MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 1263–1276 doi: 10.1111/j.1365-294X.2010.04994.x

Who are the missing parents? Grandparentage analysis identifies multiple sources of gene flow into a wild population

MARK R. CHRISTIE, MELANIE L. MARINE and MICHAEL S. BLOUIN
Department of Zoology, Oregon State University, Corvallis, OR 97331-2914, USA

Abstract

In order to increase the size of declining salmonid populations, supplementation programmes intentionally release fish raised in hatcheries into the wild. Because hatchery-born fish often have lower fitness than wild-born fish, estimating rates of gene flow from hatcheries into wild populations is essential for predicting the fitness cost to wild populations. Steelhead trout (*Oncorhynchus mykiss*) have both freshwater resident and anadromous (ocean-going) life history forms, known as rainbow trout and steelhead, respectively. Juvenile hatchery steelhead that 'residualize' (become residents rather than go to sea as intended) provide a previously unmeasured route for gene flow from hatchery into wild populations. We apply a combination of parentage and grandparentage methods to a three-generation pedigree of steelhead from the Hood River, Oregon, to identify the missing parents of anadromous fish. For fish with only one anadromous parent, 83% were identified as having a resident father while 17% were identified as having a resident mother. Additionally, we documented that resident hatchery males produced more offspring with wild anadromous females than with hatchery anadromous females. One explanation is the high fitness cost associated with matings between two hatchery fish. After accounting for all of the possible matings involving steelhead, we find that only 1% of steelhead genes come from residualized hatchery fish, while 20% of steelhead genes come from wild residents. A further 23% of anadromous steelhead genes come from matings between two resident parents. If these matings mirror the proportion of matings between residualized hatchery fish and anadromous partners, then closer to 40% of all steelhead genes come from wild trout each generation. These results suggest that wild resident fish contribute substantially to endangered steelhead 'populations' and highlight the need for conservation and management efforts to fully account for interconnected *Oncorhynchus mykiss* life histories.

Keywords: captive breeding, hatcheries, *Oncorhynchus mykiss*, parentage, reproductive success, residualized hatchery fish, steelhead

Received 5 April 2010; revision received 26 October 2010; accepted 2 November 2010

Introduction

Many salmonids have both anadromous (ocean-going) and resident (stream inhabiting) life history strategies. Understanding the connections between these diverse life history strategies and the subsequent effects upon population dynamics remains a pressing question in salmon management (Salmon Recovery Science Review Panel 2004), ecology (Quinn & Myers 2004) and conservation (Olsen *et al.* 2006; McPhee *et al.* 2007). Because 23% of pacific salmon stocks are at moderate to high risk (Augerot & Foley 2005), with most listed as threatened or endangered under the U.S. Endangered Species Act, many salmon populations have extensive recovery plans that include supplementation with hatchery fish (Waples *et al.* 2007; Kostow 2009). These supplementation efforts, in conjunction with other hatchery programmes, release more than five billion hatchery fish into northern Pacific waters each year

Correspondence: Mark R. Christie, Fax: (541) 737 0501;
E-mail: christim@science.oregonstate.edu

(Heard 1995; Augerot & Foley 2005; Naish *et al.* 2008). Such large releases of hatchery fish have far-reaching population, community and ecosystem-wide effects (McClure *et al.* 2008; Ruckelshaus *et al.* 2009).

One well-established population-level effect of captive-reared fish is that they have lower fitness in the wild than wild-born fish (Hansen 2002; McLean *et al.* 2004), and these fitness differences can be genetically based (Araki *et al.* 2007a,b, 2009). Thus, determining the amount of gene flow from hatchery into wild populations is essential for accurately predicting the effects of hatcheries on wild populations. Several studies have used parentage analysis on complete samples of returning anadromous fish to estimate the fitness of hatchery fish relative to that of wild fish (Ford *et al.* 2006; Araki *et al.* 2007a; Berntson *et al.* in preparation). However, estimates of reproductive success of hatchery fish have focused only on anadromous individuals and are thus incomplete for species with both anadromous and resident life histories. Steelhead and rainbow trout, for example, are anadromous and resident forms of the same species (*Oncorhynchus mykiss*), and they interbreed freely in many rivers (Zimmerman & Reeves 2000; McPhee *et al.* 2007). It is apparent that resident fish often mate with anadromous fish from the large fraction (up to 65%) of missing parents in parentage studies that exhaustively sampled the anadromous adults (Seamons *et al.* 2004; Araki *et al.* 2007a). Furthermore, in many rivers, it is common for released hatchery juveniles to 'residualize' (remain in freshwater), rather than go to sea as intended by the hatchery programme (Viola & Schuck 1995; McMichael *et al.* 1997; Tipping *et al.* 2003). These residualized hatchery fish could provide an additional avenue of gene flow from hatchery to wild populations (Garant *et al.* 2003). Owing to the difficulty of exhaustively sampling resident fish, there has never been an attempt to estimate rates of gene flow from hatchery to wild populations via this alternate pathway.

Using a three-generation pedigree on steelhead from the Hood River, we circumvent the problem of sampling resident fish by matching steelhead with missing parents to their hatchery grandparents (Fig. 1). Because we have complete samples of anadromous fish, we can identify nonanadromous fish with parentage analysis. If a returning adult cannot be assigned back to one of its parents after accounting for genotyping errors, then the missing parent must not have been sampled (and is therefore nonanadromous). Here, we limit our inferences to individuals that are missing a single parent, as fish that are missing both parents could be strays from other rivers. If an offspring with a missing parent can be assigned back to hatchery broodstock grandparents, then the missing parent can be identified as a residualized hatchery steelhead. If, on the other hand, an offspring with a missing parent cannot be assigned back to hatchery broodstock grandparents, then the missing parent is a resident trout.

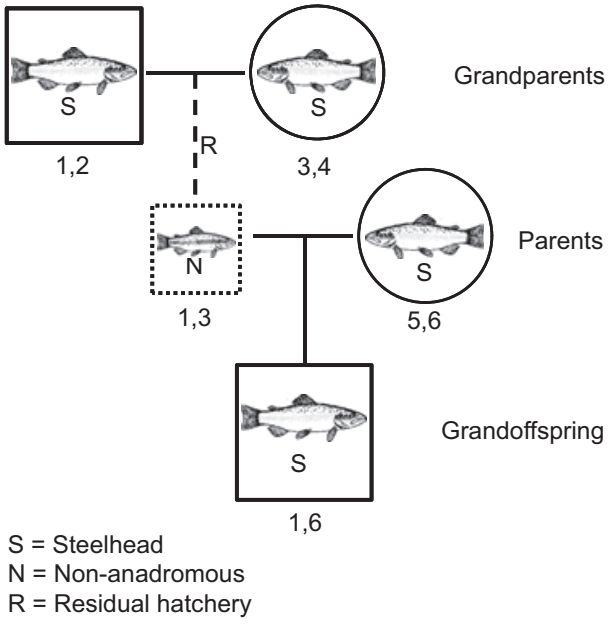

Fig. 1 A pedigree illustrating how grandparentage methods can identify reproductively successful, residualized hatchery steelhead (i.e., fish that did not go out to sea). The first step is to identify wild steelhead, labelled here as grandoffspring, that have an identified mother (solid circle) and an unsampled father (dashed square). The unsampled father must be a nonanadromous fish because all anadromous fish were sampled at the dam. If the grandoffspring can be assigned back to a pair of hatchery broodstock grandparents (dashed line) then we can infer that the missing father (dashed square) is a residualized hatchery steelhead. Numbers below the circles and squares represent alleles at a single codominant locus. At every locus, the paternal allele in the grandoffspring must match one of the four alleles present in the grandparental pair; else the trio is excluded as false. Here, paternal allele '1' is present in the grandparents, so this locus would not exclude the grandparents. The shared maternal allele (allele '6') can be eliminated from the search for shared grandparental alleles. We also performed these analyses for steelhead with identified fathers and missing mothers.

Grandparentage analysis is a powerful but underutilized tool for answering a broad array of questions in the fields of ecology and evolution. For example, while parentage analyses are routinely used to directly uncover patterns of dispersal (e.g., Waser *et al.* 2006; Planes *et al.* 2009), grandparentage methods can be used to identify patterns of gene flow by identifying the offspring of parents that dispersed from their natal populations (i.e., distinguish movement alone from realized gene flow). Grandparentage methods can also be used

to estimate the reproductive success of wild-born offspring for any captive or supplemental breeding programme (Letcher & King 2001). For example, when a small number of grandparents are used to create a large number of released parents, it may be easier to match the offspring to their grandparents rather than to their parents. Other uses for grandparentage methods include estimation of inbreeding coefficients, pedigree validation (e.g., for QTL or linkage mapping; heritability estimation), and effective captive breeding design (Allendorf & Luikart 2007).

Grandparentage analysis is, in many ways, similar to parentage analysis. While a parent shares one allele at every locus with an offspring, a grandparental pair shares one allele at every locus with a grandoffspring (i.e., one of the four alleles in the two grandparents must be identical by descent to one of the two alleles in their grandoffspring). This observation relates directly to previous work, establishing that grandparentage methods require approximately twice the number of alleles as parentage methods for the same percentage of correct assignments (Letcher & King 2001). Parentage and grandparentage analyses can be performed with either likelihood or exclusion-based methods, and each has its advantages depending upon the question being asked (reviewed by Jones & Ardren 2003; Jones *et al.* 2010). For many applications, likelihood-based methods have greater power for a given marker set because they can take into account the frequencies of shared alleles and not just the frequencies of alleles within a population (Anderson & Garza 2006; Ford & Williamson 2010). However, for this study, we employ exclusion-based methods because we could not estimate the number of unsampled grandparents in our study system. Exclusion-based methods do not need to take this information into account (Jones & Ardren 2003; Christie 2010), whereas it is required for hypothesis testing within a likelihood framework. The grandparentage methods we describe in this study do not require any of the parents of the grandoffspring to be identified or that the breeding matrix among grandparents be known, although this knowledge greatly increases exclusionary power.

In this study, we first employ detailed parentage analyses to identify fish that have one sampled (anadromous) parent and one missing (resident) parent. We next test whether the number of fish with missing fathers equals the number of fish with missing mothers. We also describe and validate new grandparentage methods, which we subsequently use to assign fish with missing parents back to candidate broodstock grandparents. This procedure allows us to identify reproductively successful residualized hatchery fish. Because hatchery fish often have lower fitness than wild fish, we further examine whether the number of offspring with a hatchery or wild resident parent is dependent upon whether the anadromous parent is hatchery or wild. We also examine the differences in age at spawning between male and female residualized hatchery fish. Lastly, we calculate the amount of gene flow into the steelhead 'population' from hatchery and wild *O. mykiss* with two anadromous parents and with one anadromous and one resident parent. These analyses reveal that wild residents contribute substantially to the anadromous steelhead gene pool.

Materials and methods

Sample collection

Samples were collected from the Hood River, Oregon, where steelhead are listed as threatened under the Endangered Species Act (Busby *et al.* 1996). Genetic samples for winter-run steelhead employed in this study were collected from run years 1991 through 2006. The number of steelhead samples analysed averaged 848 per year for a total of 12 725 samples. All samples were genotyped at eight highly polymorphic microsatellite loci (Omy 1001, Omy 1011, Omy 1191, Omy77, One108, One2, Ssa407 and Str2), which average 36 alleles per locus. These data were previously employed to determine the relative reproductive success of hatchery and wild steelhead (Araki *et al.* 2007a,b, 2009).

Resident fish in the Hood River are in low densities and very difficult to sample, which makes it impractical to use direct parentage methods to match residents to anadromous parents. All steelhead returning to spawning grounds in the Hood River must pass over the Powerdale dam, which is a complete barrier to migrating fishes. Every fish passed over the dam was individually handled, and samples of scales and fin tissue were collected for ageing and genetic analysis by staff of the Oregon Department of Fisheries and Wildlife. The staff also recorded the length, weight, gender and run-timing of every fish. Steelhead are easily categorized as hatchery or wild origin because all hatchery fish have their adipose fin removed before release. All wild fish and an approximately equal number of hatchery fish were passed over the dam each year. The winter-run hatchery fish were created using either two wild fish or one wild fish and a first-generation hatchery fish as broodstock (see Araki *et al.* 2007b for details). As per Araki *et al.* (2007b), we use 'wild' to refer to any fish spawned in the river under natural conditions, regardless of whether its parents have hatchery ancestry. We have DNA samples from all broodstock, and detailed records on broodstock pairings in the hatchery. A variety of strategies for releasing hatchery fish have been used for Hood River steelhead (see Kostow 2004 for details).

Summer-run fish present in this system are unlikely to be a source of resident fish because winter- and summer-run fish spawn in completely different forks of the river (Kostow 2004). Furthermore, winter- and summer-run fish are genetically distinct (F_{ST} ~0.01; Matala *et al.* 2009), and we cannot successfully assign anadromous winter-run fish to putative summer-run parents or vice versa (Araki *et al.* 2007c). We also compared the allele frequencies of the resident fish (i.e., alleles in steelhead that were inherited from resident parents; e.g., allele 1 from Fig. 1) to both summer- and winter-run fish. We find that the resident allele frequencies are identical to winter-run fish and very different from summer-run fish. This suggests that the resident fish that successfully mated in this system are derived from winter-run fish (See Appendix S1, Supporting information). Given these results, it is also unlikely that any of the fish identified as residents in this study are historically stocked resident trout because the stocked fish have substantially different allele frequencies (Cape Cod stock). Extensive details about management practices in this study system can be found in Olsen (2003).

Identifying steelhead with one missing parent

We examined only steelhead that were missing a single parent (Fig. 1) because (i) fish missing both parents could be strays from other rivers; (ii) the identified parent verifies the year in which the missing parent spawned; and (iii) being able to exclude half of the alleles in the grandoffspring because of a known parent greatly improves the exclusionary power for grandparentage analysis. We first performed two separate analyses to identify steelhead with known mothers and missing fathers or known fathers and missing mothers. We examined the six most recent broodyears because all of these individuals were likely to have grandparents within the 15 year data set (See Quinn 2005; Araki *et al.* 2009 for details on steelhead life cycles). Older broodyears would begin to include individuals with unsampled grandparents (i.e., grandparents that spawned before the first sample year in 1991). Because we were interested in examining mating between anadromous and resident fish, our parentage goals were twofold: (i) to definitively assign steelhead from the same broodyear to their one anadromous parent; and (ii) to be certain that their second parent really was missing. These two steps require different methodological approaches. For the first step, we used especially conservative criteria for a match, and in the second, we used relaxed criteria to ensure that the parent was not simply missing owing to genotyping or sexing errors (See Appendix S1, Supporting information).

Grandparentage methods

In order to make correct grandparentage assignments, we develop new exclusion-based methods to determine the probability of a putative grandparental pair and grandoffspring trio (hereafter: trio) sharing alleles across all loci by chance alone. We do not assume that the breeding matrix among grandparents is known, although this knowledge greatly increases exclusionary power by reducing the number of pairwise comparisons. We first present a general exclusion equation for the case where neither parent of the putative grandoffspring is known, and thus both alleles of the putative grandoffspring must be considered with equal weight. The probability of a randomly selected trio sharing an allele at a single locus equals:

$$\Pr(G) = \sum_{i=1}^{Na} 8p_i^2 - 16p_i^3 + 14p_i^4 - 6p_i^5 + p_i^6 - \sum_{i=1}^{Ng} (pq_{1i})(G_{pi}G_{qi} + G_{pq}G_{-(pq)}) \quad (1)$$

where Na equals the total number of alleles at a locus, and where p_i equals the allele frequency of allele p within the population. In the second term, Ng represents the total number of unique heterozygous genotypes shared between the samples of putative grandoffspring and grandparent pairs. The term pq_{1i} equals the frequency of the heterozygote pq_i within the sample of putative grandoffspring. The term $G_{pi}G_{qi}$ represents the proportion of grandparental pairs where one grandparent possesses allele p_i and the other grandparent possesses allele q_i. The state or order of alleles in either grandparent is inconsequential. Lastly, $G_{pq}G_{-(pq)}$ equals the proportion of grandparental pairs where one grandparent is heterozygous for pq_i and the other grandparent does not possess either allele.

We next consider the case where a putative grandoffspring has one missing parent and one identified parent. The advantage of having one identified parent is that their genotype can be used to exclude an allele from the putative grandoffspring (Fig. 1), which greatly increases exclusionary power. The first step is to exclude the known-parent allele from a putative grandoffspring and then calculate the appropriate probabilities. However, if a putative grandoffspring and its identified parent are both heterozygous for the same alleles, then it is not possible to exclude an allele from the putative grandoffspring. Thus, two equations are required: one equation for the putative grandoffspring where an allele can be excluded because of the parental contribution and one equation for the putative grandoffspring where both alleles must be considered. We first

define the equation at a single locus where 1 allele has been excluded:

$$\Pr(G') = \left[1 - \sum_{i=1}^{Ng} (pq_{1i})(pq_{2i})\right] \cdot \sum_{i=1}^{Na} 8p_i^2 - 16p_i^3 + 14p_i^4 - 6p_i^5 + p_i^6 \quad (2)$$

where pq_{2i} equals the number of pq_{i} heterozygotes within the sample of parents. Thus, $\sum_{i=1}^{Ng} (pq_{1i})(pq_{2i})$ equals the fraction of putative grandoffspring that are heterozygous for the same alleles as their one identified parent, such that 1 minus that quantity equals the fraction of putative offspring that are not heterozygous for the same alleles, and thus have one of their alleles excluded. For the occurrences where an allele cannot be excluded from a putative grandoffspring, we define:

$$\Pr(G'') = \left[\sum_{i=1}^{Ng} (pq_{1i})(pq_{2i})\right] \cdot \Pr(G_{het}) \quad (3)$$

Where Pr (G_{het}) is equivalent to eqn 1. We next calculate the probability of a randomly selected trio sharing an allele at a single locus, which equals the sum of probabilities for cases where an allele could and could not be excluded from the putative grandoffspring:

$$\Pr(G) = Pr(G') + \Pr(G'') \quad (4)$$

To generate the probability of a randomly selected trio sharing an allele across all loci, we require that all loci are in linkage equilibrium and are thus independent of one another (see Thompson & Meagher 1998 for possible solutions for linked loci). Provided that the loci are in linkage equilibrium, the probabilities from data sets with (eqn 4) or without (eqn 1) an identified parent can be multiplied across loci such that:

$$\Pr(\gamma) = \prod_{i=1}^{L} \Pr(G)_i \quad (5)$$

where L equals the total number of loci employed. This quantity can next be multiplied by the number of pairwise comparisons to generate the expected number of false trios:

$$\mathrm{FG}trios = \Pr(\gamma) \cdot n_1 \cdot n_2 \quad (6)$$

where n_1 equals the number of putative grandoffspring and n_2 equals the number of putative grandparent pairs. We can ultimately determine the probability of any putative trio being false by taking the expected number of false trios and dividing by the total number of putative trios:

$$\Pr(\varphi_G) = \frac{\mathrm{FG}trios}{N_p} \quad (7)$$

where Np equals the total number of putative grandparental pairs and grandoffspring that share at least one allele across all loci. Thus, the total number of putative trios also equals the sum of all the true and false trios. As an example, if the expected number of false trios was 10, and the total number of putative trios was 100, then the probability of any one of those putative trios being false, Pr (φ_G), would equal 0.01 (See Table 1 of Christie 2010 for explicit definitions of 'true', 'false' and 'putative'). Software for these methods is available at http://sites.google.com/site/parentagemethods/grandparentage.

We first validated eqns 1–7 with simulated data sets. Using the allele frequencies from our eight loci, we explored three scenarios: (i) data sets with an unknown breeding matrix and no sampled parents; (ii) data sets with a known breeding matrix, but with no sampled parents; and (iii) data sets with a known breeding matrix and 1 identified parent. We varied the sample size from 100 to 400 by intervals of 100, where sample size equals the number of putative grandparent pairs plus the number of putative grandoffspring. All simulated data sets had equal sample sizes of putative

Table 1 Results of parentage assignment for each broodyear. Categories are as follows: total number of steelhead from each broodyear (N_{Total}); total number of steelhead assigned to a mother (N_{Mother}) or father (N_{Father}); expected number of false parent–offspring pairs (F_{pairs}); the probability of any given putative parent–offspring pair being false [Pr (φ)]; and the number of steelhead assigned to a mother and definitively missing a father ($N_{\mathrm{No\ Father}}$) and the number of steelhead assigned to a father and definitively missing a mother ($N_{\mathrm{No\ Mother}}$). The number of fish with 1 missing parent are also expressed as a percentage of the total number of steelhead (% Missing)

Broodyear	N_{Total}	N_{Mother}	F_{pairs}	Pr (φ)	$N_{\mathrm{No\ Father}}$	% Missing
1997	793	480	4.31	0.009	224	28.25
1998	670	288	2.48	0.009	197	29.40
1999	490	156	2.94	0.019	92	18.78
2000	465	197	7.20	0.037	104	22.37
2001	391	178	5.41	0.030	87	22.25
2002	299	121	4.53	0.037	49	16.39
All years	3108	1420	26.87	0.019	753	24.23
Broodyear	N_{Total}	N_{Father}	F_{pairs}	Pr (φ)	$N_{\mathrm{No\ Mother}}$	% Missing
1997	793	263	2.68	0.010	32	4.04
1998	670	103	1.15	0.011	27	4.03
1999	490	107	1.97	0.018	26	5.31
2000	465	142	4.93	0.035	25	5.38
2001	391	138	3.83	0.028	24	6.14
2002	299	98	3.91	0.040	16	5.35
All years	3108	851	18.47	0.022	150	4.83

grandparent pairs and putative grandoffspring, which maximizes the number of pairwise comparisons. One thousand simulated data sets were created for each sample size and for each of the above three scenarios (See Appendix S1, Supporting information). For each simulated data set, theoretical estimates of the expected number of false grandparent–grandoffspring trios were calculated using eqn 6 and compared to the number of trios that shared alleles at all loci by chance alone (i.e., a false trio). Furthermore, we empirically validated our grandparentage methods by calculating the expected number of false trios at all eight loci and then subsequently measuring the actual number of false trios after genotyping all putative trios at five additional loci.

Matching steelhead to hatchery grandparents

We used all hatchery broodstock for all Hood River winter-run steelhead from 1991 to 2006 (n = 547 pairs) as the putative grandparents. We analysed each broodyear of putative grandoffspring (i.e., fish with one missing parent) separately. For each putative grandoffspring, we searched for putative grandparents among all known broodstock pairs. Assignments could only be made to the desired set of grandparents because we had excluded the alleles from the known-parent lineage from the grandoffspring. We assigned trios using Mendelian incompatibility and, at our eight highly polymorphic loci, all trios did not have more than one candidate set of grandparents (i.e., no grandoffspring matched more than one grandparental pair). Furthermore, no grandoffspring matched to putative grandparents that could not be real grandparents (e.g., putative grandparents from equivalent or more recent run years than the grandoffspring).

Using eqns 2–7 we determined the expected number of false trios (FG_{trios}) as well as the probability of any single putative trio being false, Pr (φ_G). Across broodyears, the expected number of false trios ranged from 0.08 to 2.08, and the probability of any single trio being false ranged from 0.015 to 0.353. Because many of these values were high [e.g., a Pr (φ_G) of 0.35 is interpreted as 35 of 100 trios being false], we genotyped all putative trios (n = 160 fish) at five additional polymorphic microsatellite loci (*Ogo*4, *Omm*1046, *Omy*7, *One*102 and *Ots*4; see Appendix S1, Supporting information). Using all 13 polymorphic loci, we calculated that the probability of a trio sharing alleles by chance is low ($P < 0.0006$). For all trios that did not match at the additional five loci, we rechecked the GENOTYPER files to ensure that a mismatch was not because of a binning or laboratory error. The genotyping of all putative trios at five additional loci also provided us with an opportunity to empirically validate our equations for estimating rates of false matching.

Our objective was to estimate the percentage of single-parent offspring whose missing parents were hatchery fish. The approach outlined above is conservative because it requires a match at all 13 loci. For example, some of the trios excluded above could be real trios, but mismatch at a locus owing to genotyping error. Therefore, we next performed grandparentage assignments with more relaxed criteria in order to calculate an upper bound on the number of true trios. To determine the number of loci to allow to mismatch, we assigned the offspring of known hatchery fish back to broodstock grandparents. Of a total of 210 assignments from 6 broodyears, 97% were assigned to broodstock grandparents at seven of eight loci. Therefore, we reconsidered all putative trios that shared an allele at a minimum of seven of the eight original loci. From this less conservative sample, we excluded any putative grandoffspring that were assigned to two separate paternal grandparental pairs (four occurrences), which is suggestive of a match by chance. This upper bound therefore includes all true trios, including those that may have genotyping errors. However, this upper bound may also include trios that share alleles by chance alone. Thus, the true number of trios probably lies between the conservative estimate using trios that match at exactly 13 loci and the upper bound calculated as described above.

Success of resident matings

We first compared the number of steelhead with missing fathers vs. the number of steelhead with missing mothers. For each broodyear, we performed a G-test (likelihood ratio test) to determine the goodness-of-fit between the observed and expected (50:50) number of missing fathers and mothers (Sokal & Rohlf 1995). All statistical analyses were completed in R version 2.9.1 (R Development Core Team 2009).

We next asked whether the number of offspring of a hatchery or wild resident parent is dependent upon whether the anadromous parent is hatchery or wild. Because our analyses are limited to matings that produced surviving offspring, these analyses could include a component owing to nonrandom mating and a component owing to offspring survival. Because we could not directly observe the matings, the two components cannot always be disentangled. It is important to note that while determining the exact mechanism is important, it is the number of individuals produced by each cross-type that is essential for conservation and management decisions (see Discussion). After applying our grandparentage analyses, we were able to determine

the number of offspring that resulted from one of four parental crosses: (i) residualized hatchery × hatchery anadromous; (ii) residualized hatchery × wild anadromous; (iii) wild resident (trout) × hatchery anadromous; and (iv) wild resident (trout) × wild anadromous. We constructed a two-way table and used a *G*-test of independence with a Williams' correction (Sokal & Rohlf 1995) to determine whether any of the four crosses deviate from the null expectation of independence. We performed this analysis separately for fish with missing fathers and fish with missing mothers and for both the stringent and relaxed grandparentage assignments.

Age at spawning of residualized hatchery fish

For grandparent–grandoffspring trios that were assigned at all 13 loci, we examined the age at spawning for the residualized hatchery parents. We could calculate this age because we knew the year in which the broodstock grandparents were spawned, which was also the year the missing parents were born. We also knew the broodyear of the grandoffspring from ageing the scales, which equalled the year in which the residualized hatchery fish spawned. Therefore, age at spawning of the missing parent equals the broodyear of the grandoffspring minus the year that the identified broodstock grandparents spawned. We calculated age at spawning for both residualized hatchery males and females and compared the centres of location for the two sexes using a two-tailed Mann–Whitney *U*-test.

Gene flow into the steelhead 'population'

We calculated the percentages of hatchery and wild genes in the steelhead 'population' by first using robust estimates of the percentage of Hood River steelhead with both parents, one parent, or no parents missing (Table S3 in Appendix S1, Supporting information). We employed values from Araki *et al.* (2007a) because they are calculated from the same data used in this study and because they account for the number of offspring assigned to false parents and the number of offspring that are not assigned to true parents. Similar estimates of the number of steelhead with parents missing were obtained by Seamons *et al.* (2007), suggesting that these results may apply to other systems. Because all steelhead have both, one, or no parents missing, these percentages also represent 100% of the steelhead genes. We next split these genes into resident and anadromous components (Table S3 in Appendix S1, Supporting information). For example, for fish with two anadromous or two resident parents, we doubled the genetic contribution to account for both parents. For fish with mixed parentage, we split the contribution evenly between anadromous and resident components. For anadromous genes, we calculated the gene flow from hatcheries by multiplying the fraction of hatchery fish passed over the dam (38%) by the relative reproductive success (RRS) of hatchery fish with two wild parents calculated over 6 run years (RRS = 0.848; Araki *et al.* 2007b). For resident genes (calculated via matings between resident and anadromous fish), we calculated the gene flow from hatcheries by multiplying our stringent and relaxed estimates of the percentage of missing parents that were identified to be residualized hatchery fish (6–14%) by the proportion of resident genes (Table S3 in Appendix S1, Supporting information). For steelhead that had two missing parents (most likely a cross between two resident fish), we could not determine what fraction of those genes came from wild residents vs. residualized hatchery fish (we comment further on this topic in the Discussion).

Results

Parentage results

An average of 518 fish belonged to each of the six broodyears (Table 1). For each broodyear, we were able to assign 46% to their mother at all eight loci. The probability of an incorrect assignment of a steelhead to a false mother equalled 0.019 across all broodyears (Table 1). Of the steelhead that were assigned to a mother, we next identified steelhead that were missing a father. Using our study-specific error rate of 0.0135, we determined that the probability of a putative parent–offspring pair having a genotyping error at two or more loci equalled 0.024. Thus, 97.6% of all pairwise comparisons had no errors or an error at one locus, verifying that allowing one locus to mismatch was sufficient (see also Araki *et al.* 2007b). Because many errors will not actually cause a mismatch, the probability of a putative parent–offspring pair having a mismatch-causing genotyping error at two or more loci is substantially <0.024 (See Appendix S1, Supporting information). Across broodyears, 53% of the steelhead assigned to a mother were identified as missing a father (i.e., the father was nonanadromous).

We next performed parentage analysis to identify steelhead with known fathers and missing mothers. For all broodyears, we were able to assign 27% of steelhead to a father at all eight loci. The probability of incorrectly assigning a steelhead to a false father averaged 0.022 across all broodyears (Table 1). Of the steelhead that were assigned to a father, 18% were missing a mother. The remaining unassigned individuals either were missing two parents or had sufficient probability of having

two anadromous parents that they were excluded from further analyses.

Of all of the wild steelhead, 24% (n = 753) were assigned to a mother and identified as missing a father, while only 5% (n = 150) were assigned to a father and identified as missing a mother. The percentage of fish with a missing father varied from 16% to 29% across broodyears, while the percentage of fish with a missing mother varied from 4% to 6% across broodyears (Table 1). G-tests identified that there were many more steelhead with a missing father than with a missing mother (1997–2002: $P < 0.001$). These results remained significant if the expected ratio was adjusted 64:36 (1997–2002: $P < 0.01$), which is the ratio of females to males passed over the dam, and which might have provided resident males with more opportunities to mate with a female steelhead.

Accuracy of grandparentage methods

The results from our simulations with the allele frequencies from the eight loci used in this study indicated that our equations accurately predicted the number of false trios regardless of sample size (Fig. 2). Furthermore, our methods accurately predicted the number of false trios for data sets with a known or unknown grandparental breeding matrix and for data sets with known or unknown parents. For all simulations, the grandparentage equations predicted the expected number of false trios with both low bias and high precision (see Appendix S1, Supporting information). The number of false trios increased at an increasing rate with larger sample sizes. This pattern is not surprising given that the number of pairwise comparisons increases exponentially for a linear increase in sample size. Data sets with an unknown breeding matrix and no known parents had the greatest number of false trios, followed by data sets with a known grandparental breeding matrix, but no known parents. Data sets with a known breeding matrix and one known parent had the lowest numbers of false trios for a given sample size.

From our empirical calculations with eight loci, the total expected number of false trios calculated with eqn 6 equalled 7.35 for steelhead with missing fathers and 0.86 for steelhead with missing mothers. The actual number of false trios revealed by genotyping all the putative trios at five additional loci equalled 7 and 1, respectively (Table 2). False trios were easily identified as they did not match at 4 (n = 2) or 5 loci (n = 6), while true trios matched at all five additional loci. Expected and actual number of false trios also matched closely in each individual broodyear (Table 2).

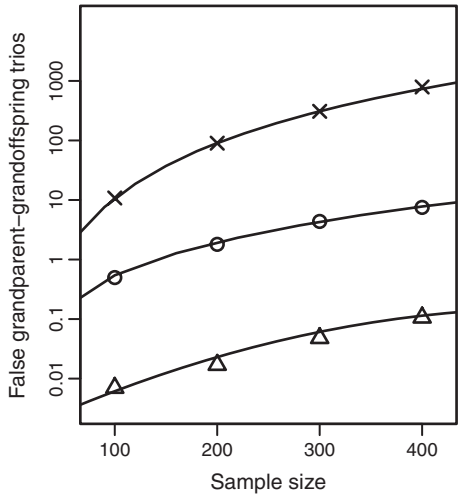

Fig. 2 Actual and predicted number of false grandparent–grandoffspring trios as measured by simulated data sets (symbols) and as predicted by the grandparentage exclusion equations (solid lines). Sample size includes the number of putative grandoffspring plus the number of putative grandparental pairs, both of which were of equal size to maximize the number of pairwise comparisons. Actual number of false grandparent–grandoffspring trios (symbols) were measured from 1000 simulated data sets that employed our empirical allele frequency estimates at all eight loci. Symbols, and the lines they are plotted on, are as follows: (×) data sets with unknown grandparental breeding matrix and no known parents; (○) data sets with a known grandparental breeding matrix, but no known parents; and (△) data sets with a known grandparental breeding matrix and one identified parent.

Identification of residualized hatchery parents

Of the 903 offspring that were missing one parent, we conservatively estimated that 57 fish (6.3%) matched a grandparental pair in the hatchery broodstock at all 13 loci and that 128 fish (14.2%) matched at seven of eight loci (Table 3). At all 13 loci, the expected number of false trios, summed over all broodyears, equalled 0.032. This probability means that only 3.2 of 100 similar samples would, on average, contain 1 false trio. The probability of any putative trio being false, calculated over all broodyears, was <0.0006 (Table 3). Thus, it is very unlikely that any of the 57 matches shared alleles by chance. To illustrate this point, we can rearrange eqn 6 to calculate the number of grandparent pairs we would have had to sample in order to observe a single false trio. Using our estimate of Pr (φ_G) averaged over broodyears, we calculate that we would have had to sample at least 985 additional grandparent pairs, on average, in order to observe a single false trio in our data set. Furthermore, we would have had to sample an additional 56 145 grandparent pairs for all of our identified trios to share alleles by chance alone. We also

Table 2 Expected number of false grandparent–grandoffspring trios calculated with individuals genotyped at eight loci. After genotyping the putative trios at five additional loci, we identified the actual number of false trios

Broodyear of Steelhead w/missing fathers	Expected # False trios	Actual # False trios
1997	1.66	2
1998	2.08	2
1999	0.94	1
2000	1.41	2
2001	0.9	0
2002	0.36	0
All years	7.35	7
Broodyear of Steelhead w/missing mothers	**Expected # False trios**	**Actual # False trios**
1997	0.17	0
1998	0.13	1
1999	0.21	0
2000	0.14	0
2001	0.13	0
2002	0.08	0
All years	0.86	1

determined that each grandparent contributed, on average, 54% of the alleles found in the grandoffspring (with the extra 4% consisting of alleles found in both grandparents), which is the Mendelian expectation (see Fig. 1).

For fish with missing fathers, analysis of the four different mating types reveal that they are not independent ($P < 0.031$; Table 4). The p-value remained low when we used the less stringent grandparentage assignments ($P < 0.041$). Because this analysis explicitly tests for independence, we can conclude that the number of offspring with residualized hatchery fathers depends on whether their anadromous mothers were hatchery or wild. By examining the expected values (the quotient of the product of row and column totals to the overall total), we observe that this pattern is largely due to there being more fish with a wild anadromous mother and a residualized hatchery father than fish with a hatchery anadromous mother and a residualized hatchery father (Table 4). For offspring with missing mothers, we could not reject the null hypothesis of independence ($P = 0.676$), possibly because of smaller sample sizes.

Age at spawning of residualized hatchery fish

The age at spawning for reproductively successful residualized hatchery steelhead differed substantially between the sexes (Fig. 3). The median age at spawning for residualized hatchery males was 1 year, while the median age at spawning for residualized hatchery females was considerably older at 5 years (Mann–Whitney *U*-test; $P < 0.001$). While the majority of males spawned at 1 year of age, low numbers of males were documented to spawn up to 9 years of age. Females, on the other hand, had a much narrower documented range of age at spawning (4–6 years).

Gene flow into the steelhead 'population'

Using our estimates that between 6% and 14% of missing parents from anadromous × resident crosses were residualized hatchery fish (Table 3) and previously published estimates of reproductive success and of missing parents (Araki *et al.* 2007a,b), we estimated the percentages of hatchery and wild gene flow into Hood River steelhead (Fig. 4). Using our stringent grandparentage assignments, only 1% of steelhead genes come from residualized hatchery steelhead via resident × anadromous matings, while 20% of steelhead genes come from wild residents. Using our relaxed estimates of the percentage of missing parents that were residualized hatchery fish (14%) increased the percentage of steelhead genes from residualized hatchery fish only to 2.6%. Thus, there is far less hatchery contribution to steelhead from anadromous × resident matings than from matings between two anadromous fish (19% hatchery genes). Up to 23% of steelhead genes come from resident by resident matings (and stray fish), but we could not estimate the proportion coming from hatcheries using these methods.

Discussion

Grandparentage analysis is a valuable tool that can address a broad suite of questions in the fields of ecology and evolution. Here, we used grandparentage methods to measure gene flow from resident wild and resident hatchery *Oncorhynchus mykiss* into the anadromous population. Using these methods, we show that a large portion of steelhead genes (at least 20%) come from wild resident fish, while a much smaller percentage of steelhead genes come from resident hatchery fish. Importantly, the grandparentage methods used in this study can be easily applied to other systems. For example, one could use these methods to determine the amount of introgression from nonlocal (introduced) conspecific fishes (e.g., Utter 2001) or to identify the source of individuals for which it is unrealistic to genotype all candidate parents (e.g., agricultural or captive breeding programmes in which a small number of grandparents lead to a large number of parents). Both simulated data sets and direct validation of our

Table 3 Number of reproductively successful residualized hatchery steelhead as determined by grandparentage analysis. Column headers are as follows: $N_{Missing}$ equals the number of steelhead having 1 missing parent and 1 identified parent. These are the potential grandoffspring; 'Lower' equals the total number of steelhead assigned to broodstock grandparents at all 13 loci; FG*trios* equals the expected number of false grandparent–grandoffspring trios at all 13 loci; Pr (φ_G) equals the probability of any putative grandparent-grandoffspring trio being false at all 13 loci; 'Upper' equals the number of grandparentage assignments using less stringent grandparentage methods (see Methods). % Assigned equals the percent of putative grandoffspring that were assigned to hatchery broodstock grandparents for lower and upper stringencies, respectively

			Residualized hatchery males				
Broodyear	$N_{Missing}$	Lower	% Assigned	FG*trios*	Pr (φ_G)	Upper	% Assigned
1997	224	10	4.5	0.0064	0.0006	27	12.1
1998	197	8	4.1	0.0080	0.0010	18	9.1
1999	92	6	6.5	0.0036	0.0006	18	19.6
2000	104	2	1.9	0.0054	0.0027	12	11.5
2001	87	7	8.1	0.0035	0.0005	13	14.9
2002	49	6	12.3	0.0014	0.0002	9	18.4
Total	753	39	5.2	0.0283	0.0007	97	12.9
			Residualized hatchery females				
Broodyear	$N_{Missing}$	Lower	% Assigned	FG*trios*	Pr (φ_G)	Upper	% Assigned
1997	32	4	12.5	0.00065	0.00016	7	21.9
1998	27	8	29.6	0.00051	0.00006	10	37.0
1999	26	3	11.5	0.00080	0.00027	5	19.2
2000	25	0	0.0	0.00056	NA	2	8.0
2001	24	2	8.3	0.00051	0.00026	5	20.8
2002	16	1	6.3	0.00029	0.00029	2	12.5
Total	150	18	12.0	0.00332	0.00018	31	20.7
			Both sexes				
Total	903	57	6.3	0.03162	0.00055	128	14.2

Table 4 Number of offspring assigned to one of four possible crosses after grandparentage analyses (expected values shown in parentheses). Contingency tests reveal that cross-type is not independent for fish with nonanadromous fathers ($P < 0.031$), which suggests that the number of returning offspring depends on whether (or how often) a nonanadromous steelhead mated with a hatchery or wild anadromous mother. We could not reject the null hypothesis of independence for nonanadromous mothers ($P = 0.68$)

		Hatchery	Wild
		Anadromous mother	
Father	Hatchery residualized	8 (14)	31 (25)
	Resident trout (Wild)	266 (260)	448 (454)
		Anadromous father	
Mother	Hatchery residualized	6 (7)	12 (11)
	Resident trout (Wild)	52 (51)	80 (81)

grandparentage methods revealed that they accurately predicted the expected number of false trios with low bias. Thus, these methods can be applied to many natural populations and are ideally suited for large natural populations where many pairwise comparisons are needed to identify grandparents and grandoffspring.

Parentage analyses revealed substantially more steelhead with missing fathers than missing mothers. A high rate of missing fathers has been documented in similar parentage studies (e.g., Seamons *et al.* 2004) and is not surprising given resident males are routinely observed sneaking fertilizations with anadromous females (McMillan *et al.* 2007). In contrast, we are unaware of any direct observations of resident females mating with steelhead males in the wild. Here, we document that up to 5% of all steelhead are the product of such matings. Because all steelhead were sampled as they passed over the dam, we are confident that the missing mothers are resident fish. Analysis of the allele frequencies of resident fish suggests that they are identical to anadromous, winter-run fish. This result suggests that resident and anadromous *O. mykiss* life histories are highly connected and that the resident fish are not summer-run or historically stocked trout. It is possible that

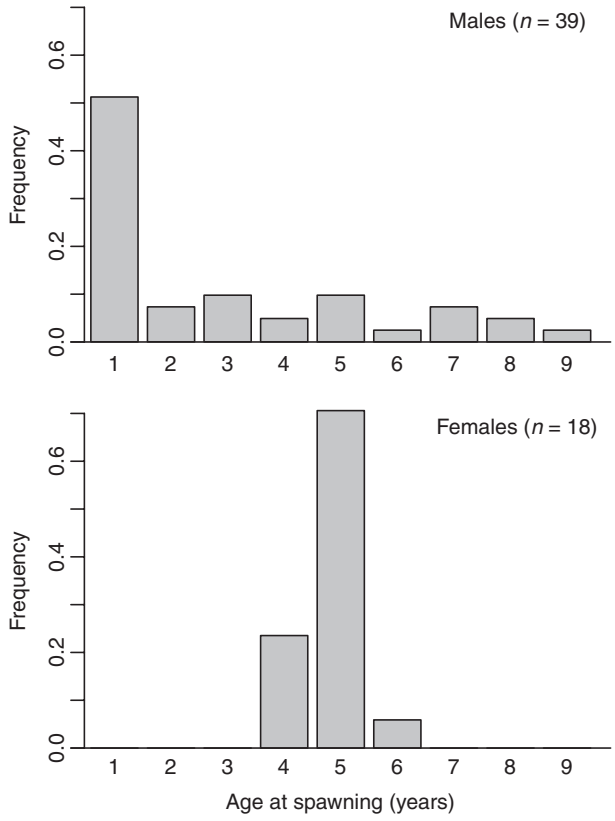

Fig. 3 Frequency distributions of the age at spawning for identified residualized hatchery steelhead. All identified residualized hatchery fish represent unique individuals because their offspring were assigned to unique grandparents. The median age at spawning for males was 1 year, while the median age at spawning for females was 5 years. A Mann–Whitney *U*-test revealed that these two distributions have very different centres of location ($P < 0.001$).

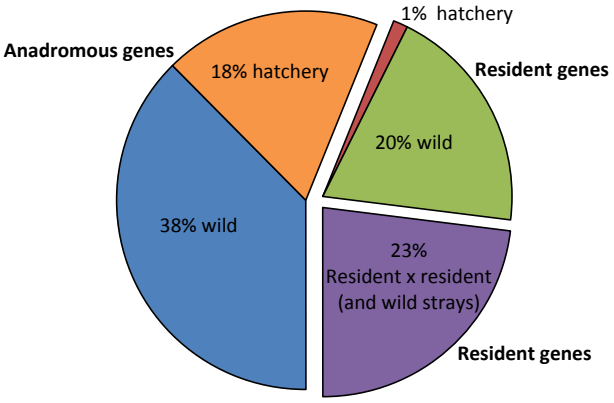

Fig. 4 Sources of gene flow into anadromous Hood River *Oncorhynchus mykiss* (steelhead). The two left slices represent the amount of hatchery and wild gene flow into the steelhead 'population' from all matings involving anadromous fish (i.e., anadromous × anadromous and anadromous × resident). The upper-right slices represent the amount of gene flow into the steelhead 'population' from residualized hatchery steelhead and wild residents as determined by matings between resident and anadromous *O. mykiss*. The lower-right slice represents the amount of gene flow from resident × resident matings (and wild strays), for which we could not estimate the proportion of hatchery gene flow.

a very small portion of the missing parents are cutthroat trout (*Oncorhynchus clarki*) (Allendorf *et al.* 2004; Matala *et al.* 2009), but not for the missing parents identified as residualized hatchery steelhead.

It has only recently become evident that resident *O. mykiss* 'populations' might play an important role in the maintenance and recovery of steelhead 'populations' (Salmon Recovery Science Review Panel 2004). In the Hood River, a minimum of 21% of the steelhead genes' come from resident fish each generation. The ability of resident fish to provide stability to steelhead populations has been previously hypothesized given the observation of constant effective population sizes in the face of wildly fluctuating census sizes (Araki *et al.* 2007c, Berntson *et al.* in preparation). Furthermore, we document that 95% of the genes from anadromous × resident matings are from wild resident *O. mykiss*. If we examined only the anadromous individuals, we would calculate that 34% of steelhead genes come from hatchery fish. By including information from anadromous × resident matings, we now calculate that 26% of steelhead genes come from hatchery fish. Thus, an additional way in which resident *O. mykiss* benefit the anadromous steelhead gene pool is by 'diluting' the genetic contribution of hatchery fish.

An additional 23% of steelhead genes come from steelhead with two missing parents. It is likely that a small portion of this percentage comes from wild strays from other rivers (notice that hatchery strays *are* included in the anadromous hatchery slice of Fig. 4). Because other studies have found that the rate of straying for wild steelhead is typically between 1% and 3% (Shapovalov & Taft 1954; Quinn 2005; Keefer *et al.* 2008), we conclude that most of this 23% comes from matings between two resident fish. Our methods could not identify what fraction of the resident by resident matings involve hatchery fish. However, if the involvement of residualized hatchery fish in resident × anadromous and in resident × resident matings is similar, then an even greater amount of steelhead genes (perhaps up to 40%) come from wild resident *O. mykiss*.

In this study, we also documented that hatchery residualized males produced more offspring with wild female steelhead than with hatchery female steelhead. This result has several possible explanations, which may not be mutually exclusive. The first explanation is mate choice: either wild steelhead females actively seek out

residualized hatchery males more often than do hatchery females or residualized hatchery males prefer to mate with wild female steelhead. The latter hypothesis seems more likely given the well-documented sneaking behaviour of resident males (McMillan *et al.* 2007). A second explanation is different encounter rates: it is possible that hatchery and wild females exhibit different spatial (Mackey *et al.* 2001) or temporal (Hansen & Mensberg 2009) behaviours on the spawning grounds. Thus, wild anadromous females may be more likely to encounter residualized hatchery males. A third explanation is differences in offspring survival. Because we are making inference based on the returning progeny of matings, it is possible that wild and hatchery females mated with nonadromous fish at equal rates, but that the progeny of matings with hatchery fish had lower survival. This hypothesis is consistent with previous data showing that anadromous hatchery fish have low fitness (Araki *et al.* 2007a,b, 2009). Whatever the mechanism, this observation provides additional evidence that wild fish have greater reproductive success than hatchery fish.

Because each grandoffspring was assigned to a unique grandparental pair, we documented a minimum of 57 reproductively successful residualized hatchery fish produced during the six broodyears we examined. It is important to note, however, that our approach focused on reproductively successful individuals and the actual rate of hatchery residualization may be considerably higher. In fact, field studies in other rivers have documented a large proportion of residents that appear to be residualized hatchery fish (McMichael & Pearsons 2001; Sharpe *et al.* 2007). The additional residualized hatchery fish not documented in this study could still negatively impact wild fish via ecological interactions even though they do not succeed in passing on their genes. For example, residualized hatchery fish could compete with wild trout for limited food resources (McMichael *et al.* 1997), shelters from predators (McMichael *et al.* 1999) or spawning habitat (Simpson *et al.* 2009).

Residualized hatchery males mated with steelhead at a median age of 1 year and a range up to 9 years, while residualized hatchery females mated with steelhead at 4–6 years of age. Thus, the sexes might have different adaptive strategies for maximizing reproductive success as residents. The older age of females could result simply because they must attain a minimum size to be of interest to the much larger male steelhead. Alternatively, it could be that older females are capable of developing eggs of greater quantity or quality (Berkeley *et al.* 2004), and so produce offspring that we were more likely to detect. The lower maximum spawning age of detected females (6 vs. 9 years for males) may be ecologically relevant, but this observation may simply be due to smaller sample sizes.

In conclusion, our grandparentage analyses revealed that only 1% of steelhead genes come from residualized hatchery fish via anadromous × resident matings vs. 20% from wild resident *O. mykiss*. These results suggests that resident *O. mykiss* play a substantial role in providing stability to an anadromous steelhead population that faces stochastic environmental variation (e.g., poor ocean conditions), anthropogenic causes of population decline (e.g., fishing pressure) and reduced reproductive success from hatchery introgression. The fact that many wild steelhead populations are continuing to decline suggests that wild resident fish may warrant further protection in populations with highly connected resident and anadromous life histories. More generally, this study underscores the need to adequately protect and appropriately manage all aspects of salmonid life history.

Acknowledgements

We thank Hitoshi Araki, William Ardren, Becky Cooper, Vindhya Amarasinghe, Jeff Stephenson and the Oregon State Center for Genome Research and Biotechnology for advice and for help with laboratory protocols and genotyping efforts. We are especially grateful to all ODFW staff that collected data and acquired tissue samples for this extensive data set. We acknowledge North Dakota Project WILD for use of their clipart. We also thank Kaitlin Bonner, Rod French, Catherine Searle, John McMillan, Lyle Curtis, Ivan Phillipsen, Michael Hansen and five anonymous reviewers for helpful comments, discussions and edits. This research was funded by grants from the Bonneville Power Administration to Michael Blouin.

References

Allendorf FW, Luikart GL (2007) *Conservation and the Genetics of Populations*. Blackwell Publishing, Singapore.

Allendorf FW, Leary RF, Hitt NP *et al.* (2004) Intercrosses and the U.S. Endangered Species Act: should hybridized populations be included as westslope cutthroat trout? *Conservation Biology*, **18**, 1203–1213.

Anderson EC, Garza JC (2006) The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, **172**, 2567–2582.

Araki H, Ardren WR, Olsen E, Cooper B, Blouin MS (2007a) Reproductive success of captive-bred steelhead trout in the wild: evaluation of three hatchery programs in the hood river. *Conservation Biology*, **21**, 181–190.

Araki H, Cooper B, Blouin MS (2007b) Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. *Science*, **318**, 100–103.

Araki H, Waples RS, Ardren WR, Cooper B, Blouin MS (2007c) Effective population size of steelhead trout: influence of variance in reproductive success, hatchery programs, and genetic compensation between life-history forms. *Molecular Ecology*, **16**, 953–966.

Araki H, Cooper B, Blouin MS (2009) Carry-over effect of captive breeding reduces reproductive fitness of wild-born descendants in the wild. *Biology Letters*, **5**, 621–624.

Augerot X, Foley DN (2005) *Atlas of Pacific salmon: the first map-based status assessment of salmon in the North Pacific*. University of California Press, Los Angeles.

Berkeley SA, Chapman C, Sogard SM (2004) Maternal age as a determinant of larval growth and survival in a marine fish, *Sebastes melanops*. *Ecology*, **85**, 1258–1264.

Busby P, Wainwright T, Bryant G *et al.* (1996) Status review of West Coast steelhead from Washington, Idaho, Oregon, and California. *Technical Memorandum NMFS-NWFSC-27*. Available from http://www.nwr.noaa.gov.

Christie MR (2010) Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. *Molecular Ecology Resources*, **10**, 115–128.

Ford M, Williamson K (2010) The aunt and uncle effect revisited-the effect of biased parentage assignment on fitness estimation in a supplemented salmon population. *Journal of Heredity*, **101**, 33–41.

Ford MJ, Fuss H, Boelts B *et al.* (2006) Changes in run timing and natural smolt production in a naturally spawning coho salmon (*Oncorhynchus kisutch*) population after 60 years of intensive hatchery supplementation. *Canadian Journal of Fisheries and Aquatic Sciences*, **63**, 2343–2355.

Garant D, Fleming I, Einum S, Bernatchez L (2003) Alternative male life-history tactics as potential vehicles for speeding introgression of farm salmon traits into wild populations. *Ecology Letters*, **6**, 541–549.

Hansen M (2002) Estimating the long-term effects of stocking domesticated trout into wild brown trout (*Salmo trutta*) populations: an approach using microsatellite DNA analysis of historical and contemporary samples. *Molecular Ecology*, **11**, 1003–1015.

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

Heard WR (1995) An estimate of total 1992 hatchery releases of the North Pacific Ocean and adjacent seas. *North Pacific Anadromous Fish Commission Doc. 154*. Auke Bay Fisheries Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service (NMFS), National Oceanic and Atmospheric Administration (NOAA), Juneau, AK.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, **12**, 2511–2523.

Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, **10**, 6–30.

Keefer ML, Caudill CC, Peery CA, Lee SR (2008) Transporting juvenile salmonids around dams impairs adult migration. *Ecological applications*, **18**, 1888–1900.

Kostow K (2009) Factors that contribute to the ecological risks of salmon and steelhead hatchery programs and some mitigating strategies. *Reviews in Fish Biology and Fisheries*, **19**, 9–31.

Kostow KE (2004) Differences in juvenile phenotypes and survival between hatchery stocks and a natural population provide evidence for modified selection due to captive breeding. *Canadian Journal of Fisheries and Aquatic Sciences*, **61**, 577–589.

Letcher B, King T (2001) Parentage and grandparentage assignment with known and unknown matings: application to Connecticut River Atlantic salmon restoration. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 1812–1821.

Mackey G, McLean JE, Quinn TP (2001) Comparisons of run timing, spatial distribution, and length of wild and newly established hatchery populations of steelhead in Forks Creek, Washington. *North American Journal of Fisheries Management*, **21**, 717–724.

Matala AP, French R, Olsen E, Ardren WR (2009) Ecotype distinctions among steelhead in Hood River, Oregon, allow real-time genetic assignment of conservation broodstocks. *Transactions of the American Fisheries Society*, **138**, 1490–1509.

McClure MM, Utter FM, Baldwin C *et al.* (2008) Evolutionary effects of alternative artificial propagation programs: implications for viability of endangered anadromous salmonids. *Evolutionary Applications*, **1**, 356–375.

McLean JE, Bentzen P, Quinn TP (2004) Does size matter? Fitness-related factors in steelhead trout determined by genetic parentage assignment. *Ecology*, **85**, 2979–2985.

McMichael GA, Pearsons TN (2001) Upstream Movement of Residual Hatchery Steelhead into Areas Containing Bull Trout and Cutthroat Trout. *North American Journal of Fisheries Management*, **21**, 943–946.

McMichael G, Sharpe C, Pearsons T (1997) Effects of residual hatchery-reared steelhead on growth of wild rainbow trout and spring Chinook salmon. *Transactions of the American Fisheries Society*, **126**, 230–239.

McMichael GA, Pearsons TN, Leider SA (1999) Behavioral interactions among hatchery-reared steelhead smolts and wild *Oncorhynchus mykiss* in natural streams. *North American Journal of Fisheries Management*, **19**, 948–956.

McMillan J, Katz S, Pess G (2007) Observational evidence of spatial and temporal structure in a sympatric anadromous (winter steelhead) and resident rainbow trout mating system on the Olympic Peninsula, Washington. *Transactions of the American Fisheries Society*, **136**, 736–748.

McPhee MV, Utter F, Stanford JA *et al.* (2007) Population structure and partial anadromy in *Oncorhynchus mykiss* from Kamchatka: relevance for conservation strategies around the Pacific Rim. *Ecology of Freshwater Fish*, **16**, 539–547.

Naish KA, Taylor JE, Levin PS *et al.* (2008) An evaluation of the effects of conservation and fishery enhancement on wild populations of salmon. In: *Advances in Marine Biology* (ed. HW Sims), pp. 61–194, Academic Press, The Netherlands.

Olsen EA (2003) *Hood River and Pelton ladder evaluation studies. Annual report 2000–2001 of the Oregon Department of Fish and Wildlife*, Oregon Department of Fish and Wildlife, Portland, OR.

Olsen JB, Wuttig K, Fleming D, Kretschmer EJ, Wenburg JK (2006) Evidence of partial anadromy and resident-form dispersal bias on a fine scale in populations of *Oncorhynchus mykiss*. *Conservation Genetics*, **7**, 613–619.

Planes S, Jones GP, Thorrold SR (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences*, **106**, 5693–5697.

Quinn TP (2005) *The Behavior and Ecology of Pacific Salmon and Trout*. University of Washington Press, Seattle.

Quinn TP, Myers KW (2004) Anadromy and the marine migrations of Pacific salmon and trout: rounsefell revisited. *Reviews in Fish Biology and Fisheries*, **14**, 421–442.

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for

Statistical Computing, Vienna, Austria. Available from http://www.R-project.org-http://www.R-project.org.

Ruckelshaus M, Essington T, Levin P (2009) Puget Sound, Washington, USA. In: *Ecosystem-Based Management for the Oceans* (eds McLeod K, Leslie H). pp. 201–226, Island Press, Washington DC.

Salmon Recovery Science Review Panel (2004) *Report for meeting held 30 August–2 September 2004*. National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle. Available from http://www.nwfsc.noaa.gov/trt/rsrpreport sept30 2004b.pdf.

Seamons T, Bentzen P, Quinn T (2004) The mating system of steelhead, *Oncorhynchus mykiss*, inferred by molecular analysis of parents and progeny. *Environmental Biology of Fishes*, **69**, 333–344.

Seamons TR, Bentzen P, Quinn TP (2007) DNA parentage analysis reveals inter-annual variation in selection: results from 19 consecutive brood years in steelhead trout. *Evolutionary Ecology Research*, **9**, 409–431.

Shapovalov L, Taft AC (1954) The life histories of the steelhead rainbow trout (*Salmo gairdneri gairdneri*) and silver salmon (*Oncorhynchus kisutch*) with special reference to Waddell Creek, California, and recommendations regarding their managemet. *California Department of Fish and Game Fish Bulletin*, **98**, 375.

Sharpe CS *et al.* (2007) Growth Modulation during Juvenile Rearing Can Reduce Rates of Residualism in the Progeny of Wild Steelhead Broodstock. *North American Journal of Fisheries Management*, **27**, 1355–1368.

Simpson WG, Kennedy BM, Ostrand KG (2009) Seasonal foraging and piscivory by sympatric wild and hatchery-reared steelhead from an integrated hatchery program. *Environmental Biology of Fishes*, **86**, 473–482.

Sokal RR, Rohlf FJ (1995) *Biometry*. W.H. Freeman and Company, New York.

Thompson EA, Meagher TR (1998) Genetic linkage in the estimation of pairwise relationship. *Theoretical and Applied Genetics*, **97**, 857–864.

Tipping JM, Gannam AL, Hillson TD, Poole JB (2003) Use of size for early detection of juvenile hatchery steelhead destined to be precocious males. *North American Journal of Aquaculture*, **65**, 318–323.

Utter F (2001) Patterns of subspecific anthropogenic introgression in two salmonid genera. *Reviews in Fish Biology and Fisheries*, **10**, 265–279.

Viola AE, Schuck ML (1995) A method to reduce the abundance of residual hatchery steelhead in rivers. *North American Journal of Fisheries Management*, **15**, 488–493.

Waples R, Ford M, Schmitt D (2007) Emperical results of salmon supplementation in the northeast pacific: a preliminary assesment. In: *Ecological and Genetic Implications of Aquaculture Activities* (ed. Bert TM), pp. 383–403, Kluwer Academic Publishers, The Netherlands.

Waser PM, Busch JD, McCormick CR, DeWoody JA (2006) Parentage analysis detects cryptic precapture dispersal in a philopatric rodent. *Molecular Ecology*, **15**, 1929–1937.

Zimmerman C, Reeves G (2000) Population structure of sympatric anadromous and nonanadromous *Oncorhynchus mykiss*: evidence from spawning surveys and otolith microchemistry. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 2152–2162.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1. The study system, parentage methods, grand-parentage methods, genotyping error rates, and estimates of gene flow.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 1080–1082

NEWS AND VIEWS

PERSPECTIVE

Grandfathering in a new era of parentage analysis

KENYON B. MOBLEY

Department of Ecology and Environmental Science, Umeå University, 90187 Umeå, Sweden

The advent of DNA fingerprinting and microsatellite techniques has revolutionized the way in which we investigate genetic pedigrees in the wild (Pemberton 2008). With large and often incomplete data sets consisting of hundreds to thousands of individuals over multiple generations becoming commonplace, new methods in parentage analysis are being developed to rise to the next generation of questions and challenges. In this issue, Christie *et al.* (2011) provide a simple yet elegant solution to the problem of identifying missing parents and assessing hybrid fitness in a mixed population of wild and hatchery steelhead trout (*Oncorhynchus mykiss*) where not all individuals can be sampled effectively. They develop a new method of grandparent analysis where parental genotypes can be reconstructed using data from candidate grandparent crosses and F2 offspring genotypes, allowing for new explorations of hybridization, migration and gene flow in wild populations.

Keywords: conservation genetics, gene flow, grandparentage analysis, microsatellites, parentage analysis, salmonidae

Received 23 November 2010; revision received 6 December 2010; accepted 10 December 2010

Threatened and endangered populations of salmonid fishes in the Pacific northwest of North America are the subject of intensive conservation efforts and stocking programmes aimed at restoring these historic populations. Often, hatchery-raised fish have lower fitness than wild-born fish raising serious concerns about how hybridization may affect the long-term fitness of these threatened populations.

To assess to what extent stocking programmes impact the fitness of wild populations, wildlife biologists have invoked a variety of molecular-based techniques in combination with parentage analysis to measure how much gene flow from hatchery-born fish ends up in wild gene pools. However, populations that display a diversity of life history strategies pose some methodological issues as differences in migration timing or whether fishes migrate at all can obfuscate the true sources of gene flow in these populations.

A threatened population of steelhead trout (*O. mykiss*) residing in the Hood River, Oregon, offers a perfect platform to elucidate multiple sources of gene flow in a population with multiple life history strategies. This population consists of wild anadromous (i.e., migrating) steelhead supplemented by an extensive stocking programme of juvenile hatchery fish. The natural anadromous life cycle of wild steelhead has been impacted for nearly a century by the construction of the Powerdale dam (Fig. 1). For nearly two decades prior to its recent decommissioning and removal in the summer of 2010, all anadromous fish returning to their spawning grounds were passed over the dam, at which point they were genetically sampled. The net result of this sampling was the creation of a rich genotypic data set of hundreds of migrating steelhead each year.

Despite the effective sampling of all migrating steelhead, the assessment of hybrid fitness and introgression in this population is hampered because a fraction of wild and hatchery steelhead trout do not complete their anadromous life cycle. Instead, these fish spend their entire life upstream of the Powerdale dam as year-round residents rather than going out to the sea (Fig. 2). Because these resident fish are in low densities and complete sampling of individuals is not an option, these fish represent a previously unmeasured avenue of gene flow between wild and hatchery fish.

So how can we measure the genetic contribution of residualized hatchery fish to the next generation of steelhead with accuracy? To piece together the genetic contributions of residualized vs. anadromous steelhead, Christie *et al.* (2011) first use parentage analysis to match offspring with one or both anadromous parent(s). If an offspring can only be assigned to one anadromous parent, then Christie *et al.* (2011) employ a previously undescribed method of grandparentage analysis to test whether that offspring matches a known hatchery stock grandparental cross. Taken together, the combined approach of parentage and grandparentage analyses shows that while almost half of the genes in steelhead originate from resident parents, only a small fraction of the unsampled parents appear to be residualized hatchery fish. Moreover, their analyses reveal asymmetric gene flow from hatchery stock into wild populations, demonstrating that hatchery males produce many more offspring with wild females rather than hatchery females. Although their analyses cannot say with certainty how many resident matings originate from hatchery stock, this represents the most complete account of gene flow within this population to date.

Christie *et al.* (2011) new method of grandparentage analysis is a logical extension of well-established exclusion principles in parentage analysis. The power of exclusionary

Correspondence: Kenyon B. Mobley, Fax: +46 (0) 907866705; E-mail: kenyon.mobley@emg.umu.se

Fig. 1 The Powerdale Dam, Hood River Oregon. Long-term, multigenerational genotype data sets such as those initiated at this site are a potential goldmine for pioneering new methods of parentage analysis. Photo courtesy of Rod French.

Fig. 2 A 'residualized' hatchery steelhead. The genetic contributions of these resident fish are difficult to assess because they spend their entire life upstream of the Powerdale Dam and do not migrate to sea. Photo courtesy of John McMillan.

methods lies in the ability to capitalize on genotypic incompatibilities between parents and offspring (in this case grandparents and grandoffspring) to reject particular parent–offspring matches (Jones & Ardren 2003). In this case of steelhead trout, knowledge of hatchery breeding stock is sufficient to exclude putative grandparent, grandoffspring pairs.

The benefit of grandparentage analysis is that neither parental genotypes need to be known. However, the power to detect true grandparent–grandoffspring matches increases exponentially with the addition of one putative parent in addition to known grandparental crosses. Therefore, the application of grandparent analysis is particularly beneficial to multigenerational data sets where at least knowledge of grandparental breeding stock exists.

Grandparentage analysis is not without its problems, as any grandoffspring that fail to share at least one allele at any particular locus with a putative grandparent cross would be rejected under strict exclusion criteria. Thus, grandparentage analysis suffers from the same Achilles heel of all parentage analyses based on exclusionary methods—namely rejection of true parent–offspring pairs based on germ-line mutations, nonamplifying (null) alleles and genotyping errors (Jones & Ardren 2003). Compounding this issue is the fact that grandparentage analysis requires at least double the amount of microsatellite loci or alleles to have comparable levels of assignment success to traditional parentage analysis (Letcher & King 2001).

To circumvent these issues, exclusion assumptions can be relaxed which may also increase the chance of falsely assigning parent–offspring pairs. Christie *et al.* (2011) route this criticism by testing the exclusionary power of their methods under both stringent and relaxed criteria. They demonstrate that hatchery X hatchery crosses are under-represented under both criteria lending support to their hypothesis that these crosses are not as successful as crosses between hatchery and wild trout. Christie *et al.* (2011) also extend recently developed methods (Christie 2010) to determine the probability of false grandparent–grandoffspring trios and, using simulated data sets, demonstrate that their methods are remarkably precise.

With the proliferation of large, multigenerational data sets, it is simply a matter of time before grandparentage analysis will be routinely conducted to help piece together multigenerational pedigrees. The application of grandparentage analysis to breeding programmes and hybrid zone dynamics is obvious, and the implementation of grandparentage analysis to various other fields such as quantitative trait loci mapping, inbreeding and heritability estimation seems to be an appropriate next step. Given the high

interest in standardized methods of parentage analysis (see Jones *et al*. 2010 for a recent review), the multitude of existing parentage computer programmes should incorporate multigenerational pedigree construction utilizing combined parentage and grandparentage analysis. Although we have just scratched the surface of what grandparentage analysis can accomplish, it will undoubtedly be a useful tool in the construction of wild pedigrees for generations to come.

References

Christie MR (2010) Parentage in natural populations: novel methods to detect partent-offspring pairs in large data sets. *Molecular Ecology Resources*, **10**, 115–128.

Christie MR, Marine ML, Blouin MS (2011) Who are the missing parents? Grandparentage analysis identifies multiple sources of gene flow into a wild population *Molecular Ecology*, **20**, 1263–1276.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, **12**, 2511–2523.

Jones AG, Small CM, Paczolt KA, Ratterman N (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, **10**, 6–30.

Letcher BH, King TL (2001) Parentage and grandparentage assignment with known and unknown matings: application to Connecticut River Atlantic salmon restoration. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 1812–1821.

Pemberton JM (2008) Wild pedigrees: the way forward. *Proceedings of the Royal Society B*, **275**, 613–621.

K.B.M. is a postdoctoral fellow at Umeå University, and his interests include various aspects of behaviour and the evolutionary ecology of fishes. Molecular techniques and parentage analysis form the backbone of his research programme.

doi: 10.1111/j.1365-294X.2010.04995.x

Supporting Information

Parentage methods

Because we were interested in examining mating between anadromous and resident fish, our parentage goals were twofold: (1) to definitively assign steelhead from the same broodyear to their one anadromous parent and (2) to be certain that their second parent really was missing. The primary challenge facing the unambiguous assignment of steelhead to one of their parents (step 1) lies in the incorrect classification of a false parent-offspring pair as a true parent-offspring pair (type I error). For our study, false assignments to a first parent could bias our subsequent findings because we would be unlikely to identify a second parent (step 2). Therefore, we conservatively eliminated all individuals with missing data and did not allow for a single locus to mismatch when assigning steelhead to the first parent. False assignments can also be caused by insufficient exclusionary power such that a certain number of pair-wise comparisons share alleles by chance alone. Therefore, we calculated the expected number of false parent-offspring pairs, *Fpairs* (Christie 2010), which incorporates allele frequencies from the samples of putative parents and offspring and fully accounts for the total number of pair-wise comparisons. The quotient of the expected number of false pairs and the total number of putative pairs (both true and false pairs) yields the probability of a randomly selected putative parent-offspring pair being false, $\Pr(\phi)$. This approach revealed that we had a suite of polymorphic makers sufficient for minimizing false assignments. Importantly, no offspring had more than 1 candidate parent of the same gender. Furthermore, because there can be error associated with aging steelhead by scales (Seamons *et al.* 2009), we ensured that all individuals assigned to 1 parent could not be assigned to any putative parents from adjacent run years.

Out of the steelhead that were definitively assigned to a single parent, we next identified the subset that were definitively missing their other parent (note that having one anadromous parent insured that we were examining the correct run year for missing parents). Here our goals for parentage assignment differed from above because we wanted to ensure that the parent was indeed missing and not simply mismatching due to genotyping errors or missing data (i.e., we now employed a liberal approach to matching). We eliminated all offspring that matched a putative 2nd parent at 7 or 8 out of the 8 loci. We next considered any comparison at a locus between an offspring and a putative parent with missing data to be a match. We then removed any offspring that "matched" a putative parent with missing data and truly matched at other loci. We further removed any putative parent-offspring pairs that mismatched at 1 locus and had missing data. Some putative parents contained missing data at 3 or more loci, which resulted in a large number of "matches" and drastically decreased our number of candidate pairs. Therefore we reprocessed (DNA extraction through scoring) these individuals (n = 28) at 4-8 microsatellite loci, which resolved this issue.

Given our study specific error rates there is a remarkably small chance that the second parent was present but mismatching at two loci owing to scoring errors (see "error rates" section below). Error rates were estimated empirically by comparing scored alleles across all loci for independent runs of the same individuals (2,224 alleles distributed evenly across all loci; see methods in Bonin *et al.* 2004). Finally, we verified that the missing parent was not actually present but erroneously categorized as the wrong sex. We re-performed the parentage analysis as above, except we examined candidate parents that were the same sex as the first assigned parent. We found no evidence of incorrectly sexed parents. Because all of the ODFW samples can be found within our database and because each fish is individually handled as it is passed over the dam, it is very unlikely that any of the missing parents are due to unsampled fish.

Lastly, because all fish are sampled at the dam, all iteroparous fish were included in our sample of candidate parents.

We next performed an analysis of allele frequencies using the alleles inherited from missing (resident) parents. We could identify the alleles from the resident fish because we knew the anadromous parent, whose alleles could be excluded from the offspring, and the resident parent, who passed an allele down to the identified offspring (e.g., allele 1 in Fig. 1). We compared these resident allele frequencies to the allele frequencies of all summer and winter run fish collected from the same broodyears using FSTAT v. 2.9.3 (Goudet 2001). If the resident fish truly were summer-run fish or hatchery stocked trout, we would expect the resident allele frequencies to differ from winter-run fish. Calculating F_{ST}, we find that these allele frequencies are identical to winter run fish (99% F_{ST} confidence interval= -0.001 to 0.001) and are substantially different from summer-run fish ($F_{ST} = 0.014$; $p < 0.01$). Excluding alleles from fish that were assigned back to hatchery broodstock did not change these results. This analysis suggests that reproductively successful resident fish in this system are derived from winter-run stock. Furthermore, historically stocked rainbow trout would have allele frequencies substantially different than the winter and summer-run fish (out-of-basin Cape Cod stock) and if these fish were passing on their genes, we would expect the resident genes to be substantially different from the winter-run fish. Also, the Cape Cod stock used in the Hood River were highly domesticated (i.e., exhibit poor natural survival) and have never been recovered in screw traps, despite intensive sampling efforts (Rod French, district manager ODFW, *personal communication*).

Grandparentage methods

We first created simulated data sets with allele frequencies from the 8 loci at which all steelhead were genotyped, and which averaged 36 alleles per locus. We created simulated data sets with (1) an unknown breeding matrix and no sampled parents; (2) with a known breeding matrix, but with no sampled parents; and (3) with a known breeding matrix and 1 identified parent. Once the locus-specific allele frequencies were determined, 100,000 genotypes were created in accordance with HWE. This pool of genotypes was randomly sampled and placed into a group of putative grandfathers, putative grandmothers, or putative grandoffspring. The above simulation process was repeated at each locus. Grandparent pairs were created by randomly pairing grandfathers with grandmothers. For simulations with a known parent, a fourth sample of simulated individuals (parents) was created as described above. True parent-offspring pairs were created by randomly sampling 1 individual from both the sample of parents and the sample of putative grandoffspring. The two individuals were aligned, locus by locus, and at each locus, a randomly chosen allele was copied from the adult to the offspring. We created 1000 simulated data sets for each combination of variables (e.g., sample size, known breeding matrix etc.). It is important to note that these simulations created loci that were independent from one another and individuals that were unrelated. In practice, the presence of related individuals and linked loci may decrease exclusionary power (Anderson & Garza 2006; Ford & Williamson 2010), and these equations do not take this information into account. For every data set, we measured the expected number of false grandparent-grandoffspring trios, the actual number of false grandparent-grandoffspring trios, the bias, mean squared error, and variance (Table S1). From the simulations, we noticed that these methods appear to have a very small negative bias, meaning that they slightly underestimate the number of false trios. Both the bias and mean

93 square error decrease with increasing sample sizes, numbers of loci, and number of alleles per
94 locus.

Table S1 Results from simulated data sets using allele frequencies for the 8 microsatellite loci used in this study. For each sample size (N), we also measured the expected number of false grandparent-grandoffspring trios (*FGtrios*), the actual number of false trios (Actual), the Bias, mean squared error (MSE), and variance (Var). Also, see Fig. 3 in main text.

Known breeding matrix and 1 identified parent					
N	*FGtrios*	Actual	Bias	MSE	Var
50	0.0013	0.0020	-1.11×10^{-06}	2.06×10^{-12}	8.24×10^{-13}
100	0.0056	0.0070	-5.75×10^{-07}	7.15×10^{-13}	3.84×10^{-13}
150	0.0130	0.0150	-3.53×10^{-07}	3.95×10^{-13}	2.70×10^{-13}
200	0.0234	0.0150	8.39×10^{-07}	8.95×10^{-13}	1.91×10^{-13}
300	0.0535	0.0480	2.43×10^{-07}	1.83×10^{-13}	1.24×10^{-13}
400	0.0952	0.1060	-2.71×10^{-07}	1.73×10^{-13}	9.95×10^{-14}
Known breeding matrix, no identified parents					
N	*FGtrios*	Actual	Bias	MSE	Var
50	0.102	0.136	-5.37×10^{-05}	4.79×10^{-09}	1.91×10^{-09}
100	0.444	0.498	-2.17×10^{-05}	1.53×10^{-09}	1.06×10^{-09}
150	1.022	1.091	-1.23×10^{-05}	7.72×10^{-10}	6.21×10^{-10}
200	1.824	1.802	2.17×10^{-06}	5.01×10^{-10}	4.96×10^{-10}
300	4.155	4.369	-9.53×10^{-06}	4.24x10-^{10}	3.33×10^{-10}
400	7.417	7.518	-2.53×10^{-06}	2.50x10-^{10}	2.44×10^{-10}
Unknown breeding matrix, no identified parents					
N	*FGtrios*	Actual	Bias	MSE	Var
50	1.22	1.36	-2.25×10^{-04}	5.26×10^{-08}	2.18×10^{-09}
100	10.80	11.23	-1.68×10^{-04}	2.93×10^{-08}	1.04×10^{-09}
150	37.65	39.95	-4.10×10^{-04}	1.69×10^{-07}	6.67×10^{-10}
200	90.01	92.78	-2.77×10^{-04}	7.72×10^{-08}	5.05×10^{-10}
300	308.50	311.78	-1.46×10^{-04}	2.16×10^{-08}	3.20×10^{-10}
400	737.98	752.80	-3.70×10^{-04}	1.37×10^{-07}	2.51×10^{-10}

Error rates

Here we determine the proportion of pair-wise comparisons between putative parents and offspring that have at least 1 error for the specific situation where 1 parent has already been identified. The identified parent is useful because it allows, in most cases, for the exclusion of an allele in a putative offspring. After excluding the known-parent allele, the sample of putative offspring is, for all intents and purposes, essentially haploid, while the sample of putative parents is diploid. First, one should perform a second analysis on a subset of genotyped samples across all loci (DNA extraction through scoring). The study-specific error rate, ε, is then defined as the quotient of the number of alleles that differ after the second analysis to the total number of alleles compared (Bonin *et al.* 2004). Given diploid samples (i.e., putative parents), one can first determine the probability of observing *i* errors in a multi-locus genotype. Here, we use a modification of Bonin *et al.'s* (2004) formula:

$$P_D = \binom{2L}{i} (\varepsilon)^i (1-\varepsilon)^{2L} \qquad \text{(S1)}$$

where L is equal to the total number of loci employed in the study. Notice that when *i* equals 0, 1-P_D equals the proportion of multilocus genotypes with at least 1 error. As Bonin *et al.* (2004) point out, this equation can also be modified for haploid markers (or putative offspring with an excluded allele), such that the probability of observing *i* errors in a multi-locus, haploid genotype equals:

115 $$P_H = \binom{L}{i} (\varepsilon)^i (1-\varepsilon)^L \quad \text{(S2)}$$

116 Therefore, the probability of observing at *i* errors in a multilocus genotype of all individuals
117 (diploid and haploid; for this study, parents and offspring) equals:
118
119 $$P_i = \frac{1}{2} \cdot (P_D + P_H) \quad \text{(S3)}$$
120
121 Notice that this equation can also account for different error rates within samples of putative
122 parents and putative offspring. The remaining calculations follow those presented in Christie
123 (2010), where we next determine the probability of observing *i* errors in a multi-locus pair-wise
124 comparison. This probability is equal to:
125
126 $$P' = 2P - P^2 \quad \text{(S4)}$$
127
128 where P' equals the proportion of pair-wise comparisons that will have *i* errors in a multi-locus
129 pair-wise comparison. Notice that if P is set to $1 - P_0$ then P' equals the proportion of pair-wise
130 comparisons with at least one error. Next, one can iteratively determine the proportion of pair-
131 wise comparisons that would have at least one error given a number of mismatching loci:
132
133 $$P_i' = 2\left(1 - P_0 - \sum_{i=1}^{M} P_i\right) - \left(1 - P_0 - \sum_{i=1}^{M} P_i\right)^2 \quad \text{(S5)}$$
134

where M equals the number of loci allowed to mismatch and must be an integer greater than 0. Thus P_i' equals the number of dyads with at least one error given that M loci are allowed to mismatch. Not all errors will cause a mismatch because the majority of pair-wise comparisons will not be parent-offspring pairs and additionally the majority of positions where an error occurs will not cause a mismatch. Thus choosing a cutoff value for P_i' is somewhat subjective and should be reported along with the number of loci allowed to mismatch. As a general rule of thumb, a P_i' between 0.05 and 0.1 is likely to result in few putative pairs with a mismatch-causing error (Christie 2010). The advantage to this method becomes quickly apparent when one notices how quickly the error rate drops by allowing a single locus to mismatch. Therefore, while a P_i' near 0.05 may be unnecessarily conservative, in most cases it is approached by simply allowing one to two loci to mismatch (Christie 2010). Here we assume that genotyping errors are distributed equally among all pair-wise comparisons and that true parent-offspring pairs are not prone to greater rates of genotyping errors.

Additional loci

To validate our grandparentage equations and verify our assignments, we genotyped all putative grandparent grandoffspring trios at 5 additional loci (*Ogo*4, *Omm*1046, *Omy*7, *One*102, and *Ots*4). Before amplification, samples were boiled for 20 minutes in a 5% Chelex100 solution. PCR reactions were performed using Qiagen Multiplex PCR kits and contained each forward and reverse primer at 0.4 μM *Ogo*4, 0.08 μM *Omm*1046, 0.72 μM *Omy*7, 1.2 μM *One*102, and 0.24 μM *Ots*4, with 1 μL of approximately 25 ng/μL template in a total reaction volume of 10 μL. The thermocycling profile consisted of an initial denature at 95°C for 15 minutes followed by 35 cycles of 30 seconds at 94°C, 90 seconds at 57°C, and 60 seconds at

158 72°C with a final extension of 30 minutes at 60°C. PCR products were screened on an ABI 3100
159 automated sequencer (Applied Biosystems). Allele sizes were determined with the fragment
160 analysis software GENOTYPER 3.7. After scoring, bins for alleles were created with
161 FLEXIBIN v. 2 (Amos *et al.* 2007). All data was scored by two independent observers.
162 Departures from Hardy-Weinberg equilibrium (HWE) were tested in GENEPOP v. 3.4 with
163 10,000 batches and 5,000 iterations per batch (Raymond & Rousset 1995). Evidence for
164 departure from linkage equilibrium was tested in GENETIX v. 4.05 with 5000 iterations across
165 all locus pairs (Belkhir *et al.* 2002). Allelic richness and the mean inbreeding coefficient (F_{IS})
166 were calculated with FSTAT v. 2.9.3 (Goudet 2001). All 5 additional loci were not significantly
167 out of Hardy-Weinberg Equilibrium (Table S3), though all loci did have a slight deficiency of
168 homozygotes. There was no evidence that any of the loci were in linkage disequilibrium. Allelic
169 richness averaged 15.71 across all loci and was correlated with the exclusionary power of the
170 locus as measured by $\Pr(G)$ (eqn. 4; Table S2).

Table S2 Characteristics of the 5 additional loci used to verify putative grandparent-grandoffspring trios. Column headings are as follows: allelic richness (A_r); mean inbreeding coefficient (F_{IS}); observed heterozygosity (H_o); expected heterozygosity (H_e); probability of a locus being out of Hardy-Weinberg equilibrium (HWE); exclusionary power (Pr(G)); and original locus description (Reference).

Locus	A_r	F_{IS}	H_o	H_e	HWE	Pr(G)	Reference
*Ogo*4	12.75	0.025	0.753	0.773	0.079	0.607	Olsen et al. (1998)
*Omm*1046	20.88	0.005	0.929	0.934	0.264	0.232	Rexroad et al. (2002)
*Omy*7	16.00	0.099	0.745	0.821	0.188	0.470	K. Gharbi, unpublished data, University of Guelph, Ontario, CA
*One*102	20.91	0.012	0.892	0.902	0.764	0.322	Olsen et al. (2000)
*Ots*4	8.00	0.016	0.734	0.746	0.883	0.662	Banks et al. (1999)

Table S3 ***Gene flow into steelhead population***

The proportion of sampled steelhead with two parents, one parent, or neither parent identified comes from Araki *et al.* (2007a). We first divided the proportion of parents indentified into resident and anadromous components (e.g., 35.2 is doubled for offspring with both parents because both parents were anadromous, 10.8 is split between resident and anadromous because these offspring had one anadromous father and one resident mother). Notice that the total resident and anadromous contributions are divided by 200, the total for both columns. We next used these values to calculate the proportion of gene flow from hatchery and wild sources. To calculate the anadromous hatchery contribution we used an RRS of 0.848 (Araki *et al.* 2007b), which is conservatively estimated from hatchery fish with two wild parents, and we also adjusted for the different numbers of hatchery and wild steelhead passed over the dam. For the anadromous wild contribution we used an RRS of 1 and again adjusted for the different numbers hatchery and wild fish passed over the dam. For hatchery resident genes, we multiplied the percentage of resident genes from anadromous x resident matings by the conservatively estimated proportion of offspring with a hatchery residualized parent (6%). For genes contributed via resident x resident matings, we could not estimate the proportion of hatchery gene flow with our methods (see Discussion).

Parents identified	Proportion sampled	Resident genes	Anadromous genes
Both Parents	35.2	0	70.4
Father Only	10.8	10.8	10.8
Mother Only	30.9	30.9	30.9
Neither Parents	23.1	46.2	0
Total (%)	100	43.95	56.05

Hatchery Anadromous Genes = 56.05% Anadromous Genes x 34.2 % Hatchery Contribution = 19.2%

Wild Anadromous Genes = 56.05% Anadromous Genes x 65.8% Wild Contribution = 36.8%

Hatchery Resident Genes = ((30.9% Mother Only+10.8% Father Only)/200)*6% (Proportion Offspring with Hatchery Residualized Parent) = 1.3

Wild Resident Genes = ((30.9% Mother Only+10.8% Father Only)/200)*94% (Proportion Offspring with a Wild Resident Parent) = 19.6

Resident x Resident Genes = 46.2/200 = 23.1

References:

Araki H, Ardren WR, Olsen E, Cooper B, Blouin MS (2007a) Reproductive Success of Captive-Bred Steelhead Trout in the Wild: Evaluation of Three Hatchery Programs in the Hood River. *Conservation Biology* **21**, 181-190.

Araki H, Cooper B, Blouin MS (2007b) Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. *Science* **318**, 100-103.

Amos W, Hoffman JI, Frodsham A*, et al.* (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes* **7**, 10-14.

Banks M, Blouin MS, Baldwin B*, et al.* (1999) Isolation and inheritance of novel microsatellites in chinook salmon (*Oncorhynchus tschawytscha*). *Journal of Heredity* **90**, 281-288.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2002) Genetix 4.03, Logiciel Sous Window TM Pour la Genetique Des Populations. *Laboratoire Genome, Populations, Interactions, CNR SUMR 5000, Universite de Montpellier II, Montpellier, France*.

Bonin A, Bellemain E, Bronken P*, et al.* (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**, 3261-3273.

Christie MR (2010) Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. *Molecular Ecology Resources* **10**, 115-128.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). *Available from* http://www.unil.ch/izea/softwares/fstat.html.

Olsen J, Bentzen P, Seeb J (1998) Characterization of seven microsatellite loci derived from pink salmon. *Molecular Ecology* **7**, 1087-1089.

Olsen, JL, Wilson SL, Kretschmer EJ, Jones KC, Seeb JE (2000) Characterization of 14 tetranucleotide microsatellite loci derived from sockeye salmon. *Molecular Ecology* **9**:2185–2187.

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

Rexroad C, Coleman R, Hershberger W, Killefer J (2002) Eighteen polymorphic microsatellite markers for rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics* **33**, 76-78.

Seamons TR, Dauer M, Sneva J, Quinn TP (2009) Use of parentage assignment and DNA genotyping to validate scale analysis for estimating steelhead age and spawning history. *North American Journal of Fisheries Management* **29**, 396-403.