MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2012) **12**, 653–660 doi: 10.1111/j.1755-0998.2012.03157.x

RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays

STEPHEN J. AMISH,* PAUL A. HOHENLOHE,† SALLY PAINTER,* ROBB F. LEARY,‡ CLINT MUHLFELD,§¶ FRED W. ALLENDORF* and GORDON LUIKART*¶**
*Fish and Wildlife Genomics Group, Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA, †Department of Biological Sciences, University of Idaho, Moscow, ID 83844-3051, USA, ‡Montana Fish, Wildlife & Parks, University of Montana, Missoula, MT 59812, USA, §U.S. Geological Survey, Northern Rocky Mountain Science Center, Glacier National Park, West Glacier, MT 59936, USA, ¶Flathead Lake Biological Station, University of Montana, Polson, MT 59860, USA, **CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, 4485-661 Vairão, Portugal

Abstract

Hybridization with introduced rainbow trout threatens most native westslope cutthroat trout populations. Understanding the genetic effects of hybridization and introgression requires a large set of high-throughput, diagnostic genetic markers to inform conservation and management. Recently, we identified several thousand candidate single-nucleotide polymorphism (SNP) markers based on RAD sequencing of 11 westslope cutthroat trout and 13 rainbow trout individuals. Here, we used flanking sequence for 56 of these candidate SNP markers to design high-throughput genotyping assays. We validated the assays on a total of 92 individuals from 22 populations and seven hatchery strains. Forty-six assays (82%) amplified consistently and allowed easy identification of westslope cutthroat and rainbow trout alleles as well as heterozygote controls. The 46 SNPs will provide high power for early detection of population admixture and improved identification of hybrid and nonhybridized individuals. This technique shows promise as a very low-cost, reliable and relatively rapid method for developing and testing SNP markers for nonmodel organisms with limited genomic resources.

Keywords: conservation genomics, hybridization, introgression, invasive species, microfluidic PCR, salmonids, SNP, trout species identification

Received 22 November 2011; revision received 14 February 2012; accepted 17 February 2012

Introduction

Rainbow trout (RBT; *Oncorhynchus mykiss*), the most widely introduced salmonid in the world (Lever 1996), produce fertile offspring when crossed with cutthroat trout (*O. clarkii*), and introgression often continues until a hybrid swarm is formed and the native cutthroat genomes are lost (Allendorf & Leary 1988). A major consequence of such interspecific hybridization may be outbreeding depression because of the break-up of co-adapted gene complexes and disruption of local adaptations (Allendorf *et al.* 2004; Muhlfeld *et al.* 2009). Introgression poses a serious threat to all subspecies of cutthroat trout in western North America owing to widespread stocking of rainbow trout and invasion by rainbow trout and hybrids into historical cutthroat trout habitats.

Currently, range-wide estimates of hybridization in many of the 12 cutthroat trout subspecies and populations are incomplete. Westslope cutthroat trout (WCT; *Oncorhynchus clarkii lewisi*), the most widely distributed subspecies of cutthroat trout, historically occupied aquatic habitats throughout the Columbia, Fraser, Missouri and Hudson Bay drainages of the United States and Canada (Behnke 2002). However, nonhybridized populations are estimated to persist in <10% of their historical range (Shepard *et al.* 2005). While over half of the population genetic samples in Shepard *et al.* (2005) found no evidence of admixture, only 30% had enough individuals sampled to detect 1% admixture at the 95% level of confidence. As a result, only 15% of the population genetic samples showed no evidence of admixture (<1%) with a high degree of confidence.

Markers detecting low amounts of admixture in populations and individuals will provide an understanding of the mechanisms causing the spread of hybridization, help protect nonhybridized populations from invasion, and aid in identifying nonhybridized populations suitable as sources for hatchery brood stocks or other conservation actions (Allendorf *et al.* 2001). Besides estimates of

Correspondence: Steve Amish, Fax: 406-243-4384;
E-mail: stephen.amish@umontana.edu

individual or population levels of admixture, the distribution and the frequency of introgressed genotypes within a population or sample can illuminate the duration and extent of hybridization (Jiggins & Mallet 2000). For example, a bimodal distribution is thought to result from selection against intermediate genotypes or assortative mating between the species (Jiggins & Mallet 2000; Weigel *et al.* 2003). Currently, the number of available species-diagnostic loci for addressing these questions in native cutthroat trout and rainbow trout is limited.

The development of additional species-diagnostic genotyping assays and high-throughput SNP genotyping systems will provide increased power for detecting hybridization at the individual level and more precisely estimating the structure of hybrid zones. Currently, 10–15 diagnostic microsatellite markers are often used to detect cutthroat and rainbow trout hybridization at the population level. If we assume that each marker sorts independently, there is no linkage disequilibrium affecting the markers, and if genotypes are distributed according to Hardy–Weinberg ratios (i.e. the samples are representative of a breeding aggregation or population), then we can calculate the likelihood of failing to detect a single RBT allele in any given fish or in a sample of unrelated fish using binomial probability (Rasmussen *et al.* 2010). The probability of detecting 1% hybridization in a population with 95% certainty and 20 individual samples requires only eight independent diagnostic markers. In contrast, to detect 1% admixture in an individual with 95% certainty will require 150 independent diagnostic markers (Table 1). Similarly, an increase in the number of diagnostic markers will also improve our ability to differentiate between parental back-crosses and later generation hybrid crosses.

Recently, we identified a large set of candidate diagnostic SNPs using restriction-site-associated DNA (RAD) sequencing (Hohenlohe *et al.* 2011). Briefly, we sequenced a single RAD library (Baird *et al.* 2008; Etter *et al.* 2011a) created from 24 fish [11 WCT, 12 coastal rainbow trout (*O. m. mykiss*, CRT), and 1 inland or redband rainbow trout (*O. m. gairdneri*, IRT)] and applied strict filtering based on observed heterozygosity and deviations from Hardy–Weinberg equilibrium to remove homeologous loci (paralogs resulting from the ancestral salmonid whole-genome duplication; Lie *et al.* 1994). That analysis produced a total of 2923 RAD markers at which there was a single candidate SNP fixed between the two species, and no other polymorphism, within the informative 48-bp RAD tag sequence (Hohenlohe *et al.* 2011). Here, we expand the list of candidate SNP markers, and we use microfluidic PCR assays to verify a subset of them for high-throughput estimates of hybridization in trout. We chose to develop a bioinformatics pipeline for this purpose because of its cost-effectiveness and because the rainbow trout genome would subsequently be available for the development of additional markers.

Table 1 The likelihood of detecting a single RBT allele in any given fish or in a population of unrelated fish using binomial probability for a number of independent species-diagnostic loci, samples and percent hybridization

Number of markers	Number of samples	%Hybridization	Probability of detection
8	1	17	95.0
8	20	1	95.0
46	1	3.1	95.0
46	4	1	97.5
77	1	1.9	95.0
96	1	1.6	95.0
150	1	1	95.0

Materials and methods

In addition to the 2923 single-SNP candidate markers from the study by Hohenlohe *et al.* (2011), we identified candidate diagnostic SNPs in RAD tags containing two putative fixed SNPs in the 48-bp sequence (an additional 643 markers) and those containing one putative fixed SNP and one additional site polymorphic within one of the taxa (an additional 1348 markers). We aligned the total set of 4914 RAD tag sequences against a published database of rainbow trout sequence contigs (Sanchez *et al.* 2009) using the program Bowtie (Langmead *et al.* 2009). We allowed up to three mismatches between the WCT or the CRT and the reference sequence. The data set from the study by Sanchez *et al.* (2009) contained 47 526 contigs ranging in size from 185 to 1978 bp, produced by 454 sequencing of a reduced representation genomic library in rainbow trout. A total of 66 (1.3%) of our candidate RAD tags aligned against one or more of the contigs from the study by Sanchez *et al.* (2009) with at least 50bp of flanking sequence on either side of the diagnostic SNP. Ten of these loci were dropped after preliminary data suggested one of the primers or probes was not amplifying. Sequences for the remaining 56 candidate markers were submitted to KBioscience for the design of KASPar SNP genotyping assays.

A total of 92 individuals from 22 populations and seven hatchery strains plus two heterozygous positive controls (F1s) were then used to validate the 56 assays on Fluidigm 96.96 microfluidic PCR chips. The individuals included two cutthroat trout species, WCT and YCT (Yellowstone cutthroat trout, *O. c. bouvieri*), as well as IRT and CRT (Table 2). All samples came from putatively nonhybridized populations, based on a current panel of seven diagnostic microsatellite loci and seven indel loci,

Table 2 Location name, species, whether the fish is of wild or hatchery origin, and the number of samples, basin and subbasin information for the screening panel used to validate species-diagnostic assays. Species are labelled as westslope cutthroat trout (WCT), Yellowstone cutthroat trout (YCT), inland or redband rainbow trout (IRT), and coastal rainbow trout (CRT)

Location name	Species	Wild/ Hatchery	N	Basin	Subbasin
Anaconda, MT	WCT	H	12	NA*	NA
Big Foot Creek	WCT	W	2	Columbia	Upper Kootenai
Copper Creek	WCT	W	2	Columbia	Flint-Rock
Cottonwood Creek	WCT	W	3	Columbia	Lower Flathead
Davis Creek	WCT	W	4	Columbia	Bitterroot
Flat Creek	WCT	W	3	Columbia	Upper Kootenai
Gillispie Creek	WCT	W	3	Columbia	Flint-Rock
Hawk Creek	WCT	W	2	Columbia	N. F. Flathead
Humbug Creek	WCT	W	2	Columbia	Blackfoot
McGinnis Creek	WCT	W	3	Columbia	Lower Clark Fork
Morrison Creek	WCT	W	3	Columbia	Middle Flathead
Ringeye Creek	WCT	W	2	Columbia	Blackfoot
Runt Creek	WCT	W	3	Columbia	Yaak
S. Fork Jocko	WCT	W	3	Columbia	Lower Flathead
Six Mile Creek	WCT	W	3	Columbia	Middle Clark Fork
Werner Creek	WCT	W	3	Columbia	N. F. Flathead
Bear Creek	WCT	W	1	Missouri	Red Rock
McClellan Creek	WCT	W	1	Missouri	Upper Missouri
McVey Creek	WCT	W	1	Missouri	Big Hole
Big Timber, MT	YCT	H	6	NA	NA
Slough Creek	YCT	W	4	Missouri	Yellowstone
Lake Koocanusa, BC	IRT	W	4	Columbia	Yaak
Yahk River, BC	IRT	W	5	Columbia	Yaak
Abbot Creek	CRT	W	2	Columbia	Middle Flathead
Arlee, MT	CRT	H	7	NA	NA
Eagle Lake, CA	CRT	H	2	NA	NA
McConaughy, NE	CRT	H	2	NA	NA
Fish Lake, UT	CRT	H	2	NA	NA
Erwin/Arlee Cross, TN	CRT	H	2	NA	NA

*Basin and Subbasin designations were not made for hatchery stocks.

except for fish from Lake Koocanusa, which appear to have both a CRT and IRT genetic component (R. Leary, personal communication), and the South Fork of the Jocko River, which appear to have a CRT component. WCT samples included two year classes from the Washoe Park State Trout Hatchery, Anaconda Montana, and samples from 18 wild populations, including three populations from the Missouri River basin east of the Continental Divide. YCT samples were from the Yellowstone River State Trout Hatchery, Big Timber, Montana, and a population in Slough Creek. IRT samples were from two populations in the Kootenai River drainage in Montana. CRT were taken from hatchery stock from across the country to account for the multitude of potential sources used currently and historically for stocking.

Results

Forty-six of fifty-six assays (82%) were diagnostic for the identification WCT, RBT and hybrids (Table 3). An assay was considered diagnostic if both heterozygous positive controls showed separation from the homozygous genotype clusters and >95% of samples had concordant genotypes (i.e. genotype agreed with expectation established by earlier microsatellite/indel data). Eight of ten assay failures were attributed to the design process, including poor quality or inadequate sequence data being used to design the assay (e.g. using sequence from a paralogous region or sequence containing errors) or errors in the primer or probe design and manufacture process. These failures included three assays where one or both probes did not amplify, and one assay appears to have amplified a homeolog. In fact, this locus had an elevated depth of sequence reads in the original RAD sequencing run, in the 95th percentile among the 2923 fixed single-SNP candidate markers (Hohenlohe *et al.* 2011), consistent with the interpretation that it represents two incorrectly assembled homeologs.

Given a total of 46 assays for detecting RBT and WCT hybridization and four samples from a population, we have a 97.5% probability of detecting 1% introgression in a hybrid swarm. In an individual fish, we have a 95% probability of detecting >3.1% introgression with the same number of assays (Table 1). However, very low

Table 3 Sequence information for the validated assays. The number of additional variable sites in the RAD tag and the reference sequence are listed, along with the call rate (% of samples assigned to a genotype cluster) and the call concordance (% of samples assigned to the correct cluster). Sequences given are for the Kbiosciences KASP assay format. RBT sequence name is the rainbow trout reference sequence (Sanchez *et al.* 2009) from the RAD loci alignments used for flanking sequence in the SNP assay design

	Additional Variable Sites						Sequence (5′-3′)			RAD GenBank Accession No.		
Assay name	RAD Tag	Ref Seq	SNP	N	% Call rate	% Concordance	Primer FAM Allele	Primer VIC Allele	Common Primer	RBT	WCT	RBT Ref Sequence GenBank Trace Archive No.
Omy_RAD_17063_31	0	0	a/c	66	100.0	100.0	GTCAGTAGGAGGTGCTATTGAGA	TCAGTAGGAGGTGCTATTGAGC	GCCTGCAGGCTGTCCAGTAGTT	JQ755432	JQ755478	gnl|ti|509018248_648370
Omy_RAD_17806_24	0	0	t/a	66	100.0	100.0	TGGGCTGTGTGAGAGACAGAGA	TGGGCTGTGTGAGAGACAGAGT	ACACCTGCAGGGCCTGTCTGAA	JQ755433	JQ755479	gnl|ti|509619938_650831
Omy_RAD_19234_51	0	0	t/c	66	100.0	98.5	TTCCTGTGTAAAGCAGTGGTGG	ACTTCCTGTGTAAAGCAGTGGTGA	GTGCTCGTACCATCCACCGTCAA	JQ755434	JQ755480	gnl|ti|509620597_651661
Omy_RAD_20663_46	0	2	t/g	66	98.5	100.0	GGAGCAAAGCATTAAAAGTGTGCTG	ATGGAGCAAAGCATTAAAAGTGTGCTT	GTTGTTGATGAGCCAGGGTCTGTTT	JQ755435	JQ755481	gnl|ti|514653699_625686
Omy_RAD_21212_50	0	0	a/t	66	93.9	100.0	TTAATAATCACTACATTTCACATAGAATTGCTT	TTAATAATCACTACATTTCACATAGAATTGCTA	GCTTAGATTGTATATTCTGCTGCTAGGTT	JQ755436	JQ755482	gnl|ti|509701993_669868
Omy_RAD_21362_28	0	0	g/t	66	100.0	100.0	GCTGCTCTGCTGACGGTTC	CTGCTCTGCTGACGGTTA	GGTTGAACTGGACGAGCCGGAA	JQ755437	JQ755483	gnl|ti|514657717_630630
Omy_RAD_21431_30	0	0	a/g	66	100.0	100.0	CTGTTCAGGGTGATGATGCTGT	CTGTTCAGGGTGATGATGCTGC	CAGGATGAGGGTTG-CCTGGTCAT	JQ755438	JQ755484	gnl|ti|509708916_678833
Omy_RAD_22111_34	0	2	t/g	66	100.0	100.0	GAACTTTGCTGGGCATGTGGG	GAACTTTGCTGGGCATGTGGT	TGCACGGATAACATGGTCTTTGATAACTT	JQ755439	JQ755485	gnl|ti|514648399_498985
Omy_RAD_23247_22	0	0	c/t	66	100.0	98.5	CCCACGGGATACTGGGTG	CCCCACGGGATACTGGGTA	CCTGCAGGAGCTGGTCAGCTAT	JQ755440	JQ755486	gnl|ti|514649789_620803
Omy_RAD_23910_14	0	0	c/t	66	100.0	100.0	TCCTGCAGGGTGTCGGCG	CTCCTGCAGGGTGTCGGCA	CGCTTTAAACAGCTGGTGGACAGTA	JQ755441	JQ755487	gnl|ti|514658405_631558
Omy_RAD_26352_38	1	0	g/a	70	94.3	100.0	ATGCACACCACTGCATCCAGAT	ATGCACACCACTGCATCCAGAC	CTACTGTTACAACCTGCAGGAGCTA	JQ755467	JQ755513	gnl|ti|509700018_667294
Omy_RAD_27337_18	0	0	c/t	66	100.0	100.0	GGTAGATTTCCGACGTAAATACGG	GGTAGATTTCCGACGTAAATACGA	GAGGCCCTGCAGGAAATAACGATT	JQ755442	JQ755488	gnl|ti|509627165_659922
Omy_RAD_28080_27	0	0	g/a	66	100.0	100.0	GATGTGTGGCTGTTGGTCAACCA	ATGTGTGGCTGTTGGTCAACCG	GCAAGACCCTCAGAATCCTCTTCAA	JQ755443	JQ755489	gnl|ti|509017279_647127
Omy_RAD_29252_34	1	0	g/t	66	100.0	100.0	GTCGTTCTTCTGGCCCAGGAC	TGTCGTTCTTCTGGCCCAGGAA	GTCAGGCTCTGACGGCCTACTT	JQ755468	JQ755514	gnl|ti|509629735_663284
Omy_RAD_29419_23	1	0	c/t	69	94.2	100.0	CCCGCCAGATGGCCAGG	CCCCGCCAGATGGCCAGA	CATGGAGGACCTGAGTGCTCTAAA	JQ755469	JQ755515	gnl|ti|514648424_499013
Omy_RAD_30378_15	1	0	a/c	69	100.0	91.3	GGTCTGTCCCCCTGTCCCGT	TCTGTCCCCCTGTCCGG	GCAGTGTGACCCTGCAGGACA	JQ755470	JQ755516	gnl|ti|509627605_660457

Table 3 (Continued)

Assay name	Additional Variable Sites		SNP	N	% Call rate	% Concordance	Sequence (5′-3′)			RAD GenBank Accession No.		RBT Ref Sequence GenBank Trace Archive No.
	RAD Tag	Ref Seq					Primer FAM Allele	Primer VIC Allele	Common Primer	RBT	WCT	
Omy_RAD_30423_10	1	0	c/t	70	94.3	100.0	ATTCTAGATTCTAGACACATGACTCC	ATTCTAGATTCTAGACACATGACTCT	TAATTCAACTAGCGGTGTGTTGTGTTGTA	JQ755471	JQ755517	gnl|ti|509010556_638514
Omy_RAD_31988_17	0	0	g/c	69	97.1	100.0	ATAATAAGATCATGCAACAGTAAGTGTTTG	ATAATAAGATCATGCAACAGTAAGTGTTTC	ATGCCCCTGCAGGCAAGCCATT	JQ755444	JQ755490	gnl|ti|514655703_628086
Omy_RAD_38141_41	0	0	t/c	69	98.6	95.6	GAACCCACCCATTTCAGTGGAC	GAACCCACCCATTTCAGTGGAT	TTCCTGGGTGAAGTAGGGGATTGAA	JQ755445	JQ755491	gnl|ti|509708189_677921
Omy_RAD_38362_46	0	0	t/c	69	97.1	100.0	AACCCTCCATTCGTCACATTTAAC	CCAACCCTCCATTCGTCACATTTAAT	CTCTTCTATCTTGTTGACGTCGACCTT	JQ755446	JQ755492	gnl|ti|509625190_657554
Omy_RAD_39958_28	0	0	a/c	69	95.7	100.0	TGGGTAATCACGAGGGTACATCT	GGTAATCACGAGGGTACATCG	CGTCCAGAGGAGCCAATGGCAT	JQ755447	JQ755493	gnl|ti|509623171_655011
Omy_RAD_40876_46	0	0	g/t	69	97.1	98.5	TCACAGTAGTCAACACTGTG	CAGACTCACAGTAGTCAACACTGTT	GTCCTGTTTGTTCATCTGGTCTCCAA	JQ755448	JQ755494	gnl|ti|509707366_676927
Omy_RAD_42014_26	0	0	t/c	69	100.0	100.0	GGTGAAAGTACAGGTAGCGCTTG	AGGTGAAAGTACAGGTAGCGCTTA	AACAGCTTACACCCAGAGCTGCTT	JQ755449	JQ755495	gnl|ti|509708189_677921
Omy_RAD_43425_10	1	0	a/g	69	95.7	98.5	ATCTGTTGACTCCCTCTCCTCT	ATCTGTTGACTCCCTCTCCTCC	TTTACCAGGCGGTGCGGCAGTT	JQ755472	JQ755518	gnl|ti|509701019_668531
Omy_RAD_44398_41	0	0	t/c	69	100.0	100.0	TGAAGAAGCCGGATGTGGAGG	TGAAGAAGCCGGATGTGGAGA	CTCACAAGCGCAGTTCGCATGTAA	JQ755450	JQ755496	gnl|ti|514649096_619853
Omy_RAD_44561_22	0	2	g/a	66	100.0	95.5	GCAGGATTCAGTCAAGAGCCCT	CAGGATTCAGTCAAGAGCCCC	TGTGGACAAGATCAGGACACGTGTT	JQ755451	JQ755497	gnl|ti|509010982_639032
Omy_RAD_44764_26	0	0	c/t	70	95.7	97.0	TGAAGAAGCCGGATGTGGAGG	TGAAGAAGCCGGATGTGGAGA	CTCACAAGCGCAGTTCGCATGTAA	JQ906728	JQ906725	gnl|ti|509008939_636117
Omy_RAD_46598_34	1	0	c/t	70	92.9	96.9	AGGTCCATCAAGTCAAAGGCG	CCAGGTCCATCAAGTCAAAGGCA	GTGGATGACCACCTGCAGGACAA	JQ755473	JQ755519	gnl|ti|509012186_640493
Omy_RAD_48301_23	1	0	c/g	66	100.0	56.1	AGGGATGAGACTCCTCTGAAC	CAGGGATGAGACTCCTCTGAAG	TCTTCCTGCTGCTGATGTTGCTGTT	JQ755466	JQ755512	gnl|ti|514658470_631642
Omy_RAD_48390_31	0	1	t/c	69	100.0	100.0	TCTGCCAGTCTGTCAGGTCG	CTCTGCCAGTCTGTCAGGTCA	GATGCTGTGTGG-GATGCAGGAGAT	JQ755452	JQ755498	gnl|ti|509620961_652168
Omy_RAD_49759_21	1	1	g/t	69	92.8	100.0	GTCTTTGTTGGAATTTATTGCCATATTC	TCTTTGTTGGAATTTATTGCCATATTA	ATATCTCACCTGCAGGTTTAAGTACCAAA	JQ755474	JQ755520	gnl|ti|509700521_667913
Omy_RAD_51740_9	0	0	c/g	69	94.2	100.0	TATCGGGTACCTGCAGGTGAC	TATCGGGTACCTGCAGGTGAG	GCCTTGACAGTACAACAGGCACTTT	JQ755453	JQ755499	gnl|ti|514655330_627614
Omy_RAD_51821_47	0	0	g/t	66	100.0	100.0	AACTCCACAAGGTCAGAGGTAAC	AACTCCACAAGGTCAGAGGTAAA	CTACTCTGCCGACATCCTATCAGAA	JQ755454	JQ755500	gnl|ti|509624106_656140

Table 3 (Continued)

Assay name	Additional Variable Sites		SNP	N	% Call rate	% Concordance	Sequence (5′-3′)			RAD GenBank Accession No.		RBT Ref Sequence GenBank Trace Archive No.
	RAD Tag	Ref Seq					Primer FAM Allele	Primer VIC Allele	Common Primer	RBT	WCT	
Omy_RAD_52968_14	0	0	c/t	69	100.0	98.6	GGTGAATTCGGTGTTGTCTGC	CGGTGAATTCGGTGTTGTCTGT	GAGTCTCATCCCTGCAGGGCTT	JQ755455	JQ755501	gnl|ti|509619107_649781
Omy_RAD_53822_13	0	0	a/c	69	100.0	100.0	AACACCGATATACATAAATGTGCTGT	AACACCGATATACATAAATGTGCTGG	GACTCAGCCTGCAGGGGTCATA	JQ755456	JQ755502	gnl|ti|509623407_655290
Omy_RAD_54126_31	0	0	t/c	69	97.1	97.0	TGGTCAATGCCATTATCAACAGC	CTGGTCAATGCCATTATCAACAGT	CTCACAGTACCAGCACGACCAATA	JQ906729	JQ906726	gnl|ti|514658810_632040
Omy_RAD_54516_35	0	0	a/t	70	92.9	100.0	TGGACTCAAACAGATCCAATAACT	ACTGGACTCAAACAGATCCAATAACA	GGTACTTCTGTGAAAACCATTTGTGTGAA	JQ755457	JQ755503	gnl|ti|514658871_632114
Omy_RAD_54584_17	0	0	c/a	69	81.2	100.0	GTACCTGCAGGGAAAGCTACTCT	TACCTGCAGGGAAAGCTACTCG	GGATCCACCAGTGTGTATGTGTAGTT	JQ755458	JQ755504	gnl|ti|509706696_676093
Omy_RAD_55391_47	0	0	c/t	69	97.1	100.0	GTCAGTTTTCCTTGTCAGGCTCC	TTGTCAGTTTTCCTTGTCAGGCTCT	GTCGAAGTCTGCCTCAACCACAATA	JQ755459	JQ755505	gnl|ti|509621087_652343
Omy_RAD_55820_28	0	0	g/a	66	100.0	100.0	GAGGCCTTACAGATTGATTGCACA	AGGCCTTACAGATTGATTGCACG	GGCACAGCAGAAGACCAATTTCCAT	JQ755460	JQ755506	gnl|ti|509010303_638105
Omy_RAD_5666_9	1	1	c/g	70	92.9	100.0	GGAGGAACCTGCAGGTGGC	GAGGAACCTGCAGGTGGG	AAAGTCAGTTAACTACACTACAGACCAATT	JQ755475	JQ755521	gnl|ti|509018236_648356
Omy_RAD_57262_34	0	0	a/g	69	97.1	97.0	CCACAGCGACCCCATCGAA	CCACAGCGACCCCATCGAG	GGTCAAATGTCAGGGTTAATCAGAAGTA	JQ755461	JQ755507	gnl|ti|509010982_639032
Omy_RAD_57673_42	1	0	c/t	66	100.0	98.5	TCTTACCACGAGCTCAGGGAC	TCTTACCACGAGCTCAGGGAT	GCTGGATCTCATGGTGGTCCAGAT	JQ755476	JQ755522	gnl|ti|509014148_643124
Omy_RAD_59515_14	0	2	g/t	69	100.0	98.6	AGGTGGTGCCAGGACAGGG	CAAGGTGGTGCCAGGACAGGT	CCAGATCCAGGCCTGCAGGTAA	JQ755462	JQ755508	gnl|ti|509707732_677382
Omy_RAD_6035_19	0	0	c/t	70	95.7	100.0	TTGGAGCGGTACTCTTTCAGG	CTTGGAGCGGTACTCTTTCAGA	GGAGTCCCTGCAGGCCAATGTA	JQ906732	JQ906731	gnl|ti|509011140_639217
Omy_RAD_60674_25	0	0	g/a	66	100.0	100.0	TGTGCTGCAGCCCACCATCAGAA	TGCTGCAGCCCACATCAGAG	TTAACCTGCAGGATGAGGAAGGCTT	JQ755463	JQ755509	gnl|ti|514655878_628332
Omy_RAD_69061_12	0	0	g/a	70	95.7	100.0	GTCTGTAGCATATACTATGTTGTCCT	TCTGTAGCATATACTATGTTGTCCC	CCTGTCTGGGAATAACAGCCGTATA	JQ755464	JQ755510	gnl|ti|509624815_657011
Omy_RAD_76689_9	1	0	c/t	66	100.0	100.0	ACCTCGTCCTGCAGGTCTG	CCACCTCGTCCTGCAGGTCTA	CCCTGCTCGACCCGTGTCTA	JQ755477	JQ755523	gnl|ti|514657197_629915
Omy_RAD_77157_46	0	0	a/c	70	98.6	100.0	TGTGTTACAGCTGCGGGTCCTT	TGTTACAGCTGCGGGTCCTG	GGTCAGGCTGCAGTGGAGGAAA	JQ755465	JQ755511	gnl|ti|509017892_647923
Omy_RAD_8436_22	0	0	g/t	70	98.6	97.1	GAGCCGTCCTTCAGGAATCGC	AGAGCCGTCCTTCAGGAATCGA	CTGCAGGAGAGGGAGGGGCTT	JQ906727	JQ906730	gnl|ti|509698940_665465

levels of introgression will still be difficult to detect at the individual level. Our probability of detecting 1% introgression in a fish using 46 markers is only 60%.

Because our preliminary screening panel was composed of a few individuals from many populations instead of many individuals from a few populations, we have little power to detect alleles at low frequency in these populations. Thus, we cannot exclude the possibility that some of the diagnostic markers may share a low-level polymorphism between WCT and RBT. In 12 species-diagnostic assays, individuals with genotypes indicative of low-level polymorphisms or RBT hybridization were detected. The SNP from RAD_49331 was not diagnostic in our screening panel. One individual was homozygous for the 'RBT allele' at this locus, and 8 others were heterozygotes. These individuals, however, did not possess any other genotypes at the other loci analysed indicative of hybridization. Thus, this locus appears to be polymorphic in WCT. WCT from northwestern Montana or south-eastern British Columbia were heterozygous in five additional assays, and the level of introgression in these populations is uncertain due in part to the potential presence of IRT alleles (Table S1). Fish from Lake Koocanusa may have low levels of introgression from CRT or WCT. In the fish from Yaak, BC and Runt Creek, these polymorphisms may represent natural levels of introgression between the sympatric WCT and IRT populations. In five of the remaining six assays, the heterozygous fish came from hatchery populations. Because hatchery brood stock samples have been extensively screened for admixture using indels and microsatellites, it is most likely that these alleles represent shared low-level polymorphisms between WCT and RBT. Testing of additional samples will be required to determine their frequency and the usefulness of the assays for species identification and admixture analysis.

Discussion

Our conversion rate of 82% for diagnostic assays suggests that RAD sequencing offers a reliable and relatively quick and inexpensive way to generate large numbers of SNP markers that does not require a large screening panel (e.g. Seeb *et al.* 2011). Conversion rates can vary widely and depend on the variability and divergence of the target species, the number of samples sequenced before designing the assays, whether the SNP is in a conserved or highly variable region (e.g. diagnostic between species or polymorphic within species) and on the number and extent of samples used to validate the assay. New sample library protocols and next-generation sequencing techniques like RADs promise to make very low-cost marker development available for most organisms (Pennisi 2011) even when no genomic resources are available.

The development of additional species-diagnostic genotyping assays provides increased power for detecting hybridization at the individual level and more precisely estimating the structure of hybrid zones. With the addition of our 46 assays to the 31 previously available SNPs (Finger *et al.* 2009; McGlauflin *et al.* 2010, Harwood & Phillips 2011; Kalinowski *et al.* 2011), the number of currently available diagnostic SNPs between WCT and RBT has increased to 77. With 77 diagnostic SNPs, we can detect 1.9% introgression with 95% certainty at the individual level. Our probability of detecting 1% introgression in a fish using 96 markers is only 85%, reaching 95% with 150 markers (Table 1). The ability to detect low levels of hybridization at the individual level increases sampling scheme flexibility, removing the requirement that aggregations of 20-30 samples be considered a population.

We developed a bioinformatic pipeline using publicly available 454 reads (Sanchez *et al.* 2009) for identifying flanking sequence required for assay design that will be easily applied to the rainbow trout genome sequence when it is published (Miller *et al.* 2011). This reduced our set of candidate loci from 4914 to 66 (1.3%). At the time of this experiment, using 454 sequencing to produce reads >100-nt reads required for SNP assay development was beyond our budget. The reference genome sequence will allow assay design for most of the SNPs identified in our RAD loci.

An alternative approach to using published long read sequence data is to generate longer contiguous sequence reads at each RAD tag using over-lapping paired-end sequencing (Etter *et al.* 2011b). This technique holds great promise for allowing assay design on the full set of candidate SNP markers for any species. In addition, this approach should have a higher validation rate, because SNP detection and flanking sequence would come from the same individuals and populations.

RAD sequencing is one of a family of approaches applying high-throughput sequencing to a reduced representation of a genome to identify and genotype large numbers of SNP markers in organisms without substantial genetic resources (Cosart *et al.* 2011; Davey *et al.* 2011). Next-generation sequencing approaches require slightly more bioinformatic effort compared with traditional marker discovery, but a number of publicly available tools are being developed to handle these types of data (Catchen *et al.* 2011; Davey *et al.* 2011). One advantage of RAD over related restriction-enzyme-reduced representation sequencing techniques in taxa with complex, repetitive genomes is that the set of markers does not depend on a fragment size selection step, so that it is more consistent across libraries (Davey *et al.* 2011). This helps reduce variation between runs and allows the compilation and re-analysis of large sequence databases across related species, populations and individuals generated using the same RAD library technique. We con-

clude that the emerging techniques for the generation and analysis of RAD sequencing data provide a relatively quick and cost-effective method for the identification of large numbers of species-diagnostic SNPs.

Acknowledgements

PAH was supported by NIH COBRE grant 5P20RR016448 (L. Forney, PI). Funding was provided in part by the Great Northern Landscape Conservation Cooperative (US Department of Interior). Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. This research was conducted in accordance with the Animal Welfare Act and its subsequent amendments. GL & FWA were supported by NSF DEB 0742181.

References

Allendorf FW, Leary RF (1988) Conservation and distribution of genetic variation in a polytypic species, the cutthroat trout. *Conservation Biology*, **2**, 170–184.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

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

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Behnke RJ (2002) *Trout and salmon of North America*. Simon and Schuster, New York.

Catchen JM, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3 Genes Genomes Genetics*, **1**, 171–182.

Cosart T, Beja-Pereira A, Chen S, Ng SB, Shendure J, Luikart G (2011) Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genomics*, **12**, 347.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011a) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: *Molecular Methods for Evolutionary Genetics* (eds Orgogozo V & Rockman MV), pp. 157–178. Humana Press, New York.

Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA (2011b) Local *de novo* assembly of RAD paired-end contigs using short sequencing reads. *PLoS ONE*, **6**, e18561.

Finger AJ, Stephens MR, Clipperton NW, May B (2009) Six diagnostic single nucleotide polymorphism markers for detecting introgression between cutthroat and rainbow trout. *Molecular Ecology Resources*, **9**, 759–763.

Harwood AS, Phillips RB (2011) A suite of twelve single nucleotide polymorphism markers for detecting introgression between cutthroat and rainbow trout. *Molecular Ecology Resources*, **11**, 382–385.

Hohenlohe P, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) RAD sequencing identifies thousands of SNPs for assessing hybridization in rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Jiggins CD, Mallet J (2000) Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*, **15**, 250–255.

Kalinowski ST, Novak BJ, Drinan DP, Jennings R, Vu NV (2011) Diagnostic single nucleotide polymorphisms for identifying westslope cutthroat trout (*Oncorhynchus clarki lewisi*), Yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) and rainbow trout (*Oncorhynchus mykiss*). *Molecular Ecology Resources*, **11**, 389–393.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**, R25.

Lever C (1996) *Naturalized Fishes of the World*. Academic Press, San Diego, California, USA.

Lie Ø, Slettan A, Lingaas F, Olsaker I, Hordvik I, Refstie T (1994) Haploid gynogenesis: A powerful strategy for linkage analysis in fish. *Animal Biotechnology*, **5**, 33–45.

McGlauflin MT, Smith MJ, Wang JT *et al.* (2010) High-Resolution Melting Analysis for the Discovery of Novel Single-Nucleotide Polymorphisms in Rainbow and Cutthroat Trout for Species Identification. *Transactions of the American Fisheries Society*, **139**, 676–684.

Miller MR, Brunelli JP, Wheeler PA *et al.* (2011) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular Ecology*, **21**, 237–249.

Muhlfeld CC, Kalinowski ST, McMahon TE *et al.* (2009) Hybridization rapidly reduces reproductive success of a native trout in the wild. *Biology Letters*, **5**, 328–331.

Pennisi E (2011) Using DNA to reveal a mosquito's history. *Science*, **331**, 1006–1007.

Rasmussen JB, Robinson MD, Heath DD (2010) Ecological consequences of hybridization between native westslope cutthroat (*Oncorhynchus clarkii lewisi*) and introduced rainbow (*Oncorhynchus mykiss*) trout: effects on life history and habitat use. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 357–370.

Sanchez CC, Smith TPL, Wiedmann RT *et al.* (2009) Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics*, **10**, 559–566.

Seeb JE, Carvalho GR, Hauser L, Naish KA, Roberts SB, Seeb LW (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, **11**, 1–8.

Shepard BB, May BE, Urie W (2005) Status and conservation of westslope cutthroat trout within the western United States. *North American Journal of Fisheries Management*, **25**, 1426–1440.

Weigel DE, Peterson JT, Spruell P (2003) The distribution of introgressive hybridization between westslope cutthroat trout and rainbow trout in the Clearwater Basin, Idaho. *Ecological Applications*, **13**, 38–50.

SJA, PAH, RFL, GL conceived and designed the project. RFL, CM supplied the samples. SJA, SP, PAH analyzed the data. SJA, PAH, RFL, CM, FWA, GL wrote the paper.

Data accessibility

SNP genotypes have been deposited at Dryad: doi: 10.5061/dryad.b31s9.

Supporting Information

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

Table S1 The sample location and number, RAD locus, species and number of heterozygous genotypes (Hets) are reported from the validated assays.

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.