Current Zoology 61 (1): 146–154, 2015

Genomics and introgression: Discovery and mapping of thousands of species-diagnostic SNPs using RAD sequencing

Brian K. HAND[1], Tyler D. HETHER[2], Ryan P. KOVACH[1,3], Clint C. MUHLFELD[3], Stephen J. AMISH[4], Matthew C. BOYER[5], Sean M. O'ROURKE[6], Michael R. MILLER[6], Winsor H. LOWE[7], Paul A. HOHENLOHE[2*], Gordon LUIKART[1,4]

[1] Flathead Lake Biological Station, Fish and Wildlife Genomics Group, University of Montana, Polson, MT 59860, USA

[2] Institute for Bioinformatics and Evolutionary Studies, Department of Biological Sciences, University of Idaho, Moscow, ID 83844, USA

[3] U. S. Geological Survey, Northern Rocky Mountain Science Center, Glacier National Park, West Glacier, Montana 59936, USA

[4] Fish and Wildlife Genomics Group, Division of Biological Sciences, The University of Montana, Missoula, Montana 59812, USA

[5] Montana Fish, Wildlife and Parks, Kalispell, Montana 59901, USA

[6] Department of Animal Science, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA

[7] University of Montana, Division of Biological Sciences, Missoula, Montana 59812, USA

Abstract Invasive hybridization and introgression pose a serious threat to the persistence of many native species. Understanding the effects of hybridization on native populations (e.g., fitness consequences) requires numerous species-diagnostic loci distributed genome-wide. Here we used RAD sequencing to discover thousands of single-nucleotide polymorphisms (SNPs) that are diagnostic between rainbow trout (RBT, *Oncorhynchus mykiss*), the world's most widely introduced fish, and native westslope cutthroat trout (WCT, *O. clarkii lewisi*) in the northern Rocky Mountains, USA. We advanced previous work that identified 4,914 species-diagnostic loci by using longer sequence reads (100 bp *vs*. 60 bp) and a larger set of individuals (n = 84). We sequenced RAD libraries for individuals from diverse sampling sources, including native populations of WCT and hatchery broodstocks of WCT and RBT. We also took advantage of a newly released reference genome assembly for RBT to align our RAD loci. In total, we discovered 16,788 putatively diagnostic SNPs, 10,267 of which we mapped to anchored chromosome locations on the RBT genome. A small portion of previously discovered putative diagnostic loci (325 of 4,914) were no longer diagnostic (i.e., fixed between species) based on our wider survey of non-hybridized RBT and WCT individuals. Our study suggests that RAD loci mapped to a draft genome assembly could provide the marker density required to identify genes and chromosomal regions influencing selection in admixed populations of conservation concern and evolutionary interest [*Current Zoology* 61 (1): 146–154, 2015].

Keywords Conservation genetics, Hybridization, Invasive species, Next generation sequencing, Salmonid fish, SNP discovery

Species invasions and resulting hybridization between native and non-native species provide exciting evolutionary experiments where natural selection on novel gene combinations can be studied in ecological time and in different environments (Fitzpatrick et al., 2010; Rieseberg, 2011). Hybridization between native and invasive species is one of the most serious threats to global biodiversity, and is a major conservation concern (Allendorf et al., 2001). The rate of hybridization for numerous taxa is increasing with habitat degradation, species translocations, and climate change (Kelly et al., 2010; Muhlfeld et al., 2014). In many cases, natural selection and dispersal by hybrids play key roles in the spread of hybridization in nature (e.g., Rhymer and Simberloff, 1996; Kovach et al., 2014). The genetic basis for mechanisms influencing introgression can now be assessed by discovering and genotyping thousands of genetic markers, thereby illuminating the genomic basis of hybridization, fitness, adaptive capacity, and dispersal (Allendorf et al., 2010; Twyford and Ennos, 2012).

Here we report on the development and mapping of thousands of diagnostic loci between rainbow trout (RBT, *Oncorhynchus mykiss*) and westslope cutthroat trout (WCT, *O. clarkii lewisi*). The RBT is the most widely introduced cold-water fish in the world (Halverson, 2010), and is one of the 100 most problematic exo-

Received Sep. 11, 2014; accepted Dec. 22, 2014.

* Corresponding author. E-mail: hohenlohe@uidaho.edu

tic invasive species (Lowe et al., 2000). WCT and all other 12 extant cutthroat trout taxa are threatened with genomic extinction due to introgressive hybridization with RBT (Allendorf et al., 2005; Trotter, 2008), and introgressive hybridization is considered the primary threat to the persistence of cutthroat trout (e.g., Shepard et al., 2005). For example, known non-hybridized WCT populations occupy only about 10% of their historic range primarily due to hybridization with introduced RBT (Shepard et al., 2005). Consequently, WCT have twice been petitioned for listing under the U.S. Endangered Species Act (USFWS, 2003; Shepard et al., 2005).

Next generation sequencing (NGS), particularly restriction-site-associated DNA (RAD) sequencing, has enabled the discovery and genotyping of thousands of single-nucleotide polymorphisms (SNPs) distributed across the genome of non-model plants and animals (Baird et al., 2008; Narum et al., 2013). The increased density of loci and genome coverage offered by RAD sequencing has enormous promise to help researchers understand the mechanisms driving hybridization and its effects on species fitness, persistence, and adaptive potential (Allendorf et al., 2010), including in the context of climate change (Muhlfeld et al., 2014).

We previously identified a set of 4,914 diagnostic loci using relatively short-read (60 bp) RAD sequencing in 24 individuals (11 WCT and 13 RBT) and *de novo* assembly of loci (Hohenlohe et al., 2011; Amish et al., 2012). We further filtered those loci using a set of RAD contigs assembled for WCT to assess patterns of introgression in populations across a large river drainage (Hohenlohe et al., 2013). Here we extend this work by applying paired-end RAD sequencing with longer reads (100 bp) to a larger set of individuals from pure RBT and WCT populations, and take advantage of a newly released reference genome assembly for RBT (Berthelot et al., 2014). Our objectives were: (i) to discover a more reliable and expanded set of diagnostic SNP loci for RBT and WCT; (ii) to physically map these SNPs to the RBT genome; and (iii) to validate previously identified diagnostic loci against this broader population sample and compare admixture estimates produced by each approach.

We show that the combination of an expanded set of populations, longer reads, and alignment to the reference genome increases the number of diagnostic SNPs detected. An increased number of species diagnostic markers aids in the discovery of putative super-invasive alleles and development of low-cost diagnostic SNP panels (e.g., the 100–200 most highly informative diagnostic SNP markers). Mapping to the RBT genome offers the additional advantage of enabling diagnostic SNP panels that are evenly distributed across chromosomes, thus increasing statistical power to detect genomic regions of reduced or elevated introgression, and to identify candidate adaptive loci within them. For the purposes of conservation and management, a mapped, well-chosen set of diagnostic loci is crucial for early detection and intervention in populations at the early stages of introgression, for choosing only non-hybridized individuals for translocations, and for prioritizing conservation and recovery programs (Allendorf et al., 2010).

1 Materials and Methods

1.1 WCT and RBT source populations

It can be difficult to find the non-hybridized populations required for discovery of species-diagnostic markers. For WCT we focused on three stream populations (South Creek, Addition Creek, Danaher Creek) of non-hybridized WCT located in the South Fork Flathead River drainage in northwestern Montana (Fig. 1). We also included samples from the Washoe Park Hatchery in Anaconda, Montana, which is the source of the mixed origin WCT M012 broodstock widely used as a source for reintroductions for conservation programs and introductions for recreational fisheries. The WCT M012 broodstock is composed of a diverse set of 21 pure stream populations, including South Creek and Addition Creek, throughout the South Fork Flathead River drainage, and two populations in the Clark Fork River drainage (Fig. 1). These populations have all been tested repeatedly for RBT admixture, and RBT introgression has never been detected (Montana Fish, Wildlife and Parks, *unpublished data*). We included 16 individual samples from each of South Creek, Danaher Creek, and Addition Creek and 18 samples from the Washoe Park Hatchery, for a total of 66 WCT samples.

We used 18 individual samples of RBT from diverse populations including three hatchery stocks (the Ennis and Arlee hatcheries in Montana, and the Rock Creek Hatchery in Oregon). Samples from the Ennis Hatchery included a mix of 6 known broodstock sources: McConaughy, Nebraska; Eagle Lake, California; Fish Lake, Utah; Erwin Hatchery, Tennessee; Shasta, California; Arlee Hatchery, Montana. Six additional individual samples from the Arlee hatchery were also included. The Arlee hatchery RBT strain was the principle source of fish stocked in the Flathead drainage until the practice ended in 1969. The Arlee strain began in 1953 with

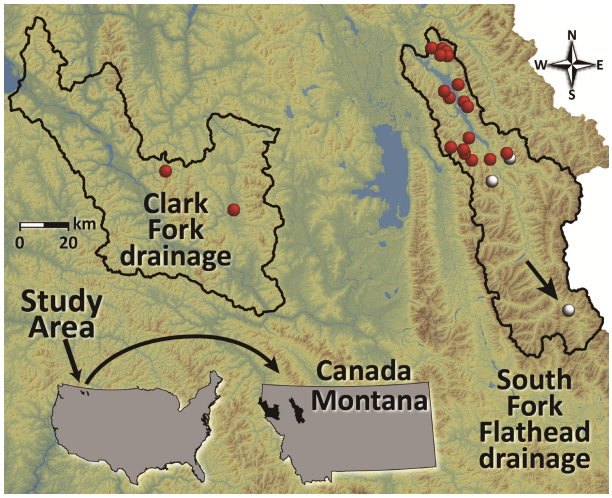

Fig. 1 Study area for westslope cutthroat trout (WCT; *Oncorhynchus clarkii lewisi*) in northwest USA

Black lines delineate the Clark Fork River drainage (left) and the South Fork Flathead River drainage (right). Individuals from three wild populations (silver dots; South Creek, Addition Creek and Danaher Creek) were sampled, in addition to the inclusion of individuals from the M012 broodstock, which represents a mixture of multiple source populations (red dots and including South Creek and Addition Creek). The arrow at bottom right denotes the Danaher Creek population from which individuals have never been used in the MO12 broodstock.

the crossing of two major strains (the Donaldson Trout strain and the Missouri Strain; Stephanie Espinoza, personal communication). Historically, the Arlee hatchery broodstrock has included coastal strains of RBT with some previous contributions from inland redband rainbow trout *Oncorhynchus mykiss gairdneri* from the Kootenai drainage. Finally, we included four individual samples from a previous study using clonal lines from the Rock Creek Hatchery on the North Umpqua River in southwestern Oregon, the Clearwater River in north central Idaho, and the Swanson River in South Central Alaska (described in Miller et al., 2012).

1.2 RAD sequencing, alignment and genotyping

We prepared libraries for RAD sequencing from genomic DNA from WCT and RBT individuals according to standard protocols using the restriction enzyme SbfI and unique 6bp barcodes for each sample (Miller et al., 2012). We multiplexed these 84 samples, along with RAD libraries (at equal concentration) for an additional 108 other salmonid samples for another study at equal concentration, in 4 sequencing lanes on an Illumina HiSeq machine. We conducted initial processing of the sequence data using several modules from the Stacks software package, version 1.19 (Catchen et al., 2013). We used *process_radtags* from Stacks to sort read pairs by barcode and remove any pairs in which the forward read did not contain both a correct barcode and the remaining six bases of the SbfI recognition sequence.

Paired end reads from the same individual were only used to identify PCR duplicates. The random shearing step in traditional RAD sequencing produces staggered paired-end reads, so that any set of read pairs that are identical across both the forward and reverse reads are likely PCR duplicates of a single original genomic DNA fragment (Davey et al., 2011). We therefore removed all read pairs that represented PCR duplicates using the Stacks program *clone_filter*. No further quality filtering was performed at this stage.

We aligned filtered read pairs from each individual to the recently published RBT reference genome (Berthelot et al., 2014). We used the alignment software bowtie2 v2.1.0 (Langmead and Salzberg, 2012), using end-to-end alignment without allowing gaps, and allowing up to an average of one high-quality nucleotide mismatch per 20bp. We retained only those read pairs that aligned uniquely to a single genomic location according to these criteria.

All further analysis considered only forward reads (i.e., sequence within the first 94 bp of each restriction enzyme cut site) to provide a set of SNP loci that could be efficiently and reliably genotyped using further single-end RAD sequencing. We assigned diploid genotypes at each nucleotide position in each individual using the bounded maximum-likelihood method described by Catchen et al. (2013), with a minimum Phred quality-score threshold of 10 at each nucleotide, the upper bound of the sequencing error rate set to 0.01, and a likelihood ratio significance level of $\alpha = 0.05$, and we calculated F_{ST} across all loci between the two species (custom software is available at http://webpages.uidaho.edu/hohenlohe/software.html). The upper bound on the sequencing error rate effectively makes the genotyping more sensitive to rare alleles and more likely to call a heterozygous genotype when two alleles are observed, making our analysis generally conservative for detecting loci that are fixed within each species.

1.3 Identifying diagnostic and polymorphic loci

We applied multiple filters to identify SNPs that were fixed for different alleles in RBT and WCT (i.e., species-diagnostic). For each putatively diagnostic locus, we required genotypes from at least 10 WCT individuals from each of the four populations and at least 12 RBT individuals. One hybrid individual from Addition Creek was present in our database. However, given the fact that all previous genetic testing indicated that this

population is composed of non-hybridized WCT, we suspected that this sample was mislabeled and from another population. We did not include this individual in further analysis, as it substantially reduced (in the thousands) the number of diagnostic SNPs when included in the dataset.

Previous work identified 4,914 putatively diagnostic SNPs (Hohenlohe et al., 2011; Amish et al., 2012) using 54 bp reads, the same restriction enzyme, and *de novo* assembly of loci. To validate this prior set of diagnostic SNPs against the new data, we first aligned these prior RAD loci to the RBT reference genome using bowtie2 (Langmead and Salzberg, 2012), providing chromosome and base pair positions for previous SNPs that exactly corresponded to the current dataset. We rejected any of the prior diagnostic SNPs that were genotyped in fewer than 10 of 84 individuals in the new dataset, or that showed a minor allele frequency (MAF) greater than 0.01 in either species (effectively removing all SNPs with more than one copy of the "wrong" allele observed in either WCT or RBT).

We also tested our newly discovered diagnostic SNPs against raw data from Hohenlohe et al. (2011). From the 24 individuals used in that previous study, we removed 7 individuals (Jocko river and Abbot Creek samples) from populations where hybridization has occurred (Hitt et al., 2003; Corsi et al., 2013). We aligned the single-end RAD data from that study to the RBT reference genome assembly using bowtie2 with parameters identical to those described above. We removed all reads that did not align uniquely to the genome, and we called diploid genotypes as described above. Finally, we removed any SNPs in the new set of diagnostic loci that were found to share a single allele (e.g., shared ancestral polymorphisms) between WCT and RBT if at least one individual was genotyped for each species.

Lastly, we compared admixture proportions of 94 individuals and population admixture in 5 hybridized populations that were previously analyzed with 3,180 SNPs (Hohenlohe et al., 2013) and 7 microsatellite loci (Boyer et al., 2008). We aligned the raw paired-end RAD data (described by Hohenlohe et al., 2013) to the RBT genome using bowtie2 and called genotypes for the forward reads, using identical parameters as described above. We extracted the genotypes for our new set of diagnostic SNPs and calculated individual-level admixture as the proportion of RBT alleles across all loci. This compares results from not only the new set of SNPs to those used previously, but also the approach of aligning sequence reads to the RBT genome vs. using WCT-based RAD contigs (Hohenlohe et al., 2013).

2 Results

2.1 Discovery of species-diagnostic SNPs

RAD sequencing of the 84 WCT and RBT samples yielded a total of 229.6 million sequence read pairs. We removed PCR duplicates and aligned the remaining 136.5 million read pairs against the recently published RBT reference genome assembly (Berthelot et al., 2014). On average, 54.5% (*SD* = 3.4%) of read pairs aligned uniquely to a single genomic location for each individual. This proportion was slightly higher (58.1%) for the RBT individuals compared to the WCT individuals (53.2%). We then estimated diploid genotypes per individual at each nucleotide position in the first 94 bp of each RAD locus (i.e., using only the forward reads), with an average of 58,995 RAD loci genotyped per individual. After filtering and alignment, the mean sequencing coverage at successfully genotyped RAD loci per individual was 14.0x (range 3.8–21.8x; *SD* = 4.4x).

Filtering based on the criteria outlined above for number of individuals genotyped per locus, and applying a minor allele frequency threshold of 0.02 across the whole dataset, produced a set of 77,773 SNPs, each genotyped at a mean of 76.2 individuals (*SD* = 4.90). Across this set of SNP loci, F_{ST} = 0.525 between the two species. Of these SNPs, 17,424 were fixed for alternative alleles between the two species. A subset of these SNPs (9,789 out of 17,424, or 56.1%, that occurred in the first 54bp of each RAD locus) was validated against the additional WCT and RBT population samples from Hohenlohe et al. (2011). We found 636 of these loci to be polymorphic within one or both species; these loci were removed, leaving a final set of 16,788 diagnostic SNP loci. Data for these diagnostic loci is available on the Dryad database (http://datadryad.org).

2.2 Mapping and validation

The newly discovered set of diagnostic SNPs were fairly evenly spaced across the RBT genome. The RBT assembly comprises 87 Mb (4%) on anchored and ordered chromosomes, along which position and orientation of contigs is known; 940 Mb (44%) on anchored chromosomes, on which local ordering and orientation of contigs is not known; and 1100 Mb (52%) on unanchored scaffolds not assigned to a chromosome (Berthelot et al., 2014). We found that putatively diagnostic SNPs were distributed in similar proportions on anchored and oriented chromosomes (5.0%; Fig. 2), overrepresented on anchored and globally ordered chromosomes (56%; Fig. 3), and underrepresented on unanchored regions (39%; Table 1).

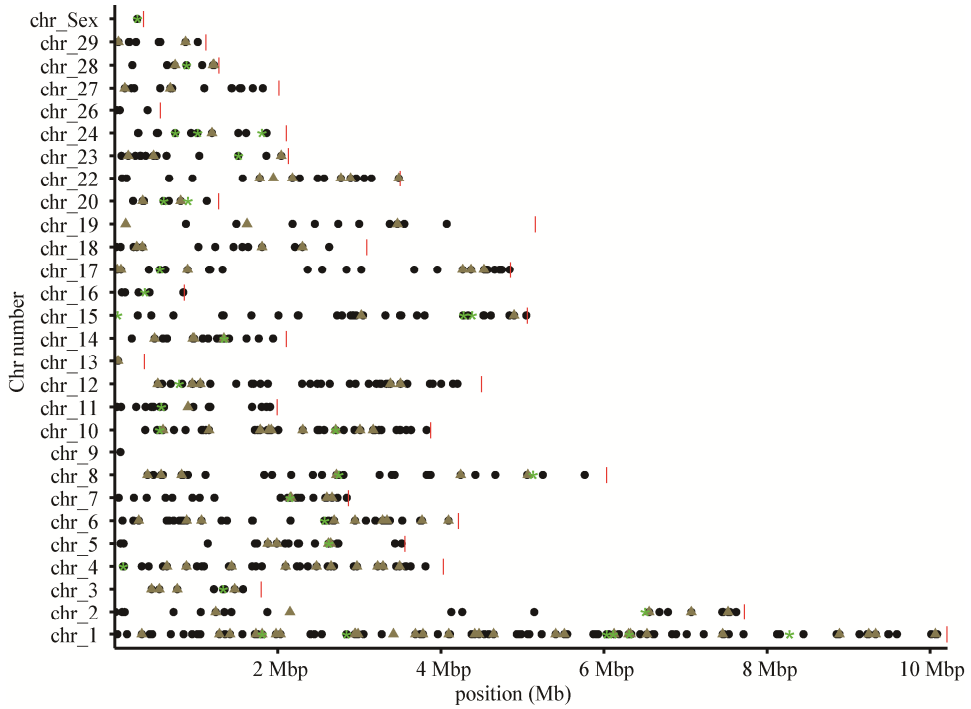

Fig. 2 Chromosomal positions of species-diagnostic single-nucleotide polymorphisms (SNPs) on anchored and ordered chromosomes in the recent assembly of the rainbow trout *Oncorhynchus mykiss* genome (Berthelot et al., 2014).

Red lines delimit the known length of each chromosome (Berthelot et al., 2014). Black dots are newly discovered diagnostic loci (832 SNPs), khaki-colored triangles are 142 SNPs shared between the new set of diagnostic loci and a previously published set (Hohenlohe et al., 2011). Green asterisks represent 40 SNPs from Hohenlohe et al. (2011) that were mapped to chromosomal regions, but were not found in the new set of diagnostic loci.

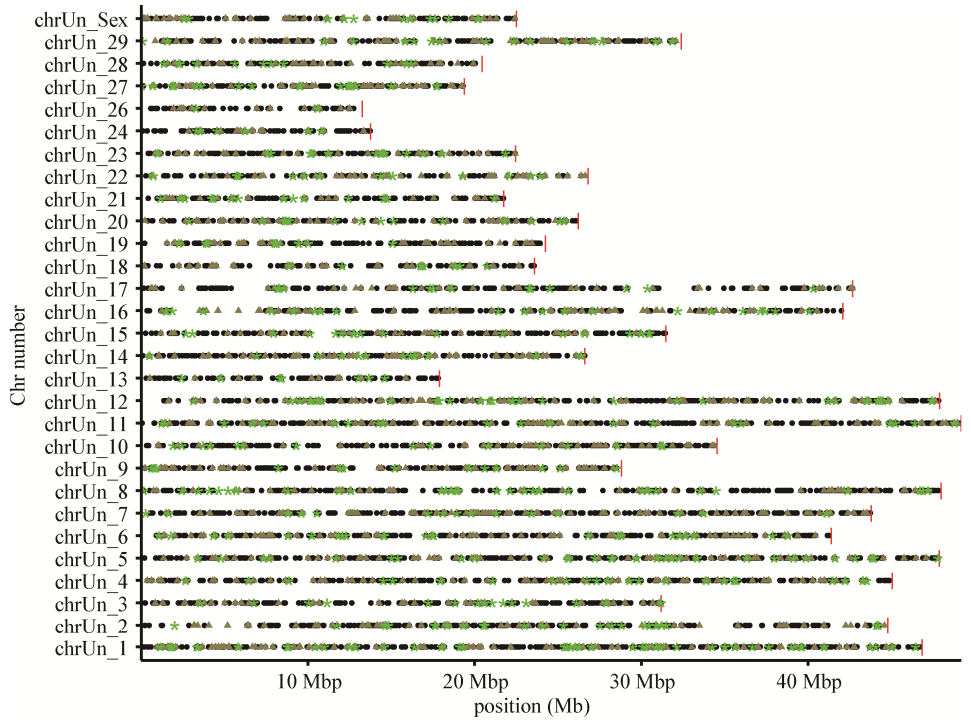

Fig. 3 Chromosomal positions of species-diagnostic single-nucleotide polymorphisms (SNPs) on globally ordered but not oriented chromosome sequences (where the order of two adjacent contiguous sequences might be ambiguous), in the recent assembly of the rainbow trout *Oncorhynchus mykiss* genome (Berthelot et al., 2014)

Red lines delimit the known length of each chromosome (Berthelot et al., 2014). Black dots are new putatively diagnostic loci (9729 SNPs), khaki-colored triangles are 1831 SNPs shared between the new set of diagnostic loci and a previously published set (Hohenlohe et al., 2011). Green asterisks represent 576 SNPs from Hohenlohe et al. (2011) that were mapped to chromosomal regions, but were not found in the new set of diagnostic loci.

Table 1 Discovery, validation, and mapping of two sets of species-diagnostic SNPs: Those from previous work (Hohenlohe et al., 2011; Amish et al., 2012), and those from the current study

Description	Anchored, ordered chromosomes	Anchored chromosomes	Unanchored scaffolds	Totals
Previous data				4914
Map to RBT genome	198	2407	1494	4099
Genotyped in new dataset	158	1831	1047	3036
Confirmed diagnostic	142	1653	916	2711
Polymorphic	16	178	131	325
Current study	877	9729	6818	17424
Confirmed diagnostic	832	9435	6521	16788
Polymorphic	45	294	297	636

The rainbow trout (RBT; *Oncorhynchus mykiss*) reference genome (Berthelot et al., 2014) is split into three groups: i) anchored and ordered chromosomes, ii) anchored chromosomes, which are globally ordered but where local order and orientation is unknown; and iii) unanchored scaffolds (not associated with any chromosome). Polymorphic sites refer to SNPs identified in a single dataset as diagnostic, but rejected here because of polymorphism detected within at least one of the species in the other dataset. Shown are numbers of SNPs in each category.

We further screened and validated previously identified loci within the RBT genome, and compared the utility of aligning against the RBT reference genome to that of our previous set of RAD contigs from WCT (Hohenlohe et al., 2013). From Hohenlohe et al. (2013), contig lengths ranged from 147 to 519 bp with most between 250 and 450 bp. A total of 3,456 out of 4,914 SNP loci (70%) aligned uniquely to a single RAD contig from Hohenlohe et al. (2013). We were able to map 4,099 (83%) of these SNPs, of which 3,036 were genotyped in a sufficient number (>10) of individuals to allow validation against the expanded set of populations. Of these, 2,711 were confirmed to be diagnostic and the remaining 325 were polymorphic in either WCT or RBT.

In our final test, we compared individual and population-level admixture proportions previously estimated with RAD-contig alignment (Hohenlohe et al., 2013) and microsatellite loci (Boyer et al., 2008) to rates of admixture proportion calculated using alignment to the RBT genome and an increased number of SNPs (16,788 *vs*. 3,180). The old and new estimates of individual admixture proportion were strongly correlated for SNP loci (r^2 = 0.987; Fig. 4, right panel). The new SNP loci were more consistent with population level estimates from microsatellite loci, despite individual estimates being more highly variable for microsatellite loci (Fig. 4, left panel). Thus both the old and new SNP loci give a more precise estimate, but the new set of SNP loci appeared to eliminate some bias that was present in the previous set of SNP loci.

3 Discussion

3.1 Discovery of species-diagnostic SNPs

We validated a previously-discovered set of 2,711 species-diagnostic SNPs between RBT and WCT (Ho-

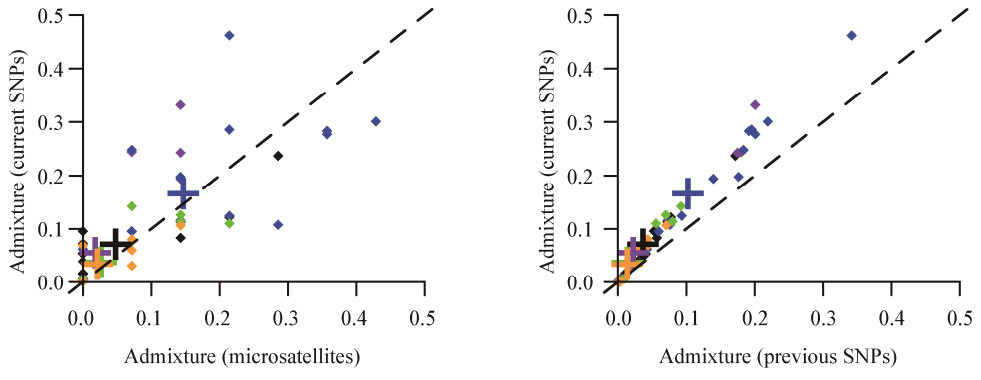

Fig. 4 Individual-level and population-mean admixture proportions (marked by +) estimated from a previous set of 7 diagnostic loci (left panel; Boyer et al. 2008) *vs*. current estimates from 16,788 SNP loci, and (right panel) using 3,180 SNPs (Hohenlohe et al., 2013) *vs*. current estimates

Estimates are for the same 94 westslope cutthroat trout individuals from five populations in the North Fork Flathead River, USA as described previously (Hohenlohe et al., 2013). Populations are Meadow (black), Nicola (green), Dutch (blue), Lower Hay (purple), and Tepee (orange). The dashed line shows the expectation of equality between the two estimates.

henlohe et al., 2013), and discovered thousands (14,077) of new diagnostic SNPs. In total, we have now identified 16,788 SNPs that can be used to describe genome-wide patterns of introgressive hybridization between these species (Figs. 2 and 3). The new discovery of thousands of SNPs is due to several factors. First, we used longer sequence reads (100 bp) compared to our previous discovery of diagnostic loci (60 bp). Second, we aligned sequence data against the newly available RBT reference genome rather than conducting a *de novo* assembly of loci, which has been shown to improve clustering of reads to form loci (Catchen et al., 2013). Third, we used paired-end sequencing, allowing us to improve sequence alignment to the reference genome and to filter out PCR duplicates. Lastly, we used genotyping parameters that were sensitive to loci with a strong excess of heterozygotes and thus more conservative in identifying diagnostic loci. All of these considerations are particularly acute in salmonid fish taxa, in which a recent ancestral genome duplication led to abundant paralogous sequence variants (PSV) that plague development of genetic markers (Seeb et al., 2011; Berthelot et al., 2014). It is worth noting that although we have identified and removed PSV from our data set, this also means that we are under-representing duplicated genomic regions. How this may influence interpretation of genome-wide patterns of admixture remains an ongoing, and poorly understood, problem in population genomics.

The available reference genome for RBT is still a work in progress, with the large majority of contigs or scaffolds either assigned to chromosomes but not locally ordered (or oriented; 44%), or not assigned to chromosomes at all (52%). Nonetheless, this resource appeared to improve SNP discovery when compared to the previous set of RAD contig loci assembled from WCT (Hohenlohe et al., 2013). Even unanchored contigs provided a valuable reference for alignment, which served to further filter reads by quality and to better assign reads to loci in comparison to *de novo* clustering. Accordingly, a higher proportion of diagnostic SNPs from the previous work aligned uniquely to the reference genome compared to the previous set of RAD contigs (83% *vs*. 70%, respectively).

The large number of species-diagnostic SNPs between RBT and WCT identified here was slightly higher than the number of diagnostic loci discovered for other hybridizing species in three recent studies (Stölting et al., 2013; Li et al., 2014; Pujolar et al., 2014). It should be noted, however, that direct (quantitative) comparisons between the studies is difficult because of several confounding factors influencing the SNP discovery process (e.g., methods, time since divergence, effective population size, genome size, sample size, etc.). We found 16,788 diagnostic SNPs out of a total of 77,773 with an overall F_{ST} = 0.525, using 84 samples (66 in WCT and 18 in RBT) and 100 bp reads. Work on European eel *Anguilla Anguilla* and American eel *A. rostrata*, with mean F_{ST} ~ 0.055–0.018 for microsatellite loci, identified 3,348 diagnostic loci (F_{ST} > 0.95) using the European eel draft genome and 25 individuals per species with RAD 100 bp sequencing (Mank and Avise, 2003; Wirth and Bernatchez, 2003; Pujolar et al., 2014). In two hybridizing tree species (*Populus alba* and *P. tremula*), with F_{ST} ~ 0.634 between species for SNP loci, a total of 5,079 fixed SNPs were found using the *Populus* reference genome, 7 samples from each species, and 70bp RAD sequencing reads (Stölting et al., 2013). Finally, in hybridizing Florida bass *Micropterus floridanus* and largemouth bass *M. salmoides*, with F_{ST} ~ 0.117 for microsatellite loci, identified 3,674 putatively diagnostic SNPs using 60 samples (20 of each species and 20 F1 hybrids) to conduct RNA sequencing and *de novo* transcriptome assembly with 100 bp single-end reads (Barthel et al. 2010; Li et al., 2014). These studies all demonstrate the power of genomic techniques to efficiently discover thousands of species-diagnostic SNPs in non-model taxa, allowing for the direct study of genomic patterns of introgressive hybridization.

3.2 Genomic studies of introgression

Ample coverage over large genomic regions is crucial for early detection of introgression when levels of hybridization are low and hard to detect with a limited number of microsatellite loci. The updated set of mapped SNPs discovered here will be used to develop a species-diagnostic SNP chip for testing thousands of individuals across the Flathead River drainage, and to more accurately detect expansion of RBT introgression and inform WCT conservation strategies range-wide (Hitt et al., 2003; Boyer et al., 2008; Muhlfeld et al., 2009; Muhlfeld et al., 2014). This panel will also be used to assess the genetic status of individual fish, improving our ability to detect even small amounts of nonnative genetic admixture (e.g., super-invasive alleles).

In this study, we used samples from three non-hybridized WCT populations in one drainage as well as a hatchery strain developed from a wide range of WCT populations. We used a smaller number of RBT individuals (though more than in Hohenlohe et al. 2011). Because introduced RBT come mainly from a few hat-

chery sources, we expected that a smaller sample size would be sufficient to capture variation in RBT ancestral polymorphic loci. However, this approach runs the risk of missing ancestral polymorphism in RBT, which could lead to underestimates of true admixture proportion in hybridized populations if putative diagnostic loci scored as pure WCT actually represent introgression from RBT. This did not appear to be the case, however, as our overall admixture estimates closely matched previous microsatellite-based estimates and were consistently higher than a previous diagnostic SNP panel (Fig. 4). The increased estimates compared to the previous SNP set may be explained by higher alignment success in RBT individuals to the RBT genome, although this difference was slight (see above). Instead, the agreement of our current estimates with the microsatellite estimates suggests the possibility that our current set of diagnostic SNPs has removed some bias toward WCT that was present in the previous SNP panel because of alignment to WCT-based RAD contigs (Hohenlohe et al., 2013).

When the source of invasive populations is not well known, a sample size of at least 30 individuals would likely allow for a high probability of detecting even rare alleles (> 95% of detecting an allele with a minor allele frequency of 0.05, or 99% with MAF = 0.1) that represent ancestral polymorphism (Allendorf et al., 2013). The degree of shared ancestral polymorphism is affected by genetic connectivity, local effective population sizes and genetic drift, so these geographic and demographic factors should also be considered in sampling design for diagnostic markers. Increasing the number of diagnostic markers in a comprehensive SNP panel may reduce the variance in admixture estimates caused by undetected ancestral polymorphism when sample sizes are low.

These diagnostic SNPs will help identify mechanisms influencing the spread of adaptive introgression into non-hybridized WCT populations. Adaptive introgression has important evolutionary consequences, but has been studied in detail only in a handful of systems, and will be greatly aided by the dense, genome-wide marker coverage we report here (Hedrick, 2013). Greater genomic coverage and density of markers is essential for identifying genomic regions under natural selection in admixed populations of RBT and WCT (Hohenlohe et al., 2013) and other cases of invasive hybridization (Fitzpatrick et al., 2010). Genome-typing a large panel of diagnostic loci across hybridized populations or through time can also help improve understanding of how environmental variation or change influences rates of introgression at neutral and adaptive loci (e.g., Fitzpatrick et al., 2010; Muhlfeld et al., 2014). This study illustrates the major leap forward that NGS allows in generating dense, genome-wide SNP coverage crucial to studying hybridization, one of the most serious threats to global biodiversity.

Acknowledgements Thanks to Morten Limborg for a helpful review of the manuscript. This work was supported by NSF (DEB-1258203), Bonneville Power Administration (199101903), and used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH (S10 Instrumentation Grants S10RR029668 and S10RR027303). BKH was supported in part by funding from the Department of the Interior Northwest Climate Science Center. BKH and GL were supported by NASA grant number NNX14AB84G. TDH and PAH received support for data analysis from NIH grant P30GM103324. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Allendorf FW, Hohenlohe PA, Luikart G, 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 11: 697–709.

Allendorf FW, Leary RF, Hitt NP, Knudsen KL, Boyer MC et al., 2005. Cutthroat trout hybridization and the U.S. Endangered Species Act: One species, two policies. Conservation Biology 19: 1326–1328.

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, Luikart G, Aitken SN, 2013. Conservation and the Genetics of Populations. 2nd edn. London, England: Wiley-Blackwell.

Amish SJ, Hohenlohe PA, Painter S, Leary RF, Muhlfeld C et al., 2012. RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays. Molecular Ecology Resources 12: 653–60.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL et al., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3 (10): e3376.

Barthel BL, Lutz-Carrillo DJ, Norberg KE, Porak WF, Tringali MD et al., 2010. Genetic relationships among populations of Florida bass. Transactions of the American Fisheries Society 139: 1615–1641.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M et al., 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nature Communications 5: 3657.

Boyer MC, Muhlfeld CC, Allendorf FW, 2008. Rainbow trout *Oncorhynchus mykiss* invasion and the spread of hybridization with native westslope cutthroat trout *Oncorhynchus clarkii lewisi*. Canadian Journal of Fisheries and Aquatic Sciences 65: 658–669.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA, 2013. Stacks: An analysis tool set for population genomics. Molecular Ecology 22: 3124–40.

Corsi MP, Eby LA, Barfoot CA, 2013. Hybridization with rainbow trout alters life history traits of native westslope cutthroat trout. Canadian Journal of Fisheries and Aquatic Sciences 70: 895–904

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

Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR et al., 2010. Rapid spread of invasive genes into a threatened native species. Proceedings of the National Academy of Sciences of the United States of America 107: 3606–3610.

Halverson A, 2010. An Entirely Synthetic Fish: How Rainbow Trout Beguiled America and Overran the World. New Haven: Yale University Press.

Hedrick PW, 2013. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology 22: 4606–4618.

Hitt NP, Frissell CA, Muhlfeld CC, Allendorf FW, 2003. Spread of hybridization between native westslope cutthroat trout *Oncorhynchus clarki lewisi* and nonnative rainbow trout *Oncorhynchus mykiss*. Canadian Journal of Fisheries and Aquatic Sciences 60: 1440–1451.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G, 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Molecular Ecology Resources 11: 117–122.

Hohenlohe PA, Day MD, Amish SJ, Miller MR, Kamps-Hughes N et al., 2013. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. Molecular Ecology 22: 3002–3013.

Kelly BP, Whiteley A, Tallmon D, 2010. The Arctic melting pot. Nature 468: 891.

Kovach RP, Muhlfeld CC, Boyer MC, Lowe WH, Allendorf FW et al., 2015. Dispersal and selection mediate hybridization between a native and invasive species. Proceedings of the Royal Society of London B 282: 20142454.

Langmead B, Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9: 357–359.

Li C, Gowan S, Anil A, Beck BH, Thongda W et al., 2014. Discovery and validation of gene-linked diagnostic SNP markers for assessing hybridization between Largemouth bass *Micropterus salmoides* and Florida bass *M. floridanus*. Molecular Ecology Resources. doi: 10.1111/1755–0998.12308.

Lowe S, Browne M, Boudjelas S, De Poorter M, 2000. 100 of the world's worst invasive alien species IUCN. SSC Invasive Species Specialist: 12.

Mank JE, Avise JC, 2003. Microsatellite variation and differentiation in North Atlantic eels. Journal of Heredity 94: 30–34.

Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad CE et al., 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Molecular Ecology 21: 237–249.

Muhlfeld CC, Kalinowski ST, Mcmahon TE, Taper ML, Painter S et al., 2009. Hybridization rapidly reduces fitness of a native trout in the wild. Biology Letters 5: 328–331.

Muhlfeld CC, Kovach RP, Jones LA, Al-Chokhachy R, Boyer MC et al., 2014. Invasive hybridization in a threatened species is accelerated by climate change. Nature Climate Change 4: 620–624.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA, 2013. Genotyping-by-sequencing in ecological and conservation genomics. Molecular Ecology 22: 2841–2847.

Pujolar JM, Jacobsen MW, Als TD, Frydenberg J, Magnussen E et al., 2014. Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. Heredity 112: 627–37.

Rhymer JM, Simberloff D, 1996. Extinction by hybridization and introgression. Annual Reivew of Ecology, Evolution and Systematics 27: 83–109.

Rieseberg L, 2011. Adaptive introgression: The seeds of resistance. Current Biology 21: R582.

Seeb JE, Pascal CE, Grau ED, Seeb LW, Templin WD et al., 2011. Transcriptome sequencing and high-resolution melt analysis advance single nucleotide polymorphism discovery in duplicated salmonids. Molecular Ecology Resources 11: 335–48.

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.

Stölting KN, Nipper R, Lindtke D, Caseys C, Waeber S et al., 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Molecular Ecology 22: 842–55.

Trotter P, 2008. Cutthroat: Native Trout of the West. Berkley: University of California Press.

Twyford AD, Ennos RA, 2012. Next-generation hybridization and introgression. Heredity 108: 179–89.

USFWS, 2003. Endangered and threatened wildlife and plants: Reconsidered finding for an amended petition to list the westslope cutthroat trout as threatened throughout its range. Federal Register 68. 46989–47009.

Wirth T, Bernatchez L, 2003. Decline of Atlantic eels: A fatal synergy? Proceedings of the Royal Society of London B 270: 681–688.