

Fishing for SNPs: A Targeted Locus Approach for Single Nucleotide Polymorphism Discovery in Rainbow Trout

A. E. SPROWLES, M. R. STEPHENS, N. W. CLIPPERTON, AND B. P. MAY*

*Genomic Variation Laboratory, University of California at Davis,
1 Shields Ave, Davis, California 95616, USA*

Abstract.—The combination of whole-genome sequencing efforts and emerging high-throughput genotyping techniques has made single nucleotide polymorphisms (SNPs) a marker of choice for molecular genetic analyses in model organisms. This class of marker holds great promise for resolving questions of phylogeny, population structure, introgression, and adaptive genetic variation. Fifty-five polymerase chain reaction primer pairs were used to target variable regions of the rainbow trout *Oncorhynchus mykiss* genome, 48 of which were designed from information found in publicly available DNA sequence databases. Forty of these primer pairs yielded sequenceable products. These sequences were compared across 1–10 individual fish from each of the following representative subspecies and strains: Sacramento redband trout *O. mykiss stonei*, California golden trout *O. mykiss aguabonita*, Little Kern golden trout *O. mykiss whitei*, coastal rainbow trout *O. mykiss irideus*, and the Mount Shasta hatchery strain. A total of 208 SNPs were identified in 37 loci, and a range of 75–128 SNPs were observed in pairwise comparisons of any two representative trout groups. As a test of high-throughput genotyping, the TaqMan 5′ nuclease assay was used to genotype 335 fish representing 14 populations at SNP *LDH-156**, enabling us to characterize allelic frequencies in larger sample sizes and additional populations of each subspecies.

Single nucleotide polymorphisms (SNPs) are rapidly emerging as the preferred genetic marker for molecular genetic analyses for a variety of reasons. Strictly defined, a SNP is a single base substitution or deletion within a sequence of DNA occurring with a population frequency of greater than 1% (Vogel and Motulsky 1979), though broader definitions now incorporate both single base nucleotide substitutions and short deletion and insertion polymorphisms (see, e.g., the dbSNP overview at <http://www.ncbi.nlm.nih.gov/projects/SNP/>). Such polymorphisms represent approximately 90% of the genetic variation in the human genome (Collins et al. 1998) and frequencies as high as 1 in 52 base pairs have been documented in drosophila (Villablanca et al. 1998). They are inherited in Mendelian fashion, are codominant, are usually biallelic, and are less susceptible to homoplasy than microsatellites (Vignal et al. 2002). Because allelic differences involve a single nucleotide, researchers can obtain information from degraded DNA that could not be analyzed with other kinds of molecular markers dependent on long stretches of intact DNA. The growing establishment of high-throughput genotyping methods such as the TaqMan fluorogenic 5′-nuclease assay and other allele-specific polymerase chain reactions (PCRs) or hybridization assays, single-base

extension-based assays, array-based assays, mass spectrometry-based systems, bead-based genotyping, Pyrosequencing, PCR-free methods, and others (reviewed in Tsuchihashi and Dracopoli 2002) has enabled rapid data collection at a relatively inexpensive cost. Such methods and data are easily standardized across studies, over time, and among laboratories.

Difficulty in obtaining species-specific DNA sequence information has precluded SNP discovery in nonmodel organisms. To overcome this hurdle, researchers have utilized the sequence homology of highly conserved DNA sequences to facilitate amplification of variable regions in related taxa. Exon-flanked intronic sequences have historically been the target for polymorphism discovery in nonmodel organisms (Palumbi and Baker 1994; Lyons et al. 1997; Bierne et al. 2000; Primmer et al. 2002; Zhang and Hewitt 2003; Aitken et al. 2004; Seddon et al. 2005). Other highly variable nuclear sequences include the 3′ untranslated region of the messenger RNA (3′ UTR) and extragenic DNA (Moran et al. 1997; Brumfield et al. 2003; Zhang and Hewitt 2003). However, because these noncoding regions are less homologous between species, it is difficult to screen these loci for variation without species-specific DNA sequence.

Genomic sequencing efforts have provided vast quantities of data that can increase the specificity of PCR primer design for an organism with little DNA sequence information. Whole-genome sequencing and assembly efforts are currently underway for 11 fish

* Corresponding author: bpmay@ucdavis.edu

Received November 18, 2005; accepted May 18, 2006
Published online November 30, 2006

TABLE 1.—Sampling locations and subspecies codes for rainbow trout populations used in the current study. The six populations used to represent subspecies during single nucleotide polymorphism (SNP) discovery are indicated with asterisks; additional localities were included in the assay for SNP LDH-156.

Rainbow trout subspecies	Sample location/strain	Subspecies code	Geographic coordinates or hatchery
McCloud River redband trout <i>O. mykiss ssp.</i> ^{ab}	Sheepheaven Creek*	SHP	41.3283°N, 121.8254°W
Sacramento redband trout <i>O. mykiss stonei</i> ^b	Davis Creek*	SAC	40.8311°N, 121.1198°W
	Edson Creek	SAC	41.3352°N, 121.1199°W
California golden trout <i>O. mykiss aguabonita</i>	Volcano Creek*	CAGT	36.3584°N, 118.3279°W
	South Fork Kern River above Templeton Barrier	CAGT	36.3347°N, 118.1911°W
	Golden Trout Creek at Mouth of Barigan Stringer	CAGT	36.4116°N, 118.2774°W
Little Kern golden trout <i>O. mykiss whitei</i> ^b	Upper Soda Spring Creek*	LKGT	36.3188°N, 118.5783°W
Eagle Lake rainbow trout <i>O. mykiss aquilarum</i> ^b	Eagle Lake ^c	ELRT	Darrah Springs Hatchery
Coastal rainbow trout <i>O. mykiss irideus</i>	North Fork American River*	RT-wild	39.2024°N, 120.6174°W
Hatchery rainbow trout <i>O. mykiss ssp.</i>	Mt. Shasta strain*	RT-hat	Mt. Shasta Hatchery
	Pitt strain	RT-hat	Crystal Lake Hatchery
	Coleman strain	RT-hat	Crystal Lake Hatchery
	Hot Creek strain	RT-hat	Hot Creek Hatchery
	Mt. Whitney strain	RT-hat	Mt. Whitney Hatchery

^a Sheepheaven Creek redband trout are currently considered members of the Sacramento redband trout subspecies but exhibit a morphological and genetic distinctiveness that warrant separate consideration and possibly unique subspecific status.

^b Subspecies designations follow Behnke (1992) but are not formally accepted taxonomy.

^c Fish collected from the wild but reared in a hatchery.

species zebra danio *Danio rerio* (also known as zebrafish), freshwater pufferfish *Tetraodon nigroviridis*, torafugu *Takifugu rubripes*, sea lamprey *Petromyzon marinus*, false killer whale *Pseudorca crassidens*, threespine stickleback *Gasterosteus aculeatus* (International Sequencing Consortium, available at: <http://www.intlgenome.org/viewDatabase.cfm>, October 2006), Japanese medaka *Oryzias latipes* (also known as ricefish) (<http://www.ensembl.org/index.html>, October 2006), Nile tilapia *Oreochromis niloticus*, and cichlid species *Haplochromis burtoni*, *Haplochromis chilotes*, and *Maylandia zebra* (T. Kocher, University of New Hampshire, Durham, personal communication). Nuclear genomic DNA is the focus of large-scale sequencing in additional fish species: channel catfish *Ictalurus punctatus*, little skate *Leucoraja erinacea*, murudai *Pagrus auriga*, spiny dogfish *Squalus acanthias*, Atlantic salmon *Salmo salar*, nurse shark *Ginglymostoma cirratum*, and southern puffer *Sphoeroides nephelus* (<http://www.intlgenome.org/viewDatabase.cfm>, October 2006). A number of smaller-scale sequencing efforts of genomic DNA, complementary DNA (cDNA), and microsatellite libraries exist for other fish species, most of which are of interest to the aquaculture industry (e.g., Benson et al. 2005 examined Atlantic salmon and rainbow trout *Oncorhynchus mykiss*). Such information has facilitated SNP discovery in various salmon subspecies (Smith et al. 2005a, 2005b, 2005c). Finally, gene cloning and genome mapping efforts have contributed smaller amounts of sequence information for a variety of species. All of this information can easily be accessed

from public genomic databases provided by the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov>; Benson et al. 2005), European Molecular Biology Laboratories-Bank (EMBL-Bank; <http://www.ebi.ac.uk/embl/>; Kanz et al. 2005), the DNA Databank of Japan (DDBJ; <http://www.ddbj.nig.ac.jp/>; Tateno et al. 2005) and the Institute for Genomic Research (TIGR; <http://www.tigr.org/>; Lee et al. 2005).

We were interested in developing a panel of SNP markers that could be used for the detection and quantification of introgressive hybridization between introduced rainbow trout and the native rainbow trout subspecies or strains listed in Table 1. McCloud River redband trout, California golden trout, and Little Kern golden trout have been classified as California State Species of Special Concern. The Little Kern golden trout has been designated as threatened under the U.S. Endangered Species Act (USFWS 1978). Single nucleotide polymorphisms that demonstrate differences in allelic frequency between these subspecies and introduced hatchery strains of rainbow trout have the potential to detect and quantify introgression in native populations and improve species management. Additionally, they could be applied to questions of phylogeny and population genetic structure. Our strategy was to identify DNA sequences within the public databases closely related to rainbow trout to facilitate a targeted locus approach to SNP discovery. The PCR primers were designed to amplify introns, 3' UTRs, and extragenic DNA within the rainbow trout genome. Primer pairs that successfully amplified

TABLE 2.—Target loci, primer design notes, and amplification conditions for the PCR primers used in this study. Target source information includes references for published and unpublished sequences and the original species or group for which the primers were used. Primer design species codes are as follows: CH = chum salmon *Oncorhynchus keta*, CS = Chinook salmon *O. tshawytscha*, MA = mammal consensus, MM = marbled murrelet *Brachyramphus marmoratus*, RT = rainbow trout *O. mykiss*, SA = salmon *Salmo* sp., TE = teleost fishes, WH = whales, ZE = zebrafish *Danio rerio*. The primer design columns provide information on primer redesign, including accession numbers (all of which are from the National Center for Biotechnology Information's GenBank database except for those designated as "TIGR," which are derived from the Institute for Genomic Research's Rainbow Trout Gene Index), sequence source type, and species identity of the sequence source used in the redesign. Sequence source type codes are as follows: EST = expressed sequence tag, GED = genomic extragenic DNA, GT = genomic sequences required for gene transcription (any combination of enhancer regions, 5' untranslated regions, exons, introns, and 3' untranslated regions), mRNA = messenger RNA, and STS = sequence tag site. The redesigned rainbow trout primers and conditions consist of the Genomic Variation Laboratory's (GVL) redesigned primer name, primer sequence, annealing temperature, and MgCl₂ concentration used in PCR amplification; asterisks indicate inability to amplify.

Target locus	Target source information		Primer design
	Reference	Original primer design species	Accession numbers for primer design
Actin alpha 1 intron	Palumbi and Baker (1994)	WH	CB489719 AF330142
Alcohol dehydrogenase 2 intron	Lyons et al. (1997)	MA	BX862123 CN331864
Alpha-tropomyosin intron	Friesen et al. (1999)	MM	AF180892 TC23655 (TIGR)
BAC clone sequence B1	R. B. Phillips, Washington State University, Vancouver, unpublished data	RT	
BAC clone sequence B4	R. B. Phillips	RT	
BAC clone sequence B9	R. B. Phillips	RT	
BAC clone sequence E1	R. B. Phillips	RT	
BAC clone sequence F5	R. B. Phillips	RT	
BAC clone sequence F9	R. B. Phillips	RT	
BAC clone sequence G9	R. B. Phillips	RT	
Beta 2 adrenergic receptor intron	Lyons et al. (1997)	MA	CR370337 CK879291 CR371409 CA036944 BV078109
Biglycan intron	Lyons et al. (1997)	MA	
BV482 intergenic sequence	Gahr et al., USDA Agrigultural Research Service, Kearneysville, West Virginia, unpublished data	RT	
c-Myc 3' UTR	Panno and McKeown (1995)	RT	S79770
Carbonyl reductase intron I	Guan et al. (1999)	RT	AF100933
Carbonyl reductase intron II	Guan et al. (1999)	RT	AF100933
Carbonyl reductase intron III	Guan et al. (1999)	RT	AF100933
Cathepsin D 3' UTR	Moran (2002)	RT	U90321
Chitinase 3' UTR	T. Wang, University of Aberdeen, UK, unpublished data	RT	AJ535688
Cholinergic receptor, nicotine, alpha1 intron	Lyons et al. (1997)	MA	CB510499 BX004803.11
Corticotropin releasing factor binding protein 3' UTR	C. Doyon et al., University of Ottawa, Ontario, unpublished data	RT	AY363677
Ependymin intron	Muller-Schmid et al. (1992)	RT	M93697
Estrogen receptor intron	Pakdel et al. (2000)	RT	AJ242740
Fibrinogen gamma polypeptide intron	Lyons et al. (1997)	MA	CA352559 BX510913.9
Forkhead box protein 3' UTR	Hidaka et al. (2004)	RT	AB164480
Glucose-6 phosphate dehydrogenase intron	Lyons et al. (1997)	MA	CA368573 BX318883

TABLE 2.— Extended.

Primer design		Redesigned rainbow trout primers and conditions			
Sequence source type	Primer design species	GVL primer name	GVL primer sequence (5' to 3')	Annealing temperature (°C)	MgCl ₂ (mM)
EST	RT	ACTCf	GCCCTGGACTTTGAGAATGAGAT	52	1.5
		ACTCr	TCAGCAATACCAGGGTACA		
EST	RT/ZE	ADH2f	CAGGTGATAAAGTCATCCC	45	1.5
		ADH2r	TGTACTCGGAGAAGGTGCTG		
EST	ZE/RT	a-Tropf	GAGTTGGATCGCGCTCAGGAGCG	58	3.0
		a-Tropr	TCAGCCTCCTCAGCGATGTGCTT		
		BAC B1f	GAATCTGGCATTCTAAGAGA	58	1.5
		BAC B1r	TGCGACGTCCAGTAATGAGAA		
		BAC B4f	GTTTGGGTAAGAACATCGGAAACA	58	1.5
		BAC B4r	GGAGAAACAATGCCCTATTACAC		
		BAC B9f	TGAATGAGTTGCAAAAATGACA	58	1.5
		BAC B9r	TCCAAAACAGAGACTAAGGTGACA		
		BAC E1f	AAACACAGGTTGGCAACTGAT	58	1.5
		BAC E1r	GAGAGAACCCAATAGGCTC		
		BAC F5f	ACACCAGGCTGTTTGATATTAGA	58	1.5
		BAC F5r	GATATCCAGCTGAGGAGAGAAAAG		
		BAC F9f	GCAGACATCCGTGGGCTACCACAG	*	*
		BAC F9r	GTTCTCACAAGTGGACTGCGATGG		
		BAC G9f	GAGCTGGGAGAGGTGACATTAAGT	58	1.5
		BAC G9r	ACGTGCGTTCTATTGTTGTACTG		
EST	RT/SA	ADRB2f	TTTAACTTCCAGAGCCTGCTGACA	*	*
		ADRB2r	TCAGCCAGCCATTTACCAACA		
EST	RT/SA	BGNf	ACCTGGTGACTCAATCGTTC	*	*
		BGNr	TTATTCTTCAAAGCCACTGTTCT		
STS	RT	BV482f	GAGGGGTGTCAGGGTGTGGAG	62	1.5
		BV482r	AGGCTATGTGGCTGTGGCTTTACG		
mRNA	RT	mycf1216	CCGGAGGTGGCTAACAAATG	51	1.5
		mycr1682	TGCCCTGAACCTCCTGACAAAT		
GT	RT	CRB0826	TTTACCGGGGATGTTATTCTTACT	51	1.5
		CRB1883	CACCTCTGGCATTGGTCTGA		
GT	RT	CRB2358	TGATGGGGCAGTTTGTATTGG	50	1.5
		CRB2581	TGCCCGGGTCTTAGTCAG		
GT	RT	CRB2677	GACTCCTACCTATCTGGCACTT	50	1.5
		CRB2995	TTTTATTCTTTACATCATTTCAAC		
mRNA	RT	CTSDf1203	GACAACAACAGGGTGGG	51	1.5
		CTSDr1776	GACAAGAGGTCCATTGC		
mRNA	RT	CHITUTRf	ACCTCGGCTCGTCTCAAATA	60	1.5
		CHITUTRr	GTTGCCCAATACTGCCTAAATG		
EST	SA/ZE	CHRNA1f	TGCTACCAGTTGGGGGAAGA	*	*
		CHRNA1r	GTCCATCACTGTTTTTAT		
mRNA	RT	CRFBPUTRf	CATACCGGCTGCCAAATCTACTA	60	1.5
		CRFBPUTRr	CTCGCTCCCATCCACTCC		
mRNA	RT	EPD1f101	CCGTCGCTGCCCTCTCC	*	*
		EPD1r762	GCCACTGTCTTAGACAC		
mRNA	RT	ERf839	GTCACAATGACTACATGTG	*	*
		ERr1306	CTGGTACTTTCTTAGCCC		
EST	RT/ZE	FGGf	CAGCCCATGAAGACAGACTC	45	2.0
		FGGr	TATTTGTCTTCCCTTATTCC		
mRNA	RT	FOXUTRf	ACGCCAGAGAACCCATAAACGAA	60	1.5
		FOXUTRr	CGAGTAGCCTAACCCGGATAAATG		
EST	RT	G6PDr	CTGGCAAGGAGATGGTG	55	3.0
		G6PDF	TCACCCAGCAACAAGGAG		

TABLE 2.—Continued.

Target locus	Target source information		Primer design
	Reference	Original primer design species	Accession numbers for primer design
Glucosidase, beta; acid intron	Lyons et al. (1997)	MA	BI883119 CR375502
Glyceraldehyde-3-phosphate dehydrogenase intron	Hassan et al. (2002)	TE	AF027130 AF027130
Gonadotropin hormone beta subunit intron	Moran (1997)	SA	CR375538 BX912330
Growth hormone 1, intron D	Docker and Heath (2002)	CS	CX719599 CX719599
Growth hormone 2, intron C	Docker and Heath (2002)	RT	503747
Heat shock protein 70 3' UTR	Ojima et al. (2005)	RT	AB176854
Heat shock protein 71 intron 1	Kothary et al. (1984)	RT	S85730
Heat shock protein 71 intron 5	Kothary et al. (1984)	RT	S85730
HoxD intron	Lyons et al. (1997)	MA	CN022672 BX885496
Immunoglobulin H intron	Lee et al. (1993)	SA	X67713
Inhibitor of DNA binding/differentiation 1 3' UTR	Moran (2002)	RT	Y08368
Inhibitor of DNA binding differentiation 1C 3' UTR	Gahr et al. (2004)	RT	AY325276
Lactate dehydrogenase B intron	Friesen et al. (1999)	MM	BX862830 TC9994 (TIGR)
Lysozyme intron II	Mitra et al. (2003)	RT	AF452171
Lysozyme intron III	Mitra et al. (2003)	RT	AF452171
Menkes disease syndrome, ATP7A intron	Lyons et al. (1997)	MA	BX0074784.8 BX913443
Major histocompatibility complex class I variant, <i>OmyUAA*101</i> intron	Shum et al. (1999)	RT	AF091779
Myelin proteolipid protein intron	Friesen et al. (1999)	MM	BX890326 BX878994
Myeloperoxidase intron	Lyons et al. (1997)	MA	CK990624 BX884906
Novel immune-type receptor 2 intron	Yoder et al. (2002)	RT	AY082616
<i>OmyP9-B</i> extragenic sequence	Iturra et al. (1998)	RT	AF323613
Recombination activating gene 3' UTR 0331–864	Hansen (1997)	RT	U73750
Recombination activating gene 3' UTR 0917–1503	Hansen (1997)	RT	U73750
Recombination activating gene 3' UTR 1175–1760	Hansen (1997)	RT	U73750
Recombination activating gene 3' UTR 1564–1957	Hansen (1997)	RT	U73750
RAPD (OPC-02) marker sequence	L.G. Jin et al., Pukyong National University, Korea, unpublished data	RT	AF243431
Somatolactin intron	Ford (2000)	CS	D10638
Tc1-like extragenic element	Radice et al. (1994)	RT	L12209
Urotensin I 3' UTR	D.A. Lovejoy et al., University of Manchester, UK, unpublished data	RT	AJ005264

hatchery rainbow trout DNA were used to generate rainbow trout subspecies-specific DNA sequences. The size and sequence of the PCR products were compared and evaluated for the presence of SNPs.

One of the attractions of using SNPs is the ability to perform high-throughput genotyping without post-PCR processing. We genotyped several hundred samples for one of the SNPs discovered in this study by using the

TABLE 2.—Extended.

Primer design		Redesigned rainbow trout primers and conditions			
Sequence source type	Primer design species	GVL primer name	GVL primer sequence (5' to 3')	Annealing temperature (°C)	MgCl ₂ (mM)
EST	RT/ZE	GBAf	GCTTCAATAACTGCAGGTGTAGAT	45	2.0
mRNA	RT	GBAr	TAGGCCATACGCGATAGGTC	55	3.0
		GPD2f	GCCATCAACGATCCCCTTCATC		
EST	RT	GPD2r	CTCACCTCACCTTGTAAACGG	*	*
		GTH2Bf520	CATTGTCTAGGACAGAGTCA		
EST	RT	GTH2Br394	TAGTACTCTTTGGCGTAATGC	62	1.5
		GH1Df	CTCCGAAACATCCCCAGGTG		
GT	RT	GH1Dr	ACGCAGGTATGTGTTTGTGTCTCA	58	4.0
		GH2Cff	ATCGTGAGCCCAATCGACAAGCAG		
mRNA	RT	GH2Cr	GGGTACTCCCAAGGATTCAATCAGG	60	1.5
		HSP70UTRf	CAGCTCCAAGGCCCAACCATT		
GT	RT	HSP70UTRr	ACAACCCAACCAAGCCGCACTG	60	1.5
		HSPINT1f	ACTGCGTTTGTGTTGATTCTT		
GT	RT	HSPINT1r	TGTGTTGCAGGGGTTTCAT	60	1.5
		HSPINT5f	TGTCCAGGCAGCCATACTCTC		
EST	RT/ZE	HSPINT5r	TGGCCCTCTCACCTCATACA	50	2.0
		HOXDf	TCACGAACAATCACGGGAAACT		
GT	RT	HOXDr	TCCACTTCATCCTTCGGTTCT	50	1.5
		IgHf	ACCGTTTGGACTACTGG		
GT	RT	IgHr	CCCGCCTTCGTCATTCC	58	1.5
		Id1Cf	ATCATTTCCATTCCACATCA		
mRNA	RT	Id1CR	ACCCCTGCCCTGCCGAAATAGAAA	47	2.0
		Id1f448	GTGGAGAACGGATGCTC		
EST	RT	Id1r1002	AAACACCAGAAGTACATTG	55	3.0
		LDBf	TCAAACGTAAAGGGGAGATGATGGA		
GT	RT	LDBr	TTCCTCTGGACCAGGTTGAGCCTGCTCTC	58	1.5
		LYSIIAf	CCCGGGGGTAAGAACG		
GT	RT	LYSIIAr	ATGCCTGAAAAATATGAAGAGTGGT	58	1.5
		LYSIII f	AGGCCTACTTCTCTATTCTAAACC		
EST	ZE/RT	LYSIII r	GCGCCAAGCCACCTAAA	*	*
		MNKf	TCACGATGACTTGTAAATTCTG		
GT	RT	MNKr	TCAGGCAGGAAGGCATCA	*	*
		OmyUAAf	TCATTTCATTACGAATCTGGATGCATCC		
EST	RT	OmyUAAr	TAGAACCCTGTACATGACATGTGA	53	2
		MPPf	TTTAACACCTGGACCACGGAGCGCAGGAACCT		
EST	SA/RT	MPPr	TGACTGTGTGGAGAGCAGGTTGGAGCAGT	51	2.0
		MPOf	TCCACTTCATCCTTCGGTTCT		
GT	RT	MPOr	GGCTGATCTCCTCTCTATAAAGA	60	1.5
		NITR2f	GGAGGGCATGGACTACAAG		
GT	RT	NITR2r	TACAACCACTGTGAGAGCCTGATA	62	1.5
		OmyP9-B1f	GGCAGCAGGCTACCATAAACAT		
GT	RT	OmyP9-B1r	TCTACTGGCCTTACCTACTCAGC	53	2.5
		RAG336f	CACGAAAAGCCAGATGCCA		
GT	RT	RAG864r	GGCAGCCTGAAACGAAGACA	53	2.5
		RAG917f	GGGGCCTGGGTGACTTTGA		
GT	RT	RAG1503r	TACACCGCTGATGAGATTC	45	1.5
		RAG1175f	GAGCCTTGAGGGGAAGTAAAT		
GT	RT	RAG1700r	GAAACAATAAAGGCCAATAAAACC	45	2.0
		RAG1564f	TTGCCAAAAACACTCACTCA		
GED	RT	RAG1957r	TAGGCTAATACTGCTTGGACTG	60	1.5
		RAPDf	TTAAGCGGGGAAATGGAG		
GT	CH	RAPDr	TAAGGGCGTTGATAAAGGAGACTA	*	*
		SOMAf	ACAAGTGGCTCTCCACTCTGT		
GED	RT	SOMAr	GATAACCACAAACGTAAGTGTGCC	50	1.5
		TelF375	CATGGGCAAATCAAAGAAATCAG		
mRNA	RT	TelR961	CAGCAAAGCACCCACAACA	60	1.5
		UROUTRf	CATTGCCCGTAGGTTTCTGTGATA		
		UROUTRr	AGGGGTTTTGCAATGAGTGTTAG		

fluorogenic 5'-nuclease (TaqMan) assay. This screen was performed as a preliminary effort to investigate SNP frequencies in additional populations and to assess the utility of the 5'-nuclease assay for high-throughput genotyping.

Methods

Samples and DNA extractions.—Caudal fin clips of the following resident rainbow trout populations were obtained by California Department of Fish and Game personnel: redband trout at Sheepheaven, Edson, and

Davis creeks; California golden trout at Volcano Creek, South Fork Kern River above Templeton barrier, and Golden Trout Creek at Mouth of Barigan Stringer; Little Kern golden trout at the upper Soda Spring Creek; and Eagle Lake rainbow trout at Eagle Lake (Table 1). In addition, fin clips were collected from the Mt. Shasta, Mt. Whitney, Hot Creek, Pitt, and Coleman hatchery strains. Genomic DNA was extracted using the Promega Wizard 96-well extraction kit, quantified using PicoGreen double-stranded DNA Quantitation Reagent (Molecular Probes), and standardized to concentrations ranging from 5 to 10 ng/ μ L. A total of 77 different fish were utilized during SNP discovery, one to seven individuals being sampled per subspecies group.

PCR primer design.—We searched NCBI's Entrez database for rainbow trout DNA sequences. Over 200,000 sequences were recovered, only 1,700 of which were affiliated with a specific gene. All others reflect extragenic DNA, microsatellite loci, or expressed sequence tags (ESTs). Both nuclear genomic sequences and ESTs were utilized to design a total of 48 primer pairs that were able to amplify three different classes of DNA sequence from the rainbow trout genome: introns ($N = 31$ primer pairs), 3' UTRs ($N = 13$), and extragenic DNA ($N = 4$). Twenty-two of the primer pairs were redesigned from existing primers shown to amplify polymorphic regions in other species (Palumbi and Baker 1994; Lyons et al. 1997; Moran et al. 1997; Friesen et al. 1999; Ford 2000; Docker and Heath 2002; Hassan et al. 2002; Table 2). These published primer sequences were submitted into the GenBank BLAST algorithm (Altschul et al. 1997), and teleost matches were used as a template for changes to the initial primer sequence. The remaining primer sets were designed from sequences of rainbow trout previously demonstrated to have variation and rainbow trout sequences in the public databases with an unknown level of nucleotide variation (see Table 2 for references and NCBI accession numbers for all sequences).

The programs Primer Select (DNASTAR 1999) and Primer3 (version 1.0, Rozen and Skaletsky 2000) were used to design PCR primers from rainbow trout exon, intron, 3' UTR, extragenic, and EST sequences in GenBank. Of the 48 primer pairs, 37 were based on rainbow trout sequences for both forward and reverse primers, 9 contained one primer derived from rainbow trout and one from a species other than rainbow trout, and 2 had both primers derived from a species other than rainbow trout (Table 2). Seven additional primer sets designed to amplify extragenic regions derived from bacterial artificial chromosome (BAC) libraries and containing known nucleotide variation were

obtained and utilized (R. B. Phillips, Washington State University, Vancouver, personal communication).

PCR protocols and product purification.—The general reaction protocol for primer optimization was based on a 25- μ L reaction volume with an average of 25 ng Mount Shasta strain genomic DNA, 400 nM of each primer, 0.5 units of Promega *Taq* DNA polymerase, 2.5 μ L of Mg^{2+} -free Promega PCR buffer, and 1.5–2.5 mM of $MgCl_2$ per reaction, depending on the primer pair (Table 2). Reactions were run on DNA Engine Dyad thermocyclers (MJ Research) at 30-s annealing times and 1-min extension times for a total of 40 cycles. Initial annealing temperatures were chosen from the predicted melting temperatures of the primers. Possible optimization procedures included annealing temperature gradients, touchdown amplifications, increased template concentration (up to 100 ng of genomic DNA), and $MgCl_2$ titrations and pH titrations. Final conditions for each primer pair are reported in Table 2.

McCloud River redband trout from Sheepheaven Creek, Sacramento redband trout from Davis Creek, California golden trout from Volcano Creek, and Little Kern golden trout from upper Soda Spring Creek were selected as representative populations for each of our subspecies of interest, based on previous morphometric and genetic studies (Behnke 1992; Bagley and Gall 1998; Cordes et al. 2006) and historical stocking records showing no evidence of rainbow trout influence (C. McGuire and M. Dean, California Department of Fish and Game, personal communication). Upon successful optimization of primer conditions on Mt. Shasta Hatchery strain rainbow trout, one individual from each of these populations and one from the North Fork American River resident rainbow population were amplified. A fraction of the PCR products were resolved on a 5% polyacrylamide gel via electrophoresis and visualized on a FluorImager (Molecular Dynamics) using a SYBR green agarose overlay (Rodzen et al. 1998). Products consisting of a single amplicon were purified with the QIAquick PCR Purification Kit (QIAGEN).

Many of the PCR primer pairs amplified more than one product. In these cases, common optimization techniques such as Mg^{2+} , temperature, or DMSO titration were applied to eliminate nonspecific products. In cases where multiple bands remained, we assumed bands of the same size were generated from the same nuclear locus. For primers designed to amplify known rainbow trout sequences, the product closest to the expected size was used for sequence analysis. For primers designed to amplify product of unknown length, the most prominent band was chosen for purification. If possible, the band was cut from a

1% low melt agarose gel and purified with the QIAquick Gel Extraction Kit (QIAGEN). If this method was unable to resolve multiple fragments, the band was cut from a 5% acrylamide gel and incubated overnight in 200 μ L of H₂O. The eluted DNA was used as template for a second amplification of 25 cycles. The PCR products were purified with a QIAquick column, quantified, and a small fraction of each was run on a 5% acrylamide gel to confirm band purity. Primer pairs generating multiple bands that could not be separated by either agarose gel purification or excision from acrylamide and reamplification were removed from the screening process.

Sequencing DNA and SNP identification.—Single-direction sequencing of the purified PCR products was performed with either the forward or reverse PCR primer. Sequences were generated on an Applied Biosystems (ABI) 3730 Capillary Electrophoresis Genetic Analyzer with ABI BigDye Terminator Version 3.1 Cycle Sequencing chemistry at the University of California—Davis Division of Biological Science Sequencing Facility. Each sequence chromatogram was inspected visually via Sequencher software (version 4.5, Gene Codes Corp.) for sequence quality and accurate base pair determination. Sequences were submitted to the GenBank BLAST algorithm (Altschul et al. 1997) to verify that the target loci had been amplified by the new primers.

For each locus, sequences from each of the representative subspecies were aligned and multiple sequence alignments searched for nucleotide variants. Chromatograms were manually inspected to confirm polymorphisms at variable sites. In instances of weak or questionable sequence, amplicons were resequenced in the opposite direction for confirmation. Because of their diagnostic potential to identify populations or subspecies, we assigned special priority to single nucleotide polymorphisms occurring between two individuals, such that each represented subspecies was homozygous for a different allele. When such a polymorphism was discovered between a native population and the Mount Shasta strain, one to seven additional individuals from each population were amplified and sequenced. If all individuals in each population remained fixed for their respective nucleotide variant, the SNP was considered a candidate diagnostic between those populations.

TaqMan assay.—The 5'-nuclease TaqMan SNP Genotyping Assay (ABI) was chosen as our method of high-throughput genotyping based on its simple protocol and real time scoring. (Chen and Sullivan 2003). A 307 base-pair sequence from the lactose dehydrogenase (*LDH*) locus was submitted to ABI's Assays-By-Design service. A TaqMan assay was de-

signed to distinguish between two alleles (*C or *T) located at SNP *LDH-156*. The primer-probe pairs that were generated are as follows: forward primer, 5'-GTT TTGAAACCAGTTTAAAGGTTGATTGC-3'; reverse primer, 5'-ACGGCATAGTCTGGACAGAGAT-3'; VIC-labeled probe, 5'-CCATTTAGACGTTTTTTT-3'; FAM-labeled probe, 5'-CCATTTAGATGTTTTTTT-3'. Each probe bore a minor groove binder and nonfluorescent quencher on the 3' end. The reaction was optimized to individuals of known genotype (based on sequence data), including homozygotes, true heterozygotes, and composite heterozygotes (generated by combining DNA from known homozygotes in ratios of 3:1, 1:1, and 1:3). Each of these controls was included on every plate of samples analyzed.

The TaqMan assay was used to genotype 335 fish from 14 populations: Shasta, Pitt, Coleman, Hot Creek, and Mt. Whitney hatchery strains; coastal rainbow trout from the North Fork American River; Eagle Lake rainbow trout; McCloud River redband trout from Sheepheaven; and Sacramento redband trout from Edson and Davis creeks; California golden trout from Volcano Creek, the South Fork Kern River above Templeton barrier, and Golden Trout Creek at Mouth of Barigan Stringer; and Little Kern golden trout from upper Soda Spring Creek. Assays were performed using the Chromo4 Real-Time PCR Detector (MJ Research/Bio-Rad Laboratories, Inc.). Reactions were carried out in 96-well microplates at a 10- μ L volume. Omega reagents were used at the following concentrations: 20 units/mL *Taq* polymerase, 0.2 mM each dNTP, 5 mM of MgCl, 50 mM KCl, 10 mM tris-HCl, 0.1% Triton X-100, 540 nM each primer, 120 nM each probe, and 10–20 ng template DNA. The reaction protocol began with an initial denaturation of 94°C for 5 min, followed by 40 cycles of 92°C for 15 s, and 62°C for 1 min.

Results

PCR Primer Pairs

Of the 55 primer pairs attempted, 46 resulted in the amplification of rainbow trout DNA, of which 40 (73% of total designed) yielded distinct, robust amplicons suitable for sequencing. The proportion of sequenceable loci (number sequenced out of number attempted) varied by category: 92% (11/12) for 3' UTR loci, 73% (8/11) for extragenic loci, and 54% (13/24) for primers designed to amplify introns (Table 3; Figure 1). It is likely the lower rate of success for primers designed to amplify introns is reflective of our redesign strategy because many of these either failed to amplify rainbow trout DNA or amplicon sequence failed to match targeted loci (see "DNA Sequencing and SNP Discovery" section below). Many of these primer pairs

TABLE 3.—Analysis of PCR amplification for each locus and summary of nucleotide variation in the sequenced amplicon. Loci are organized alphabetically by category of sequence type. Sequences that did not show homology with the target design sequence are listed in the “unknown” category, with a new locus name identifier given in italics parenthetically.

Locus	PCR amplification		Nucleotide variation	
	Bands amplified	Base pairs sequenced	SNPs discovered	Base pairs/SNP
Introns				
Alpha-tropomyosin intron	3	392	0	
Beta 2 adrenergic receptor intron	0			
Biglycan intron	0			
Carbonyl reductase intron I	1	945	5	189
Carbonyl reductase intron II	4	163	2	82
Carbonyl reductase intron III	1	251	8	31
Cholinergic receptor, nicotine, alpha1 intron	0			
Ependymin intron	2	0		
Estrogen receptor intron	0			
Glyceraldehyde-3-phosphate dehydrogenase intron	4	218	0	
Gonadotropin hormone beta subunit intron	0			
Growth hormone 1, intron D	1	0		
Growth hormone 2, intron C	2	510	3	170
Heat shock protein 71 intron 1	1	224	0	
Heat shock protein 71 intron 5	1	151	3	50
Immunoglobulin H intron	2	878	26	34
Lactate dehydrogenase B intron	4	303	3	101
Lysozyme intron II, 1st half ^a	1	417	7	60
Lysozyme intron II 2nd half ^a	1	379	8	47
Lysozyme intron III	1	535	11	49
Menkes disease syndrome, ATP7a intron	0			
Major histocompatibility complex class I variant, <i>OmyUAA*101</i> intron	0			
Myelin proteolipid protein intron	1	0		
Novel immune-type receptor 2 intron	5	442	18	25
Somatolactin intron	0			
Total		5,808	94	62
3' untranslated regions				
c-Myc 3' UTR	1	406	4	102
Cathepsin D 3' UTR	2	521	17	31
Chitinase 3' UTR	1	239	1	239
Corticotropin releasing factor binding protein 3' UTR	1	347	1	347
Forkhead box protein 3' UTR	1	0		
Heat shock protein 70 3' UTR	1	370	1	370
Inhibitor of DNA binding/differentiation 1C 3' UTR	3	361	3	120
Recombination activating gene 3' UTR 0336–864	4+	503	2	252
Recombination activating gene 3' UTR 0917–1503	2	512	6	85
Recombination activating gene 3' UTR 1175–1700	1	465	3	155
Recombination activating gene 3' UTR 1564–1957 ^b	1	270	3	90
Urotensin I 3' UTR	1	424	3	141
Total		4,418	44	100
Extragenic sequences				
BAC clone sequence B1	1	382	4	96
BAC clone sequence B4	1	446	3	149
BAC clone sequence B9	1	443	11	40
BAC clone sequence E1	1	364	9	40
BAC clone sequence F5	1	361	3	120
BAC clone sequence F9	0			
BAC clone sequence G9	1	581	8	73
BV482 extragenic sequence	1	0		
<i>Omyp9b1F</i> genomic sequence	3	300	3	100
RAPD (OPC-02) marker sequence	5	334	3	111
Tc1-like element sequence	5+	0		
Total		3,211	44	73
Unknown sequences				
Actin alpha 1 (<i>Omy_a1</i>)	6	307	5	61
Alcohol dehydrogenase 2 (<i>Omy_a2</i>)	2	622	3	207
Fibrinogen gamma polypeptide (<i>Omy_f1</i>)	5 +	580	4	145
Glucose-6 phosphate dehydrogenase (<i>Omy_g1</i>)	3	263	1	263
Glucosidase, beta; acid (<i>Omy_g2</i>)	2	376	5	75
HoxD (<i>Omy_h1</i>)	2	340	2	170

TABLE 3.—Continued.

Locus	PCR amplification		Nucleotide variation	
	Bands amplified	Base pairs sequenced	SNPs discovered	Base pairs/SNP
Inhibitor of DNA binding/differentiation 1 (<i>Omy_il</i>)	3	322	5	64
Myeloperoxidase (<i>Omy_ml</i>)	10 +	277	1	277
Total		3,087	26	119
Grand total		16,524	208	79

^a Single locus assigned two separate accession numbers; forward and reverse sequences yielded insufficient data to confidently align at an 11–12 base-pair poly-T region.

^b Sequence truncated to account for overlap between recombination activating gene 3' UTR 1175–1700 and RAG1564–1957.

originated from sets previously reported for other species and were redesigned to increase their specificity to rainbow trout DNA. The majority of these primer pairs were redesigned with teleost EST sequences rather than teleost clones of the gene in which the polymorphic intron was originally described. The ESTs utilized as templates for primer redesign were mostly derived from rainbow trout; however, sequences from Atlantic salmon and zebrafish were also represented (Table 2). Most successful primer pairs required additional optimization efforts.

After optimization to produce consistently robust amplicons, the persistent presence of multiple products was an issue for many of the primer pairs (Table 3). In a few cases, these bands were eliminated through modifications in annealing temperature, salt concentration, and the addition of DMSO. For the majority, however, these standard optimizing techniques failed to eliminate multiple amplicons. The products from three representative primer pairs—lactate dehydrogenase B intron, alpha-tropomyosin intron, and glyceraldehyde-3 phosphate dehydrogenase (GAPDH)—were chosen for further evaluation to characterize the underlying DNA sequence of the additional bands. All bands generated from each of these primer sets were purified and sequenced (data not shown). Sequence data confirmed that the multiple products all contained regions of

homologous DNA sequences, further evidence for gene duplication or possibly a consequence of the amplification of paralogous copies of these genes (Lim et al. 1975). To increase the likelihood that sequences compared between individuals were all derived from the same targeted locus, one band of identical size was chosen from each individual for sequence analysis. The band was selected based on predicted amplicon size and robustness of the product.

Sequencing DNA and SNP Discovery

High-quality DNA sequence data were generated from 40 primer pairs (Table 3). When submitted in the NCBI BLAST algorithm, 32 of these sequences demonstrated homology with the locus to which the primers had been targeted. Seven of the eight sequences that did not demonstrate homology with the intended locus were from primer pairs that had been designed from two separate EST sequences. These eight sequences were reclassified as unknown sequence source type and renamed in Table 3.

Of the 40 loci, 37 demonstrated sequence variation between subspecies, a total of 208 SNPs being identified (Table 4). GenBank dbSTS for all sequences and dbSNP accession numbers for SNPs are provided in Table 4 with detailed information on discovery population numbers and SNP allele frequencies available through individual dbSNP submissions. On average, a SNP was discovered every 79 base pairs, with different, though not statistically significant (analysis of variance [ANOVA]; $F_{2, 26} = 2.39, P = 0.11$), rates of variation being found in each target category (the unknown category not included). By category, SNPs were found in 77% of sequenced intronic loci and 100% of 3' UTRs and extragenic DNA loci (Table 3). Additionally, all of the “unknown” sequences that did not demonstrate homology with the loci for which they were targeted also contained SNPs. The number of SNPs apparent in pairwise comparisons of subspecies ranged from 75 to

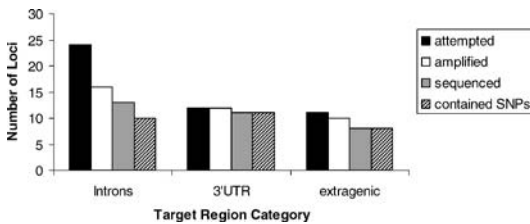


FIGURE 1.—Degree of success in the amplification and single nucleotide polymorphism (SNP) discovery in rainbow trout, by region category: intron, 3' untranslated region (3' UTR), and extragenic DNA; a fourth category, unknown (i.e., sequences not homologous to the targeted region), was not included.

TABLE 4.—Complete list of SNPs discovered among trout populations. The total number of individuals sequenced (*N*), National Center for Biotechnology Information (NCBI) accession numbers for DNA sequences (STS), and number of base pairs sequence per locus are given for each targeted region, along with individual SNP name and NCBI accession numbers (dbSNP). Locus name formats are as given in Table 3.

Targeted region	<i>N</i>	STS	Base pairs/locus	SNP	dbSNP
Actin alpha 1 (<i>Omy_al</i>)	8	BV678702	307	<i>Omy_al</i> 091	ss52084297
				<i>Omy_al</i> 105	ss52084298
				<i>Omy_al</i> 115	ss52084299
				<i>Omy_al</i> 134	ss52084300
				<i>Omy_al</i> 1156	ss52084301
Alcohol dehydrogenase 2 (<i>Omy_a2</i>)	4	BV677387	622	<i>Omy_a2</i> 221	ss46566284
				<i>Omy_a2</i> 249	ss46566285
				<i>Omy_a2</i> 269	ss46566286
BAC clone sequence B1	15	DQ465748	382	B1 039	ss52084314
				B1 102	ss52084315
				B1 257	ss52084316
				B1 266	ss52084317
				B4 170	ss52084326
BAC clone sequence B4	6	DQ465749	446	B4 193	ss52084327
				B4 198	ss52084328
BAC clone sequence B9	23	DQ465750	443	B9 016	ss52084303
				B9 037	ss52084304
				B9 090	ss52084305
				B9 164	ss52084306
				B9 181	ss52084307
				B9 182	ss52084308
				B9 202	ss52084309
				B9 330	ss52084310
				B9 335	ss52084311
				B9 355	ss52084312
				B9 388	ss52084313
BAC clone sequence E1	21	DQ465751	364	E1 029	ss52084329
				E1 030	ss52084330
				E1 100	ss52084331
				E1 147	ss52084332
				E1 159	ss52084333
				E1 239	ss52084334
				E1 241	ss52084335
				E1 289	ss52084336
				E1 300	ss52084337
				F5 090	ss52084338
				F5 306	ss52084339
BAC clone sequence F5	19	DQ465752	361	F5 323	ss52084340
				G9 125	ss52084318
				G9 168	ss52084319
BAC clone sequence G9	20	DQ465753	581	G9 221	ss52084320
				G9 465	ss52084321
				G9 502	ss52084322
				G9 504	ss52084323
				G9 520	ss52084324
				G9 547	ss52084325
				CRB 0826 102	ss38350754
				CRB 0826 110	ss38350755
CRB 0826 152	ss38350756				
Carbonyl reductase intron I	9	BV677367	945	CRB 0826 172	ss38350757
				CRB 0826 756	ss38350758
				CRB 2358 020	ss38350752
				CRB 2358 120	ss38350753
				CRB 2677 106	ss46566266
Carbonyl reductase intron II	27	BV677371	163	CRB 2677 117	ss46566267
				CRB 2677 126	ss46566268
				CRB 2677 129	ss46566269
				CRB 2677 166	ss46566270
				CRB 2677 168	ss46566271
				CRB 2677 203	ss46566272
				CRB 2677 226	ss46566273
				CTSD 033	ss46566519
				CTSD 097	ss46566520
				CTSD 152	ss46566521
				CTSD 158	ss46566522
				CTSD 173	ss46566523
				Cathepsin D 3' UTR	15

TABLE 4.—Continued.

Targeted region	N	STS	Base pairs/locus	SNP	dbSNP
				CTSD 174	ss46566524
				CTSD 182	ss46566525
				CTSD 183	ss46566526
				CTSD 187	ss46566527
				CTSD 193	ss46566528
				CTSD 226	ss46566529
				CTSD 271	ss46566530
				CTSD 305	ss46566531
				CTSD 309	ss46566532
				CTSD 339	ss46566533
				CTSD 398	ss46566534
				CTSD 431	ss46566535
Chitinase 3' UTR	16	BV677374	239	CHIT 80	ss38350751
Corticotropin releasing factor binding protein 3' UTR	16	BV677377	347	CRFBP 74	ss38508078
Fibrinogen gamma polypeptide (<i>Omy_fl</i>)	13	BV677362	580	<i>Omy_fl</i> 259	ss46566536
				<i>Omy_fl</i> 260	ss46566537
				<i>Omy_fl</i> 509	ss46566538
				<i>Omy_fl</i> 570	ss46566539
Glucose-6 phosphate dehydrogenase (<i>Omy_g1</i>)	19	BV677370	263	<i>Omy_g1</i> 103	ss38350764
Glucosidase, beta; acid (<i>Omy_g2</i>)	5	BV677363	376	<i>Omy_g2</i> 084	ss38350759
				<i>Omy_g2</i> 138	ss38350760
				<i>Omy_g2</i> 192	ss38350761
				<i>Omy_g2</i> 224	ss38350762
				<i>Omy_g2</i> 330	ss38350763
Growth hormone 2, intron C	22	BV677389	510	GH2C 323	ss46566274
				GH2C 324	ss46566275
				GH2C 501	ss46566276
Heat shock protein 70 3' UTR	6	BV677376	370	HSP70UTR 221	ss38508079
Heat shock protein 71 intron 5	6	BV677383	151	HSC715 38	ss38508081
				HSC715 56	ss38508082
				HSC715 98	ss38508083
HoxD (<i>Omy_h1</i>)	12	BV677364	340	<i>Omy_h1</i> 075	ss46566277
				<i>Omy_h1</i> 170	ss46566278
Immunoglobulin H intron	23	BV677379	878	IGH 108	ss46566549
				IGH 216	ss46566550
				IGH 220	ss46566551
				IGH 222	ss46566552
				IGH 241	ss46566553
				IGH 262	ss46566554
				IGH 266	ss46566555
				IGH 275	ss46566556
				IGH 280	ss46566557
				IGH 295	ss46566558
				IGH 366	ss46566559
				IGH 404	ss46566560
				IGH 465	ss46566561
				IGH 511	ss46566562
				IGH 546	ss46566563
				IGH 547	ss46566564
				IGH 727	ss52084287
				IGH 734	ss52084288
				IGH 750	ss52084289
				IGH 754	ss52084290
				IGH 757	ss52084291
				IGH 777	ss52084292
				IGH 792	ss52084293
				IGH 819	ss52084294
				IGH 839	ss52084295
				IGH 844	ss52084296
Inhibitor of DNA binding/differentiation 1 (<i>Omy_il</i>)	6	BV677365	322	<i>Omy_il</i> 038	ss46566279
				<i>Omy_il</i> 101	ss46566280
				<i>Omy_il</i> 201	ss46566281
				<i>Omy_il</i> 232	ss46566282
				<i>Omy_il</i> 254	ss46566283
Inhibitor of DNA binding differentiation 1C	22	BV677373	361	ID1C 77 ^a	ss46566540
				ID1C 160	ss46566546
				ID1C 241	ss46566547
				ID1C 262	ss46566548

TABLE 4.—Continued.

Targeted region	N	STS	Base pairs/locus	SNP	dbSNP
Lactate dehydrogenase B	17	BV677369	303	LDH 156 LDH 201 LDH 212	ss46565746 ss46565747 ss46565748
Lysozyme intron II	18	BV677381	379	LysII3 18 LysII3 102 LysII3 120 LysII3 123 LysII3 138 LysII3 193 LysII3 252 LysII3 340	ss46566572 ss46566573 ss46566574 ss46566575 ss46566576 ss46566577 ss46566578 ss46566579
Lysozyme intron II	16	BV677380	417	LysII5 128 LysII5 178 LysII5 186 LysII5 199 LysII5 215 LysII5 237 LysII5 269	ss46566565 ss46566566 ss46566567 ss46566568 ss46566569 ss46566570 ss46566571
Lysozyme intron III	12	BV677382	535	LysIII 039 LysIII 062 LysIII 069 LysIII 082 LysIII 154 LysIII 212 LysIII 237 LysIII 246 LysIII 264 LysIII 377 LysIII 466	ss46566580 ss46566581 ss46566582 ss46566583 ss46566584 ss46566585 ss46566586 ss46566587 ss46566588 ss46566589 ss46566590
Myc 3' UTR	6	BV677375	406	MYC 077 MYC 125 MYC 158 MYC 232	ss38350747 ss38350748 ss38350749 ss38350750
Myeloperoxidase (<i>Omy_ml</i>)	5	DQ201133	277	<i>Omy_ml</i> 117	ss52084302
Novel immune-type receptor 2	18	BV377384	442	NITR 013 NITR 025 NITR 078 NITR 083 NITR 146 NITR 170 NITR 178 NITR 293 NITR 296 NITR 302 NITR 404 NITR 406 NITR 411 NITR 414 NITR 415 NITR 416 NITR 419 NITR 421	ss52049627 ss52049628 ss52049629 ss52049630 ss52049631 ss52049632 ss52049633 ss52049634 ss52049635 ss52049636 ss52049637 ss52049638 ss52049639 ss52049640 ss52049641 ss52049642 ss52049643 ss52049644
<i>OmyP9b</i>	14	BV677385	300	<i>OmyP9</i> 128 <i>OmyP9</i> 180 <i>OmyP9</i> 270	ss52049648 ss52049649 ss52049650
<i>OmyP9b</i>	18	BV677386	334	RAPD 132 RAPD 167 RAPD 277	ss38508087 ss38508088 ss38508089
RAPD intergenic sequence (OPC-02)	15	BV677359	503	R0336 062 R0336 129	ss38508084 ss38508085
RAPD intergenic sequence (OPC-02)	15	BV677360	512	R0917 125 R0917 230 R0917 262 R0917 391 R0917 401 R0917 455	ss52049618 ss52049619 ss52049620 ss52049621 ss52049622 ss52049623
RAPD intergenic sequence (OPC-02)	15	BV677359	503	R0336 062 R0336 129	ss38508084 ss38508085
Recombination activating gene 3'UTR 0336–864	15	BV677359	503	R0336 062 R0336 129	ss38508084 ss38508085
Recombination activating gene 3'UTR 0917–1503	15	BV677360	512	R0917 125 R0917 230 R0917 262 R0917 391 R0917 401 R0917 455	ss52049618 ss52049619 ss52049620 ss52049621 ss52049622 ss52049623
Recombination activating gene 3'UTR 1175–1700	17	BV677361	465	R1175 009 R1175 137 R1175 147	ss52049645 ss52049646 ss52049647

TABLE 4.—Continued.

Targeted region	N	STS	Base pairs/locus	SNP	dbSNP
Recombination activating gene 3'UTR 1564–1957	6	BV677366	270	R1564 121 R1564 272 R1564 325	ss52049624 ss52049625 ss52049626
Urotensin I 3' UTR	17	BV677378	424	URO 188 URO 217 URO 373	ss38508090 ss38508091 ss38509092
Non-variable loci					
Alpha tropomyosin	12	BV678701	392		
Glyceraldehyde-3-phosphate dehydrogenase	3	BV678700	218		
Heat shock protein 71 intron 1	5	BV677388	224		

^a Represents a 6 base pair insertion/deletion.

128 SNPs, and 4–52 candidate diagnostic SNPs were found between subspecies (Table 5, Appendix 1).

TaqMan Assay

Because sequence data were only generated for 1–10 individuals per population for each locus, it was of interest to assess how a SNP appearing to have fixed allelic frequencies between subspecies might perform on a larger sample of individuals. Additionally, we were curious to know how the frequency of a specific polymorphism might be reflected in additional populations of the same subspecies and in other hatchery strains of rainbow trout. By preliminary sequence comparisons, the frequency of SNP *LDH-156** was a candidate diagnostic between the Mount Shasta strain (“fixed” for the T allele) and Sheepheaven Creek and Volcano Creek fish (“fixed” for the C allele). Both alleles were observed in North Fork American River and Davis Creek fish (Table 6). The TaqMan assay enabled us to increase the sample size for populations in the SNP discovery panel to levels at which we would expect to see rare alleles emerge. The allelic frequencies for the *LDH-156** derived by sequencing were generally maintained. However, the C allele, which was previously not detected in the Mt. Shasta discovery sample, was observed at a frequency of 0.15 in the larger sample (Table 6). When additional populations were analyzed, *LDH-156** allele frequencies were

similar to related populations in the discovery panel. When genotypes were compared between the hatchery strains, all showed allele frequencies of 0.18 or lower for the C allele, except for the Mount Whitney Hatchery fish, which had a frequency of 0.53. Intermediate frequencies were also observed for Eagle Lake (0.48) and Davis Creek (0.49) populations. The Sheepheaven Creek, Volcano Creek, South Fork Kern River (above Templeton Barrier), Golden Trout Creek (at Mouth of Barigan Stringer), and Upper Soda Spring Creek populations showed frequencies of 0.93 or higher for the C allele; the California golden trout subspecies was completely fixed for the C allele. The only native population with a low level of the C allele was the Edson Creek fish.

Discussion

The sport and commercial importance of rainbow trout have driven extensive genomic research in the species, including the complete sequencing of the mitochondrial genome (Zardoya et al. 1995), linkage maps of the nuclear genome (Young et al. 1998; Sakamoto et al. 2000; Nichols et al. 2003), the cloning of multiple rainbow trout genes and mRNA (Benson et al. 2005), and over 239,516 EST submissions to date (Benson et al. 2005). We were able to harness this information to implement a targeted locus approach for SNP discovery between native subspecies and hatchery

TABLE 5.—Total number of SNPs found for each pairwise comparison of subspecies (above diagonal) and candidate diagnostic SNPs (below diagonal). The percentage of SNPs that are candidate diagnostics for each pairwise comparison is given in parentheses. Subspecies codes are given in Table 1.

Subspecies	RT-hat	RT-wild	SHP	SAC	CAGT	LKGT
RT-hat		113	117	99	122	124
RT-wild	6 (5%)		114	109	111	119
SHP	23 (20%)	27 (24%)		102	111	128
SAC	4 (4%)	11 (10%)	21 (21%)		92	111
CAGT	24 (20%)	25 (23%)	52 (47%)	34 (37%)		75
LKGT	21 (17%)	19 (16%)	49 (38%)	30 (27%)	14 (19%)	

TABLE 6.—Taqman-generated *LDH 156** SNP genotype data for multiple individuals from additional populations. The frequency of the C allele is given in the third column. The original frequency from the SNP discovery process is included for reference, with number of individuals in parentheses. Subspecies codes are given in Table 1.

Locality (N)	Subspecies code	Frequency	Original frequency
Hot Creek strain (20)	RT-hat	0.05	
Pitt strain (20)	RT-hat	0.10	
Mount Shasta strain (38)	RT-hat	0.15	0.00 (3)
Coleman strain (20)	RT-hat	0.18	
Mount Whitney strain (19)	RT-hat	0.53	
North Fork American River (8)	RT-wild	0.94	0.88 (4)
Eagle Lake (20)	ELRT	0.48	
Davis Creek (34)	SAC	0.49	0.50 (4)
Edson Creek (40)	SAC	0.00	
Sheepheaven Creek (42)	SHP	0.96	1.00 (3)
Upper Soda Spring Creek (22)	LKGT	1.00	
Volcano Creek (16)	CAGT	1.00	1.00 (3)
South Fork Kern River, above Templeton barrier (16)	CAGT	1.00	
Golden Trout Creek, at Mouth of Barigan Stringer (20)	CAGT	1.00	

strains of rainbow trout. Of 55 PCR primer pairs screened, 40 generated sequenceable amplicons. Within these, 208 SNPs were identified among six representative populations: McCloud River redband trout from Sheepheaven Creek, Sacramento redband trout from Davis Creek, California golden trout from Volcano Creek, Little Kern golden trout from upper Soda Spring Creek, coastal rainbow trout from the North Fork American River, and the Mount Shasta Hatchery strain.

Expressed sequence tags represent the majority of known DNA sequence information for most organisms. We attempted to utilize these sequences to redesign primer pairs that amplify genes from other taxa that have not yet been cloned in rainbow trout. It was not clear from the outset whether this strategy would be successful for two primary reasons. First, although the NCBI BLAST algorithm was able to align the primer sequences with highly similar teleost ESTs, never did both primers in the pair align with same sequence. Therefore, primer pairs were designed from two separate ESTs, one for the forward oligonucleotide and a second for the reverse. Second, 10 of the 16 EST-based primer sets were derived from EST sequences of non-rainbow trout organisms. Of the 16 primer pairs designed from EST sequences, 6 utilized rainbow trout sequence for both the forward and reverse primer, and 10 contained one or more non-rainbow trout derived primers. Of the 16 pairs, 11 amplified rainbow trout DNA and 9 produced sequenceable amplicons. These amplification and sequencing success rates were comparable to those in the three targeted-region categories and did not differ by species used for primer redesign. However, given that only two of the nine sequences demonstrated homology with the targeted

locus, it appears the utilized redesigning method may have actually increased amplification of nontarget loci.

Our final total of 208 SNPs was generated by visual inspection of sequences from 40 primer pairs. Computer programs such as PolyPhred, PolyBayes, and InSNP are available to facilitate SNP discovery for larger volumes of sequencing data (Nickerson et al. 1997; Marth et al. 1999; Manaster et al. 2005). Thirty-seven loci contained at least one polymorphism, and 100% of 3' UTR, extragenic, and unknown DNA sequences contained SNPs, as did 77% of intronic regions. Given the lack of statistically significant differences for base pairs per SNP by category, our results fail to dispute the null hypothesis of no difference demonstrated by Zhang and Hewitt (2003).

The total number of SNPs apparent in pairwise comparisons of rainbow trout subspecies ranged from 75 to 128, and candidate diagnostics ranged from 4 to 52 (Table 5; Appendix 1). However, the allelic frequencies observed between groups are based on only 1–10 individuals per population. The minimum sample size necessary to be confident that 95% of the time, at least one (diploid) individual will be detected for an allele present at a proportions of at least 1% can be calculated as

$$n = \frac{\log_e(1 - \beta) / \log_e(1 - P)}{4},$$

where β is the desired probability or confidence level (0.95) and P is the frequency of the allele (0.01) (adapted from Schwager et al. 1990 as cited by Grewe et al. 1993). Therefore, a minimum sample of approximately 75 individuals would be required. In general, it seems reasonable that a minimum of at least 40–50 individuals should be screened before an allele

is considered as fixed or diagnostic between populations.

Although frequencies in the relatively small discovery panel were generally maintained in the larger sample size genotyped for *LDH-156**, some changes did occur. The greatest change in frequency of the C allele was seen in the Mt. Shasta Strain population, which went from 0.00 in three individuals (i.e., 0 of 6 alleles) assayed to 0.15 in 38 individuals (Table 6). When additional populations of California golden trout and hatchery rainbow trout were genotyped at *LDH-156**, allelic frequencies similar to those of the relevant representative populations were obtained. California golden trout remained fixed for the C allele for the two additional populations (South Fork Kern River at Templeton Barrier and Golden Trout Creek at Mouth of Barigan Stringer). Little Kern golden trout from Upper Soda Spring Creek, for which sequence data were not previously collected, were also fixed for the C allele. The additional hatchery strains (Pitt, Coleman, and Hot Creek hatcheries) showed similar frequencies (0.05–0.18) to the Mt. Shasta Hatchery strain. Only the Mt. Whitney Hatchery strain appeared to deviate from other hatchery strains (intermediate frequency of 0.53). This is probably attributable to the mixed origins of this latter strain, which Busack and Gall (1980) suggest may contain a mixture of coastal rainbow trout from Kings River, steelhead (anadromous rainbow trout) from Klamath and Eel rivers, and perhaps Lahontan cutthroat trout *O. clarkii henshawii*.

The McCloud and Sacramento redband trout yielded somewhat mixed results for the *LDH-156** locus. Frequencies for the C allele were high for Sheepheaven Creek and intermediate for Davis Creek. The intermediate frequency in Davis Creek is somewhat surprising for a putative native redband population. This suggests that it either contains a higher degree of coastal rainbow trout ancestry (given its geographic proximity to rainbow trout source populations) or, alternatively, contemporary gene flow with introduced rainbow stocks. The observation of a high (fixed) level of what was generally characterized as a hatchery rainbow allele was somewhat unexpected for the Edson Creek population, given that microsatellite analyses show a close relationship between Edson and Sheepheaven populations (Nielsen et al. 1999). This too suggests either hatchery influence or a possible reflection of ancestry with coastal rainbow trout. Lastly, the Eagle Lake rainbow trout, which was not previously examined in the discovery panel, had an intermediate frequency at the *LDH-156** locus, which may also be indicative of its potentially mixed origins from coastal rainbow and redband trout (Behnke 1992:198). The insight gained from the analysis of additional loci will

be necessary for a more accurate evaluation of these relationships, and interpretation of relationships based on a single SNP locus is clearly provisional.

Polymorphisms with apparent fixed allelic frequencies between groups clearly have the greatest utility for studying introgression between native and introduced trout. Additionally, they will have the greatest power for inferring individual population assignments. In our study, 4–20% of the total number of polymorphisms discovered between the Mount Shasta Hatchery strain and native populations of interest and 10–24% between the wild (North Fork American River) and native populations had apparent fixed allelic frequencies. The limited number of individuals screened in our discovery panel is clearly not sufficient to designate a SNP as diagnostic for different subspecies groups; we expect that as sample size and number of populations screened are increased for the other apparently fixed loci in our SNP panel, the total number of fixed alleles will be reduced. However, the majority of the polymorphisms occur as nonfixed allelic frequency differences between subspecies. Our discovery method did allow the detection of alleles with larger frequency differentials between groups (>50%), which are likely to have significant power for detecting and quantifying introgression between groups (e.g., population-specific alleles, per Shriver et al. 1997).

It is generally understood that a larger number of SNPs (40–100 or more) are required to match the power of a smaller panel (10–15) of microsatellite markers, particularly in applications such as individual identification, parentage, and relatedness analyses (reviewed in Morin et al. 2004). The highly polymorphic nature of microsatellites (due to their higher mutation rate of 1×10^{-5} versus 1×10^{-8} for SNPs; Li et al. 1981; Kruglyak et al. 1998; Martinez-Arias et al. 2001) increases the likelihood that informative (i.e., unique) alleles exist. However, even highly polymorphic microsatellite loci can be limited in studies of admixture because of a paucity of alleles unique to any one group or even significant frequency differences between groups. This is particularly problematic in recently diverged groups that share common ancestry, like the relatively young, shallowly diverged rainbow trout subspecies groups (Behnke 1992; Bagley and Gall 1998). In a study of California golden trout introgression, only 9 of 95 alleles at six microsatellite loci could be classified as known rainbow trout microsatellite alleles that were not present in California golden trout reference populations (Cordes et al. 2006). Utilizing a smaller panel of SNPs with a higher degree of per-locus information content may be a viable alternative to using more variable, but less informative, microsatellite loci.

We observed an average of five SNPs per locus. Though it is assumed that each SNP represents a different mutational event, presumed linkage between SNPs in a given locus may require analysis of only a subset or, alternatively, of a composite haplotype for a given locus (Chakraborty et al. 1999; Fries et al. 2001). Nearly all targeted regions exhibited some linkage between one or more SNPs within each region, typically between adjacent SNPs (e.g., *Omy_al-115**, *-134**, and *-156**, but not *Omy_al-091** or *-105**). However, all loci except two showed evidence of recombination in the observed genotypes. Only heat shock protein 71 intron 5 (see below) and alcohol dehydrogenase 2 intron showed total linkage among all SNP loci within each respective region. Evaluation of linkage disequilibrium among regions will require greater sampling for each population.

Because our panel of SNPs was generated by screening small numbers of individuals from each subspecies, our SNP markers are biased toward higher-frequency polymorphisms. An important concern is the reduction of taxonomic complexity within subspecies as a result of the discovery bias of our approach toward differences between subspecies (Pearson et al. 2004). Given the ascertainment bias associated with this panel of SNPs, careful use of these markers for appropriate applications is required (Morin et al. 2004). Use of our SNPs in studies of hybridization between native and nonnative rainbow trout groups presents less of a conflict, for example, than does use in population structure or phylogenetic applications. The use of appropriate statistical methods in data analysis has the potential to overcome some of these problems (Wakeley et al. 2001; Brumfield et al. 2003; Nielsen and Signorovitch 2003; Nielsen et al. 2004). This panel could also be corrected for other applications through the sequencing of additional individuals from a wider geographic range.

Other methods of SNP discovery include high-throughput sequencing of genomic cDNA libraries and shotgun clones (random genomic libraries) and mining sequences within a multiple sequence alignment of homologous DNA sequences (reviewed in Vignal et al. 2002). Arguably, these methods are equally efficient ways to discover nucleotide variation. However, valuable information can be lost when polymorphisms are evaluated between cloned fragments rather than by amplified genomic DNA. This is of particular concern for researchers interested in species with a history of genomic duplication and polyploidy. By amplifying fragments directly from genomic DNA, paralogous copies may be identified by the persistent presence of multiple amplicons, if there is size variation between copies. By restricting sequence comparisons to ampli-

cons of similar length, one can reduce the likelihood of mistaking nuclear variation between paralogous copies for allelic variation at a single locus. As summarized in Table 3, the presence of multiple bands made it clear that many of our PCR primer pairs targeted multiple sites within the genome. One of our loci, *HSP-71** intron 5, exhibited heterozygotic SNPs between every individual screened, a phenomenon indicative of paralogous copies (Ohno et al. 1969; Smith et al. 2005b). With the knowledge of locus duplication, a researcher will be better prepared should results of a genotyping assay fail to correlate with sequencing data or deviations from Hardy–Weinberg equilibrium be encountered. Inclusion of positive controls of known DNA sequence and scrutiny of TaqMan endpoint fluorescence plots may further assist in detecting potentially paralogous loci. Despite these precautions, definitive confirmation by southern blotting or a real-time copy number assay (Ginzinger 2002) may still be required to ascertain whether the amplified sequence is generated from multiple nuclear loci.

One byproduct of the SNP discovery process is the detection of insertions and deletions (indels) greater than a single base pair in size. Several of these were encountered during our screening of individuals, and a few are candidate diagnostics for a particular subspecies (Genomic Variation Laboratory, unpublished data). Clear, fixed size differences of alleles between two subspecies makes this marker type advantageous for detecting the occurrence of introgression between subspecies. However, difficulties in developing TaqMan assays for some indels will require that they be screened using a gel-based or other detection system. Additionally, some of the indels we originally encountered exhibited multiple bands that could not be eliminated through primer redesign, and some indels disappeared after primer redesign; both scenarios are suggestive of possible gene or chromosome duplication, which would not be unlikely given the polyploid derived origin of the study species. Further investigation of these issues is required for some of the discovered candidate diagnostic indels to determine their utility as potential markers.

For researchers interested in converting to SNPs, genotyping technology is an important concern. Our use of the TaqMan assay for genotyping multiple individuals at SNP *LDH-156** generated a significant amount of high-quality data in a very short time (1.5 h, 96-well plate) at an inexpensive cost (US\$0.25/genotype; this cost can be further reduced to about \$0.10/genotype by purchasing larger quantities of probe). However, it is unlikely that the TaqMan assay will work as effectively for every SNP within our panel. By nature of design, the TaqMan protocol

requires the development of an assay for each individual SNP. Assay design can be a challenge if other SNPs occur near the locus to be evaluated or the surrounding sequence contains nucleotide repetition. Because this is the case for many of the polymorphic loci, alternative methods for genotyping will need to be investigated. Costs per assay range from \$310 (ABI Assay's-By-Design, 1 nM of each probe) to \$500 (primers and 6 nM of each probe ordered individually). Though sufficient reagents are provided for the genotyping of thousands of individuals, the cost is fixed; for a researcher with only a few samples, it may not be an economical choice. For assays that require optimization, the expense of the primer-probe combinations may prohibit testing multiple combinations of oligonucleotides. In short, the TaqMan assay works well for genotyping a large number of samples at a single point mutation. For studies that require the genotyping of only a few individuals at loci with indels, multiple polymorphisms, or repetitive sequence, alternative options should be considered (see Chen and Sullivan 2003 for a review of methods).

Use of SNP markers holds great promise for addressing questions of phylogeny, population structure, introgression, and adaptive genetic variation in native populations. With whole genomic sequencing efforts underway for nearly a dozen fish species and large-scale efforts for many others, there is ample DNA sequence data available to facilitate SNP discovery in fish. We hope the work described in this manuscript is encouraging for researchers interested in SNP marker development in other fish species.

Acknowledgments

This project was supported by agreement numbers P0380035 and P0480052 from California Department of Fish and Game, grant C0310004-02 from the Cantara Trustee Council, and U.S. Fish and Wildlife Service agreement 101814M429. We thank Ruth Phillips for her gift of PCR primers to regions of variation in the rainbow trout genome that she discovered; Michael Dean, Christy McGuire, and Stanley Stephens for helpful history, discussion, and comments; John Pedroia and Jessica Petersen for technical assistance; and Benjamin Sacks and Holly Ernest for manuscript critique.

References

- Aitken, N., S. Smith, C. Schwarz, and P. A. Morin. 2004. Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology* 13(6):1423–1431.
- Altschul, S., T. L. Madden, A. A. Schaffer, J. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25(17):3389–3402.
- Bagley, M. J., and G. A. E. Gall. 1998. Mitochondrial and nuclear DNA sequence variability among populations of rainbow trout (*Oncorhynchus mykiss*). *Molecular Ecology* 7(8):945–961.
- Behnke, R. J. 1992. Native trout of western North America. American Fisheries Society, Monograph 6, Bethesda, Maryland.
- Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler. 2005. GenBank. *Nucleic Acids Research* 33:D34–D38.
- Bierne, N., S. A. Lehnert, E. Bedier, F. Bonhomme, and S. S. Moore. 2000. Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. *Molecular Ecology* 9(2):233–235.
- Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* 18:249.
- Busack, C. A., and G. A. E. Gall. 1980. Ancestry of artificially propagated California rainbow trout strains. *California Fish and Game* 66(1):17–24.
- Chakraborty, R., D. N. Stivers, B. Su, Y. Zhong, and B. Budowle. 1999. The utility of short tandem repeat loci beyond human identification: applications for development of new DNA typing systems. *Electrophoresis* 20:1682–1696.
- Chen, X., and P. F. Sullivan. 2003. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost, and throughput. *Pharmacogenomics Journal* 3:77–99.
- Collins, F. S., L. D. Brooks, and A. Chakravarti. 1998. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research* 8:1229–1231.
- Cordes, J. F., M. R. Stephens, M. A. Blumberg, and B. P. May. 2006. Identifying introgressive hybridization in native populations of California golden trout based on molecular markers. *Transactions of the American Fisheries Society* 134:110–128.
- Docker, M. F., and D. D. Heath. 2002. PCR-based markers detect genetic variation at growth- and immune function-related loci in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology Notes* 2(4):606–609.
- Ford, M. J. 2000. Effects of natural selection on patterns of DNA sequence variation at the transferrin, somatolactin, and p53 genes within and among Chinook salmon (*Oncorhynchus tshawytscha*) populations. *Molecular Ecology* 9(7):843–855.
- Fries, R., D. Ewald, G. Thaller, and J. Buitkamp. 2001. Assessment of the nucleotide sequence variability in the bovine T-cell receptor alpha delta joining gene region. *Animal Biotechnology* 12(1):29–49.
- Friesen, V. L., B. C. Congdon, M. G. Kidd, and T. P. Birt. 1999. Polymerase chain reaction (PCR) primers for the amplification of five nuclear introns in vertebrates. *Molecular Ecology* 8(12):2147–2149.
- Gahr, S. A., Y. Palti, and C. E. Rexroad. 2004. Genomic characterization of a novel pair of ID genes in the rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics* 35(4):317–321.
- Ginzinger, D. G. 2002. Gene quantification using real-time

- quantitative PCR: an emerging technology hits the mainstream. *Experimental Hematology* 30:503–512.
- Grewe, P. M., C. C. Krueger, C. F. Aquadro, E. Bermingham, H. Kincaid, and B. May. 1993. Mitochondrial DNA variation among lake trout strains (*Salvelinus namaycush*) stocked into Lake Ontario. *Canadian Journal of Fisheries and Aquatic Sciences* 50:2397–2403.
- Guan, G., M. Tanaka, T. Todo, G. Young, M. Yoshikuni, and Y. Nagahama. 1999. Cloning and expression of two carbonyl reductase-like 20[β]-hydroxysteroid dehydrogenase cDNAs in ovarian follicles of rainbow trout (*Oncorhynchus mykiss*). *Biochemical and Biophysical Research Communications* 255(1):123–128.
- Hansen, J. D. 1997. Characterization of rainbow trout terminal deoxynucleotidyl transferase structure and expression: TdT and RAG1 co-expression define the trout primary lymphoid tissues. *Immunogenetics* 46(5):367–375.
- Hassan, M., C. Lemaire, C. Fauvelot, and F. Bonhomme. 2002. Seventeen new exon-primed intron-crossing polymerase chain reaction amplifiable introns in fish. *Molecular Ecology Notes* 2(3):334–340.
- Iturra, P., J. F. Medrano, M. Bagley, M. Lam, N. Vergara, and J. C. Marin. 1998. Identification of sex chromosome molecular markers using RAPDs and fluorescent in situ hybridization in rainbow trout. *Genetica* 101:209–213.
- Kanz, C., P. Aldebert, N. Althorpe, W. Baker, A. Baldwin, K. Bates, P. Browne, A. van den Broek, M. Castro, G. Cochrane, K. Duggan, R. Eberhardt, N. Faruque, J. Gamble, F. G. Diez, N. Harte, T. Kulikova, Q. Lin, V. Lombard, R. Lopez, R. Mancuso, M. McHale, N. Nardone, V. Silventoinen, S. Sobhany, P. Stoehr, M. A. Tuli, K. Tzouvara, R. Vaughan, D. Wu, Z. Zhu, and R. Apweiler. 2005. The EMBL nucleotide sequence database. *Nucleic Acids Research* 33:D29–D33.
- Kothary, R. K., D. Jones, and E. P. Candido. 1984. 70-Kilodalton heat shock polypeptides from rainbow trout: characterization of cDNA sequences. *Molecular and Cellular Biology* 4(9):1785–1791.
- Kruglyak, S., R. T. Durrett, M. D. Schug, and C. F. Aquadro. 1998. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences of the USA* 95:10774–10778.
- Lee, M. A., E. Bengten, A. Daggfeldt, A. S. Rytting, and L. Pilstrom. 1993. Characterization of rainbow trout cDNAs encoding a secreted and membrane-bound Ig heavy chain and the genomic intron upstream of the first constant exon. *Molecular Immunology* 30(7):641–648.
- Lee, Y., J. Tsai, S. Sunkara, S. Karamycheva, G. Perte, R. Sultana, V. Antonescu, A. Chan, F. Cheung, and J. Quackenbush. 2005. The TIGR gene indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Research* 33:D71–D74.
- Li, W. H., T. Gojbori, and M. Nei. 1981. Pseudogenes as a paradigm of neutral evolution. *Nature (London)* 292:237–239.
- Lim, S. T., R. M. Kay, and G. S. Baily. 1975. Lactate dehydrogenase isozymes of salmonid fish: evidence for unique and rapid functional divergence of duplicated H-4 lactate dehydrogenases. *Journal of Biological Chemistry* 250(5):1790–1800.
- Lyons, L. A., T. F. Laughlin, N. G. Copeland, N. A. Jenkins, J. E. Womack, and S. J. O'Brien. 1997. Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics* 15(1):47–56.
- Manaster, C., W. Zheng, M. Teuber, S. Wachter, F. Doring, S. Schreiber, and J. Hampe. 2005. InSNP: a tool for automated detection and visualization of SNPs and InDels. *Human Mutation* 26(1):11–19.
- Marth, G. T., I. Kort, M. D. Yandell, R. T. Yeh, Z. Gu, H. Zakeri, N. O. Stitzel, L. Hillier, P. Kwok, and W. R. Gish. 1999. A general approach to single nucleotide polymorphism discovery. *Nature Genetics* 23(4):452–456.
- Martinez-Arias, R., F. Calafell, E. Mateau, D. Comas, A. Andres, and J. Bertranpetit. 2001. Sequence variability of a human pseudogene. *Genome Research* 11:1071–1085.
- Mitra, A., J. Foster-Frey, C. E. Rexroad III, K. D. Wells, and R. J. Wall. 2003. Molecular characterization of lysozyme type II gene in rainbow trout (*Oncorhynchus mykiss*): evidence of gene duplication. *Animal Biotechnology* 14(1):7–12.
- Moran, P. 2002. Current conservation genetics: building an ecological approach to the synthesis of molecular and quantitative genetic methods. *Ecology of Freshwater Fish* 11:30–55.
- Moran, P., D. A. Dightman, R. S. Waples, and L. K. Park. 1997. PCR–RFLP analysis reveals substantial population-level variation in the introns of Pacific salmon (*Oncorhynchus* spp.) *Molecular Marine Biology and Biotechnology* 6(4):315–327.
- Morin, P. A., G. Luikart, and R. K. Wayne. 2004. SNPs in ecology, evolution, and conservation. *Trends in Ecology and Evolution* 19(4):208–216.
- Muller-Schmid, A., H. Rinder, F. Lottspeich, and E. M. Gertzen. 1992. Ependymins from the cerebrospinal fluid of salmonid fish: gene structure and molecular characterization. *Gene* 118(2):189–192.
- Nichols, K. M., W. P. Young, R. G. Danzmann, B. D. Robinson, C. Rexroad, M. Noakes, R. B. Phillips, P. Bentzen, I. Spies, K. Knudsen, F. W. Allendorf, B. M. Cunningham, J. Brunelli, H. Zhang, S. Ristow, R. Drew, K. H. Brown, P. A. Wheeler, and G. H. Thorgaard. 2003. A consolidated linkage map for rainbow trout (*Oncorhynchus mykiss*). *Animal Genetics* 34(2):102–115.
- Nickerson, D. A., V. O. Tobe, and S. L. Talyor. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research* 25(14):2745–2751.
- Nielsen, J. J., K. D. Crow, and M. C. Fountain. 1999. Microsatellite diversity and conservation of a relic trout population: McCloud River redband trout. *Molecular Ecology* 8:S129–S142.
- Nielsen, R., M. J. Hubisz, and A. G. Clark. 2004. Reconstituting the frequency spectrum of ascertained single nucleotide polymorphism data. *Genetics* 168(4):2373–2382.
- Nielsen, R., and J. Signorovitch. 2003. Correcting for ascertainment bias when analyzing SNP data: applications to the estimation of linkage disequilibrium. *Theoretical Population Biology* 63:245–255.
- Ohno, S. M., J. Muramoto, J. Klein, and N. B. Atkin. 1969.

- Diploid–tetraploid relationship in clupeoid and salmonid fish. *Chromosomes Today* 2:139–147.
- Ojima, N., M. Yamashita, and S. Watabe. 2005. Comparative expression analysis of two paralogous Hsp70s in rainbow trout cells exposed to heat stress. *Biochimica et Biophysica Acta* 1681(2–3):99–106.
- Pakdel, F., R. Metivier, G. Flouriot, and Y. Valotaire. 2000. Two estrogen receptor (ER) isoforms with different estrogen dependencies are generated from the trout ER gene. *Endocrinology* 141(2):571–580.
- Palumbi, S. R., and C. S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. *Molecular Biology and Evolution* 11:426–435.
- Panno, J. P., and B. A. McKeown. 1995. Cloning and expression of a myc family member from the pituitary gland of the rainbow trout, *Oncorhynchus mykiss*. *Biochimica et Biophysica Acta* 1264(1):7–11.
- Pearson, T., J. D. Busch, J. Ravel, T. D. Read, S. D. Rhoton, J. M. U'Ren, T. J. Simonson, S. M. Kachur, R. R. Leadmen, M. L. Cardon, M. N. Van Ert, L. Y. Huynh, C. M. Fraser, and P. Keim. 2004. Phylogenetic discovery bias in *Bacillus anthracis* using single nucleotide polymorphisms from whole-genome sequencing. *Proceedings of the National Academy of Sciences of the USA* 101:13536–13541.
- Primmer, C. R., T. Borge, J. Haavie, and G.-P. Sætre. 2002. Single nucleotide polymorphism (SNP) characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology* 11:603–612.
- Radice, A. D., B. Bugaj, D. H. Fitch, and S. W. Emmons. 1994. Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. *Molecular and General Genetics* 244(6):606–612.
- Rodzen, J. A., J. J. Agresti, G. Tranah, and B. May. 1998. Agarose overlays allow simplified staining of polyacrylamide gels. *Biotechniques* 25(4):584.
- Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* 132:365–386.
- Sakamoto, T., R. G. Danzmann, K. Gharbi, P. Howard, A. Ozaki, S. K. Khoo, R. A. Woram, N. Okamoto, M. M. Ferguson, L. E. Holm, R. Guyomard, and B. Howheim. 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155(3):1331–1345.
- Seddon, J. M., H. G. Parker, E. A. Ostrander, and H. Ellegren. 2005. SNPs in ecological and conservation studies: a test in the Scandinavian wolf population. *Molecular Ecology* 14:503–511.
- Shriver, M. D., M. W. Smith, L. Jin, A. Marcini, J. M. Akey, R. Deka, and R. E. Ferrell. 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. *American Journal of Human Genetics* 60:957–964.
- Shum, B. P., R. Rajalingam, K. E. Magor, K. Azumi, W. H. Carr, B. Dixon, R. J. M. Stet, M. A. Adkison, R. P. Hedrick, and P. Parham. 1999. A divergent, nonclassical class I gene conserved in salmonids. *Immunogenetics* 49(6):479–490.
- Smith, C. T., J. Baker, L. Park, L. W. Seeb, C. Elfstrom, S. Abe, and J. E. Seeb. 2005a. Characterization of 13 single nucleotide polymorphism markers for chum salmon. *Molecular Ecology Notes* 5:259–262.
- Smith, C. T., C. M. Elfstrom, L. W. Seeb, and J. E. Seeb. 2005b. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. *Molecular Ecology* 14:4193–4203.
- Smith, C. T., W. D. Templin, J. E. Seeb, and L. W. Seeb. 2005c. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of U.S. and Canadian Chinook salmon caught in Yukon River fisheries. *North American Journal of Fisheries Management* 25:944–953.
- Tateno, Y., N. Saitou, K. Okubo, H. Sugawara, and T. Gojobori. 2005. DDBJ in collaboration with mass-sequencing teams on annotation. *Nucleic Acids Research* 33:D25–D28.
- Tsuhiihashi, Z., and N. C. Dracopoli. 2002. Progress in high-throughput SNP genotyping methods. *Pharmacogenomics Journal* 2:103–110.
- USFWS (U.S. Fish and Wildlife Service). 1978. Listing of the Little Kern golden trout as a threatened species with critical habitat. *Federal Register* 43 (13 April 1978):15427–15429.
- Vignal, A., D. Milan, M. M. SanCristoball, and A. Eggen. 2002. A review of SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution* 34:275–305.
- Villablanca, F. X., G. K. Roderick, and S. R. Palumbi. 1998. Invasion genetics of the Mediterranean fruit fly: variation in multiple nuclear introns. *Molecular Ecology* 7(5):547–560.
- Vogel, F., and A. G. Motulsky. *Human genetics: problems and approaches*. Springer, New York.
- Wakeley, J., R. Nielsen, S. N. Liu-Cordero, and K. Ardlie. 2001. The discovery of single nucleotide polymorphisms and inferences about human demographic history. *American Journal of Human Genetics* 69:1332–1347.
- Yoder, J. A., M. G. Mueller, K. M. Nichols, S. S. Ristow, G. H. Thorgaard, T. Ota, and G. W. Litman. 2002. Cloning novel immune-type inhibitory receptors from the rainbow trout, *Oncorhynchus mykiss*. *Immunogenetics* 54(9):662–670.
- Young, W. P., P. A. Wheeler, V. H. Coryell, P. Keim, and G. H. Thorgaard. 1998. A detailed linkage map of rainbow trout produced using double haploids. *Genetics* 148(2):839–850.
- Zardoya, A., A. Garrido-Pertierra, and J. M. Bautista. 1995. The complete sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*. *Journal of Molecular Evolution* 41(6):942–951.
- Zhang, D. X., and G. M. Hewitt. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems, and prospects. *Molecular Ecology* 12:563–584.

Appendix: SNPs Found in Pairwise Comparisons of Subspecies

TABLE A.1.—List of SNPs found in pairwise comparisons of subspecies. Candidate diagnostic SNPs exhibiting fixed allele frequencies in the discovery subspecies panel are designated by boldface type. Subspecies codes as given in Table 1.

RT-hatchery versus					RT-wild versus		
RT-wild	SHP	SAC	CAGT	LKGT	SHP	SAC	CAGT
<i>Omy_al</i> 091	<i>Omy_al</i> 091	<i>Omy_a2</i> 221	<i>Omy_al</i> 091	<i>Omy_al</i> 091	<i>Omy_al</i> 091	<i>Omy_a2</i> 221	<i>Omy_al</i> 091
<i>Omy_al</i> 105	<i>Omy_al</i> 105	<i>Omy_a2</i> 249	<i>Omy_al</i> 105	<i>Omy_al</i> 105	<i>Omy_al</i> 105	<i>Omy_a2</i> 269	<i>Omy_al</i> 105
<i>Omy_al</i> 115	<i>Omy_al</i> 115	<i>Omy_a2</i> 269	<i>Omy_al</i> 115	<i>Omy_al</i> 115	<i>Omy_al</i> 134	B1 102	<i>Omy_al</i> 134
<i>Omy_al</i> 134	<i>Omy_al</i> 134	B1 102	<i>Omy_al</i> 134	<i>Omy_al</i> 134	B1 102	B1 266	B1 102
<i>Omy_al</i> 156	<i>Omy_al</i> 156	B1 257	<i>Omy_al</i> 156	<i>Omy_al</i> 156	B1 266	B1 39	B1 266
<i>Omy_a2</i> 221	B1 102	B9 181	B1 102	B1 102	B1 39	B9 181	B1 39
<i>Omy_a2</i> 249	B1 257	B9 330	B1 257	B1 257	B4 170	B9 335	B9 335
<i>Omy_a2</i> 269	B4 170	B9 335	B1 266	B9 16	B4 193	B9 355	B9 37
B1 102	B4 193	B9 355	B9 181	B9 164	B4 198	B9 37	B9 388
B1 257	B4 198	B9 90	B9 330	B9 181	B9 181	B9 90	B9 90
B1 266	B9 181	E1 239	B9 335	B9 182	B9 202	E1 100	E1 100
B1 39	B9 202	E1 241	B9 355	B9 330	B9 335	E1 108	E1 108
B9 181	B9 330	E1 289	B9 388	B9 335	B9 355	E1 109	E1 109
B9 330	B9 335	E1 300	B9 90	B9 355	B9 37	E1 110	E1 110
B9 335	B9 355	G9 168	E1 239	E1 100	E1 100	E1 111	E1 111
B9 355	E1 100	G9 465	E1 241	E1 108	E1 108	E1 239	E1 239
B9 37	E1 147	CRB 2358 020	E1 289	E1 109	E1 109	E1 241	E1 241
E1 100	E1 159	CRB 2358 120	E1 300	E1 110	E1 110	E1 289	E1 300
E1 108	E1 239	CRB 2677 106	F5 306	E1 111	E1 111	E1 300	F5 306
E1 109	E1 241	CRB 2677 117	F5 090	E1 147	E1 147	F5 306	F5 323
E1 110	E1 289	CRB 2677 126	G9 168	E1 239	E1 159	F5 323	F5 090
E1 111	E1 029	CRB 2677 129	G9 221	E1 241	E1 239	F5 090	G9 125
E1 239	E1 030	CRB 2677 168	G9 465	E1 289	E1 241	G9 125	G9 168
E1 241	E1 300	CRB 2677 203	G9 520	E1 300	E1 29	G9 168	G9 221
E1 289	F5 090	CTSD 097	G9 547	F5 306	E1 30	G9 465	G9 465
E1 300	G9 168	CHIT 80	CRB 2358 020	F5 090	E1 300	CRB 0826 102	G9 520
F5 306	G9 465	CRFBP 74	CRB 2358 120	G9 168	F5 306	CRB 0826 110	G9 547
F5 323	G9 502	<i>Omy_fl</i> 259	CTSD 033	G9 465	F5 323	CRB 0826 152	CRB 0826 110
F5 090	G9 504	<i>Omy_fl</i> 260	CTSD 097	G9 547	F5 090	CRB 0826 756	CRB 0826 756
G9 125	CRB 2358 020	<i>Omy_g2</i> 138	CTSD 158	CRB 2358 020	G9 125	CRB 2358 020	CRB 2358 020
G9 168	CRB 2358 120	<i>Omy_g2</i> 192	CTSD 173	CRB 2358 120	G9 168	CRB 2358 120	CRB 2358 120
G9 465	CRB 2677 106	<i>Omy_g2</i> 224	CTSD 174	CRB 2677 117	G9 465	CRB 2677 106	CTSD 033
CRB 2677 106	CRB 2677 117	<i>Omy_g2</i> 84	CTSD 339	CRB 2677 166	G9 502	CRB 2677 117	CTSD 097
CRB 2677 117	CRB 2677 126	GH2C 323	CTSD 431	CRB 2677 203	G9 504	CRB 2677 126	CTSD 158
CRB 2677 126	CRB 2677 129	GH2C 324	CHIT 80	CTSD 033	CRB 0826 110	CRB 2677 129	CTSD 173
CRB 2677 129	CRB 2677 168	GH2C 501	<i>Omy_fl</i> 259	CTSD 097	CRB 0826 756	CRB 2677 168	CTSD 174
CRB 2677 168	CRB 2677 203	HSPINT5 38	<i>Omy_fl</i> 260	CTSD 158	CRB 2358 020	CRB 2677 203	CTSD 339
CRB 2677 203	CTSD 033	HSPINT5 56	<i>Omy_fl</i> 509	CTSD 173	CRB 2358 120	CRB 2677 226	CTSD 398
CRB 2677 226	CTSD 097	HSPINT5 98	<i>Omy_g1</i> 103	CTSD 174	CRB 2677 106	CTSD 033	CTSD 431
CTSD 033	CTSD 158	<i>Omy_h1</i> 75	<i>Omy_g2</i> 138	CTSD 339	CRB 2677 117	CTSD 097	CHIT 80
CTSD 097	CTSD 173	IGH 262	<i>Omy_g2</i> 192	CTSD 431	CRB 2677 126	CHIT 80	<i>Omy_fl</i> 259
CTSD 158	CTSD 339	IGH 275	<i>Omy_g2</i> 224	CHIT 80	CRB 2677 129	CRFBP 74	<i>Omy_fl</i> 260
CTSD 173	CTSD 398	IGH 280	<i>Omy_g2</i> 330	<i>Omy_fl</i> 259	CRB 2677 168	<i>Omy_fl</i> 259	<i>Omy_fl</i> 509
CTSD 339	CHIT 80	IGH 295	<i>Omy_g2</i> 84	<i>Omy_fl</i> 260	CRB 2677 203	<i>Omy_fl</i> 260	<i>Omy_fl</i> 570
CTSD 398	CRFBP 74	IGH 366	GH2C 323	<i>Omy_g1</i> 103	CRB 2677 226	<i>Omy_g2</i> 192	<i>Omy_g1</i> 103
CHIT 80	<i>Omy_fl</i> 259	IGH 465	GH2C 324	HSPINT5 38	CTSD 033	<i>Omy_g2</i> 84	<i>Omy_g2</i> 192
<i>Omy_g2</i> 138	<i>Omy_fl</i> 260	IGH 754	GH2C 501	HSPINT5 56	CTSD 097	GH2C 323	<i>Omy_g2</i> 330
<i>Omy_g2</i> 224	<i>Omy_g1</i> 103	IGH 757	HSPINT5 38	HSPINT5 98	CTSD 152	GH2C 324	<i>Omy_g2</i> 84
GH2C 323	<i>Omy_g2</i> 138	IGH 777	HSPINT5 56	<i>Omy_h1</i> 170	CTSD 158	GH2C 501	GH2C 501
GH2C 324	<i>Omy_g2</i> 224	IGH 792	HSPINT5 98	<i>Omy_h1</i> 75	CTSD 339	HSP70UTR 221	HSP70UTR 221
GH2C 501	GH2C 323	IGH 839	<i>Omy_h1</i> 170	IGH 108	CTSD 398	HSPINT5 38	HSPINT5 38
HSP70UTR 221	GH2C 324	<i>Omy_il</i> 232	<i>Omy_h1</i> 75	IGH 216	CHIT 80	HSPINT5 56	HSPINT5 56
HSPINT5 38	GH2C 501	<i>Omy_il</i> 254	<i>Omy_h1</i> 108	IGH 220	CRFBP 74	HSPINT5 98	HSPINT5 98
HSPINT5 56	HSPINT5 38	ID1C 160	IGH 734	IGH 222	<i>Omy_fl</i> 259	<i>Omy_h1</i> 170	<i>Omy_h1</i> 170
HSPINT5 98	HSPINT5 56	ID1C 241	IGH 750	IGH 262	<i>Omy_fl</i> 260	<i>Omy_h1</i> 75	<i>Omy_h1</i> 75
<i>Omy_h1</i> 170	HSPINT5 98	ID1C 262	IGH 754	IGH 266	<i>Omy_g1</i> 103	IGH 275	IGH 108
<i>Omy_h1</i> 75	<i>Omy_h1</i> 75	LDH 156	IGH 757	IGH 275	<i>Omy_g2</i> 138	IGH 280	IGH 734
IGH 262	IGH 108	LDH 201	IGH 777	IGH 280	GH2C 501	IGH 295	IGH 750
IGH 275	IGH 216	LDH 212	<i>Omy_il</i> 232	IGH 295	HSP70UTR 221	IGH 734	IGH 757
IGH 280	IGH 220	LysII3 018	<i>Omy_il</i> 254	IGH 366	HSPINT5 38	IGH 750	IGH 777
IGH 295	IGH 222	LysII3 120	ID1C 160	IGH 734	HSPINT5 56	IGH 754	IGH 839
IGH 734	IGH 241	LysII3 123	ID1C 241	IGH 750	HSPINT5 98	IGH 757	<i>Omy_il</i> 038
IGH 750	IGH 262	LysII3 138	ID1C 262	IGH 754	<i>Omy_h1</i> 170	IGH 777	<i>Omy_il</i> 101
IGH 754	IGH 275	LysII3 193	LDH 156	IGH 757	<i>Omy_h</i> 175	IGH 792	<i>Omy_il</i> 201
IGH 757	IGH 280	LysII3 252	LysII3 018	IGH 777	IGH 108	IGH 839	<i>Omy_il</i> 232

TABLE A1.—Extended.

RT-wild versus		SHP versus		SAC versus		CAGT versus
LKGT	SAC	CAGT	LKGT	CAGT	LKGT	LKGT
<i>Omy_al</i> 091	B4 170	<i>Omy_al</i> 091	<i>Omy_al</i> 091	B1 102	<i>Omy_a2</i> 221	<i>Omy_al</i> 091
<i>Omy_al</i> 105	B4 193	<i>Omy_al</i> 105	<i>Omy_al</i> 105	B1 266	<i>Omy_a2</i> 249	<i>Omy_al</i> 105
<i>Omy_al</i> 134	B4 198	<i>Omy_al</i> 134	<i>Omy_al</i> 134	B9 181	<i>Omy_a2</i> 269	<i>Omy_al</i> 134
<i>Omy_a2</i> 221	B9 181	B1 102	B4 170	B9 355	B9 16	B1 102
<i>Omy_a2</i> 249	B9 202	B1 266	B4 193	B9 388	B9 164	B1 266
<i>Omy_a2</i> 269	B9 355	B4 170	B4 198	B9 90	B9 181	B9 16
B1 102	B9 90	B4 193	B9 16	E1 239	B9 182	B9 164
B1 266	E1 100	B4 198	B9 164	E1 241	B9 335	B9 182
B1 39	E1 147	B9 181	B9 181	E1 289	B9 355	B9 335
B9 16	E1 159	B9 202	B9 182	F5 306	B9 90	B9 388
B9 164	E1 239	B9 355	B9 202	F5 090	E1 100	B9 90
B9 182	E1 241	B9 388	B9 335	G9 221	E1 108	E1 100
B9 335	E1 289	B9 90	B9 355	G9 465	E1 109	E1 108
B9 37	E1 29	E1 100	E1 100	G9 520	E1 110	E1 109
E1 100	E1 30	E1 147	E1 108	G9 547	E1 111	E1 110
E1 108	F5 090	E1 159	E1 109	CRB 0826 102	E1 147	E1 111
E1 109	G9 168	E1 29	E1 110	CRB 0826 110	E1 239	E1 147
E1 110	G9 465	E1 30	E1 111	CRB 0826 152	E1 241	F5 306
E1 111	G9 502	F5 306	E1 147	CRB 0826 756	E1 289	F5 090
E1 147	G9 504	F5 090	E1 159	CTSD 033	F5 306	G9 221
E1 239	CRB 0826 102	G9 168	E1 29	CTSD 097	F5 090	G9 520
E1 241	CRB 0826 110	G9 221	E1 30	CRFBP 74	G9 465	G9 547
F5 306	CRB 0826 152	G9 502	F5 306	Omy_fl 509	G9 547	CRB 0826 110
F5 323	CRB 0826 756	G9 504	G9 168	Omy_fl 570	CRB 0826 102	CRB 0826 172
F5 090	CRB 2677 106	G9 520	G9 502	Omy_g1 103	CRB 0826 110	CTSD 033
G9 125	CRB 2677 117	G9 547	G9 504	Omy_g2 192	CRB 0826 152	CTSD 097
G9 168	CRB 2677 126	CTSD 033	G9 547	Omy_g2 330	CRB 0826 172	CTSD 158
G9 465	CRB 2677 129	CTSD 152	CRB 0826 110	Omy_g2 84	CRB 2677 106	CTSD 173
G9 547	CRB 2677 168	CTSD 158	CRB 0826 172	GH2C 323	CRB 2677 117	CTSD 174
CRB 0826 110	CRB 2677 203	CTSD 173	CRB 2677 106	GH2C 324	CRB 2677 126	CTSD 187
CRB 2358 020	CTSD 033	CTSD 174	CRB 2677 126	HSPINT5 38	CRB 2677 129	CTSD 193
CRB 2358 120	CTSD 097	CTSD 182	CRB 2677 129	HSPINT5 56	CRB 2677 166	CTSD 226
CRB 2677 106	CHIT 80	CTSD 183	CRB 2677 166	HSPINT5 98	CRB 2677 168	CTSD 271
CRB 2677 117	CRFBP 74	CTSD 187	CRB 2677 168	Omy_h1 170	CRB 2677 203	CTSD 339
CRB 2677 126	Omy_g1 103	CTSD 193	Omy_h1 75	Omy_h1 75	CTSD 033	CTSD 431
CRB 2677 129	Omy_g2 138	CTSD 226	CTSD 097	IGH 108	CTSD 097	CHIT 80
CRB 2677 168	Omy_g2 192	CTSD 271	CTSD 152	IGH 734	CHIT 80	Omy_fl 509
CRB 2677 203	Omy_g2 84	CTSD 305	CTSD 158	IGH 750	CRFBP 74	Omy_fl 570
CRB 2677 226	GH2C 323	CTSD 309	CTSD 173	IGH 754	Omy_g1 103	HSPINT5 38
CTSD 033	GH2C 324	CTSD 339	CTSD 174	IGH 757	HSPINT5 38	HSPINT5 56
CTSD 097	GH2C 501	CTSD 398	CTSD 182	IGH 777	HSPINT5 56	HSPINT5 98
CTSD 158	HSPINT5 38	CTSD 431	CTSD 183	IGH 792	HSPINT5 98	Omy_h1 170
CTSD 173	HSPINT5 56	CHIT 80	CTSD 187	IGH 839	Omy_h1 170	Omy_h1 75
CTSD 174	HSPINT5 98	CRFBP 74	CTSD 193	Omy_il 160	Omy_h1 75	IGH 108
CTSD 339	Omy_h1 75	Omy_fl 509	CTSD 226	ID1C 241	IGH 108	IGH 734
CTSD 398	IGH 108	Omy_fl 570	CTSD 271	ID1C 262	IGH 216	IGH 750
CTSD 431	IGH 216	Omy_g2 138	CTSD 305	LDH 156	IGH 220	IGH 757
CHIT 80	IGH 220	Omy_g2 192	CTSD 309	LDH 201	IGH 222	IGH 777
E1 300	IGH 222	Omy_g2 330	CTSD 339	LDH 212	IGH 262	LysII3 102
Omy_fl 259	IGH 241	Omy_g2 84	CTSD 398	LysII3 018	IGH 266	LysII5 186
Omy_fl 260	IGH 262	GH2C 501	CTSD 431	LysII3 102	IGH 275	LysIII 039
Omy_g1 103	IGH 275	HSPINT5 38	CRFBP 74	LysII3 138	IGH 280	LysIII 154
HSP70UTR 221	IGH 280	HSPINT5 56	Omy_h1 170	LysII3 193	IGH 295	LysIII 237
HSPINT5 38	IGH 295	HSPINT5 98	Omy_h1 170	LysII3 340	IGH 366	LysIII 246
HSPINT5 56	IGH 366	Omy_h1 170	HSPINT5 98	LysII5 237	IGH 465	LysIII 264
HSPINT5 98	IGH 404	Omy_h1 75	Omy_h1 170	LysIII 039	IGH 511	LysIII 377
<i>Omy_h1</i> 170	IGH 465	IGH 108	<i>Omy_h1</i> 75	LysIII 069	IGH 546	LysIII 466
<i>Omy_h1</i> 75	IGH 511	IGH 727	IGH 108	LysIII 154	IGH 547	MYC 158
IGH 108	IGH 546	IGH 734	IGH 216	LysIII 237	IGH 734	MYC 232
IGH 216	IGH 547	IGH 750	IGH 220	LysIII 246	IGH 750	Omy_ml 117
IGH 220	IGH 727	IGH 754	IGH 222	LysIII 264	IGH 754	<i>OmyP9</i> 128
IGH 222	IGH 754	IGH 757	IGH 241	LysIII 377	IGH 757	<i>OmyP9</i> 180
IGH 262	IGH 757	IGH 777	IGH 262	LysIII 466	IGH 777	<i>OmyP9</i> 270
IGH 266	IGH 777	IGH 792	IGH 266	MYC 158	IGH 792	R0336 129
	IGH 792	IGH 819	IGH 275	MYC 232	IGH 839	R0917 230

TABLE A.1—Continued.

RT-wild	RT-hatchery versus				RT-wild versus		
	SHP	SAC	CAGT	LKGT	SHP	SAC	CAGT
IGH 777	IGH 295	LysII3 340	LysII3 102	<i>Omy_il</i> 232	IGH 216	<i>Omy_il</i> 038	ID1 254
IGH 839	IGH 366	LysII5 128	LysII3 120	<i>Omy_il</i> 254	IGH 220	<i>Omy_il</i> 101	LDH 156
<i>Omy_il</i> 038	IGH 404	LysII5 199	LysII3 123	ID1C 160	IGH 222	<i>Omy_il</i> 201	LysII3 018
<i>Omy_il</i> 101	IGH 465	LysII5 215	LysII3 138	ID1C 241	IGH 241	<i>Omy_il</i> 232	LysII3 102
<i>Omy_il</i> 201	IGH 777	LysII5 237	LysII3 252	ID1C 262	IGH 262	<i>Omy_il</i> 254	LysII3 123
<i>Omy_il</i> 232	IGH 792	LysIII 062	LysII3 340	LysII3 018	IGH 275	ID1C 160	LysII3 138
<i>Omy_il</i> 254	IGH 819	LysIII 069	LysII5 128	LysII3 120	IGH 280	ID1C 241	LysII3 252
ID1C 160	IGH 844	LysIII 082	LysII5 199	LysII3 123	IGH 295	ID1C 262	LysII3 340
ID1C 241	ID1 232	LysIII 154	LysII5 215	LysII3 138	IGH 727	LDH 156	LysII5 178
ID1C 262	ID1 254	LysIII 212	LysII5 237	LysII3 252	IGH 734	LDH 201	LysII5 199
LDH 156	ID1C 160	LysIII 237	LysIII 039	LysII3 340	IGH 750	LDH 212	LysII5 215
LysII3 018	ID1C 241	LysIII 246	LysIII 062	LysII5 128	IGH 754	LysII3 018	LysIII 039
LysII3 120	ID1C 262	LysIII 264	LysIII 082	LysII5 186	IGH 757	LysII3 123	LysIII 062
LysII3 123	LDH 156	LysIII 377	LysIII 154	LysII5 199	IGH 777	LysII3 138	LysIII 082
LysII3 138	LysII3 018	LysIII 466	LysIII 212	LysII5 215	IGH 792	LysII3 193	LysIII 154
LysII3 252	LysII3 120	MYC125	LysIII 237	LysII5 237	IGH 819	LysII3 252	MYC 077
LysII3 340	LysII3 123	NITR 025	LysIII 039	LysIII 039	IGH 839	LysII3 340	MYC 158
LysII5 128	LysII3 138	NITR 13	LysIII 264	LysIII 062	IGH 844	LysII5 178	MYC 232
LysII5 178	LysII3 252	NITR 146	LysIII 377	LysIII 082	<i>Omy_il</i> 038	LysII5 199	Omy_m1 117
LysII5 199	LysII3 340	NITR 293	LysIII 466	LysIII 154	<i>Omy_il</i> 101	LysII5 215	NITR 146
LysII5 215	LysII5 128	NITR 302	MYC 158	LysIII 212	<i>Omy_il</i> 201	LysII5 237	NITR 178
LysII5 237	LysII5 178	NITR 406	MYC 232	LysIII 237	<i>Omy_il</i> 232	LysIII 062	NITR 296
LysIII 062	LysII5 199	NITR 421	Omy_m1 117	LysIII 246	<i>Omy_il</i> 254	LysIII 069	NITR 404
LysIII 082	LysII5 215	NITR 78	NITR 146	LysIII 264	ID1C 262	LysIII 082	NITR 411
LysIII 154	LysII5 237	NITR 83	NITR 293	LysIII 377	LDH 156	LysIII 154	NITR 414
MYC 077	LysII5 269	RAPD 132	NITR 296	LysIII 466	LysII3 018	MYC 077	NITR 415
NITR 025	LysIII 062	RAPD 277	NITR 302	NITR 146	LysII3 123	MYC125	NITR 416
NITR 13	LysIII 082	R0336 129	NITR 404	NITR 296	LysII3 138	NITR 025	NITR 419
NITR 146	LysIII 154	R0336 62	NITR 406	NITR 302	LysII3 252	NITR 13	NITR 83
NITR 170	LysIII 212	R0917 262	NITR 411	NITR 404	LysII3 340	NITR 146	<i>OmyP9</i> 128
NITR 178	LysIII 237	R1175 9	NITR 414	NITR 406	LysII5 178	NITR 170	<i>OmyP9</i> 180
NITR 293	LysIII 246	R1564 121	NITR 415	NITR 411	LysII5 199	NITR 178	<i>OmyP9</i> 270
NITR 302	LysIII 264	R1564 325	NITR 416	NITR 414	LysII5 215	NITR 293	RAPD 167
NITR 406	NITR 025	URO 188	NITR 419	NITR 415	LysII5 269	NITR 83	R0336 129
NITR 421	NITR 13		NITR 421	NITR 416	LysIII 062	RAPD 132	R0336 62
NITR 78	NITR 146		NITR 78	NITR 419	LysIII 082	RAPD 167	R0917 125
NITR 83	NITR 293		NITR 83	NITR 421	LysIII 154	RAPD 277	R0917 230
RAPD 132	NITR 302		<i>OmyP9</i> 128	NITR 78	MYC 077	R0336 129	R0917 391
RAPD 167	NITR 406		<i>OmyP9</i> 180	NITR 83	NITR 170	R0336 62	R0917 401
RAPD 277	NITR 421		<i>OmyP9</i> 270	<i>OmyP9</i> 128	NITR 178	R0917 262	R1175 137
R0336 129	NITR 78		RAPD 132	<i>OmyP9</i> 180	NITR 83	R1175 147	R1175 147
R0336 62	NITR 83		RAPD 277	<i>OmyP9</i> 270	RAPD 167	R1175 9	R1564 121
R0917 125	RAPD 132		R0336 129	RAPD 132	R0336 129	R1564 121	R1564 272
R0917 262	RAPD 167		R0336 62	RAPD 277	R0336 62	R1564 325	R1564 325
R1175 147	RAPD 277		R0917 125	R0336 129	R0917 262		URO 217
R1564 121	R0336 129		R0917 230	R0336 62	R1175 147		URO373
R1564 325	R0336 62		R0917 262	R0917 125	R1175 9		
URO 188	R0917 262		R0917 391	R0917 230	R1564 121		
	R1175 9		R0917 401	R0917 262	R1564 325		
	R1564 121		R1175 137	R0917 391			
	R1564 325		R1175 147	R0917 401			
	URO 188		R1564 121	R0917 455			
			R1564 272	R1175 137			
			R1564 325	R1175 147			
			URO 188	R1564 121			
			URO 217	R1564 325			
			URO373	URO 188			
				URO 217			
				URO373			

TABLE A1.—Extended.

RT-wild versus		SHP versus		SAC versus		CAGT versus
LKGT	SAC	CAGT	LKGT	CAGT	LKGT	LKGT
IGH 280	IGH 819	IGH 844	IGH 280	MYC125	ID1C 160	R0917 391
IGH 295	IGH 839	ID1C 262	IGH 295	Omy_m1 117	ID1C 241	R0917 401
IGH 734	IGH 844	LysII3 018	IGH 366	NITR 146	ID1C 262	R0917 455
IGH 750	ID1C 160	LysII3 102	IGH 404	NITR 178	LysII3 018	R1175 137
IGH 757	ID1C 241	LysII3 138	IGH 465	NITR 296	LysII3 138	R1175 147
IGH 777	ID1C 262	LysII3 252	IGH 511	NITR 404	LysII3 193	R1564 121
IGH 839	LDH 156	LysII3 340	IGH 546	NITR 411	LysII3 340	R1564 272
<i>Omy_il</i> 038	LDH 201	LysII5 178	IGH 547	NITR 414	LysII5 186	R1564 325
<i>Omy_il</i> 101	LDH 212	LysII5 199	IGH 727	NITR 415	LysII5 237	URO 217
<i>Omy_il</i> 201	LysII3 138	LysII5 215	IGH 734	NITR 416	LysIII 039	URO373
<i>Omy_il</i> 232	LysII3 193	LysII5 269	IGH 750	NITR 419	LysIII 069	
<i>Omy_il</i> 254	LysII3 252	LysIII 039	IGH 754	NITR 83	LysIII 154	
LysII3 018	LysII3 340	LysIII 062	IGH 757	<i>OmyP9</i> 128	LysIII 237	
LysII3 123	LysII5 178	LysIII 082	IGH 777	<i>OmyP9</i> 180	LysIII 246	
LysII3 138	LysII5 199	LysIII 154	IGH 792	<i>OmyP9</i> 270	LysIII 264	
LysII3 252	LysII5 215	LysIII 212	IGH 819	RAPD 132	LysIII 377	
LysII3 340	LysII5 237	LysIII 237	IGH 844	RAPD 277	LysIII 466	
LysII5 178	LysII5 269	LysIII 246	ID1C 262	R0917 230	MYC125	
LysII5 186	LysIII 062	LysIII 264	LysII3 018	R0917 262	NITR 146	
LysII5 199	LysIII 069	LysIII 377	LysII3 138	R0917 391	NITR 178	
LysII5 215	LysIII 082	MYC 158	LysII3 252	R0917 401	NITR 296	
LysIII 039	LysIII 154	MYC 232	LysII3 340	R1175 137	NITR 404	
LysIII 062	LysIII 212	Omy_m1 117	LysII5 178	R1175 147	NITR 411	
LysIII 082	LysIII 237	NITR 146	LysII5 186	R1175 9	NITR 414	
LysIII 154	LysIII 246	NITR 296	LysII5 199	R1564 272	NITR 415	
MYC 077	LysIII 264	NITR 404	LysII5 215	URO 217	NITR 416	
NITR 146	LysIII 377	NITR 411	LysII5 269	URO373	NITR 419	
NITR 170	MYC125	NITR 414	LysIII 039		NITR 83	
NITR 178	NITR 025	NITR 415	LysIII 062		<i>OmyP9</i> 128	
NITR 296	NITR 13	NITR 416	LysIII 069		<i>OmyP9</i> 180	
NITR 404	NITR 146	NITR 419	LysIII 082		<i>OmyP9</i> 270	
NITR 411	NITR 178	NITR 83	LysIII 154		RAPD 132	
NITR 414	NITR 293	<i>OmyP9</i> 128	LysIII 212		RAPD 277	
NITR 415	RAPD 132	<i>OmyP9</i> 180	LysIII 237		R0336 129	
NITR 416	RAPD 167	<i>OmyP9</i> 270	LysIII 246		R0917 230	
NITR 419	RAPD 277	RAPD 167	LysIII 264		R0917 262	
NITR 83	R0336 129	R0917 230	LysIII 377		R0917 391	
<i>OmyP9</i> 128		R0917 262	NITR 146		R0917 401	
<i>OmyP9</i> 180		R0917 391	NITR 296		R0917 455	
<i>OmyP9</i> 270		R0917 401	NITR 404		R1175 137	
RAPD 167		R1175 137	NITR 411		R1175 147	
R0336 129		R1175 147	NITR 414		R1175 9	
R0336 62		R1175 9	NITR 415		R1564 121	
R0917 125		R1564 272	NITR 416		R1564 325	
R0917 230		URO 217	NITR 419		URO 217	
R0917 391		URO373	NITR 83		URO373	
R0917 401			<i>OmyP9</i> 128			
R0917 455			<i>OmyP9</i> 180			
R1175 137			<i>OmyP9</i> 270			
R1175 147			RAPD 167			
R1564 121			R0336 129			
R1564 325			R0917 230			
URO 217			R0917 262			
URO373			R0917 391			
			R0917 401			
			R0917 455			
			R1175 137			
			R1175 147			
			R1175 9			
			R1564 121			
			R1564 325			
			URO 217			
			URO373			