NRC Research Press

ARTICLE

A validation of parentage-based tagging using hatchery steelhead in the Snake River basin

Craig A. Steele, Eric C. Anderson, Michael W. Ackerman, Maureen A. Hess, Nathan R. Campbell, Shawn R. Narum, and Matthew R. Campbell

Abstract: Parentage-based tagging (PBT) is a promising alternative to traditional coded-wire tag (CWT) methodologies for monitoring and evaluating hatchery stocks. This approach involves the genotyping of hatchery broodstock and uses parentage assignments to identify the origin and brood year of their progeny. In this study we empirically confirmed that fewer than 100 single nucleotide polymorphisms (SNPs) were needed to accurately conduct PBT, we demonstrated that our selected panel of SNPs was comparable in accuracy to a panel of microsatellites, and we verified that stock assignments made with this panel matched those made using CWTs. We also demonstrated that when sampling of spawners was incomplete, an estimated PBT rate for the offspring could also be predicted with fewer than 100 SNPs. This study in the Snake River basin is one of the first large-scale implementations of PBT in salmonids and lays the foundation for adopting this technology more broadly in the region, thereby allowing the unprecedented ability to mark millions of smolts and an opportunity to address a variety of parentage-based research and management questions.

Résumé : Le marquage basé sur l'ascendance (PBT) est une solution de rechange intéressante aux méthodes traditionnelles reposant sur les micromarques magnétisées codées (CWT) pour la surveillance et l'évaluation des stocks issus d'écloseries. Cette approche comprend le génotypage des géniteurs d'écloserie et le recours à l'assignation parentale pour déterminer l'origine et l'année d'éclosion de la progéniture. Dans la présente étude, nous avons confirmé de manière empirique que moins de 100 polymorphismes mononucléotidiques (SNP) étaient nécessaires pour assurer l'exactitude du PBT, démontré que le lot de SNP que nous avons sélectionné donnait une exactitude semblable à celle obtenue avec un lot de microsatellites et vérifié que les assignations parentales obtenues avec ce lot correspondaient à celles obtenues par la méthode des CWT. Nous avons également démontré que, dans les cas d'échantillonnage incomplet des géniteurs, un taux de PBT estimé pour la progéniture pouvait être prédit avec moins de 100 SNP. Cette étude dans le bassin de la rivière Snake est une des premières applications à grande échelle du PBT à des salmonidés et jette les bases de l'adoption élargie de cette technique dans la région, permettant ainsi, pour la première fois, le marquage de millions de saumoneaux et offrant la possibilité d'aborder de multiples questions touchant à la recherche axée sur l'ascendance et à la gestion. [Traduit par la Rédaction]

Introduction

One of the main goals for fisheries managers is to track the movement and harvest of their resource. This task can be challenging for species that inhabit expansive ranges or difficult-to-sample environments (Parker et al. 1990). It is even more difficult when the resource comprises mixed stocks that are difficult to differentiate. The predominant approach to tracking has been to mark a portion of a stock and then, upon recapture, use information from the tag to determine origin and age. There are a variety of physical tags available to accomplish this goal (Bergman et al. 1992; McKenzie et al. 2012), but the most widely utilized is the coded-wire tag (CWT), which has been used by researchers and managers to monitor harvest and escapement of salmonids in the Columbia River basin for over 40 years (Johnson 2004).

Despite the predominance of CWTs, this technology has become limited in its ability to provide sufficient data to managers. Originally, salmonids with CWTs also received an adipose fin clip in many hatcheries. This externally visible mark allowed CWT fish to be easily identified and sampled in a fishery. Since the early 1990s, changes in marking policies resulted in adipose clipping most hatchery fish originating from the Columbia River basin;

however, only a small proportion of these fish are also tagged with CWTs. This approach facilitated mark-selected fisheries, in which only adipose-clipped hatchery fish can be harvested, but it inherently encumbered the ability to recover CWTs by requiring many adipose-clipped fish to be screened before encountering a CWT. The resulting small sample size of recovered CWTs greatly reduces confidence in estimates of stock contributions because the precision of these estimates are directly related to the number of CWTs recovered (Pacific Salmon Commission 2005).

While the ability to collect sufficient data from CWTs has deteriorated, the demand for data by managers has grown and comprehensive reviews of tagging methodologies have suggested exploring alternative tagging technologies (Pacific Salmon Commission 2005; Independent Scientific Review Panel and Independent Scientific Advisory Board 2009). One emerging technology that can provide the same information as CWTs, but without the limitation of small sample sizes, is parentage-based tagging (PBT; Anderson and Garza 2005). PBT involves the annual genotyping of hatchery broodstock, creating a database of parental genotypes from each hatchery. This approach could be applied to any species of hatchery broodstock and would permanently

C.A. Steele and M.W. Ackerman. Pacific States Marine Fisheries Commission, Eagle Fish Genetics Laboratory, 1800 Trout Road, Eagle, ID 83616, USA.
E.C. Anderson. Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, 110 Shaffer Road, Santa Cruz CA 95060, USA.
M.A. Hess, N.R. Campbell, and S.R. Narum. Columbia River Inter-tribal Fisheries Commission, Hagerman Fish Culture Experiment Station, 3059-F National Fish Hatchery Road, Hagerman, ID 83332, USA.
M.R. Campbell. Idaho Department of Fish and Game, Eagle Fish Genetics Laboratory, 1800 Trout Road, Eagle, ID 83616, USA.
Corresponding author: Craig A. Steele (e-mail: craig.steele@idfg.idaho.gov); corresponding author for Supplementary Material: Eric C. Anderson (e-mail: eric.anderson@noaa.gov).

2

Can. J. Fish. Aquat. Sci. Vol. 70, 2013

and noninvasively mark all offspring. Progeny can also be non-lethally sampled at any life stage from fingerling to adult and assigned back to their parents using parentage analysis, thus identifying their hatchery of origin and brood year (i.e., age). One advantage PBT has over CWTs is a substantial increase in the number of tagged fish. When all parental broodstock are genotyped, then every offspring is genetically "tagged", which is similar to a 100% CWT rate and higher than the current basin-wide CWT tagging rates of 7%–21% over the past 10 years (RMIS, online database updated continuously since 1977). A parentage-based approach also allows origin to be determined even when low levels of genetic differentiation among stocks prohibits the use of other genetic identification techniques, such as genetic stock identification (Shaklee et al. 1999). While theoretically appealing (Anderson and Garza 2005, 2006), PBT still needs to be empirically validated, and large-scale evaluations of the technology have been recommended (Pacific Salmon Commission 2005; Pacific Fishery Management Council 2008; Independent Scientific Review Panel and Independent Scientific Advisory Board 2009).

PBT could be implemented with any variable molecular marker, and microsatellites are the marker of choice for many parental studies because of their variability (Webster and Reichart 2005). However, there is widespread interest in the use of single nucleotide polymorphisms (SNPs), especially for large-scale parentage studies. SNPs are gaining favor because of the speed at which they can be screened, their low frequency of genotyping errors, and the ease of standardization among laboratories (Morin et al. 2004). The utility of SNPs for parentage analysis has been explored (Krawczak 1999; Gill 2001; Glaubitz et al. 2003), and modeling indicates PBT can be conducted with as few as 60–100 SNPs (Anderson and Garza 2006). Comparable accuracy of parentage assignments between SNPs and microsatellites has been demonstrated using 80 SNPs within a single parental population (Hauser et al. 2011), but empirically evaluating differently sized SNP panels is necessary, especially when PBT is to be implemented across multiple populations, each with hundreds of contributing parents.

As with CWTs, PBT employs a stock-specific tagging rate — the fraction of fish whose parents have been genotyped. In the controlled setting of a hatchery, the PBT rate can be well-estimated by tracking the fraction of spawned males and females that are sampled and genotyped. Outside the hatchery environment, the fraction of sampled parents is usually not accurately known, making the estimation of a PBT rate for wild stocks more difficult. Despite preliminary efforts (Nielsen et al. 2001) and an implementation suitable for only small populations (Koch et al. 2008), little progress has been made in the past decade on estimating tagging rates from genetic data, and most software packages require that it be known or assumed (see Supplementary Material[1]). In sparsely sampled parental populations, such as wild stocks, it will be necessary to determine the tagging rate not just in terms of the probability that both of a fish's parents are sampled and genotyped, but also in terms of the probability that just one of its parents was sampled and genotyped.

In this study we test a number of differently sized SNP panels to empirically determine a sufficient number of SNPs needed to accurately conduct PBT across multiple hatchery steelhead (*Oncorhynchus mykiss*) broodstocks within the Snake River basin (Fig. 1). We compare assignment accuracy of our selected panel of 95 SNPs with a panel of 17 microsatellites and demonstrate that assignments made with PBT match those using CWTs. Finally, we introduce a statistical framework to estimate the PBT rate of offspring from an incomplete parental sample so as to employ a PBT approach in wild populations. This work lays the foundation for

the implementation of PBT in the region and the opportunity to conduct future parentage-based projects.

Methods

Sampling of hatchery broodstock and known-origin juveniles

Beginning in 2008, fin tissue was sampled from nearly all adult steelhead broodstock returning to Snake River hatcheries in Idaho, Oregon, and Washington (Steele et al. 2011). Progeny from all hatchery steelhead spawned in Idaho were raised at rearing facilities, including the Magic Valley rearing hatchery in southern Idaho (Fig. 1), where they were reared for a year prior to smoltification and released for downstream migration. Progeny from the Cottonwood Creek stock were reared at Lyons Ferry Hatchery in Washington State. Offspring of known crosses of the hatcheries were sampled when they reached a size at which fin clips could be taken nonlethally (~100 mm TL). Fin tissue was stored in 100% nondenatured ethanol prior to DNA extraction.

Laboratory procedures and marker selection

Genomic DNA extractions were carried out using a Nexttec DNA isolation kit according to the manufacturer's instructions (http://www.nexttec.biz). Samples were genotyped with a panel of 188 SNPs (Table S1[1]). The 188 steelhead SNP markers were selected from available loci because previous genotyping indicated the assays were robust, exhibited variation in hatchery steelhead populations in the Snake River basin, and conformed to Hardy–Weinberg and linkage equilibrium (HWE) expectations (Hess et al. 2011).

Prior to DNA amplification of SNP loci using primer–probe sets (fluorescent tags), an initial polymerase chain reaction (PCR) "pre-amp" was implemented using whole genomic DNA to jumpstart SNP amplification via increased copy number of the targeted DNA regions. The PCR conditions for the pre-amp step were as follows: an initial mixing step of 95 °C for 15 min, followed by 14 cycles of 95 °C for 15 s and 60 °C for 4 min, ending with a final 4 °C holding step. Genotyping was performed using Fluidigm 96.96 Dynamic Array IFCs (chips). For each genotyping run, 96 samples (including an extraction negative control, a PCR negative control, and a PCR positive control) and 96 TaqMan SNP assays were loaded onto the 96.96 chips. One assay, a diagnostic sex-determining marker (Campbell et al. 2012), is included in the 96-SNP panel but is not included in subsequent PBT analysis. Sample cocktail and SNP assay cocktail recipes are available by request. Each 96.96 chip was pressurized to load the DNA and SNP assays into the array using a Fluidigm IFC Controller HX. Amplification of SNPs on the 96.96 chips were performed using either an Eppendorf Mastercycle thermal cycler (protocol: thermal mixing step of 50 °C for 2 min, 70 °C for 30 min, and 25 °C for 10 min; a hotstart of 50 °C for 2 min and 95 °C for 10 min, followed by 50 cycles of 95 °C for 15 s and 60 °C for 60 s; and a final cool down step of 25 °C for 10 min) or a Fluidigm FC1 Fast-cycler (protocol: thermal mixing step of 70 °C for 30 min and 25 °C for 10 min; a hot-start of 95 °C for 60 s, followed by 50 cycles of 95 °C for 5 s and 58 °C for 25 s; and a final cool down step of 25 °C for 10 s). Chips were imaged on a Fluidigm EP1 system and analyzed and scored using the Fluidigm SNP Genotyping Analysis software version 3.1.1. Genotypes were imported and organized in a Progeny database (Progeny Software, South Bend, Indiana, USA).

Samples were also genotyped with a panel of 17 microsatellites (Table S2[1]), 12 of which are from a standardized series developed to address genetic questions for steelhead throughout the Columbia River basin (Stephenson et al. 2008). Of the 17 loci, 15 (Ots100, Omy1001, Ogo4, Omy07, Oke4, Oki23, Omy1011, Ots3, Ssa407,

---

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2012-0451.

Fig. 1. Map of Snake River basin indicating origin of hatchery stocks. Only hatcheries (FH), rearing facilities, and trapping facilities referenced in this study are represented. ESU, evolutionarily significant unit.
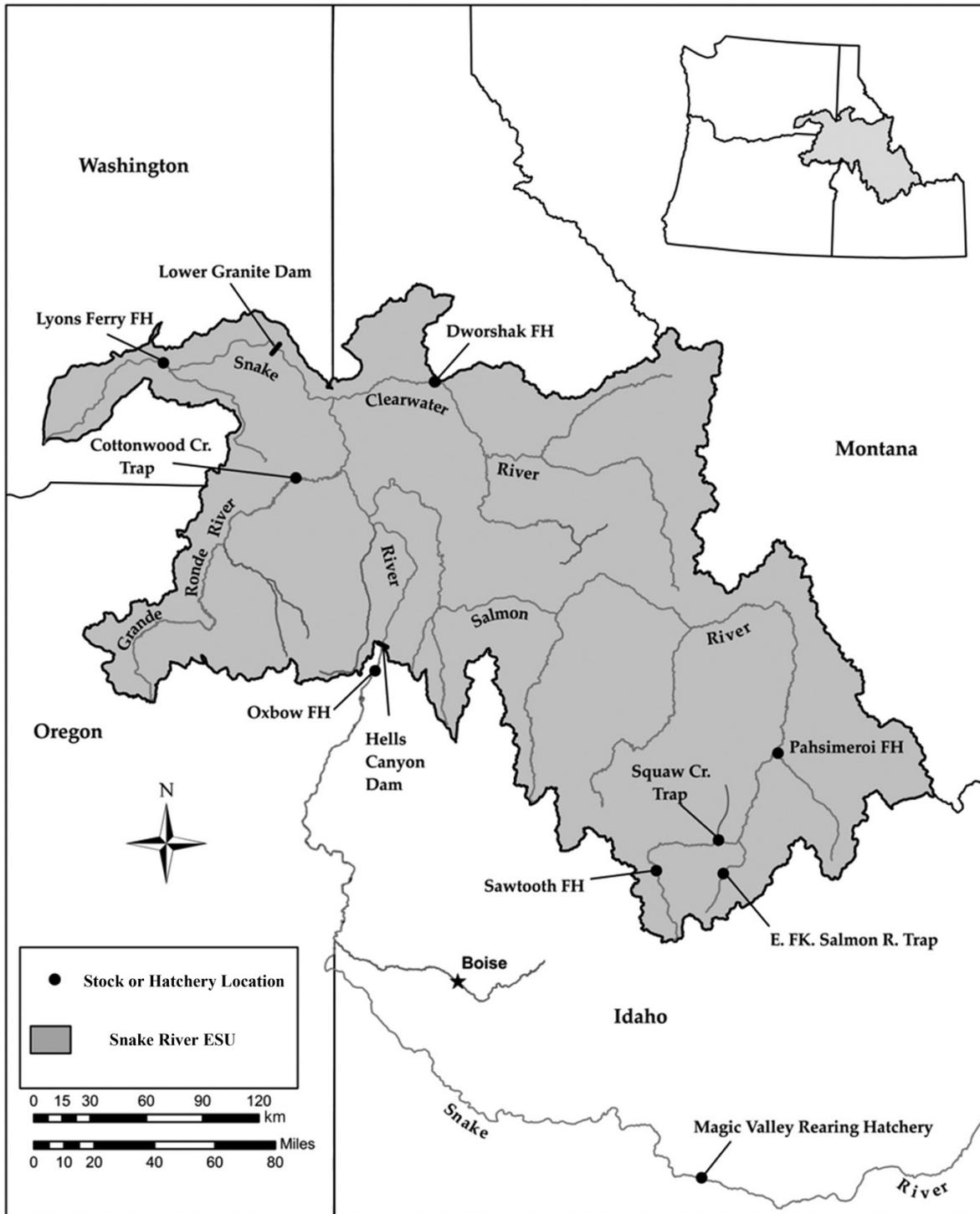
Ssa408, Ogo1a, Omy27, Oneu14, Oneu8, and Ots4) were amplified in three multiplex PCRs (three runs on an ABI 3100 fragment analyzer). The remaining two loci (Omy325 and Ssa289) were amplified in single PCR reactions and were each run alone on an ABI 3100. Summaries of genetic diversity and deviations from HWE for the sample groups were calculated using the GenAlEx 6.4 add-on for Microsoft Excel (Peakall and Smouse 2006). For significance testing, a P value threshold of 0.05 was used and then adjusted for multiple comparisons using the modified false discovery rate (B-Y FDR) procedure as suggested by Narum (2006).

Assessment of SNP power

To select a SNP panel with a sufficient number of loci to answer future management questions and to empirically test the predictions of Anderson and Garza (2006) that 60–100 SNPs provide adequate power for PBT applications, we assessed the ability of

SNP panels with variable numbers of loci to correctly assign off-spring of known parentage tagged by PBT. We first ranked all 188 SNPs based on their minor allele frequencies (MAFs) by pooling all the parental hatchery populations, computing MAFs for the pooled population, and then ranking the SNPs by their MAF. We then iteratively conducted parentage assignments using SNP panels with sequentially larger numbers of the top-ranked SNPs (i.e., SNPs with the highest MAF). For our evaluation of differently sized SNP panels, we chose the top-ranked tiers of SNPs rather than randomly selecting loci because this more accurately reflects the process of SNP selection for parentage studies, such that when presented with a large number of potential loci, one would not randomly select loci to construct a panel but rather would select the most informative loci.

Parentage assignment using SNP genotypes was performed using the program SNPPIT (Anderson 2010). While additional information such as gender of parental samples, cross records, and spawn day could be included in SNPPIT to reduce the number of parent–offspring trios considered, we purposefully exclude this information from analysis to examine the resolving ability of the SNP genotypes without this additional information. We allowed up to 10% missing genotype data for a sample within each 95-SNP panel before excluding the sample from consideration in parentage. We used an estimated SNP genotyping error rate of 1% or a per allele rate of 0.5%. SNPPIT assesses the confidence of parentage assignments using a false discovery rate (FDR), and we only accepted assignments with a very stringent FDR threshold of <1%.

We quantified the false negative rate and the unsampled-parent false positive rate to evaluate the assignment ability of the differently sized SNP panels. The false negative rate is the proportion of offspring that did not assign to their parental pair despite the parental genotypes being present in the dataset. The unsampled-parent false positive rate represents the proportion of offspring whose parents were not sampled but assigned incorrectly to a nonparental pair. This was evaluated by including samples of offspring originating from an unsampled broodstock at Lyons Ferry Hatchery, Washington (Fig. 1). Because the Lyons Ferry offspring did not have parents in this dataset, the assigned proportion of these offspring was used to calculate an unsampled-parent false positive rate for the differently sized SNP panels. This test using Lyons Ferry fish was intended to demonstrate that offspring from unsampled broodstock would not be erroneously assigned to fish in the parental database. However, if the Lyons Ferry population had different allele frequencies than the populations sampled for the parental database, then offspring from Lyons Ferry would have been inherently less likely to misassign to any parents in the database than were fish from the sampled populations, and false positive rates derived from that test might not have accurately reflected actual error rates. Thus, to further examine the potential for false positives in our methodology, we attempted to assign the entire collection of broodstock genotypes collected in 2008 ($N = 5107$) to the broodstock genotypes collected in 2009 ($N = 5672$) using a panel of 95 SNPs (see below) in SNPPIT. Because it was physically impossible for the 2009 broodstock to be parents of the previous year's broodstock, any parental assignment in this exercise would be known to be incorrect.

Assignment accuracy compared with microsatellites

Based on performance results of the differently sized SNP panels (below), we selected a final panel of 95 SNPs for use in subsequent analyses. In several cases, SNPs of lower rank, based on MAF, were chosen instead of SNPs of higher rank (Table S1[1]) because of their higher quality and consistently scorable genotype plots. To compare the assignment accuracy of the 95-SNP panel with that of microsatellites, we genotyped the known-origin offspring and their parents with the panel of 17 microsatellites. Parentage assignments using microsatellite genotypes were made with a maximum likelihood approach using CERVUS 3.0.3

(Kalinowski et al. 2007; Marshall et al. 1998) and assuming unknown gender of the parental samples to allow for unrestricted matings. We determined accuracy for parental assignments of the offspring by comparing the assignments with stock and parental cross records recorded at the hatchery. An average error rate of 0.44% (a per allele rate of 0.0022 was used in SNPPIT) was calculated for the 95 SNPs by regenotyping a subset of samples at the 95 loci.

Comparison of known-origin CWT adults in the Snake River harvest program

In addition to collecting parental genotypes from several hatchery populations for the pilot study described above, we also sampled and genotyped the majority of steelhead broodstocks spawned during 2008 in the Snake River basin (Steele et al. 2011). A portion of offspring from these broodstocks was marked with CWTs (~13%). During the winter of 2010–2011, creel clerks from the Lower Snake River Compensation Plan's (LSRCP) harvest monitoring program recovered snouts from CWT-detected adults captured in the Snake River fishery. Because hatchery steelhead typically smolt after 1 year in fresh water, any offspring from the 2008 brood year that returned as one-ocean adults during this season are expected to be tagged by PBT, while all older adult offspring would be precluded from PBT assignment. Snouts were sent to the Idaho Fish and Game CWT Lab in Nampa, Idaho, for processing. During CWT removal and decoding, muscle tissue was sampled from a subsample of 186 snouts for DNA extraction and genotyping with the 95-SNP panel using the same procedures described previously. Parentage assignment and hatchery of origin were determined using PBT conducted in SNPPIT. Hatchery origin, based on CWTs, was then compared with stock assignment provided from the PBT approach.

Estimating unsampled parents in wild populations

To assess the potential of using multilocus SNP genotypes for PBT applications in wild populations, we propose a statistical framework for estimating the PBT rate of offspring under incomplete sampling scenarios when only a subset of parental pairs is sampled or when only a proportion of a single parental gender is sampled (see Supplementary Material[1]). We then used this framework to estimate the tagging rate in simulated data. Specifically, we simulated a population of 500 males and 500 females that mated monogamously to produce 1000 offspring. The expected number of offspring produced by each pair of parents was equal, and the actual number was multinomially distributed. Using this single simulated population, the PBT rate for offspring was estimated. The software program SNPPIT was developed for large-scale PBT projects using SNP markers (Anderson 2010) and uses a false discovery rate correction to account for scenarios in which the tagging rate is not known (as will be the case in most mixed-stock samples). However, SNPPIT's current formulation allows neither an accurate estimate of the false negative rate nor identification of single parents of offspring; it focuses entirely on parental pairs. For PBT to be useful in wild populations will require that the PBT rate be estimated in terms of (i) the fraction of offspring with both parents in the database, g_{pair}, (ii) the fraction of offspring whose father is in the database, but not their mother, g_S, (iii) the fraction of offspring whose mother is in the database, but not their father, g_D, and (iv) the fraction of offspring with neither parent in the database, $g_{none} = 1 - g_{pair} - g_S - g_D$. In each simulation, 100 complete mated pairs were randomly sampled from the parental generation (true $g_{pair} = 100/500 = 0.2$). An additional 50 males were sampled without their mates (true $g_S = 50/500 = 0.1$), and 25 females were sampled without their mates (true $g_D = 25/500 = 0.05$). All 1000 offspring were sampled. Our method was used to estimate g_{pair}, g_S, and g_D using L SNP loci, each with a minor allele frequency of 0.25. At each value of L between 15 and 100, 10 separate datasets were simulated and analyzed.

Table 1. Percentage of samples that accurately assigned, misassigned, or were unassigned (false negative) using different numbers of single nucleotide polymorphisms (SNPs) for parentage-based tagging (PBT).

No. of loci used	% Accurate assignment	% Misassignment of assigned fish	False negative	False positive	No. of parent–offspring trios evaluated
36	38.5	7.9	53.6	0.0	76 432
48	87.6	2.2	10.2	1.1	21 251
72	95.7	0.0	4.3	0.0	1 036
96	96.1	0.0	3.9	0.0	759
120	96.2	0.0	3.8	0.0	707
144	97.3	0.0	2.7	0.0	849
168	97.1	0.0	2.9	0.0	867
188	97.1	0.0	2.9	0.0	680

Note: Percentage of false positives represents the portion of offspring known to not have parents represented in the dataset (Lyons Ferry) that assigned to a parental pair regardless. The number of parent–offspring trios evaluated refers to the number of potential parental assignments that could not be excluded based on Mendelian incompatibilities and were subsequently evaluated with ML in SNPPIT. SNPs for the differently sized panels were ranked and selected based on the minor allele frequencies within six hatchery broodstocks in the Snake River basin.

Results

Assessment of SNP power

Assignment rates of known-origin offspring to correct stock were above 95% for all sizes of SNP panels, except the smallest panels of 36 and 48 SNPs (Fig. 2; Table 1). Correct assignment rates remained high as the number of SNPs decreased until a threshold of too few SNPs was reached, after which the correct assignment rate dropped sharply. No offspring were misassigned to an incorrect stock except when the two smallest panels were used. The false negative rate (sampled offspring that should have assigned to a parental pair, but did not) increased gradually as fewer SNPs were used in the parentage assignments, but increased sharply only with the two smallest SNP panels (Table 1). There were no unsampled-parent false positives from the Lyons Ferry offspring except with the 48-SNP panel when one Lyons Ferry offspring was assigned to a parental pair. Additionally, no unsampled-parent false positives were observed when we attempted to assign the 2008 broodstock to the 2009 broodstock using the 95-SNP panel despite a total of >17 × 10⁹ possible parent–offspring trios being evaluated.

### Assignment accuracy of SNPs compared with microsatellites

Assignment accuracy to correct stock and parentage was similar between microsatellites and SNPs (Table 2). For microsatellites, a high percentage of genotyped offspring (98.6%) received assignment to a hatchery stock, and in every case the stock assignment was correct. None of the Lyons Ferry offspring received parentage assignment using the 17 microsatellites. Assignment rates to the parental pairs identified in cross records ranged from a low of 70.7% for Dworshak-origin offspring to a high of 98.9% for Grande Ronde-origin offspring. The average mismatch rate of microsatellite alleles between parents and assigned offspring was 0.068 mismatches/assignment.

For SNPs, 97.3% of the genotyped offspring received an assignment to a hatchery stock, and in every case the assignment was also to the correct stock. None of the Lyons Ferry offspring assigned to a stock with the 95-SNP panel. Assignment rates to the parental pair identified in the cross records ranged from a low of 71.0% for Dworshak-origin offspring to a high of 98.9% for Grande Ronde-origin offspring. The average mismatch rate of SNP alleles between parents and assigned offspring was 0.15 mismatches/assignment.

Assignment accuracy of PBT compared with CWTs

A total of 61 one-ocean aged hatchery fish (1 year at sea) that were subsampled from the Idaho steelhead fishery in the winter of 2010–2011 had CWTs indicating their origin was from a Snake River basin hatchery broodstock sampled for PBT in 2008. Of

Fig. 2. Relationship between accuracy of parentage-based tagging (PBT) assignment to stock of origin and number of single nucleotide polymorphisms (SNPs) used in the assignment. SNPs for the differently sized panels were ranked and selected based on the minor allele frequencies within six hatchery broodstocks in the Snake River basin.
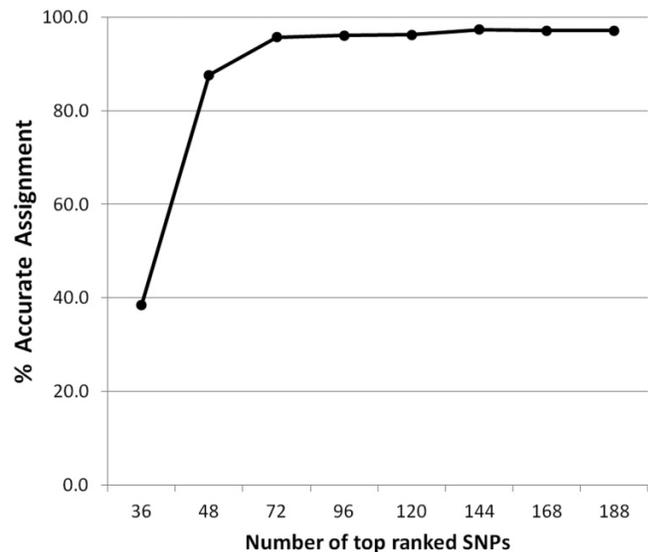
those 61 samples, 59 were successfully genotyped and 52 (88.1%) were assigned with PBT to a hatchery stock (Table 3). Recovery information for CWTs in this fishery indicated that ~14% of harvested fish had CWTs (T. Curet, B. Esselman, M. White, M. Biggs, J. Hansen, and B. Beller, Idaho Department of Fish and Game, unpublished report). All assignments made to stock with PBT matched the stock identified with CWT. Based on CWT information, the seven samples that did not assign using PBT were composed of five fish originating from the Oxbow Hatchery and two from the Dworshak Hatchery.

Direct estimation of tagging rate with genetic data

In all, we analyzed 860 simulated datasets. It is clear that given the conditions of our simulations, g_{pair} can be estimated with high accuracy using 40 or more SNPs (Fig. 3a). In similar fashion, g_S and g_D can be accurately estimated with 60 or more SNPs (Figs. 3b and 3c). In fact, with just 40 or 60 SNPs, the observed distribution of estimates falls largely within the 95% confidence intervals that would be expected of the estimates when parentage is unambiguously known.

Table 2. Comparisons of parentage and stock assignment accuracy between 17 microsatellites (Micros) and 95 SNPs using offspring of known origin from six upper Snake River hatchery stocks.

Hatchery stock	Sample size		Samples genotyped (adults, juveniles)		Juveniles assigned		No. assigned to correct stock		No. (%) matching spawn records	
	Parents (♂, ♀)	Offspring	Micros	SNPs	Micros	SNPs	Micros	SNPs	Micros	SNPs
Squaw Creek	15, 20	92	35, 91	35, 92	91	92	91	92	79 (86.8)	79 (85.9)
Sawtooth	118, 116	93	232, 93	231, 93	91	91	91	91	86 (94.5)	86 (94.5)
Dworshak[*]	184, 209	93	393, 92	393, 93	92	93	92	93	65 (70.7)	66 (71.0)
East Fork Salmon River	83, 36	94	118, 93	118, 94	87	85	87	85	83 (95.4)	80 (94.1)
Pahsimeroi[†]	121, 134	93	255, 92	252, 93	92	88	92	88	—	—
Cottonwood Creek	85, 84	93	169, 91	167, 93	91	91	91	91	90 (98.9)	91 (100.0)
Lyons Ferry	—	93	—, 93	—, 93	0	0	—	—	—	—
Total	606, 599	651	1202, 645	1196, 651	544	540	544	540	403	405

Note: Broodstock from Lyons Ferry were not sampled in order to estimate a proportion of false positive assignments of their offspring using a 95-SNP panel.

[*]Cross information not available for all broodstock; 27 offspring assigned to parents with missing cross information, thereby lowering the number of possible matches to spawn records.

[†]Cross information not available.

Table 3. Results of PBT-based stock assignments for known-origin one-ocean fish with CWTs that were genetically tagged through the genotyping of the 2008 Snake River basin parental broodstock.

CWT-based origin	Samples	Genotyped (failed to amplify)	Assigned	Assignment rate (%)	PBT rate (%)
Dworshak	9	9 (0)	7	77.8	85.1
Pahsimeroi	30	30 (0)	30	100	95.8
Oxbow	16	15 (1)	10	66.7	93.3
Sawtooth	4	3 (1)	3	100	98.2
Pahsimeroi (Squaw Creek)	1	1 (0)	1	100	100
Lyons Ferry (Cottonwood Creek)	1	1 (0)	1	100	100
Total	61	59 (2)	52	88.1	96.8

Fig. 3. Distribution of estimates of the tagging rate via (*a*) parental pairs, (*b*) single fathers, and (*c*) single mothers, using different numbers of SNPs. Ten datasets were simulated for each number of SNPs from 15 to 100, and each dataset is represented by a single dot on each figure. The true tagging rates (g_{pair}, g_S, and g_D, respectively) are indicated on each plot by the black dashed lines. The gray dashed lines represent the 2.5% and 97.5% quantiles of a binomial random variable of 1000 trials with the success probability given by g_{pair}, g_S, or g_D. These are the 95% confidence intervals that would be expected if parentage could be assigned without any error to the sampled offspring in the simulations.
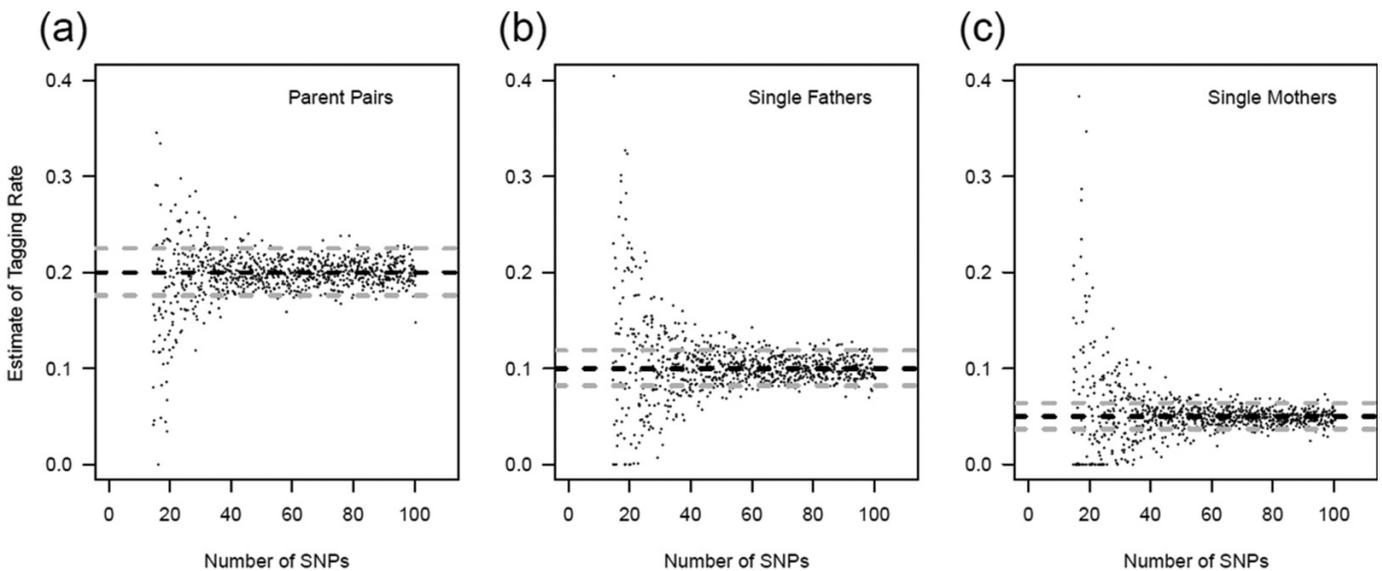
Discussion

Implementing PBT is likely to provide managers a more efficient, versatile, and powerful tool for tagging hatchery fish than CWTs. Tagging with PBT carries a number of advantages over using CWTs. Most prominently, because every juvenile is tagged when its two parents are genotyped, the cost of PBT is low enough to tag nearly every hatchery offspring compared with historical rates of 5%–10% with physical CWTs (RMIS, online database updated continuously since 1977). Higher tagging rates allow larger numbers of fish tagged by PBT to be recovered than fish with CWTs, which can improve error associated with stock assessments. Additionally, pedigrees generated through genotyping multiple generations of broodstock provide opportunities for addressing management issues associated with relative reproduc-

tive success of hatchery fish, trait heritability of broodstocks, and reforms in hatchery management. Few programs have initiated PBT as a large-scale tagging strategy (Denson et al. 2012). Our study provides one of the first empirical validations of PBT using SNPs, resulting in the unprecedented ability to mark millions of steelhead smolts and provide opportunities for parentage-based research.

Assessment of SNP power

Because of the advantages SNPs have over other molecular markers (Morin et al. 2004), they are becoming the marker of choice for large-scale collaborative parentage projects. The number of SNPs needed to sufficiently conduct PBT was theoretically estimated to be 60–100 (Anderson and Garza 2006). This guideline was met with skepticism by some agencies (Pacific Salmon Commission 2005), perhaps because the bialleleic nature of SNPs inherently reduces the resolving power of a single SNP compared with that of a multi-allelic microsatellite. We empirically confirmed Anderson and Garza's (2006) theoretical prediction and demonstrate that accurate parentage (>95%) can be achieved with as few as 72 SNPs (Fig. 2). Even though 72 SNPs provided accurate parentage assignments, a panel of 95 SNPs for subsequent analyses was selected because the current technology of the 96.96 Fluidigm allows up to 96 SNPs on a single run, and the additional SNPs should also provide an abundance of power in the most limiting scenarios. The minimum number of SNPs needed for accurate parentage assignment will depend on the MAF of the markers such that fewer loci with higher MAFs can provide comparable power as more loci with lower MAFs (Anderson Garza 2006). There is also a diminishing return between the resolving power of a SNP locus and increases in MAF such that more power is gained as MAF increases from 0.2 to 0.3 than from 0.4 to 0.5 (Anderson and Garza 2006). Our final panel of 95 SNPs has relatively high MAFs ranging from 0.155 to 0.486 (Table S1[1]). This suggests that if these markers are used for PBT with additional hatchery stocks in the Columbia River basin, the number of SNPs needed to accurately conduct PBT can remain the same even if MAF is reduced at some loci within other populations.

Comparison with microsatellites

When study systems transition from microsatellites to SNPs, it is often desirable or necessary to confirm similar results between the two kinds of markers and to demonstrate the resolving power of SNPs (Hauser et al. 2011). Assignment results between microsatellites and SNPs were comparable but not identical, and inconsistencies appear to be due to differences in the genotyping completeness of samples by each marker set. While every attempt was made to sample all broodstock contributing to the study, we suspect there were unsampled parents. Nonassignment of offspring is attributed to either an unsampled parent or incomplete genotypes from a sampled parent. For microsatellites and SNPs, respectively, 544 and 540 of the 558 offspring received a PBT assignment resulting in a false negative rate of 2.5% and 3.2% for each dataset. These false negative values are low but are also inflated because ungenotyped parents preclude some assignments. Unsampled or ungenotyped parents within the SNP dataset likely account for the two unassigned Sawtooth Hatchery offspring and for six of the nine unassigned East Fork Salmon offspring (Table 2). The remaining three unassigned East Fork Salmon offspring had an FDR above 1% (but below 5%). Unassigned offspring from the Cottonwood Creek stock and Pahsimeroi Hatchery were traced to parents that had been sampled but failed to be genotyped. When complete SNP genotypes for parents and offspring were available, a correct assignment was always made, indicating that the assignment ability of PBT with SNPs was ultimately restricted by the completeness of sampling and genotyping of the parental broodstock and not by limitations in the molecular markers or algorithms in the assignment software.

For both marker sets, a proportion of parentage assignments did not match the cross records from the hatcheries. However, all parentage assignments using SNPs matched those made with microsatellites, even when the parentage assignment did not match the cross records. This suggests that some hatchery-recorded cross information was incorrect, and the error was identified using parentage assignments with the independent datasets. Despite potential for such errors, a record of individual parents used in spawning can be valuable information to include in parentage assignments by reducing the number of possible parental combinations and thereby reducing computation time. In general, including additional data, such as cross information or sex of parental samples, can improve parental assignments if a small number of loci are used or if the loci have low power of resolution. However, if these data are not recorded accurately it can have an adverse effect by inadvertently precluding true parents from being considered as possible mates, thereby decreasing the number of correct assignments. Initial analyses of our data confirmed this result and showed a slightly higher proportion of unassigned offspring (data not shown) when cross information of the spawners, along with errors, was included. Even if cross information is not used to help with parentage assignments, the basic information about broodstock samples ("spawn year" and "hatchery stock") can help to greatly reduce the number of possible parental matings in a large multiyear PBT dataset. Another strategy to reduce the number of potential crosses is "day-binning" (Anderson and Garza 2005), which is less error prone than recording specific cross information. Collecting cross information from broodstock remains important because it allows both members of a parental pair to be identified when a genotype is missing from one parent. In such cases, single-parentage assignment can identify the parent for which data has been collected and cross information can identify the other parent with which it was mated.

Comparison with CWTs

All PBT assignments matched the stock identified with CWT, but not all CWT fish received a PBT assignment (Table 3). Nonassignment of two CWT samples from Dworshak Hatchery was likely due to unsampled parental broodstock. A portion of the parental crosses were not sampled for PBT in 2008, yielding an estimated PBT rate of 85.1% for the stock overall. If we assumed equal production of offspring across families, we would expect to assign 85.1% of Dworshak-origin offspring to their two parents. We had a slightly lower, but not significantly different, assignment rate of 77.8% for Dworshak-origin offspring (binomial test, $P = 0.79$) and concluded that the nonassigned fish were offspring from the unsampled Dworshak parents. The nonassignment of five CWT samples from Oxbow Hatchery resulted in an assignment rate (66.7%) lower than the expected PBT rate (93.3%) for this stock. Currently, SNPPIT is only able to assign offspring if both parents' genotypes have been collected. To determine if at least one parent was included in the dataset, the unassigned Oxbow Hatchery-origin offspring were analyzed separately in CERVUS using exclusion-based procedures. Single parent assignments resulted in assigning 14 of the 15 samples and an assignment rate of 93.3%, identical to the expected PBT rate. Of the five previously unassigned Oxbow Hatchery offspring, four received a single-parent assignment with no allelic mismatches, and one individual was assigned equally to two different parents with no allelic mismatches. A check of the Oxbow Hatchery cross records indicated that all four of the parents were crossed with one parent that did not genotype. In addition, one of the two possible parents identified for the final individual was also crossed with a parent that failed to be genotyped. This indicates that the initial assignment rate for Oxbow Hatchery-origin samples was largely driven by the failure to genotype one individual from a parental pair and not by unsampled broodstock or analytical limitations of the software. It also emphasizes the need to develop methods that account for the

pattern of mating and variability in individual reproductive success when estimating uncertainty in the PBT rate.

Estimating the PBT rate with genetic data

PBT was originally proposed as an alternative tagging methodology that provides advantages over CWTs (Anderson and Garza 2005). Our results from PBT assignments demonstrated the same stock assignments as known-origin CWT fish, but the results also highlighted current software limitations. Genotypes from just a single parent will not result in a parent–offspring assignment in SNPPIT. Although CERVUS is capable of making single-parentage assignments, the computational time required for such an analysis on a large multiyear PBT dataset can be impractical. Given some improvements in software, however, it may become possible to conduct PBT on wild-spawning stocks if an ample, albeit unknown, fraction of the parents are sampled. For example, if a wild population could be nonlethally sampled at a weir or by sampling carcasses after spawning, and a representative sample of outmigrating smolts was genotyped and used to estimate the fraction of sampled parents, then that fraction, along with the known number of sampled fish or carcasses, would yield an estimate of the abundance of spawners. This would also yield a PBT rate for the wild stock, which might make it possible to use PBT to estimate the number of wild fish harvested in a fishery.

One way to deal with unassigned offspring due to missing parental genotypes is to estimate the PBT rate and then use it to extrapolate assignments to the total number of unassigned fish in a sample. A PBT rate can be estimated by multiplying the proportions of successfully genotyped male and female broodstock. Using this simple approach for calculating the proportion of tagged individuals is only possible if the proportion of successfully genotyped adults is known. This is often the case within hatchery settings, but PBT can also be applied to wild populations in which the proportion of sampled adults is not known with great certainty. Our simulations reveal that with sufficient numbers of SNPs (i.e., >40) there is substantial power to estimate the fraction of offspring with both parents or with just a single parent in a parental database. This suggests that with carefully designed sampling protocols and advances in statistical tools, PBT may be successfully applied to the management of wild salmon populations (though doing so may require intensive sampling of adults and possibly of some outmigrating juveniles).

The simulations also demonstrate that a mixture formulation to infer parentage with PBT, like that proposed in the Supplementary Material[1], would likely carry a number of advantages over the allocation method implemented in the current version of SNPPIT. For example, with only 40 markers and a false negative rate of 10%, the method of Anderson and Garza (2006) would assign roughly 10% of offspring to the wrong pair of parents. However, despite that level of inaccuracy in individual assignment, the tagging rate can still be estimated accurately. A similar phenomenon is seen in mixed stock analysis of fisheries whereby individual fish cannot all be assigned to their population without error, but the fraction of fish in the sample from any one population can still be estimated accurately (Koljonen et al. 2005). This suggests that it is worthwhile to develop software that can handle large-scale parentage analysis for PBT in the context of a mixture model that allows the estimation of tagging rates. Ideally, such an analysis would be combined with genetic stock identification using baseline samples from different wild populations, so that parentage assignments, population assignments, tagging rates, and mixture proportions can be simultaneously estimated using all the genotyped individuals, regardless if their parents appear in the parent database. A number of challenges will need to be addressed, in particular the problems posed by nonparental relatives and by the scale of PBT datasets, and will require novel solutions.

Implementation of PBT

This project demonstrates the feasibility of implementing a large-scale PBT project and the accuracy of SNPs for parentage analysis. This work lays the foundation for the creation of parental baselines for hatchery stocks in the Snake River basin and the utilization of these baselines in answering a variety of parentage-based research questions. We estimated the implementation of the PBT program in the Snake River basin has already resulted in the genetic tagging of ~95% of steelhead and Chinook salmon (*Oncorhynchus tshawytscha*) in the Snake River basin, which corresponds to ~67% of outmigrating hatchery steelhead smolts and ~55% of returning hatchery adults in the entire Columbia River basin. As a result, fish management agencies are continuing annual sampling and genotyping of broodstock throughout the Snake River basin with plans to expand throughout the rest of the Columbia River, thereby creating parental databases that will allow for tagging and parentage analysis of hatchery steelhead originating from the region.

The implications of this study have clear utility for managing and monitoring hatchery stocks within the region. Screening of additional hatchery stocks with this SNP panel is underway within the Columbia basin to determine its applicability in a larger and comprehensive Columbia River basin-wide PBT program. The results also have implications beyond the study system and indicate that any large-scale captive rearing program can use a moderately sized panel of SNPs to evaluate the contribution of hatchery efforts to harvest or supplementation programs or to trace an aquaculture product through production, even when multiple closely related broodstocks have contributed. The statistical framework described in the Supplementary Material[1], which lays the foundation for applying a PBT-style approach in wild populations, also has far-reaching applications and the potential for implementation in wild populations of fish or non-fish species. PBT clearly has the potential to provide many opportunities for addressing management and research questions.

References

Anderson, E.C. 2010. Computational algorithms and user-friendly software for parentage-based tagging of Pacific salmonids [online]. Final report submitted to the Pacific Salmon Commission's Chinook Technical Committee (US Section). Available from http://swfsc.noaa.gov/textblock.aspx?Division=FED&ParentMenuId=54&id=16021.

Anderson, E.C., and Garza, J.C. 2005. A description of full parentage genotyping [online]. Report submitted to the Pacific Salmon Commission, Vancouver, British Columbia. Available from http://swfsc.noaa.gov/publications/FED/00675.pdf.

Anderson, E.C., and Garza, J.C. 2006. The power of single-nucleotide polymor-

phisms for large-scale parentage inference. Genetics, **172**(4): 2567–2582. doi:10.1534/genetics.105.048074. PMID:16387880.

Bergman, P.K., Haw, F., Blankenship, H.L., and Buckley, R.M. 1992. Perspectives on design, use, and misuse of fish tags. Fisheries, **17**(4): 20–25. doi:10.1577/1548-8446(1992)017<0020:PODUAM>2.0.CO;2.

Campbell, M.R., Kozfkay, C.C., Copeland, T., Schrader, W.C., Ackerman, M.W., and Narum, S.R. 2012. Estimating abundance and life history characteristics of threatened wild Snake River steelhead stocks by using genetic stock identification. Trans. Am. Fish. Soc. **141**: 1310–1327. doi:10.1080/00028487.2012.690816.

Denson, M.R., Brenkert, K., IV, Jenkins, W.E., and Darden, T.L. 2012. Assessing red drum juvenile stocking in a South Carolina estuary using genetic identification. N. Am. J. Fish. Manage. **32**(1): 32–43. doi:10.1080/02755947.2011.649577.

Gill, P. 2001. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int. J. Leg. Med. **114**(4–5): 204–210. doi:10.1007/s004149900117.

Glaubitz, J.C., Rhodes, O.E., and Dewoody, J.A. 2003. Prospects for inferring pairwise relationships with single nucleotide polymorphisms. Mol. Ecol. **12**(4): 1039–1047. doi:10.1046/j.1365-294X.2003.01790.x. PMID:12753222.

Hauser, L., Baird, M., Hilborn, R., Seeb, L.W., and Seeb, J.E. 2011. An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. Mol. Ecol. Res. **11**(Suppl. 1): 150–161. doi:10.1111/j.1755-0998.2010.02961.x. PMID:21429171.

Hess, J., Campbell, N., Matala, A., and Narum, S. 2011. Annual report: Genetic assessment of Columbia River stocks. US Dept. of Energy Bonneville Power Administration Report No. 2008-907-00.

Independent Scientific Review Panel and Independent Scientific Advisory Board. 2009. Tagging Report — a comprehensive review of Columbia River Basin fish tagging technologies and programs. ISRP/ISAB 2009-1. Northwest Power and Conservation Council, Portland, Oregon.

Johnson, K.J. 2004. Regional overview of coded wire tagging of anadromous salmon and steelhead in northwest America. Regional Mark Processing Center, Pacific States Marine Fisheries Commission.

Kalinowski, S.T., Taper, M.L., and Marshall, T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. **16**(5): 1099–1106. doi:10.1111/j.1365-294X.2007.03089.x. PMID:17305863.

Koch, M., Hadfield, J.D., Sefc, K.M., and Sturmbauer, C. 2008. Pedigree reconstruction in wild cichlid fish populations. Mol. Ecol. **17**: 4500–4511. doi:10.1111/j.1365-294X.2008.03925.x. PMID:18986496.

Koljonen, M.-L., Pella, J.J., and Masuda, M. 2005. Classical individual assignments versus mixture modeling to estimate stock proportions in Atlantic salmon (*Salmon salar*) catches from DNA microsatellite data. Can. J. Fish. Aquat. Sci. **62**(9): 2143–2158. doi:10.1139/f05-128.

Krawczak, M. 1999. Informativity assessment for biallelic single nucleotide polymorphisms. Electrophoresis, **20**(8): 1676–1681. doi:10.1002/(SICI)1522-2683(19990101)20:8<1676::AID-ELPS1676>3.3.CO;2-4. PMID:10435431.

Marshall, T.C., Slate, J., Kruuk, L.E.B., and Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. **7**(5): 639–655. doi:10.1046/j.1365-294x.1998.00374.x. PMID:9633105.

McKenzie, J., Parsons, B., Sietz, A., Kopf, R.K., Mesa, M., and Phelps, Q. (*Editors*). 2012. Advances in Fish Tagging and Marking Technology. American Fisheries Society Symposium 76, American Fisheries Society, Bethesda, Maryland.

Morin, P.A., Luikart, G., Wayne, R.K., and the SNP Workshop Group. 2004. SNPs in ecology, evolution and conservation. Trends Ecol. Evol. 19(4): 208–216. doi:10.1016/j.tree.2004.01.009.

Narum, S.R. 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. Conserv. Genet. **7**(5): 783–787. doi:10.1007/s10592-005-9056-y.

Nielsen, R., Mattila, D.K., Clapham, P.J., and Palsbøll, P.J. 2001. Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. Genetics, **157**(4): 1673–1682. PMID:11290722.

Pacific Fishery Management Council. 2008. Research and data needs [online]. Availablie from http://www.pcouncil.org/wp-content/uploads/Res_Data_Needs_2008_Final_OCT08.pdf.

Pacific Salmon Commission. 2005. Report of the Expert Panel on the Future of the Coded Wire Tag Program for Pacific Salmon. 2005. Pacific Salmon Comm. Tech. Rep. No. 18 [online]. Availablie from http://www.rmpc.org/files/psctr18_CWT_Expert_Panel_Report.pdf.

Parker, N.C., Giorgi, A.E., Heidinger, R.C., Jester, D.B., Jr., Prince, E.D., and Winans, G.A. (*Editors*). 1990. Fish-marking techniques. American Fisheries Society Symposium 7, American Fisheries Society, Bethesda, Maryland.

Peakall, R., and Smouse, P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes, **6**(1): 288–295. doi:10.1111/j.1471-8286.2005.01155.

RMIS. Updated continuously since 1977. Regional Mark Information System Database [online]. Regional Mark Processing Center, Pacific States Marine Fisheries Commission, Portland, Ore. Available from http://www.rmpc.org.

Shaklee, J.B., Beacham, T.D., Seeb, L., and White, B.A. 1999. Managing fisheries using genetic data: case studies from four species of Pacific salmon. Fish. Res. **43**(1–3): 45–78. doi:10.1016/S0165-7836(99)00066-1.

Steele, C.A., Campbell, M.R., Ackerman, M., McCane, J., Hess, M., Campbell, N., and Narum, S. 2011. Parentage based tagging of Snake River hatchery steelhead and Chinook salmon. US Dept. of Energy Bonneville Power Administration Report No. 2010-031-00.

Stephenson, J.J., Campbell, M.R., Hess, J.E., Kozfkay, C.C., Matala, A.P., McPhee, M.V., Moran, P., Narum, S.R., Paquin, M.M., Schlei, O., Small, M.P., Van Doornik, D.M., and Wenburg, J.K. 2008. A centralized model for creating shared, standardized, microsatellite data that simplifies inter-laboratory collaboration. Conserv. Genet. **10**(4): 1145–1149. doi:10.1007/s10592-008-9729-4.

Webster, M.S., and Reichart, L. 2005. Use of microsatellites for parentage and kinship analyses in animals. Methods Enzymol. **395**: 222–238. doi:10.1016/S0076-6879(05)95014-3. PMID:15865970.