Report to the Northern Endowment Fund on project NF-2007-I4

Chinoook Salmon SNP Development

Kristi M. Miller, Terry D. Beacham, and Tobi J. Ming

Pacific Biological Station

Nanaimo, B.C.

Sept 11, 2008

Abstract

Eleven new SNPs for Chinook salmon were developed and applied to a baseline that included 20 stocks from California, Washington, BC, and Alaska. The range in stock differentiation (Fst) was similar to that observed with SNP loci in the existing CTC baseline. Two of the newly identified SNPs were among the most powerful uncovered thus far, and showed evidence of positive selection. These SNPs are now ready for addition to the CTC baseline.

Introduction

Single nucleotide polymorphisms (SNPs) are a class of genetic marker based upon DNA-level variation where a single base pair difference between alleles at a defined position in the genome constitutes a polymorphism. The four DNA bases are adenine (A), guanine (G), cytosine (C), and thymine (T). A deletion may also give rise to a SNP, where a deletion is the absence of the defined base or bases in an individual due to mutation. Identifying positions in the genome, or within targeted genetic markers, where SNPs occur involves initially developing amplifying and sequencing the DNA from a broad range of stocks. Once SNPs have been identified, assays can be developed to genotype individuals without the need of direct sequencing.

Microsatellite and SNP loci have the potential to resolve different aspects of population structure. Most microsatellite loci are considered selectively neutral, hence the patterns of genetic variation most closely reflect the demographics (or linkage history) of populations. Alternately, SNPs can be developed in coding genes potentially under selection, which enables the resolution of adaptive differentiation among populations. We have previously shown through our extensive research on MHC variation that adaptive variation can result in much higher levels of differentiation among proximate populations, hence providing powerful markers for stock ID (Beacham et al. 2004). Hence, in this proposal, we attempted to find genes for SNP development that carried a high probability of being under selection. To do this, we used data from our gene array program on sockeye migration physiology, and chose genes that were differentially regulated along the migration path, some of which were also differentially regulated with respect to different stocks and that were associated with fate in the river. We developed primers from these genes using sequences from Atlantic salmon and rainbow trout, the two species with large expressed sequence databases, and used them to amplify the sequences in a wide range of Chinook salmon samples. SNPs were detected through sequencing.

The SNPs developed in the current project can be used to update existing baselines. We are ultimately looking towards developing a stock ID baseline that is not only easily transportable between laboratories and faster and cheaper to use for both in and out of season applications, but is also economical to develop. It is important that the costs associated with building new baselines include both the cost of marker development and the cost of running the markers on fish in the baseline. SNPs have proven notoriously difficult and expensive to develop, in part due to the tetraploid nature of salmon (i.e. many loci are duplicated). However, judicious choice of the most powerful SNPs combined with a set of

powerful, easy to score microsatellite loci could increase baseline transportability and resolution (Beacham et al. 2005).

Methods

We used three sources to identify SNPs in Chinook salmon that carried a high probability of being under selection. First, ninety-two genes that were differentially expressed in gill and brain tissue during spawning migration of Fraser River sockeye salmon were screened in Chinook salmon for SNPs. Second, thirty-two EST sequences were identified as possible SNP targets from the GRASP Chinook EST database (kindly provided by Ben Koop, UVIC). Third, three genes identified in a Loma infection study in Chinook salmon were screened for SNPs, for a total of 130 genes screened (Table 1).

For each of the differentially expressed genes from the migration study, sequences from Atlantic salmon and Rainbow trout were obtained through BLAST searches of the EST databases from "The Institute for Genomic Research" (TIGR) and the "Genomic Research on Atlantic Salmon Project" (GRASP) and aligned in Sequencher. Because the duplication of the salmon genome through tetraploidization occurred prior to the split between Atlantic and Pacific salmon species, we attempted to identify the duplicated copies of genes through alignment of sequences from both species, with the assumption that there would only be two copies of each gene. As such, the orthologs (same gene copy) in each species tended to contain higher sequence identity than the homologs (duplicated copies) within a species. Using AlleleID 4.0 software (Premier Biosoft International), we attempted to design primers that would amplify only one of the duplicated copies of the gene, that which reflected the copy that was differentially expressed during migration. Primers were generally designed to amplify cDNA products of 500-1,000 bases. Because these were to eventually be applied to genomic DNA, which contains additional non-coding introns of variable size, we also designed some smaller primer sets to apply if genes contained large introns. In most cases, we did not have the intron-exon organization of the gene available for initial primer design. Hence, we designed roughly two to four primer sets per gene with the expectation that some would not, by chance, be placed over an intron.

To screen for SNPs, we used separate pooled samples of cDNA and DNA and amplified each with primers from the targeted markers. The cDNA pool was comprised of 34 Chinook salmon representing Washington stocks from the Columbia and Snake Rivers and coastal Washington, and BC stocks from Vancouver Island (Table 2). The gDNA pool was comprised of 94 individuals representing California, Washington, British Columbia, Yukon Territory, and Alaska. For a subset of the genes/primer sets that amplified small enough product sizes (<600 bases), perpendicular denaturing gradient gel electrophoresis was performed on pooled amplified products to determine the presence or absence of SNPs therein. Alternately, if the PCR yielded the amplification of a single band of the correct size, pooled amplification products were directly sequenced. If the sequence contained insertions or deletions that made it difficult to sequence through, products were cloned into a TA vector and 8 clones were sequenced. Sequences were obtained in both directions from each gene of interest and analyzed with SEQUENCHER 4.5 (GeneCodes Corporation). Once potential SNPs were identified, sequences

were sent to the Assay-By-Design service from Applied Biosystems for the development of TaqMan assays.

All SNP assays were performed in 384-well reaction plates. Each 6 uL reaction contained 20 ng template DNA, 900 nM of each PCR primer, 200 nM of each probe, and 1X Taqman Universal PCR Master Mix, No AmpErase UNG or TaqMan Genotyping Master Mix (Applied Biosystems, New Jersey, USA). The DNA was aliquoted using a Beckman Coulter Biomek FX Liquid Handler (Fullerton, CA, USA). Each 384-well plate contained at least 2 no template controls, and 1 positive control for each allele. Thermal cycling was performed on an ABI 7900 real-time sequence detection system. The thermal cycling protocol contained an initial 10 min incubation at 95 degrees, followed by 50 cycles of 92 degrees for 15 sec., and an annealing-extension temperature of 60 for 90 sec. Scoring of individuals was performed using the ABI 7900 Sequence Detection System 2.3 which generates scatter plots denoting which samples had which combination of alleles.

In addition to the SNPs designed, one assay was developed for the exon 1 sequence of the MHC class 1 gene for which we have extensive sequence information. A six base pair deletion is present in half of the known chinook alleles and was treated statistically as a SNP assay. A PCR based assay was designed to distinguish the presence or absence of this deletion, with a primer for the deletion containing alleles labelled with VIC, and another for the insertion labelled with FAM. Amplifications were performed with 0.9mM dNTP, 1.8nmol each primer, 0.18U HotStar Taq (Qiagen), 1X Taq buffer (Qiagen), 20ng DNA in a 6uL total volume. The thermal cycling protocol was an initial 15 minute incubation at 95º to activate the enzyme, the 35 cycles of 30s at 95º, 30s at 50º, 30s at 72º. The amplified products were pooled and run on an AB 3730 automated sequencer, and scored using GeneMapper v. 3.7 (Applied Biosystems, New Jersey, USA)

Individual SNP assays were tested on a sample of fish representing the 20 stocks laid out in the proposal, with a minimum of 94 individuals per stock surveyed, and 2008 fish surveyed in total (Table 4).

GENEPOP v3.2a (Raymond and Roussett 1997) was used to calculate Fst (Weir and Cockerham 1984), expected (He) and observed (Ho) heterozygosities, the inbreeding coefficient (Fis) and P-values for departures from Hardy-Weinburg equilibrium (HWE). Dendrograms were constructed using GDA (Lewis and Zaykin, 2001). Outlier analysis of SNPs for balancing, directional, or neutral selection was performed with Lositan (Beaumont and Nichols, 1996; Antao, et. al., 2008)

Results

SNP Discovery

130 genes total were surveyed for SNPs (Table 1). Of these, 31% amplified in genomic DNA and 72% amplified in cDNA. Most of the amplifications that worked for genomic DNA also worked using cDNA. When both genomic and cDNA were amplified, we sequenced or separated on DGGE products from both to find SNPs. This approach broadened the range of individuals surveyed to include a

wider sampling of stocks. Alternately, when only cDNA amplified (likely because primers were placed over introns or products spanned large introns) our survey for SNPs included mostly Washington state stocks, as this was the only cDNA available (Table 2).

Overall, 28% of the amplified genes contained SNPS and 56% of the genes with SNPs amplified in both cDNA and gDNA. Twenty-eight assays were designed, of which 17% were confounded by duplicated gene copies, despite the fact that we took considerable time to try to develop single gene assays. An additional 31% of the assays did not work well, and 14% were confounded by the presence of introns. Thus far 8% of all genes and 11% of amplified genes yielded functional SNP assays. SNPS were observed at a frequency of one in every 896 base pairs.

SNP Application

94 to 142 samples from each of the twenty stocks identified in the proposal were screened with each of the SNP assays designed (Table 3). Expected heterozygosities over all SNP loci ranged from 0.099 to 0.428 (Table 4). Most stocks contained both alleles of each SNP locus. Exceptions included EST 1363, which was monomorphic in the Feather and Johnson stocks, and EST 1803, monomorphic in the Johnson, Situk, and Twisp stocks. To gain a measure of the range of variation among stocks, we calculated the maximum frequency difference among stocks of the subdominant allele, and obtained a range in frequencies of 0.136 (MHC A1) to 1.0 (EST 1363), with a mean of 0.613.

Overall Fst values ranged from 0.084 (Phos) to 0.452 (GCSH) (Table 3). Outlier analysis revealed that EST 1363 and GCSH, which contained the highest Fst values of the 11 markers, were under positive selection. Four SNPs were under neutral selection, and the final five were under balancing selection (Figure 1, Table 2).

UPGMA clustering based on the 11 SNP loci resolved three clusters of populations, but showed no distinct regional groupings. The most highly differentiated cluster contained 5 stocks from a wide range of locations, including two from Washington (Twisp and Johnson), one from Skeena (Bulkley), and two from northern British Columbia or Alaska (LittleTatsamenie and Situk) (Figure 2). Examination of the allele frequencies showed that the these five stocks were distinct at multiple SNP loci (Table 5).

Fst ranges of the SNP loci reported here were comparable to the CTC database for the four populations which overlapped (Bulkley, Kateen, Little Tatsemanie and Nanaimo). The CTC SNP locus Fsts ranged from 0.001 to 0.493, and the new assays reported here range from 0.084 to 0.452 (Table 3).

Discussion

This SNP discovery project yielded 11 informative SNPs, one more than in the original proposal. Using genes differentially expressed in migration improved the frequency of SNP discovery. Smith et al (2005) found that SNPS are present in the salmon genome at a frequency of approximately 1 in 4300 base pairs, whereas we were able to observe SNPS at a frequency of 1 in 896. This number is comparable to the frequency observed in Atlantic salmon (1 in 962; Hayes et al. 2007). This is promising for future SNP discovery in genes of migratory interest.

While we were able to skip the sequencing step in the survey of SNPs from EST databases, only 9% of these putative SNPs provided functional assays compared to 13% amplified migration genes and sequenced in lab. However, there was no apparent bias towards SNPs under positive selection from the migration study, as there was one each SNP locus under positive selection identified using each screening method (EST 1363, GCSH). The higher percentage of migration genes which failed to amplify is most likely due to the lack of Chinook salmon-specific sequence for these genes. As the Chinook salmon sequence database is enlarged, it is likely that SNP discovery will become more efficient.

The low percentages of SNPs identified to genes screened may have been due to the restricted stocks for which we had cDNA (Table 4). Only 31% of all genes amplified in genomic DNA, but this small proportion provided 64% of functional assays, possibly reflecting the wider range of stocks surveyed. Most of the genes for which we only had cDNA sequence contained no SNPs. However, these were also shorter sequences, given the lack of introns. Screening genes with a group of cDNA samples representing a wider geographical area would have presumably illuminated more SNPs.

Class II MHC SNPs have been previously developed in sockeye salmon by ourselves (see SEF SNP report for sockeye in 2007) and Elfstrom et al. (2006). Our Class II MHC sockeye SNP was reasonably powerful, with an Fst of 0.122, and Elfstrom's Fsts ranged from 0.22 to 0.38. The finding that the Class I MHC SNP developed here carried one of the lowest Fst's (0.031) contradicts our genotyping data that encompasses all exon 2 and 3 sequence variants, whereby MHC class I and II are more powerful, on average, than most microsatellite loci. Unfortunately, it is not possible to design SNP assays for each allelic variant due to the low GC content and high frequency of SNPs in this gene (Miller et al. 2001).

The presence of duplicated loci, as previously shown in our sockeye survey, was a significant hurdle to overcome in the design process. Denaturing gradient gel electrophoresis was not always diagnostic in determining if multiple loci were amplified. The salmon genome was duplicated through tetraploidization in the common ancestor of extant salmonids (Allendorf and Thorgaard, 1984). As a result, most genes contain two duplicated copies that can, at times, be difficult to differentiate. Hence, if one is not careful in identifying the sequence of both copies of a gene, many identified SNPs can arise not from allelic variation within a single gene but from variation between duplicated copies of the gene

(Smith et al. 2005). Those that arise from duplicated copies of genes will be monomorphic (meaning that each individual carries both copies) and uninformative whereas allelic SNPs are polymorphic (individuals vary in the copies they inherit from their parents). Despite our best efforts, duplicated copies of genes were frequently amplified, lowering the percentage of successful assays.

In conclusion, the deliverables for this SNP discovery study were all met, and we have added 11 SNPs to the pool of Chinook SNPs available. Two of these are among the most powerful detected thus far. Our technique of targeting genes putatively under selection was fruitful, but did not yield a higher than average number of genes under positive selection. One reason for this may have been that the discovery of these genes was done in a different species (sockeye salmon). Overall, while developed SNPs are easy to use and quick to survey, SNP discovery remains a time consuming, expensive process. Without an extensive sequence database in the species from which SNPs are required, this will not change. While it appears that the Atlantic salmon genome is now slated for full sequencing, even this will not directly aid in the discovery of new SNPs for Pacific salmon, but it may enable better annotation of EST sequences. The availability of annotation is useful for identifying intron/exon boundaries and can aid in the development of primers that will work for SNP discovery in genomic DNA.

References

Allendorf, F.W. and Thorgaard, G.H. 1984. Tetraploidy and the evolution of salmonoid fishes. In: Turner, J.B. (Ed.), Evolutionary Genetics of Fishes. Plenum Press, New York, pp. 1-53.

Antao, T., et. al. 2008. Lositan: A workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics. 9:323.

Beacham, T.D., M. Lapointe, J.R. Candy, B. McIntosh, C. MacConnachie, A. Tabata, K. Kaukinen, L. Deng, K.M. Miller, and R.E. Withler. 2004. Stock identification of Fraser River sockeye salmon using microsatellites and major histocompatibility complex variation. Transactions of the American Fisheries Society 133: 1117-1137.

Beamont, M. A. and Nichols, R.A. Evaluating loci for use in the genetic analysis of population structure. Proceedings: Biological Sciences, Vol. 263, No. 1377 1619-1626.

Hayes, B., J.K. Laerdahl, S. Lien, T. Moen, P. Berg, K. Hindar, W.S. Davidson, B.F. Koop, A. Adzhubei, and B. Hoyheim. 2007. An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (Salmo salar) expressed sequences. Aquaculture 265: 82-90.

Lewis, P.O. and Zaykin, D. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from http://lewis.eeg.uconn.edu/lewishome/software.html

Miller, K.M., K.H. Kaukinen, T.D. Beacham, and R.E. Withler. 2001. Geographic heterogeneity in natural selection of an MHC locus in sockeye salmon. Genetica 111: 237-257.

Smith, CT, CM Elfstrom, LW Seeb, and JE Seeb. 2005. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Mol. Ecol. 14(13): 4193-203.

Figures and Tables

Table 1. Genes targeted for SNP development.

SSMS = Sockeye salmon migration study, GRASP = Genome Research on Atlantic Salmon Project

Gene name	sequence source	Primer name	#BP SURVEYED	cDNA amplification	genomic DNA amplification	SNPS	Assay design	Results
Working SNP Assays								
1 Unknown	GRASP	Est 803	659	659	659	1	AB Assays-by-design	GOOD
2 Unknown	GRASP	Est 1363	672	672	672	1	AB Assays-by-design	GOOD
3 Receptor expression-enhancing protein 5	GRASP	Est 740	737	737	737	1	AB Assays-by-design	GOOD
4 Class1 MHC exon 2	MHC	MHC A1	252	252	252	6 bp indel	Allele ID 4.0	GOOD
5 Phospholemman	SMMS	Phos	469	469	0	1	AB Assays-by-design	GOOD
6 Isopeptidase T	SMMS	IsoT	532	0	532	1	AB Assays-by-design	GOOD
7 Thioredoxin	SMMS	Thio	649	649	0	1	AB Assays-by-design	GOOD
8 Cold-inducible RNA binding protein A	SMMS	CirpA	362	362	362	1	AB Assays-by-design	GOOD
9 Glycine cleavage system H protein	SMMS	GCSH	504	504	0	1	AB Assays-by-design	GOOD
10 Osteopetrosis associated transmembrane protein 1	SMMS	Ostm1	464	464	464	3	AB Assays-by-design	GOOD
11 Phosphatidylethanolamine N-methyltransferase	SMMS	PEMT	415	415	0	1	AB Assays-by-design	GOOD
SNP Assays Designed								
12 Unknown	GRASP	EST 628	810	950	0	1	AB Assays-by-design	poor distinction
13 Unknown	GRASP	EST 941	722	850	100	2 bp indel	AB Assays-by-design	poor distinction
14 Unknown	GRASP	EST 1590	707	800	0	1	AB Assays-by-design	poor distinction
15 Vacuolar ATP synthase subunit C 1-A	SMMS	vatpc1a	478	478	0	4	AB Assays-by-design	duplicated
16 Cold-inducible RNA binding protein C	SMMS	CirpC	251	251	251	3	AB Assays-by-design	intron interference
17 FK506-binding protein 2 precursor	SMMS	Fkbp2	585	185	585	1	AB Assays-by-design	duplicated
18 Unknown	SMMS	EST 628	810	810	0	1	AB Assays-by-design	duplicated
19 Proteasome alpha 5	SMMS	PSMA5	369	0	281	4	AB Assays-by-design	duplicated
20 Activating transcription factor 7-interacting protein 1	SMMS	atf7ip	574	574	0	7	AB Assays-by-design	duplicated
21 Heat shock protein 90	SMMS	Hsp90	673	673	673	1	AB Assays-by-design	intron interference
22 Tropomodulin	SMMS	Tropo	497	0	500	3	AB Assays-by-design	intron interference
23 Unknown	SMMS	EST 941	722	722	722	1	AB Assays-by-design	intron interference
24 Unknown	Loma study	Ci L43	325	325	0	3	AB Assays-by-design	poor distinction
25 Transglutaminase	SMMS	BTG1	762	350	750	1	AB Assays-by-design	poor distinction
26 Ectonucleoside triphosphate diphosphohydrolase 2	SMMS	Etd2a	625	231	625	1	AB Assays-by-design	poor distinction
27 Actin-related protein	SMMS	L124	426	426	0	1	AB Assays-by-design	poor distinction
28 Unknown	SMMS	EST 1590	707	707	0	1	AB Assays-by-design	poor distinction
29 Unknown	Loma study	Ci L124	426	426	0	1	AB Assays-by-design	poor distinction
Amplified but no SNPs or obviously duplicated genes								
30 Unknown	Loma study	Ci L121	434	434	0	1		SNP too close to end
31 SPARC protein	GRASP	EST 434	850	850	0	indel		
32 Unknown	GRASP	EST 921	691	700	0	0		
33 Unknown	GRASP	EST 952	851	1000	0	0		
34 Unknown	GRASP	EST 1095	776	1000	1000	0		
35 Unknown	GRASP	EST 1184	684	0	750	0		
36 Unknown	GRASP	EST 1752	650	650	650	0		
37 Unknown	GRASP	EST 2061	600	600	600	0		
38 Serum albumin 2 precursor	SMMS	alb2	374	374	0	0		
39 Ankyrin repeat and SOCS box protein 12	SMMS	Asb12	399	399	0	0		
40 Methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial precursor	SMMS	Aldh6a1	437	437	1000	0		
41 Pyruvate dehydrogenase kinase isozyme 1	SMMS	PDK1	489	489	0	0		
42 Isoleucyl-tRNA synthetase, cytoplasmic	SMMS	IARS	300	300	0	0		
43 Ectonucleoside triphosphate diphosphohydrolase 2	SMMS	Etd2a		595	0	0		
44 Caspase 8	SMMS	Casp8	606	606	0	0		
45 Heat shock protein 70	SMMS	Hsp70	300	300	600	0		
46 Aldo-keto reductase family1	SMMS	AKID1	575	575	0	0		
47 Ependymin	SMMS	EpdRP	595	600	1800	0		
48 FUS interacting serine-arginine rich protein	SMMS	Fus2	442	442	442	0		
49 Selenoprotein	SMMS	Sepp1B	625	625	625	0		
50 Selenoprotein	SMMS	Sepp1a	648	200	500	0		
51 Cathepsin S	SMMS	CtsS	625	625	0	0		
52 Cathepsin H	SMMS	CtsH	800	800	0	0		
53 CXXC finger protein 1	SMMS	CFP-1	550	550	0	0		
54 Myeloid specific peroxidase	SMMS	MPX	184	184	184	0		
55 Inhibitor of nuclear factor K B A	SMMS	Chuk	250	250	450	0		
56 Slow myosin heavy chain	SMMS	sMHC	204	0	204	0		
57 Cathepsin L	SMMS	CtsL	300	0	300	0		
58 Cathepsin X	SMMS	CtsX FL2/RL2	753	753	0	0		
59 CD74	SMMS	CD74	241	241	0	0		
60 Trifunctional enzyme subunit beta, mitochondrial precursor	SMMS	HadHBa	508	508	0	0		
61 Thioredoxin	SMMS	ThioC	500	500	1100	0		
62 Succinate dehydrogenase cytochrome b560 subunit, mitochondrial precursor	SMMS	SDHC	400	400	0	0		
63 Ornithine aminotransferase, mitochondrial precursor	SMMS	OAT	534	534	534	indel		
64 Keratin, type I cytoskeletal 13	SMMS	KRT13	575	575	0	indel		
65 3-ketoacyl-CoA thiolase, mitochondrial	SMMS	Accaa2a	558	558	0	duplicated		
66 NAD(P) transhydrogenase, mitochondrial precursor	SMMS	NNT	435	435	0	duplicated		
67 Complement C4-B precursor	SMMS	C4B	500	500	0	duplicated		
68 Creatine kinase M-type	SMMS	CKM	598	598	450	duplicated		
69 Protein-arginine deiminase type-2	SMMS	Padi2	750	400	750	indel		

Table 1 continued. Genes targeted for SNP development.

SSMS = Sockeye salmon migration study, GRASP = Genome Research on Atlantic Salmon Project

	Gene name	sequence source	Primer name	#BP SURVEYED	cDNA amplification	genomic DNA amplification	SNPS	Assay design	Results
70	FAD-dependent oxidoreductase domain-containing protein 1	SMMS	Foxred	0	doublet	0	duplicated		
71	Unknown	GRASP	EST 1190	0	850	0	duplicated		
72	Unknown	GRASP	EST 1370	0	750	750	duplicated		
73	Unknown	GRASP	EST 113	0	multiples	1300	duplicated		
74	Glucose phosphate isomerase a	GRASP	EST 1080	0	900	multiples			
75	Unknown	GRASP	EST 1102	0	450	multiples			
76	Unknown	GRASP	EST 1185	0	600	multiples			
77	Beta-actin like protein	GRASP	EST 1584	0	700	multiples			
78	Unknown	GRASP	EST 1810	0	400	multiples			
79	Unknown	GRASP	EST 1745	0	1000	multiples			
80	Unknown	GRASP	EST 473	0	800	multiples			
81	Unknown	GRASP	EST 504	0	700	multiples			
82	Unknown	GRASP	EST 611	0	850	multiples			
83	Unknown	GRASP	EST 1647	0	700	multiples			
84	Keratin	GRASP	EST 1652	0	800	multiples			
85	Unknown	GRASP	EST 1242	0	750	multiples			
86	Unknown	GRASP	EST 962	0	850	2000	messy seq		
87	Unknown	GRASP	EST 1503	0	950	0	duplicated		
88	Unknown	GRASP	EST 2122	0	0	600	messy seq		
89	Tubulin 6 alpha	GRASP	EST 1905	0	multiples				
90	Unknown	GRASP	EST 615	0	1000	0	duplicated		
91	Dickkopfs-3	SMMS	DKK3	0	multiples	0	duplicated		
92	Prion protein	SMMS	PRP31	0	550		duplicated		
93	Stathmin	SMMS	Stath	0	multiples	150	duplicated		
94	Cathepsin B1	SMMS	CtsB1	0	multiples	0	duplicated		
95	Cathepsin B2	SMMS	CtsB2a	0	700	0	duplicated		
96	Cathepsin B2	SMMS	Cts B2b	0	350	350	duplicated		
97	TGFB inducible early gene	SMMS	Tieg	0	multiples	multiples	duplicated		
98	Slow myosin heavy chain	SMMS	SMHC	0	0	multiples	duplicated		
99	Cathepsin D	SMMS	Ctsd	0	multiples	0	duplicated		
100	Fibre cell intrinsic membrane protein	SMMS	MP19	0	0	multiples	duplicated		
101	Proteasome regulatory 6	SMMS	PSMR6	0	0	multiples	duplicated		
No amplified product									
102	S-arrestin	SMMS	SAG	0	0	0			
103	Proproteinase E precursor	SMMS	PPE	0	0	0			
104	Mitochondrial import inner membrane translocase subunit Tim9	SMMS	timm9	0	0	0			
105	Growth hormone alpha	SMMS	GTHa	0	0	0			
106	Collagen	SMMS	Coll	0	0	0			
107	Heat shock protein 47	SMMS	HSP47	0	0	0			
108	Heat shock protein 10	SMMS	HSP10	0	0	0			
109	Rhogap	SMMS	Rhogap	0	0	0			
110	Rhodopsin	SMMS	Rhodopsin	0	0	0			
111	Cathepsin L	SMMS	CtsL1	0	0	0			
112	Myosin heavy chain	SMMS	fMHC	0	0	0			
113	Proteasome subunit n3	SMMS	PSMN3	0	0	0			
114	Cytochrome	SMMS	P450	0	0	0			
115	Ostm-1	SMMS	OSTM-1	0	0	0			
116	Oocyte protease inhibitor	SMMS	OP1-1	0	0	0			
117	Caspase 8	SMMS	Casp8_A	0	0	0			
118	Caspase 8	SMMS	Casp8_AB	0	0	0			
119	N-formylpeptide receptor-like 2	SMMS	FPR2	0	0	0			
120	Ubiquitin E1	SMMS	Ube1c	0	0	0			
121	Actin	SMMS	actB	0	0	0			
122	Cytochrome C	SMMS	CYTC	0	0	0			
123	Tubulin 2 alpha chain	SMMS	tba2	0	0	0			
124	Cathepsin B	SMMS	CtsB	0	0	0			
125	Proteasome subunit b3	SMMS	PSMB3	0	0	0			
126	Angiogenin2	SMMS	Ang2	0	0	0			
127	Chloride channel 2	SMMS	CLC2	0	0	0			
128	Gonadotropin releasing hormone	SMMS	GnRH 80xT (L1)	0	0	0			
129	Glycogen phosphorylase, muscle form	SMMS	Pygm	0	0	0			
130	Sar1 homolog B	SMMS	Sar1B	0	0	0			

DNA	Region	N
cDNA	Columbia River	14
cDNA	Snake River	16
cDNA	Vancouver Island	1
cDNA	Coastal Washington	3
genomic	Lower Fraser River	12
genomic	Alaska	30
genomic	Puget Sound	6
genomic	Columbia River	12
genomic	Skeena River	6
genomic	Yukon Territory	6

Table 2. Samples pooled for initial SNP screening

Population	Region	N
Bulkley (replaces Unuk)	Skeena River	94
Tahini River	SE Alaska	142
Situk River	SE Alaska	117
Nanaimo	Vancouver Island	94
Lower Kalum (Kitsumkalum)	Skeena River	95
Harrison	Fraser River	95
Little Tatsamenie	Alaska	94
Kateen	Nass River	94
Hanford Reach	Upper Columbia River	96
Spring Creek Hatchery	Mid Columbia River	94
Johnson Creek	Snake River	96
Cowlitz Hatchery	Lower Columbia River	95
Nestucca falls	Oregon	96
N. Umpqua	Oregon	94
Feather Hatchery	California	132
Eel River	California	96
Soos Creek Hatchery	Puget Sound	96
Lyons Ferry Hatchery	Snake River	96
Twisp (replaces Cle Elum)	Upper Columbia River	96
Solduc (replaces Forks Creek)	Coastal Washington	96

Table 3. Populations screened with SNP assays

Gene	Abbreviations	Function	Primer and Probe sequences	n	He	Ho	Fis	Q	Fst	Selection
Unknown	Est 803	Unknown	F:ACCACAGAATTGCCAATGGAAATAGA R:CCTCCTGAGCACTCTTACAATCCTA rep1:AAACGTAGTACTTACAGAAAT rep2:ACGTAGTACTTAAAGAAAT	1952	0.255	0.255	0.000	0.473	0.159	neutral
Unknown	Est 1363	Unknown	F:GGTGATTTTGCCACAGAGTAGAGAT R:AGTGTTAAATGTAACTTCAGTATACAGGCAAT rep1:CCATCCTGTCTTGTCTG rep2:CATCCTGTCATGTCTG	1966	0.225	0.221	0.021	1.000	0.492	positive
Receptor expression-enhancing protein 5	Est 740	odorant response	F:GGACTCGTGCTTGAGGAAGATG R:TGCATGGCTCCAACTCCTT rep1:TCTGGATGGAACCGTTAG rep2:CTGGATGGAGCCGTTAG	1955	0.417	0.413	0.009	0.583	0.104	balancing
Class1 MHC exon 2	MHC A1	immunity	F1: GTACTGTGGCAGTGTTTCTGC F2: TACTGTGGGTCTGCTGCC R: CCTGACTCACGCCCTGAAGTA	1738	0.408	0.546	-0.343	0.136	0.031	balancing
Phospholemman-like protein	Phos	osmoregulation	F:GGGCATACTTTACCACAGCTAATGA R:CCGTGTTGTTGGTCTGTCTATGG rep1:CAGGATGAAGAGCGTG rep2:AGGATGAAAAGCGTG	1931	0.200	0.249	-0.243	0.367	0.084	balancing
Isopeptidase T	IsoT	ubiquitin cycle	F:GACTCAGGTAAGGAAACATCAATGTCA R:GAAAGCAAAGCATTTTATCCACCACTA rep1:AACCAGTAGAATAACC rep2:CAGTGGAATAACC	1947	0.099	0.090	0.091	0.365	0.196	neutral
Thioredoxin	Thio	regulation of cellular processes	F:TTTTAAAAATGGAGATAAACTCCTGACCTGAA R:AATACCAAACCATGCCACTAAATCCT rep1:CAGTGTATTAGTCATTCTTA rep2:CAGTGTATTAGTCGTTCTTA	1927	0.428	0.450	-0.053	0.575	0.109	balancing
Cold-inducible RNA binding protein A	CirpA	salinity responsive	F:GCTGTGATTGTGCTCTAAAGACATG R:CTCCCACTTAGCATTCCTACCTT rep1:AATGCATTACAGAACTGA rep2:AATGCATTACAAAACTGA	1866	0.359	0.362	-0.009	0.778	0.233	neutral
Glycine cleavage system H protein	GCSH	glycine catabolism	F:GTTCTTTTTAATGATGACTACAGGTCTTTCAC R:GCTACTTTACATAATACCATTTGAGCTGAGA rep1:TATCTGGGCGGGCTG rep2:CTATCTGGACGGGCTG	1894	0.263	0.259	0.015	0.955	0.452	positive
Osteopetrosis associated transmembrane protein 1	Ostm1	osteoclast resorption	F:CCAGCCCCGTAACACACAT R:GAGAGGAAGCAGAAAGGTCGTTTAA rep1:CCGTGGTATTGTTTCAA rep2:CCGTGGTATTCTTTCAA	1925	0.409	0.438	-0.072	0.564	0.090	balancing
Phosphatidylethanolamine N-methyltransferase	PEMT	phosphatidylcholine biosynthesis	F:AGAGCATTCAATTTAAAAGCTGAAAACGA R:CTTTGATCCCTGCTTGCAGTATTTT rep1:CAAGTCTTAGCAATGCA rep2:CAAGTCTTAACAATGCA	1859	0.384	0.307	0.203	0.949	0.162	neutral
						Average		0.613	0.192	

Table 4. Statistical analysis of SNPs. **Abbreviations: N=sample size, A=# alleles, He=Expected heterozygosity, Ho=Observed heterozygosity, Fis=inbreeding coefficient, Q=maximum range in frequency of subdominant allele, Fst=divergence among stocks**

Table 5. Allele frequencies over stocks of interest.

SNP assay	Stock	Region	Allele A	Allele B
EST 1363	Bulkley	Skeena River	0.553	0.447
EST 1363	Cowlitz	Lower Columbia R.	0.817	0.183
EST 1363	EelRiver	California	0.968	0.032
EST 1363	Feather	California	1	0
EST 1363	HanfordReach	Upper Columbia R.	0.615	0.385
EST 1363	Harrison	Fraser River	0.691	0.309
EST 1363	Johnson	Snake River	0	1
EST 1363	Kateen	Nass River	0.84	0.16
EST 1363	Kitsumkalum	Skeena River	0.779	0.221
EST 1363	LittleTatsamenie	Taku River	0.043	0.957
EST 1363	LyonsFerry	Snake River	0.957	0.043
EST 1363	Nanaimo	Vancouver Island	0.909	0.091
EST 1363	Nestucca	Oregon	0.879	0.121
EST 1363	NUmpqua	Oregon	0.823	0.177
EST 1363	Situk	Alaska	0.139	0.861
EST 1363	Solduc	Coastal Washington	0.989	0.011
EST 1363	SoosCreek	Puget Sound	0.3	0.7
EST 1363	SpringCreek	Mid Columbia River	0.245	0.755
EST 1363	Tahini	Alaska	0.964	0.036
EST 1363	Twisp	Upper Columbia R.	0.718	0.282
EST 1740	Bulkley	Skeena River	0.569	0.431
EST 1740	Cowlitz	Lower Columbia R.	0.483	0.517
EST 1740	EelRiver	California	0.821	0.179
EST 1740	Feather	California	0.674	0.326
EST 1740	HanfordReach	Upper Columbia R.	0.629	0.371
EST 1740	Harrison	Fraser River	0.716	0.284
EST 1740	Johnson	Snake River	0.469	0.531
EST 1740	Kateen	Nass River	0.569	0.431
EST 1740	Kitsumkalum	Skeena River	0.679	0.321
EST 1740	LittleTatsamenie	Taku River	0.59	0.41
EST 1740	LyonsFerry	Snake River	0.582	0.418
EST 1740	Nanaimo	Vancouver Island	0.404	0.596
EST 1740	Nestucca	Oregon	0.59	0.41
EST 1740	NUmpqua	Oregon	0.685	0.315
EST 1740	Situk	Alaska	0.841	0.159
EST 1740	Solduc	Coastal Washington	0.258	0.742
EST 1740	SoosCreek	Puget Sound	0.776	0.224
EST 1740	SpringCreek	Mid Columbia River	0.387	0.613
EST 1740	Tahini	Alaska	0.837	0.163
EST 1740	Twisp	Upper Columbia R.	0.468	0.532
EST 1803	Bulkley	Skeena River	0.473	0.527
EST 1803	Cowlitz	Lower Columbia R.	0.244	0.756

EST 1803	EelRiver	California	0.047	0.953
EST 1803	Feather	California	0.368	0.632
EST 1803	HanfordReach	Upper Columbia R.	0.471	0.529
EST 1803	Harrison	Fraser River	0.452	0.548
EST 1803	Johnson	Snake River	0	1
EST 1803	Kateen	Nass River	0.188	0.812
EST 1803	Kitsumkalum	Skeena River	0.112	0.888
EST 1803	LittleTatsamenie	Taku River	0.092	0.908
EST 1803	LyonsFerry	Snake River	0.376	0.624
EST 1803	Nanaimo	Vancouver Island	0.18	0.82
EST 1803	Nestucca	Oregon	0.016	0.984
EST 1803	NUmpqua	Oregon	0.124	0.876
EST 1803	Situk	Alaska	0	1
EST 1803	Solduc	Coastal Washington	0.121	0.879
EST 1803	SoosCreek	Puget Sound	0.208	0.792
EST 1803	SpringCreek	Mid Columbia River	0.101	0.899
EST 1803	Tahini	Alaska	0.227	0.773
EST 1803	Twisp	Upper Columbia R.	0	1
MHC A1	Bulkley	Skeena River	0.677	0.323
MHC A1	Cowlitz	Lower Columbia R.	0.689	0.311
MHC A1	EelRiver	California	0.778	0.222
MHC A1	Feather	California	0.496	0.504
MHC A1	HanfordReach	Upper Columbia River	0.5	0.5
MHC A1	Harrison	Fraser River	0.704	0.296
MHC A1	Johnson	Snake River	0.62	0.38
MHC A1	Kateen	Nass River	0.478	0.522
MHC A1	Kitsumkalum	Skeena River	0.543	0.457
MHC A1	LittleTatsamenie	Taku River	0.651	0.349
MHC A1	LyonsFerry	Snake River	0.556	0.444
MHC A1	Nanaimo	Vancouver Island	0.476	0.524
MHC A1	Nestucca	Oregon	0.549	0.451
MHC A1	NUmpqua	Oregon	0.658	0.342
MHC A1	Situk	Alaska	0.68	0.32
MHC A1	Solduc	Coastal Washington	0.67	0.33
MHC A1	SoosCreek	Puget Sound	0.512	0.488
MHC A1	SpringCreek	Mid Columbia River	0.658	0.342
MHC A1	Tahini	Alaska	0.485	0.515
MHC A1	Twisp	Upper Columbia R.	0.541	0.459
CirpA	Bulkley	Skeena River	0.812	0.188
CirpA	Cowlitz	Lower Columbia R.	0.185	0.815
CirpA	EelRiver	California	0.758	0.242
CirpA	Feather	California	0.331	0.669
CirpA	HanfordReach	Upper Columbia R.	0.154	0.846
CirpA	Harrison	Fraser River	0.489	0.511
CirpA	Johnson	Snake River	0.932	0.068

CirpA	Kateen	Nass River	0.608	0.392
CirpA	Kitsumkalum	Skeena River	0.71	0.29
CirpA	LittleTatsamenie	Taku River	0.793	0.207
CirpA	LyonsFerry	Snake River	0.159	0.841
CirpA	Nanaimo	Vancouver Island	0.313	0.688
CirpA	Nestucca	Oregon	0.602	0.398
CirpA	NUmpqua	Oregon	0.576	0.424
CirpA	Situk	Alaska	0.659	0.341
CirpA	Solduc	Coastal Washington	0.5	0.5
CirpA	SoosCreek	Puget Sound	0.435	0.565
CirpA	SpringCreek	Mid Columbia River	0.206	0.794
CirpA	Tahini	Alaska	0.772	0.228
CirpA	Twisp	Upper Columbia R.	0.787	0.212
GCSH	Bulkley	Skeena River	0.952	0.048
GCSH	Cowlitz	Lower Columbia R.	0.113	0.887
GCSH	EelRiver	California	0.608	0.392
GCSH	Feather	California	0.038	0.962
GCSH	HanfordReach	Upper Columbia R.	0.319	0.681
GCSH	Harrison	Fraser River	0.431	0.569
GCSH	Johnson	Snake River	0.911	0.089
GCSH	Kateen	Nass River	0.787	0.213
GCSH	Kitsumkalum	Skeena River	0.728	0.272
GCSH	LittleTatsamenie	Taku River	0.941	0.059
GCSH	LyonsFerry	Snake River	0.032	0.968
GCSH	Nanaimo	Vancouver Island	0.5	0.5
GCSH	Nestucca	Oregon	0.685	0.315
GCSH	NUmpqua	Oregon	0.452	0.548
GCSH	Situk	Alaska	0.987	0.013
GCSH	Solduc	Coastal Washington	0.635	0.365
GCSH	SoosCreek	Puget Sound	0.081	0.919
GCSH	SpringCreek	Mid Columbia River	0.104	0.896
GCSH	Tahini	Alaska	0.561	0.439
GCSH	Twisp	Upper Columbia R.	0.916	0.084
IsoT	Bulkley	Skeena River	0.995	0.005
IsoT	Cowlitz	Lower Columbia R.	0.989	0.011
IsoT	EelRiver	California	1	0
IsoT	Feather	California	1	0
IsoT	HanfordReach	Upper Columbia R.	0.988	0.012
IsoT	Harrison	Fraser River	0.956	0.044
IsoT	Johnson	Snake River	0.803	0.197
IsoT	Kateen	Nass River	0.966	0.034
IsoT	Kitsumkalum	Skeena River	0.944	0.056
IsoT	LittleTatsamenie	Taku River	0.872	0.128
IsoT	LyonsFerry	Snake River	1	0
IsoT	Nanaimo	Vancouver Island	0.989	0.011
IsoT	Nestucca	Oregon	1	0

IsoT	NUmpqua	Oregon	1	0
IsoT	Situk	Alaska	0.635	0.365
IsoT	Solduc	Coastal Washington	1	0
IsoT	SoosCreek	Puget Sound	1	0
IsoT	SpringCreek	Mid Columbia River	1	0
IsoT	Tahini	Alaska	0.713	0.287
IsoT	Twisp	Upper Columbia R.	0.943	0.057
Ostm1	Bulkley	Skeena River	0.085	0.915
Ostm1	Cowlitz	Lower Columbia R.	0.364	0.636
Ostm1	EelRiver	California	0.368	0.632
Ostm1	Feather	California	0.406	0.594
Ostm1	HanfordReach	Upper Columbia R.	0.649	0.351
Ostm1	Harrison	Fraser River	0.625	0.375
Ostm1	Johnson	Snake River	0.401	0.599
Ostm1	Kateen	Nass River	0.319	0.681
Ostm1	Kitsumkalum	Skeena River	0.314	0.686
Ostm1	LittleTatsamenie	Taku River	0.239	0.761
Ostm1	LyonsFerry	Snake River	0.615	0.385
Ostm1	Nanaimo	Vancouver Island	0.541	0.459
Ostm1	Nestucca	Oregon	0.415	0.585
Ostm1	NUmpqua	Oregon	0.419	0.581
Ostm1	Situk	Alaska	0.236	0.764
Ostm1	Solduc	Coastal Washington	0.462	0.538
Ostm1	SoosCreek	Puget Sound	0.313	0.688
Ostm1	SpringCreek	Mid Columbia River	0.271	0.729
Ostm1	Tahini	Alaska	0.296	0.704
Ostm1	Twisp	Upper Columbia R.	0.146	0.854
PEMT	Bulkley	Skeena River	0.648	0.352
PEMT	Cowlitz	Lower Columbia R.	0.426	0.574
PEMT	EelRiver	California	0.022	0.978
PEMT	Feather	California	0.246	0.754
PEMT	HanfordReach	Upper Columbia R.	0.328	0.672
PEMT	Harrison	Fraser River	0.473	0.527
PEMT	Johnson	Snake River	0.881	0.119
PEMT	Kateen	Nass River	0.436	0.564
PEMT	Kitsumkalum	Skeena River	0.529	0.471
PEMT	LittleTatsamenie	Taku River	0.554	0.446
PEMT	LyonsFerry	Snake River	0.256	0.744
PEMT	Nanaimo	Vancouver Island	0.261	0.739
PEMT	Nestucca	Oregon	0.402	0.598
PEMT	NUmpqua	Oregon	0.332	0.668
PEMT	Situk	Alaska	0.57	0.43
PEMT	Solduc	Coastal Washington	0.314	0.686
PEMT	SoosCreek	Puget Sound	0.443	0.557
PEMT	SpringCreek	Mid Columbia River	0.267	0.733
PEMT	Tahini	Alaska	0.657	0.343

PEMT	Twisp	Upper Columbia River	0.971	0.029
Phos	Bulkley	Skeena River	0	1
Phos	Cowlitz	Lower Columbia R.	0.044	0.956
Phos	EelRiver	California	0.213	0.787
Phos	Feather	California	0.115	0.885
Phos	HanfordReach	Upper Columbia R.	0.011	0.989
Phos	Harrison	Fraser River	0.169	0.831
Phos	Johnson	Snake River	0.172	0.828
Phos	Kateen	Nass River	0.139	0.861
Phos	Kitsumkalum	Skeena River	0.025	0.975
Phos	LittleTatsamenie	Taku River	0.09	0.91
Phos	LyonsFerry	Snake River	0.01	0.99
Phos	Nanaimo	Vancouver Island	0.367	0.633
Phos	Nestucca	Oregon	0.258	0.742
Phos	NUmpqua	Oregon	0.071	0.929
Phos	Situk	Alaska	0.241	0.759
Phos	Solduc	Coastal Washington	0.073	0.927
Phos	SoosCreek	Puget Sound	0.195	0.805
Phos	SpringCreek	Mid Columbia River	0.132	0.868
Phos	Tahini	Alaska	0.054	0.946
Phos	Twisp	Upper Columbia R.	0.25	0.75
Thio	Bulkley	Skeena River	0.723	0.277
Thio	Cowlitz	Lower Columbia R.	0.782	0.218
Thio	EelRiver	California	0.446	0.554
Thio	Feather	California	0.651	0.349
Thio	HanfordReach	Upper Columbia R.	0.636	0.364
Thio	Harrison	Fraser River	0.456	0.544
Thio	Johnson	Snake River	0.207	0.793
Thio	Kateen	Nass River	0.59	0.41
Thio	Kitsumkalum	Skeena River	0.543	0.457
Thio	LittleTatsamenie	Taku River	0.527	0.473
Thio	LyonsFerry	Snake River	0.495	0.505
Thio	Nanaimo	Vancouver Island	0.645	0.355
Thio	Nestucca	Oregon	0.494	0.506
Thio	NUmpqua	Oregon	0.643	0.357
Thio	Situk	Alaska	0.728	0.272
Thio	Solduc	Coastal Washington	0.606	0.394
Thio	SoosCreek	Puget Sound	0.552	0.448
Thio	SpringCreek	Mid Columbia River	0.634	0.366
Thio	Tahini	Alaska	0.239	0.761
Thio	Twisp	Upper Columbia R.	0.226	0.774

Figure 1. SNP assays under selection

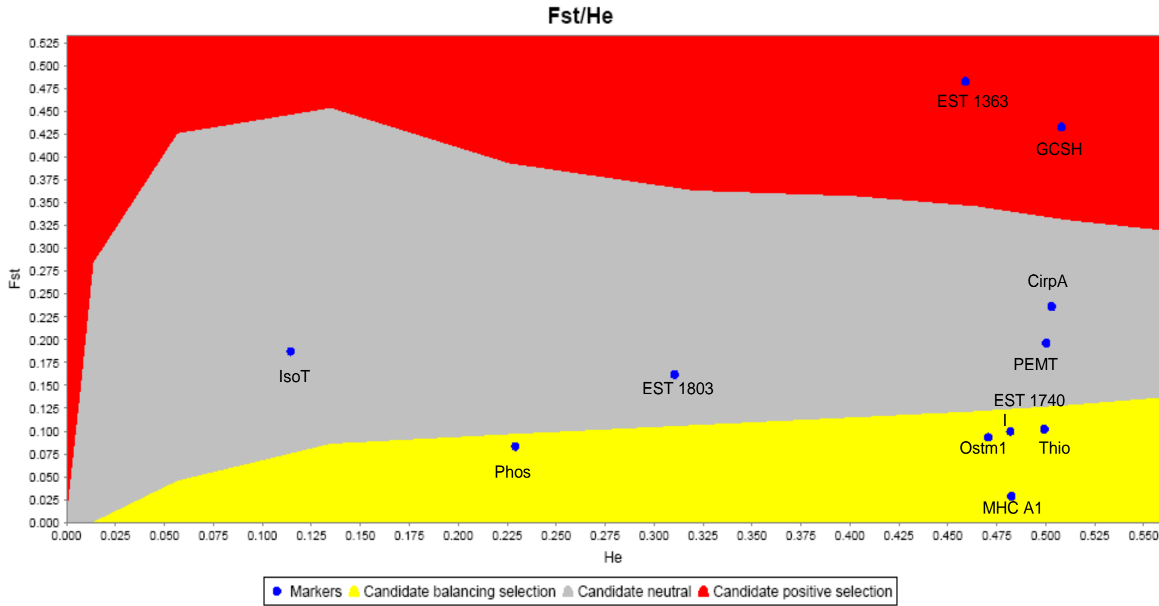

Figure 2. UPGMA dendrogram of screened stocks