Proceedings, 10th World Congress of Genetics Applied to Livestock Production

The Development and Characterization of a 57K SNP Chip for Rainbow Trout

Y. Palti[1], G. Gao[1], T. Moen[2], S. Liu[1], M.P. Kent[3], S. Lien[3], M.R. Miller[4] and C.E. Rexroad III[1]
[1]NCCCWA-ARS-USDA, Leetown, WV, USA; [2]AquaGen, Aas, Norway; [3]Centre of Integrative Genetics (CIGENE), Aas, Norway; [4]University of California, Davis, USA

ABSTRACT: In this paper we describe the development and characterization of the first high density SNP chip for rainbow trout. The SNPs are distributed throughout the genome with good representation in all 29 chromosomes. The genotyping quality was high and validation rate was close to 90%. This is comparable to other farm animals and is much higher than previous smaller scale SNP validation studies in rainbow trout. The chip is more useful for rainbow trout aquaculture populations with more than 83% polymorphic markers per population, but even in wild populations the number of polymorphic markers was greater than 10,000. The SNP chip is publically available and it has already been used in a proof- of concept study to demonstrate its utility for genome enabled selection in rainbow trout presented in a companion WCGALP 2014 paper by Vallejo et al.
Keywords: Rainbow trout; SNP Chip; Genetic diversity; Linkage analysis

Introduction

The development of high density and high throughput SNP genotyping assays has radically changed genetic and genome analyses of complex traits in farm animals. Various high density SNP chips developed in recent years include cow (Matukumalli et al., 2009), pig (Ramos et al., 2009), chicken (Groenen et al., 2011) and Atlantic salmon (Houston et al., 2014). The high density SNP assays can be used to capture population-wide linkage disequilibrium for genome-wide association studies or to increase the accuracy of breeding program through genomic selection (Chen et al., 2011; Cole et al., 2011; Goddard et al., 2011). The ancestor of the salmonids which include rainbow trout and Atlantic salmon has undergone a whole genome duplication event between 25 and 100 million years ago, which complicates the discovery of true bi-allelic SNPs in those species. To overcome this technical difficulty, we recently used next generation sequencing of restriction-site associated DNA (RAD) tags for SNP discovery in a panel of rainbow trout doubled haploid (DH) lines to produce a large dataset from which SNPs can be selected for a whole genome high-density SNP chip assays (Palti et al., 2013). In the current study, we developed, validated and characterized a high-density SNP genotyping assay for rainbow trout.

Materials and Methods

Fish and DNA samples: Fin clips and DNA were collected from fish representing 19 DH lines as previously described (Palti et al., 2013). For the rainbow trout populations' survey, fin clips were collected from 265 unrelated fish representing 18 populations with the number of fish per population between 5 and 26 (Table 1). Full-sib families from the NCCCWA QTL mapping population with 39-90 offspring per family were sampled for pedigree validation of the SNPs and for genetic linkage analysis. Samples from other *Oncorhynchus* species included cutthroat trout (N=5), Chinook salmon (N=3) and Coho salmon (N=4). A total of 960 samples were included in the genotype validation panel.

Source of SNPs: SNP information from previous SNP discovery projects (Boussaha et al., 2012; Castano-Sanchez et al., 2009; Palti et al., 2013; Salem et al., 2012; Sanchez et al., 2011) was evaluated with the primary source coming from the DH RADs database (Palti et al., 2013). The combined dataset of SNPs from previous projects was termed "USDA". In addition, we generated a new large dataset of putative SNPs from random Illumina re-sequencing of 16 fish from the AquaGen (Norway) breeding nucleus and alignment of the reads to the rainbow trout draft genome assembly (animalgenome.org). The putative SNPs from the re-sequencing project were filtered from suspected multi-site variants (MSVs) using tests for excess hetrozygosity and overrepresentation of one allele, and grouped into categories based on level of sequence coverage from all 16 animals and whether the sequence provided a unique genome hit Vs. multiple hits .

Selection of SNPs: All SNPs selected for the chip had high p-convert score which is an *in silico* prediction of the probability of conversion to a reliable SNP assay based on Affymetrix algorithms. Ranking criteria for SNPs within each dataset included priority to sequences from transcribed regions; uniqueness of hit to the draft genome assembly; genetic map information from previous studies; and minor allele frequency (MAF) information from previous studies. All the 20,716 SNPs shared by the USDA and Aqua Gen datasets were selected. In addition, ~20,000 SNPs unique to the Aqua Gen dataset were selected and ~17,000 SNPs unique to the USDA dataset were selected. The latter included ~10,000 SNPs that do not match the current draft genome assembly as approximately 30% of the genome are not represented in the current draft. For filtering of back-

ground hybridization signal we identified 5,000 non-redundant and monomorphic conservative transcriptome sequences of which 2,500 were strategically placed on the oligonucleoties array,

SNP genotyping and linkage analysis: Samples were genotyped by a commercial service provider (Geneseek, Inc., Lincoln, NE) according to the Axiom genotyping procedures described by Affymetrix. Genotyping calls and quality control analyses were conducted according to the Affymetrix recommended workflow using Genotyping Console and SNPolisher software packages. Between 40 and 60 SNPs per family were filtered out due to highly significant distortion from the expected Mendelian segregation (Bonferonni adjustment to $P<0.05$). Two-point linkage analyses were conducted using MULTIMAP and CRI-MAP as we have previously described (Palti et al., 2011), and consensus RAD SNPs that are also present on the SNP-chip were used to anchor linkage groups to the rainbow trout chromosomes.

Results and Discussion

Genotypes quality assessment: The total number of putative SNPs we have placed on the chip was 57,501. A total of 49,468 SNPs (86%) were categorized as high quality and polymorphic and an additional 654 SNPs (1.1%) as high quality but monomorphic, using the default quality filtering of the Affymetrix SNPolisher software (Houston et al., 2014). Of the 960 samples we used, 924 (96%) passed the genotyping quality filtering and 97% call rate threshold. One sample was repeated 11 times to assess genotyping reliability, but three of that sample replicates failed. The error rate among the eight passing replicates was 0.58%. We used the genotypes of 19 DH lines to identify potential paralogous sequence variants (PSVs). A SNP marker with heterozygous genotype in two DH lines is likely a PSV (Palti et al., 2013). A total of 756 markers (1.5%) were heterozygous in at least two DH samples. We also used the 27,061 SNPs with RAD and SNP chip genotypes from the same DH samples to evaluate consistency between the two methods. A total of 628 SNPs (2.3%) had at least one unmatched genotype, but for the individual samples only 923 of 297,671 genotypes (0.3%) did not match between the two methods.

Polymorphism in populations: Of the 265 samples from the populations' survey, 249 passed the CR>97% threshold. The percent of polymorphic markers and average MAF per population are listed in Table 1. Of the 18 populations we sampled, 10 were from commercial breeders or aquaculture research programs, six were from wild rainbow or steelhead populations and two were hatchery strains that were used for QTL mapping of Whirling disease resistance (Baerwald et al., 2010). Our data clearly show that this SNP chip is more informative for the aquaculture populations with 83%-93% polymorphic markers. For the wild populations polymorphism was only found in 21%-49% of the markers. However, these are likely under-estimates for the wild populations affected by the small sample size of only 4-11 fish for each population. The average MAF was similar for all populations in the range of 0.22-0.28. We also evaluated the utility of the SNP chip for genotyping or transferring SNP information to other Pacific salmon and trout species. None of the "other" species samples passed the CR>97% threshold. For the Yellowstone or Westslope cutthroat trout samples 40-47K SNP genotypes were called per sample, but only 2,500-4,500 were polymorphic. For Chinook and Coho salmon the number of marker genotypes called was between 33K and 38K and 5K-6K were polymorphic per sample.

Table 1. Number and percent of polymorphic SNP markers per population.

Pop[a]	N[b]	No. SNPs[c]	No. Poly	% Poly	MAF[d]
1	13	49468	45928	93%	0.26
2	14	49468	43425	88%	0.25
3	25	49468	40917	83%	0.24
4	24	49468	44467	90%	0.24
5	25	49468	45428	92%	0.25
6	20	49468	43435	88%	0.25
7	22	49468	44494	90%	0.25
8	21	49468	44262	89%	0.25
9	12	49468	31935	65%	0.25
10	12	49468	42046	85%	0.25
11	11	49299	10577	21%	0.22
12	6	49375	14720	30%	0.24
13	14	49468	45120	91%	0.26
14	12	49468	41786	84%	0.26
15	4	49462	19420	39%	0.27
16	4	49464	20095	41%	0.26
17	5	49455	17268	35%	0.28
18	5	49468	24330	49%	0.26

a. Populations: 1) NCCCWA Bacterial cold water disease QTL mapping source population; 2) NCCCWA growth select broodstock; 3) Troutlodge, Inc. Kamloop strain; 4) Troutlodge, Inc. Jumpers strain; 5) Troutlodge, Inc. November Steelhead strain; 6) Troutlodge, Inc. February Steelhead strain; 7) Clear Springs Food broodstock; 8) USDA-ARS Hagerman, Idaho broodstock; 9) UC Davis Hofer strain; 10) UC Davis Colorado River strain; 11) UC Davis California golden trout; 12) UC Davis Little Kern golden trout; 13) INRA Sy population; 14) INRA Prt population; 15) Hale and Nichols, Little Sheep Creek, Oregon rainbow trout; 16) Hale and Nichols, Little Sheep Creek, Oregon steelhead; 17) Hale and Nichols, Sashin Creek, Alaska rainbow trout; and 18) Hale and Nichols, Sashin Creek, Alaska steelhead.

b. Number of unrelated fish that were genotyped at CR>97% from each population.
c. Number of markers that were genotyped at CR>97% for all the fish from each population.
d. Average MAF from all the polymorphic markers in each population.

Pedigree validation and genomic distribution: Approximately 46,500 SNPs were informative for linkage analysis in at least one of the 10 pedigreed families we genotyped. On average, 25,134 SNPs were informative per family (range of 22,800 to 26,800), and ~21,400 SNPs were in the combined category of monomorphic or failed pedigree check. In most families the number of SNPs that failed pedigree check was less than 1,000, and in the families were it was higher, it balanced out with reduction in the number of monomorphic markers. Hence, it is likely that the main cause of higher pedigree-check failure was from false heterozygote genotype calls in monomorphic markers. A total of 44,991 SNPs were assigned to linkage groups with the number per chromosome ranging from 740 on Omy26 to 2,875 on Omy5. We have identified eight pairs of homeologous chromosome arms; which is in good agreement with published studies (Naish et al., 2013; Phillips et al., 2009).

Literature Cited

Baerwald, M. R., J. L. Petersen, R. P. Hedrick, G. J. Schisler, and B. May. 2010. A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (Oncorhynchus mykiss). Heredity 106: 920-926.

Boussaha, M., R. Guyomard, C. Cabau, D. Esquerre, and E. Quillet. 2012. Development and characterisation of an expressed sequence tags (EST)-derived single nucleotide polymorphisms (SNPs) resource in rainbow trout. BMC Genomics 13: 238.

Castano-Sanchez, C. et al. 2009. Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. BMC Genomics 10: 559.

Chen, C. Y. et al. 2011. Genome-wide marker-assisted selection combining all pedigree phenotypic information with genotypic data in one step: An example using broiler chickens. Journal of Animal Science 89: 23-28.

Cole, J. et al. 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics 12: 408.

Goddard, M. E., B. J. Hayes, and T. H. E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. Journal of Animal Breeding and Genetics 128: 409-421.

Groenen, M. et al. 2011. The development and characterization of a 60K SNP chip for chicken. BMC Genomics 12: 274.

Houston, R. et al. 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). BMC Genomics 15: 90.

Matukumalli, L. K. et al. 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS ONE 4: e5350.

Naish, K. A. et al. 2013. Comparative Genome Mapping Between Chinook Salmon (Oncorhynchus tshawytscha) and Rainbow Trout (O. mykiss) Based on Homologous Microsatellite Loci. G3: Genes|Genomes|Genetics 3: 2281-2288.

Palti, Y. et al. 2013. A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. Molecular Ecology Resources E-published: Online.

Palti, Y. et al. 2011. A first generation integrated map of the rainbow trout genome. BMC Genomics 12: 180.

Phillips, R. et al. 2009. Assignment of Atlantic salmon (Salmo salar) linkage groups to specific chromosomes: Conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (Oncorhynchus mykiss). BMC Genetics 10: 46.

Ramos, A. M. et al. 2009. Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. PLoS ONE 4: e6524.

Salem, M. et al. 2012. RNA-Seq Identifies SNP Markers for Growth Traits in Rainbow Trout. PLoS ONE 7: e36264.

Sanchez, C. et al. 2011. Generation of a reference transcriptome for evaluating rainbow trout responses to various stressors. BMC Genomics 12: 626.