modules. (c) LD between nearest adjacent SNPs in the modules. Red and blue colours indicate the outlier and the baseline modules, respectively.

References

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C. & Miller, M. R. (2016). RAD capture (rapture): flexible and efficient sequence-based genotyping. *Genetics* **202,** 389–400. doi: 10.1534/genetics.115.183665

Allendorf, F. W., Hohenlohe, P. A. & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics* **11,** 697–709. doi: 10.1038/nrg2844

Angilletta, M. J., Steel, E. A., Bartz, K. K., Kingsolver, J. G., Scheuerell, M. D., Beckman, B. R. & Crozier, L. G. (2008). Big dams and salmon evolution: changes in thermal regimes and their potential evolutionary consequences. *Evolutionary Applications* **1,** 286–299. doi: 10.1111/j.1752-4571.2008.00032.x

Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23,** 1294–1296. doi: 10.1093/bioinformatics/btm108

Aykanat, T., Johnston, S. E., Orell, P., Niemela, E., Erkinaro, J. & Primmer, C. R. (2015). Low but significant genetic differentiation underlies biologically meaningful phenotypic divergence in a large *S. salar* population. *Molecular Ecology* **24,** 5158–5174. doi: 10.1111/mec.13383

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemela, E., Nome, T., Naesje, T. F., Orell, P., Romakkaniemi, A., Saegrov, H., Urdal, K., Erkinaro, J., Lien, S. & Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528,** 405–408. doi: 10.1038/nature16062

Bourret, V., Kent, M. P., Primmer, C. R., Vasemagi, A., Karlsson, S., Hindar, K., McGinnity, P., Verspoor, E., Bernatchez, L. & Lien, S. (2013*a*). SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology* **22,** 532–551. doi: 10.1111/mec.12003

Bourret, V., Dionne, M., Kent, M. P., Lien, S. & Bernatchez, L. (2013*b*). Landscape genomics in Atlantic salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. *Evolution* **67,** 3469–3487. doi: 10.1111/evo.12139

Bybee, S. M., Bracken-Grissom, H., Haynes, B. D., Hermansen, R. A., Byers, R. L., Clement, M. J., Udall, J. A., Wilcox, E. R. & Crandall, K. A. (2011). Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biology and Evolution* **3,** 1312–1323. doi: 10.1093/gbe/evr106

Campbell, N. R., Amish, S. J., Pritchard, V. L., McKelvey, K. S., Young, M. K., Schwartz, M. K., Garza, J. C., Luikart, G. & Narum, S. R. (2012). Development and evaluation of 200 novel SNP assays for population genetic studies of westslope cutthroat trout and genetic identification of related taxa. *Molecular Ecology Resources* **12,** 942–949. doi: 10.1111/j.1755–0998.2012.03161.x

Campbell, N. R., Harmon, S. A. & Narum, S. R. (2015). Genotyping-in-thousands by sequencing (GT-seq): a cost effective SNP genotyping method based on custom amplicon sequencing. *Molecular Ecology Resources* **15,** 855–867. doi: 10.1111/1755-0998.12357

Candy, J. R., Campbell, N. R., Grinnell, M. H., Beacham, T. D., Larson, W. A. & Narum, S. R. (2015). Population differentiation determined from putative neutral and divergent adaptive genetic markers in eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources* **15,** 1421–1434. doi: 10.1111/1755-0998.12400

Carroll, S. P., Jorgensen, P. S., Kinnison, M. T., Bergstrom, C. T., Denison, R. F., Gluckman, P., Smith, T. B., Strauss, S. Y. & Tabashnik, B. E. (2014). Applying evolutionary biology to address global challenges. *Science* **346,** 1245993. doi: 10.1126/science.1245993

Cavileer, T. D., Hunter, S. S., Olsen, J., Wenburg, J. & Nagler, J. J. (2015). A sex-determining gene (sdy) assay shows discordance between phenotypic and genotypic sex in wild populations of Chinook salmon. *Transactions of the American Fisheries Society* **144,** 423–430. doi: 10.1080/00028487.2014.993479

Chao, S. & Lawley, C. (2015). Use of the Illumina GoldenGate assay for single nucleotide polymorphism (SNP) genotyping in cereal crops. *Methods in Molecular Biology* **1245,** 299–312. doi: 10.1007/978-1-4939-1966-6_22

Chaput, G. (2012). Overview of the status of Atlantic salmon (*Salmo salar*) in the North Atlantic and trends in marine mortality. *ICES Journal of Marine Science* **69,** 1538–1548. doi: 10.1093/icesjms/fss013

Davidson, W. S., Koop, B. F., Jones, S. J., Iturra, P., Vidal, R., Maass, A., Jonassen, I., Lien, S. & Omholt, S. W. (2010). Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology* **11,** 403. doi: 10.1186/gb-2010-11-9-403

Elphinstone, M. S., Hinten, G. N., Anderson, M. J. & Nock, C. J. (2003). An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Molecular Ecology Notes* **3,** 317–320. doi: 10.1046/j.1471-8286.2003.00397.x

Excoffier, L., Dupanloup, I., Huerta-Sanchez, E., Sousa, V. C. & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics* **9,** e1003905. doi: 10.1371/journal.pgen.1003905

Finger, A. J., Anderson, E. C., Stephens, M. R., May, B. P. & Taylor, E. (2011). Application of a method for estimating effective population size and admixture using diagnostic single nucleotide polymorphisms (SNPs): implications for conservation of threatened Paiute cutthroat trout (*Oncorhynchus clarkii seleniris*) in Silver King Creek, California. *Canadian Journal of Fisheries and Aquatic Sciences* **68,** 1369–1386. doi: 10.1139/f2011-075

Gabriel, S., Ziaugra, L. & Tabbaa, D. (2009). SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Current Protocols in Human Genetics* **Chapter 2**, Unit 2.12. doi: 10.1002/0471142905.hg0212s60

Gidskehaug, L., Kent, M., Hayes, B. J. & Lien, S. (2011). Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. *Bioinformatics* **27,** 303–310. doi: 10.1093/bioinformatics/btq673

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5,** 184–186. doi: 10.1111/j.1471-8278.2004.00828.x

Gutierrez, A. P., Lubieniecki, K. P., Davidson, E. A., Lien, S., Kent, M. P., Fukui, S., Withler, R. E., Swift, B. & Davidson, W. S. (2012). Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5K SNP array. *Aquaculture* **358–359**, 61–70. doi: 10.1016/j.aquaculture.2012.06.017

Gutierrez, A. P., Lubieniecki, K. P., Fukui, S., Withler, R. E., Swift, B. & Davidson, W. S. (2014). Detection of quantitative trait loci (QTL) related to grilsing and late sexual

maturation in Atlantic salmon (*Salmo salar*). *Marine Biotechnology* **16,** 103–110. doi: 10.1007/s10126-013-9530-3

Gutierrez, A. P., Yanez, J. M., Fukui, S., Swift, B. & Davidson, W. S. (2015). Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *PLoS One* **10,** e0119730. doi: 10.1371/journal.pone.0119730

Harrisson, K. A., Pavlova, A., Telonis-Scott, M. & Sunnucks, P. (2014). Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications* **7,** 1008–1025. doi: 10.1111/eva.12149

Hayden, M. J., Nguyen, T. M., Waterman, A. & Chalmers, K. J. (2008). Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping. *BMC Genomics* **9,** 80. doi: 10.1186/1471-2164-9-80

Hendry, A. P., Kinnison, M. T., Heino, M., Day, T., Smith, T. B., Fitt, G., Bergstrom, C. T., Oakeshott, J., Jorgensen, P. S., Zalucki, M. P., Gilchrist, G., Southerton, S., Sih, A., Strauss, S., Denison, R. F. & Carroll, S. P. (2011). Evolutionary principles and their practical application. *Evolutionary Applications* **4,** 159–183. doi: 10.1111/j.1752-4571.2010.00165.x

Hill, W. G. & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38,** 226–231. doi: 10.1007/BF01245622

Hoffmann, A., Griffin, P., Dillon, S., Catullo, R., Rane, R., Byrne, M., Jordan, R., Oakeshott, J., Weeks, A., Joseph, L., Lockhart, P., Borevitz, J. & Sgrò, C. (2015). A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses* **2.** 1. doi: 10.1186/s40665-014-0009-x

Johnston, S. E., Orell, P., Pritchard, V. L., Kent, M. P., Lien, S., Niemela, E., Erkinaro, J. & Primmer, C. R. (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology* **23,** 3452–3468. doi: 10.1111/mec.12832

Jonsson, N. & Jonsson, B. (2007). Sea growth, smolt age and age at sexual maturation in Atlantic salmon. *Journal of Fish Biology* **71,** 245–252. doi: 10.1111/j.1095-8649.2007.01488.x

Jonsson, N., Hansen, L. P. & 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. doi: 10.2307/5423

Jonsson, B., Jonsson, N. & Albretsen, J. (2016). Environmental change influences the life history of salmon (*Salmo salar*) in the North Atlantic Ocean. *Journal of Fish Biology* **88,** 618–637. doi: 10.1111/jfb.12854

Kaplinski, L., Andreson, R., Puurand, T. & Remm, M. (2005). MultiPLX: automatic grouping and evaluation of PCR primers. *Bioinformatics* **21,** 1701–1702. doi: 10.1093/bioinformatics/bti219

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D. & Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications* **7,** 355–369. doi: 10.1111/eva.12128

Lien, S., Gidskehaug, L., Moen, T., Hayes, B. J., Berg, P. R., Davidson, W. S., Omholt, S. W. & Kent, M. P. (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. doi: 10.1186/1471-2164-12-615

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., Leong, J. S., Minkley, D. R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R. A., von Schalburg, K., Rondeau, E. B., Di Genova, A., Samy, J. K., Olav Vik, J., Vigeland, M. D., Caler, L., Grimholt, U., Jentoft, S., Vage, D. I., de Jong, P., Moen, T., Baranski, M., Palti, Y., Smith, D. R., Yorke, J. A., Nederbragt, A. J., Tooming-Klunderud, A., Jakobsen, K. S., Jiang, X., Fan, D., Hu, Y., Liberles, D. A., Vidal, R., Iturra, P., Jones, S. J., Jonassen, I., Maass, A., Omholt, S. W. & Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* **533,** 200–205. doi: 10.1038/nature17164

Moen, T., Hoyheim, B., Munck, H. & Gomez-Raya, L. (2004). A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Animal Genetics* **35,** 81–92. doi: 10.1111/j.1365-2052.2004.01097.x

Moore, J. S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., Chaput, G. & Bernatchez, L. (2014). Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Molecular Ecology* **23,** 5680–5697. doi: 10.1111/mec.12972

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R. & Hohenlohe, P. A. (2013). Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22,** 2841–2847. doi: 10.1111/mec.12350

Nussberger, B., Greminger, M. P., Grossen, C., Keller, L. F. & Wandeler, P. (2013). Development of SNP markers identifying European wildcats, domestic cats and their admixed progeny. *Molecular Ecology Resources* **13,** 447–460. doi: 10.1111/1755-0998.12075

Ogden, R., Gharbi, K., Mugue, N., Martinsohn, J., Senn, H., Davey, J. W., Pourkazemi, M., McEwing, R., Eland, C., Vidotto, M., Sergeev, A. & Congiu, L. (2013). Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology* **22,** 3112–3123. doi: 10.1111/mec.12234

Otero, J., Jensen, A. J., L'Abee-Lund, J. H., Stenseth, N. C., Storvik, G. O. & Vollestad, L. A. (2012). Contemporary ocean warming and freshwater conditions are related to later sea age at maturity in Atlantic salmon spawning in Norwegian rivers. *Ecology and Evolution* **2,** 2192–2203. doi: 10.1002/ece3.337

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K. & Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends in Genetics* **26,** 177–187. doi: 10.1016/j.tig.2010.01.001

Ozerov, M., Vasemagi, A., Wennevik, V., Diaz-Fernandez, R., Kent, M., Gilbey, J., Prusov, S., Niemela, E. & Vaha, J. P. (2013). Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. *PLoS ONE* **8,** e82434. doi: 10.1371/journal.pone.0082434

Ozerov, M. Y., Himberg, M., Aykanat, T., Sendek, D. S., Hagerstrand, H., Verliin, A., Krause, T., Olsson, J., Primmer, C. R. & Vasemagi, A. (2015). Generation of a neutral FST baseline for testing local adaptation on gill raker number within and between European whitefish ecotypes in the Baltic Sea basin. *Journal of Evolutionary Biology* **28,** 1170–1183. doi: 10.1111/jeb.12645

Paetkau, D., Calvert, W., Stirling, I. & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4,** 347–354. doi: 10.1111/j.1365-294X.1995.tb00227.x

Parrish, D. L., Behnke, R. J., Gephard, S. R., McCormick, S. D. & Reeves, G. H. (1998). Why aren't there more Atlantic salmon (*Salmo salar*)? *Canadian Journal of Fisheries and Aquatic Sciences* **55,** 281–287. doi: 10.1139/cjfas-55-S1-281

Piou, C. & Prevost, E. (2013). Contrasting effects of climate change in continental *vs*. oceanic environments on population persistence and microevolution of Atlantic salmon. *Global Change Biology* **19,** 711–723. doi: 10.1111/gcb.12085

Primmer, C. R. (2009). From conservation genetics to conservation genomics. *Annals of the New York Academy of Sciences* **1162,** 357–368. doi: 10.1111/j.1749-6632.2009.04444.x

Pritchard, V. L., Erkinaro, J., Kent, M. P., Lien, S., Orell, P., Niemela, E. & Primmer, C. R. (2016). SNPs to discriminate wild Atlantic salmon and aquaculture escapees. *Evolutionary Applications* (in press).

Pujolar, J. M., Jacobsen, M. W., Als, T. D., Frydenberg, J., Magnussen, E., Jonsson, B., Jiang, X., Cheng, L., Bekkevold, D., Maes, G. E., Bernatchez, L. & Hansen, M. M. (2014). Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. *Heredity* **112,** 627–637. doi: 10.1038/hdy.2013.145

Rannala, B. & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 9197–9201. doi: 10.1073/pnas.94.17.9197

Rohland, N. & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* **22,** 939–946. doi: 10.1101/gr.128124.111

Russello, M. A., Kirk, S. L., Frazer, K. K. & Askey, P. J. (2012). Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications* **5,** 39–52. doi: 10.1111/j.1752-4571.2011.00206.x

Shafer, A. B., Wolf, J. B., Alves, P. C., Bergstrom, L., Bruford, M. W., Brannstrom, I., Colling, G., Dalen, L., De Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill,

J. A., Hoezel, A. R., Hoglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., Krutzen, M., McKay, J. K., Norman, A. J., Ogden, R., Osterling, E. M., Ouborg, N. J., Piccolo, J., Popovic, D., Primmer, C. R., Reed, F. A., Roumet, M., Salmona, J., Schenekar, T., Schwartz, M. K., Segelbacher, G., Senn, H., Thaulow, J., Valtonen, M., Veale, A., Vergeer, P., Vijay, N., Vila, C., Weissensteiner, M., Wennerstrom, L., Wheat, C. W. & Zielinski, P. (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30,** 78–87. doi: 10.1016/j.tree.2014.11.009

Thompson, J., Charpentier, A., Bouguet, G., Charmasson, F., Roset, S., Buatois, B., Vernet, P. & Gouyon, P. H. (2013). Evolution of a genetic polymorphism with climate change in a Mediterranean landscape. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 2893–2897. doi: 10.1073/pnas.1215833110

Vähä, J. P., Erkinaro, J., Niemelä, E. & Primmer, C. R. (2007). Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology* **16,** 2638–2654. doi: 10.1111/j.1365-294X.2007.03329.x

Vähä, J. P., Erkinaro, J., Niemela, E., Primmer, C. R., Saloniemi, I., Johansen, M., Svenning, M. & Brors, S. (2011). Temporally stable population-specific differences in run timing of one-sea-winter Atlantic salmon returning to a large river system. *Evolutionary Applications* **4,** 39–53. doi: 10.1111/j.1752-4571.2010.00131.x

Vesterinen, E. J., Ruokolainen, L., Wahlberg, N., Pena, C., Roslin, T., Laine, V. N., Vasko, V., Saaksjarvi, I. E., Norrdahl, K. & Lilley, T. M. (2016). What you need is what you eat? Prey selection by the bat *Myotis daubentonii*. *Molecular Ecology* **25,** 1581–1594. doi: 10.1111/mec.13564

Weir, B. S. & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population-structure. *Evolution* **38,** 1358–1370. doi: 10.2307/2408641

Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R. & Guiguen, Y. (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. doi: 10.1111/eva.12032

You, F. M., Huo, N., Gu, Y. Q., Luo, M. C., Ma, Y., Hane, D., Lazo, G. R., Dvorak, J. & Anderson, O. D. (2008). BatchPrimer3: a high throughput web application for PCR and sequencing primer design. *BMC Bioinformatics* **9,** 253. doi: 10.1186/1471-2105-9-253

Electronic Reference

ICES (2013). Report of the Working Group on North Atlantic salmon (WGNAS). *ICES Document CM 2013/ACOM:09*. Available at http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2013/WGNAS/wgnas_2013.pdf/