Assessment of genetic differentiation and genetic assignment
of commercial rainbow trout strains using a SNP panel

Sixin Liu ^{a,*}, Yniv Palti ^a, Kyle E. Martin ^b, James E. Parsons ^b, Caird E. Rexroad III ^a

^a National Center for Cool and Cold Water Aquaculture, USDA/ARS, 11861 Leetown Rd., Kearneysville, WV 25430, USA

^b Troutlodge, Inc., 12000 McCutcheon Rd., Sumner, WA 98390, USA

* Corresponding author.

E-mail address: sixin.liu@ars.usda.gov

Abstract

Rainbow trout (*Oncorhynchus mykiss*) is the most widely cultured cold freshwater fish in the world. Troutlodge, Inc., one of the largest commercial rainbow trout egg producers in the world, has eight breeding populations of rainbow trout. Assessment of population genetic differentiation is critical for selective breeding, and genetic assignment is often required to address production issues arising on fish farms. Previously, we developed a SNP (single nucleotide polymorphism) panel for parentage assignment in rainbow trout. The objectives of this study were: (1) to characterize the genetic differentiation of the eight Troutlodge breeding populations; and (2) to evaluate the accuracy of genetic assignment to identify Troutlodge fish using the same SNP panel previously developed for parentage assignment. A total of 1,732 breeders of the eight Troutlodge breeding populations were genotyped with the SNP panel. The global F_{ST} over all SNPs was 0.13 and the pairwise F_{ST} between any two breeding populations ranged from 0.056 to 0.195. Both phylogenetic tree and structure analyses revealed that the odd year and even year populations for the same strain were closely related to each other. Also, the eight populations were clustered into two groups. Based on the results of self-assignment, 97.1% of the Troutlodge breeders were correctly assigned to the population of origin. To further evaluate the accuracy of genetic assignment, we also genotyped 280 egg production fish from three Troutlodge populations of known origin, 49 fish from a Canadian farm, 70 fish from a farm in Idaho and 188 fish from four known non-Troutlodge strains. Among the 280 production fish, 98.2% fish were correctly assigned to the Troutlodge strain of origin. Consistent with the purchasing records, the fish from the two farms were also correctly assigned to two Troutlodge strains. Based on the SNP genotypes, 185 out of 188 fish (98.4%) from the four non-Troutlodge strains could be excluded as Troutlodge fish. In conclusion, our previously developed SNP panel for parentage assignment is also useful to characterize genetic differentiation and has sufficient power for genetic assignment of commercial rainbow trout strains used in this study.

Keywords: Genetic assignment; Individual assignment; Population assignment; Genetic differentiation; SNP; Rainbow trout

1. Introduction

Rainbow trout (*Oncorhynchus mykiss*) is the most widely cultured cold freshwater fish in the world, with production on every continent except Antarctica. In 2014, the global rainbow trout production was 812,940 tons with an economic value of 3.9 billion U.S. dollars (FAO, 2016). Troutlodge, Inc., founded in 1945, is one of the largest rainbow trout egg producers in the world with annual sale of over 500 million eyed eggs in over 70 countries. To produce and ship eggs every week of the year without the use of photoperiod manipulation, Troutlodge has developed four strains of rainbow trout with distinct spawning times. These four strains are named according to their peak spawning months, February, May, August and November, respectively. Since it takes two years for rainbow trout to reach sexual maturity, there are two populations per strain, odd year and even year populations. As a result, a total of eight rainbow trout populations are maintained in the breeding program.

Characterization of population genetic differentiation is fundamental for selective breeding of rainbow trout. Unfortunately, there are only a few reports to characterize domesticated rainbow trout populations using molecular markers. Silverstein et al. (2004) used nine microsatellites to characterize three domesticated rainbow trout strains, and concluded that all three strains were unique with large genetic variation within each strain. Similarly, Johnson et al. (2007) used 12 microsatellites to genotype rainbow trout broodstocks and detected subpopulation structure. Using 10 microsatellites to genotype 12 rainbow trout strains from northern and Eastern Europe, Gross et al. (2007) reported that only 0.9% of the total genetic variation was due to the origin of countries, 13.7% was due to differentiation among the strains within the countries and 85.5% was due to variation within strains. Glover (2008) reported that the pairwise F_{ST} values ranged from 0 to 0.127 among farmed rainbow trout in Norway. To our knowledge, the genetic differentiation of the eight Troutlodge rainbow trout breeding populations has not been characterized using molecular markers.

Because two or more strains of fish are often cultured simultaneously on the same farm, one of the challenges to address production issues is to identify the population of origin of individual fish. Visual inspection of morphological traits are often inconclusive. Because every fish has a unique genetic makeup, use of molecular markers is the most logical approach to distinguish rainbow trout strains. Genetic assignment, which uses genetic information to ascertain population membership of individuals or groups of individuals, has been widely used in wildlife management and conservation (Manel et al., 2005). However, there are only few studies of genetic assignment of farmed rainbow trout. Using 10 microsatellites, Gross et al. (2007) assigned rainbow trout individuals to their strain of origin with an accuracy ranged from 63% to 100%. To identify rainbow trout escaped from farms, 12 microsatellites were used to genotype farmed rainbow trout and 82.5% fish were assigned to the correct farms (Glover, 2008).

Single nucleotide polymorphisms (SNPs) are becoming increasingly popular in many species because they are highly abundant, amenable for high throughput genotyping. Next-generation sequencing has led to the discovery of a large number of SNPs in rainbow trout (Palti et al., 2014; Sanchez et al., 2009), and a 57K SNP genotyping array (Palti et al., 2015a) was developed for high-throughput SNP genotyping in rainbow trout. Regardless of the high genotyping cost

using the 57K SNP genotyping array, there is no need to use such a large number of SNPs for genetic assignment. It has been reported that a set of 90 to 100 SNPs is sufficient for genetic assignment in several species (Clemento et al., 2014; Matala et al., 2014; Negrini et al., 2009; Ramos et al., 2009; Starks et al., 2016). Although SNPs under diversifying selection can improve the accuracy of genetic assignment (Ackerman et al., 2011), correct assignment rates ranged from 79.5% to 97.7% were achieved with 91 random SNPs derived from expressed sequence tags in coho salmon (*Oncorhynchus kisutch*) (Starks et al., 2016). Previously, we developed a panel of 96 SNPs for parentage assignment in rainbow trout (Liu et al., 2016). These SNPs distributed across all 29 chromosomes of the rainbow trout genome were selected for the panel because either their minor allele frequencies were greater than 0.45 among 249 fish from 18 diverse rainbow trout populations or they were associated with traits of interest in our previous studies (Liu et al., 2015; Palti et al., 2015b; Salem et al., 2012). To expand the utilities of the SNP panel, the objectives of this study were: (1) to characterize the genetic differentiation of the eight Troutlodge breeding populations; and (2) to evaluate the accuracy of genetic assignment to identify Troutlodge fish using the same SNP panel previously developed for parentage assignment.

2. Materials and methods

2.1 Fish samples and SNP genotyping

Most of the fish used in this study (Table 1) were from Troutlodge including 1,732 breeders of the eight breeding populations (four strains, two year-class populations per strain) and 280 production fish from three populations of known origin. The relationships between breeders, production fish and farmed fish are illustrated in Fig. 1. To test whether the farmed fish were from one of the Troutlodge populations, we included 47 fish from one farm in Canada and 70 fish from another farm in Idaho, USA. Four strains of known non-Troutlodge fish (47 fish per strain) including two American strains and two European strains (Table 1) were also used in this study to test the power of the SNP panel to exclude the non-Troutlodge fish.

Fin clips were collected and stored in 95% ethanol. DNA was extracted using the AutoGenprep 965 automated DNA extractor (Autogen, Holliston, MA, USA) following the manufacturer's instructions. The DNA samples were genotyped with the same SNP panel and protocol as described in our previous study (Liu et al., 2016). Briefly, DNA samples were pre-amplified, and the pre-amplified products were diluted and used for genotyping with 96.96 Dynamic Array IFCs (Integrated Fluidic Circuits). The arrays were read using EP1 system, and genotypes were called automatically using Fluidigm SNP genotyping analysis software 4.1 with a confidence threshold 85. The genotype clusters were examined for each assay, any wrong calls or no calls were corrected manually.

2.2 Statistical analyses

Program GENEPOP v4.4.3 (Rousset, 2008) was used to calculate the F_{ST} values with the method of Weir and Cockerham (1984). Based on the pairwise F_{ST} matrix, program MEGA6 (Tamura et al., 2013) was used to construct the UPGMA phylogenetic tree of the eight Troutlodge breeding populations. We also used the model-based clustering method implemented in program

STRUCTURE v2.3 (Pritchard et al., 2000) to identify the genetic relationships among the eight Troutlodge breeding populations. We evaluated the number of populations K from 1 to 10, and each K was run five iterations. The admixture model and correlated allele frequencies (Falush et al., 2003) were used for structure analyses, which were conducted with a run length of 100 000 Markov chain Monte Carlo steps following a burn-in of 100 000 steps. Structure Harvester (Earl and Vonholdt, 2012) was used to determine the most appropriate K value following the method of Evanno et al. (2005). CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) was used to calculate the mean cluster membership coefficients over the five runs of STRUCTURE for each K , and the clusters were visualized using program Distruct v1.1 (Rosenberg, 2004).

Program GeneClass2 (Piry et al., 2004) was used to assign individual fish to the population of origin. The Rannala and Mountain (1997) criterion for computation was used for likelihood estimation. The assignment probability of each individual fish to each reference population was estimated with the simulation algorithm of Paetkau et al. (2004), and the Type I error threshold was set to 0.01. The individual fish was assigned to the reference population with the highest assignment probability. We performed three sets of genetic assignment tests including self-assignment of the eight Troutlodge breeding populations and genetic assignment tests of production fish and farmed fish. The SNP genotypes of the eight Troutlodge breeding populations were used as reference populations for all genetic assignment tests. We also used the program GeneClass2 to exclude non-Troutlodge fish. The program was run with the same parameters as genetic assignment tests except that Type I error threshold was set to 0.05.

3. Results and discussion

3.1 Genetic differentiation among the eight Troutlodge breeding populations

Among the 96 SNPs used to genotype the 1,732 fish from the eight Troutlodge breeding populations, all SNPs were informative except SNP P096. All fish had the GG genotype for SNP P096. The global F_{ST} over all SNPs was 0.13 and F_{ST} for each SNP ranged from 0.023 to 0.516 (Table S1). The population pairwise F_{ST} values are presented in Table 2, and the pairwise F_{ST} between any two breeding populations ranged from 0.056 (between population 11TLUM and 12TLUM) to 0.195 (between population 11TLUA and 12TLUM).

F_{ST} provides a measure of population genetic differentiation (Holsinger and Weir, 2009). The global $F_{ST} = 0.13$ indicated that 13% of the genetic variation among the Troutlodge breeding populations was accounted by genetic difference among populations. It has been reported that both marker types and marker contents can influence the estimate of population genetic differentiation (DeFaveri et al., 2013; Karlsson et al., 2011; Glover et al., 2010). The SNPs influenced by directional selection show a larger genetic differentiation than neutral SNPs, and SNPs that are subject to balancing selection show a lower genetic differentiation. We did not purposely select SNPs with high or low F_{ST} values to develop the SNP panel, and the global $F_{ST} = 0.13$ reported in this study is consistent with previous studies. Based on analysis of isozymes, most of the genetic variation (92%) in rainbow trout was within populations (Hershberger, 1992). Studying the genetic differentiation of rainbow trout strains in northern and Eastern Europe, Gross et al. (2007) reported that 13.7% of the total genetic variation was due to

differentiation among 12 rainbow trout strains from northern and Eastern Europe. A pairwise $F_{ST} = 0.127$ was observed among farmed rainbow trout in Norway (Glover, 2008). Silverstein et al. (2004) reported an overall $F_{ST} = 0.089$ among three domesticated U.S. rainbow trout strains.

The UPGMA phylogenetic tree of the eight Troutlodge breeding populations is presented in Fig. 2. The odd year and even year populations for the same strain were closely related to each other. Also, the eight populations were clustered into two groups. One group included the populations of February and May strains and the other group included the populations of August and November strains. Since this is the first report of the genetic relationships among the eight Troutlodge breeding populations, no previous results are available for comparison. Furthermore, due to the lack of documentation of founder populations of the eight Troutlodge breeding populations, it is hard to speculate why the eight populations were clustered into two spawning groups. It would be interesting to test whether the same two spawning groups would remain true with a broader collection of rainbow trout populations.

We also used the Bayesian clustering approach implemented in program STRUCTURE to investigate the genetic relationships among the eight Troutlodge breeding populations. Following the method of Evanno et al. (2005), the most appropriate number of clusters was $K = 2$ (Fig. 3C). Consistent with the result of the phylogenetic tree (Fig. 2), the eight Troutlodge breeding populations were clustered into the same two groups although admixture individuals were clearly detected in all populations (Fig. 3A). The February and May populations formed one cluster, and the other cluster included the August and November populations. This result is not surprising because STRUCTURE can accurately detect the uppermost hierarchical level of population structure (Evanno et al., 2005). Although the most appropriate number of clusters was $K = 2$, visual inspection of the clusters at higher number of K also revealed the genetic relationships among the eight Troutlodge breeding populations. Consistent with the phylogenetic tree, the two year-class populations of each strain were clustered into one group (Fig. S1A) at $K = 4$, and each year-class population was assigned to its own cluster (Fig. S1B) at $K = 8$.

3.2 Self-assignment of breeders of the eight Troutlodge breeding populations

The results of self-assignment of the eight Troutlodge breeding populations are presented in Table 3. All 1,732 fish were assigned to one of the eight Troutlodge breeding populations. The correct assignment rates ranged from 92.2% to 100%, and the overall correct assignment rate was 97.1%. Most of the incorrectly assigned fish were assigned to the correct strain but incorrect year-class population. For example, 200 out of 217 11TLUM breeders were correctly assigned, and 14 out of the 17 incorrectly assigned fish were assigned to the population 12TLUM, the even year population of the May strain. This result was not totally unexpected because the two year-class populations of the same strain were closely related to each other as indicated by both phylogenetic and structure analyses.

3.3 Genetic assignment of the Troutlodge production fish of known origin

All 94 production fish of population 2012TLUN were correctly assigned to the breeding population 10TLUN. Also, all 94 production fish of population 2015TLUF were correctly assigned to the breeding population 11TLUF. Among the 92 production fish of population

2013TLUN, 87 fish were correctly assigned to the breeding population 11TLUN, and 5 fish were assigned incorrectly (three fish were assigned to the breeding population 10TLUN, one fish was assigned to the breeding population 12TLUF, and another fish was assigned to the breeding population 10TLUA). Thus, 98.2% of production fish were assigned to the correct breeding populations even though production fish are one generation behind their corresponding breeding population. It is noteworthy that the production fish of population 2015TLUF were derived from the breeding population 13TLUF, which was not included in the eight Troutlodge breeding populations used in this study. Nonetheless, production fish of population 2015TLUF were assigned to the breeding population 11TLUF, a temporal replicate (the same strain and the same year-class population) of breeding population 13TLUF. This result indicated that the genetic differentiation between temporal replicate populations was small, and the SNP genotypes of the eight Troutlodge breeding populations can also be used as reference populations for genetic assignment of temporal replicate populations.

3.4 Genetic assignment of farmed fish

All 49 fish from the Canadian farm were assigned to the February strain of even year breeding population (12TLUF). This result was consistent with the purchasing record which showed that the eyed eggs were purchased from Troutlodge in April 2014. Among the 70 fish from the Idaho farm, 33 fish were assigned to the breeding population 11TLUA and 37 fish were assigned to the breeding population 10TLUA. Based on the purchasing record, the eyed eggs were purchased from Troutlodge in September 2014. Thus, it was reasonable that these fish were assigned to the Troutlodge August strain. Since 37 fish were assigned to the even year population and 33 fish were assigned to the odd year population, we suspected that the eyed eggs were from both two-year-old females and three-year-old females of the August strain.

3.5 Genetic exclusion of non-Troutlodge fish

To test the ability of the SNP panel to exclude non-Troutlodge fish, fish from four non-Troutlodge strains (47 fish per strain) including UW (University of Washington), Shasta and two European strains were genotyped with the SNP panel. Among 188 non-Troutlodge fish, only three (1.6%) fish could not be excluded as Troutlodge fish. These three fish were from the Shasta strain and were assigned incorrectly to the breeding population 11TLUF. Therefore, the SNP panel is powerful to exclude non-Troutlodge fish. As might be expected, the exclusion power of the SNP panel depends on the genetic similarity between Troutlodge strains and non-Troutlodge strains. It would be challenging to exclude fish closely related to the Troutlodge strains. As shown in the Fig. S2, the four non-Troutlodge strains used in this study were not closely related to Troutlodge populations. Thus, the exclusion power might be overestimated in this study.

The SNP panel has sufficient power for genetic assignment of commercial rainbow trout strains used in this study. Based on the results of self-assignment, 97.1% Troutlodge breeders were correctly assigned to the population of origin. Among 280 production fish from three populations of known origin, 98.2% fish were correctly assigned to the Troutlodge breeding populations of origin. Consistent with the purchasing records, the fish collected from the Canadian farm were correctly assigned to the February strain of even year breeding population 12TLUF, and fish

from the Idaho farm were correctly assigned to the Troutlodge August strain. Furthermore, 185 out of 188 (98.4%) fish from the four non-Troutlodge strains were excluded as Troutlodge fish. This is the first report to use SNPs for genetic assignment in domesticated rainbow trout populations, and the rates of correct genetic assignment reported in this study exceeded those achieved with microsatellites reported in previous studies (Gross et al., 2007; Glover, 2008). Since this SNP panel was initially developed for parentage assignment in rainbow trout, we did not deliberately select SNPs with high discriminatory power for genetic assignment. Thus, we suggest further validation of the accuracy of genetic assignment before using this SNP panel for rainbow trout populations not examined in this study.

4. Conclusions

The global $F_{ST} = 0.13$ indicated that 13% of the genetic variation among the eight Troutlodge breeding populations was accounted by genetic difference among populations. The odd year and even year populations for the same strain were closely related to each other, and the eight populations were clustered into two groups. The February and May strains were clustered into one group, and the August and November strains were clustered into another group. Based on the results of self-assignment, 97.1% of the Troutlodge breeders were correctly assigned to the population of origin. Both production fish and fish from two farms were also correctly assigned to Troutlodge strains. Based on the SNP genotypes, 185 out of 188 fish (98.4%) from the four non-Troutlodge strains could be excluded as Troutlodge fish. Therefore, we conclude that our previously developed SNP panel for parentage assignment is also useful to characterize genetic differentiation and has sufficient power for genetic assignment of commercial rainbow trout strains used in this study.

Figures and legends

Fig. 1. Flow chart of Troutlodge rainbow trout breeding and production systems.

Fig. 2. UPGMA phylogenetic tree of the eight Troutlodge breeding populations of rainbow trout. The year-class of each population is coded by the two-digit prefix and the strain name is coded by the last letter of the population name.

Fig. 3. Population structure ($K = 2$) of the eight Troutlodge breeding populations, which are separated by black lines. The cluster membership coefficients were averaged over five replicates of cluster analyses. A, individual level; B, population level; C, delta K plot to determine the most appropriate K value following the method of Evanno et al. (2005).

Supplementary material

Table S1. F -statistics for each SNP calculated with the method of Weir and Cockerham (1984).

Fig. S1. Population structure of the eight Troutlodge breeding populations, which are separated by black lines. The cluster membership coefficients were averaged over five replicates of cluster analyses. Individual level is shown at the top row and the population level is shown at the bottom row. A, $K = 4$; B, $K = 8$.

Fig. S2. UPGMA phylogenetic tree of the eight Troutlodge breeding populations and four non-Troutlodge strains of domesticated rainbow trout.

Acknowledgments

We would like to thank Roseanna Long, Kristy Shewbridge and Bryce Williams for their technical assistance, and Drs. Tim Leeds and Beth Cleveland for providing fish samples used in this study. This work was supported by Agricultural Research Service CRIS Project 8082-31000-012 “Integrated Research Approaches for Improving Production Efficiency in Salmonids”. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture (USDA). USDA is an equal opportunity provider and employer.

References

- Ackerman, M.W., Habicht, C. & Seeb, L.W. (2011). Single-Nucleotide Polymorphisms (SNPs) under Diversifying Selection Provide Increased Accuracy and Precision in Mixed-Stock Analyses of Sockeye Salmon from the Copper River, Alaska. *Transactions of the American Fisheries Society*, 140: 865-881.
- Clemento, A.J., Crandall, E.D. & Garza, J.C. (2014). Evaluation of a single nucleotide polymorphism baseline for genetic stock identification of Chinook Salmon (*Oncorhynchus tshawytscha*) in the California Current large marine ecosystem. *Fish B-Noaa*, 112: 112-130.
- Defaveri, J., Viitaniemi, H., Leder, E. & Merila, J. (2013). Characterizing genic and nongenetic molecular markers: comparison of microsatellites and SNPs. *Mol Ecol Resour*, 13: 377-392.
- Earl, D.A. & Vonholdt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*, 4: 359-361.
- Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*, 14: 2611-2620.
- Falush, D., Stephens, M. & Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164: 1567-1587.
- FAO (2016). Global aquaculture production 1950-2014. <http://www.fao.org/fishery/statistics/global-aquaculture-production/en>
- Glover, K.A. (2008). Genetic characterisation of farmed rainbow trout in Norway: intra- and inter-strain variation reveals potential for identification of escapees. *BMC Genetics*, 9.
- Glover, K.A., Hansen, M.M., Lien, S., Als, T.D., Hoyheim, B. & Skaala, O. (2010). A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics*, 11.
- Gross, R., Lulla, P. & Paaver, T. (2007). Genetic variability and differentiation of rainbow trout (*Oncorhynchus mykiss*) strains in northern and Eastern Europe. *Aquaculture*, 272: S139-S146.
- Hershberger, W.K. (1992). Genetic variability in rainbow trout populations. *Aquaculture*, 100: 51-71.
- Holsinger, K.E. & Weir, B.S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . *Nature Reviews Genetics*, 10: 639-650.
- Jakobsson, M. & Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23: 1801-1806.
- Johnson, N.A., Rexroad, C.E., Hallerman, E.M., Vallejo, R.L. & Palti, Y. (2007). Development and evaluation of a new microsatellite multiplex system for parental allocation and management of rainbow trout (*Oncorhynchus mykiss*) broodstocks. *Aquaculture*, 266: 53-62.
- Karlsson, S., Moen, T., Lien, S., Glover, K.A. & Hindar, K. (2011). Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Mol Ecol Resour*, 11: 247-253.

- Liu, S., Palti, Y., Gao, G. & Rexroad, C.E. (2016). Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture*, 452: 178-182.
- Liu, S., Vallejo, R.L., Gao, G., Palti, Y., Weber, G.M., Hernandez, A. & Rexroad, C.E., 3rd (2015). Identification of single-nucleotide polymorphism markers associated with cortisol response to crowding in rainbow trout. *Mar Biotechnol (NY)*, 17: 328-37.
- Manel, S., Gaggiotti, O.E. & Waples, R.S. (2005). Assignment methods: matching biological questions techniques with appropriate. *Trends Ecol Evol*, 20: 136-142.
- Matala, A.P., Ackerman, M.W., Campbell, M.R. & Narum, S.R. (2014). Relative contributions of neutral and non-neutral genetic differentiation to inform conservation of steelhead trout across highly variable landscapes. *Evol Appl*, 7: 682-701.
- Negrini, R., Nicoloso, L., Crepaldi, P., Milanese, E., Colli, L., Chegdani, F., Pariset, L., Dunner, S., Leveziel, H., Williams, J.L. & Marsan, P.A. (2009). Assessing SNP markers for assigning individuals to cattle populations. *Anim Genet*, 40: 18-26.
- Paetkau, D., Slade, R., Burden, M. & Estoup, A. (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol Ecol*, 13: 55-65.
- Palti, Y., Gao, G., Liu, S., Kent, M.P., Lien, S., Miller, M.R., Rexroad, C.E. & Moen, T. (2015a). The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Mol Ecol Resour*, 15: 662-72.
- Palti, Y., Gao, G., Miller, M.R., Vallejo, R.L., Wheeler, P.A., Quillet, E., Yao, J., Thorgaard, G.H., Salem, M. & Rexroad, C.E., 3rd (2014). A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Mol Ecol Resour*, 14: 588-96.
- Palti, Y., Vallejo, R.L., Gao, G., Liu, S., Hernandez, A.G., Rexroad, C.E., 3rd & Wiens, G.D. (2015b). Detection and Validation of QTL Affecting Bacterial Cold Water Disease Resistance in Rainbow Trout Using Restriction-Site Associated DNA Sequencing. *Plos One*, 10: e0138435.
- Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L. & Estoup, A. (2004). GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, 95: 536-539.
- Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155: 945-959.
- Ramos, A.M., Crooijmans, R.P., Affara, N.A., Amaral, A.J., Archibald, A.L., Beever, J.E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., Hansen, M.S., Hedegaard, J., Hu, Z.L., Kerstens, H.H., Law, A.S., Megens, H.J., Milan, D., Nonneman, D.J., Rohrer, G.A., Rothschild, M.F., Smith, T.P., Schnabel, R.D., Van Tassell, C.P., Taylor, J.F., Wiedmann, R.T., Schook, L.B. & Groenen, M.A. (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.
- Rannala, B. & Mountain, J.L. (1997). Detecting immigration by using multilocus genotypes. *P Natl Acad Sci USA*, 94: 9197-9201.
- Rosenberg, N.A. (2004). DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes*, 4: 137-138.
- Rousset, F. (2008). GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour*, 8: 103-106.
- Salem, M., Vallejo, R.L., Leeds, T.D., Palti, Y., Liu, S., Sabbagh, A., Rexroad, C.E., 3rd & Yao, J. (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *Plos One*, 7: e36264.
- Sanchez, C.C., Smith, T.P.L., Wiedmann, R.T., Vallejo, R.L., Salem, M., Yao, J.B. & Rexroad, C.E. (2009). Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics*, 10.

- Silverstein, J.T., Rexroad, C.E. & King, T.L. (2004). Genetic variation measured by microsatellites among three strains of domesticated rainbow trout (*Oncorhynchus mykiss*, Walbaum). *Aquac Res*, 35: 40-48.
- Starks, H.A., Clemento, A.J. & Garza, J.C. (2016). Discovery and characterization of single nucleotide polymorphisms in coho salmon, *Oncorhynchus kisutch*. *Mol Ecol Resour*, 16: 277-287.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, 30: 2725-2729.
- Weir, B.S. & Cockerham, C.C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38: 1358-1370.

Table 1. Fish samples used in this study. The strains of Troutlodge breeders and production fish are coded by the last letter of the population names, which are prefixed with two-digit or four-digit of year-class of the populations.

Group	Population	Strain	No. of fish	Source
Breeders	10TLUA	August	216	Troutlodge
	10TLUN	November	234	Troutlodge
	11TLUF	February	216	Troutlodge
	11TLUM	May	217	Troutlodge
	11TLUA	August	213	Troutlodge
	11TLUN	November	234	Troutlodge
	12TLUF	February	196	Troutlodge
	12TLUM	May	206	Troutlodge
Production fish	2012TLUN	November	94	Troutlodge
	2013TLUN	November	92	Troutlodge
	2015TLUF	February	94	Troutlodge
Farmed fish	Farm 1		49	Canada
	Farm 2		70	Idaho, USA
Non-Troutlodge fish	UW		47	NCCCWA
	Shasta		47	NCCCWA
	European 1		47	Europe
	European 2		47	Europe

UW: University of Washington; NCCCWA: National Center for Cool and Cold Water Aquaculture.

Table 2. Pairwise F_{st} values of the eight Troutlodge breeding populations.

	10TLUA	10TLUN	11TLUF	11TLUM	11TLUA	11TLUN	12TLUF
10TLUN	0.097						
11TLUF	0.132	0.110					
11TLUM	0.155	0.161	0.108				
11TLUA	0.086	0.112	0.156	0.192			
11TLUN	0.087	0.069	0.127	0.158	0.104		
12TLUF	0.162	0.149	0.102	0.120	0.163	0.126	
12TLUM	0.158	0.156	0.104	0.056	0.195	0.158	0.110

Table 3. Results of self-assignment of the eight Troutlodge breeding populations. Numbers in bold represent the number of fish assigned to the correct populations.

[illegible]

Fig. 1. Flow chart of Troutlodge rainbow trout breeding and production systems.

Fig. 2 UPGMA phylogenetic tree of the eight Troutlodge breeding populations of rainbow trout. The year-class of each population is coded by the two-digit prefix and the strain name is coded by the last letter of the population name.

Fig. 3 Population structure ($K = 2$) of the eight Troutlodge breeding populations, which are separated by black lines. The cluster membership coefficients were averaged over five replicates of cluster analyses. A, individual level; B, population level; C, delta K plot to determine the most appropriate K value following the method of Evanno et al. (2005).