Genetic Management Of Broodstock Populations With DNA Markers In Rainbow Trout

R.L. Vallejo[*], C.E. Rexroad III[*], S. Liu[*] and Y. Palti[*]

Introduction

DNA markers are very useful for aquaculture and fisheries broodstock management. They have been used for parentage assignment when spawning families share common environments, and for evaluating genetic parameters in broodstock populations. Previously, we developed a multiplex system (MS) of 10 co-amplifying microsatellite markers to reduce costs and improve efficiency of parentage assignments and estimation of population structure in rainbow trout (Johnson et al. 2007). The selective breeding program at the National Center for Cool and Cold Water Aquaculture (NCCCWA) provides a unique resource for assessing accuracy of estimation of pair-wise relationship coefficients and population *F*-statistics with pedigree and molecular data. The NCCCWA even- and odd-year strains were initiated in 2002 and 2003, respectively (Silverstein et al. 2004), maintaining a pedigree database system with every generation of breeders genotyped with the MS. We plan to use real and simulated marker genotype data to determine the number of markers and sample size needed for accurate prediction of pair-wise relationship coefficients, population *F*-statistics and structure with molecular data. We need this information as pair-wise relationships among the NCCCWA broodstock founders are unknown. This information is also useful for industry breeding programs lacking pedigree information for evaluating relatedness within/between hatchery stocks and wild populations. Overall the number of simulated marker genotypes needed for accurate prediction of relationship coefficients is still cost prohibitive, but anticipated future developments in DNA marker technology, will likely make it a feasible option for aquaculture and fisheries broodstock management.

Materials and methods

A multiplex of eight microsatellite loci were typed in individuals sampled from BY2005 (n = 143), BY2006 (n = 243), BY2007 (n = 177), and BY2008 (n = 171) following procedures described by Johnson et al. (2007). Marker allele frequency, heterozygosity, and exact test for HWE were performed using 10,000 permutations with SAS Proc ALLELE (SAS 2007).

Relationship coefficients from pedigree and marker data. Pair-wise relationship coefficients were estimated using marker genotype data from sampled individuals in BY2007 (n = 221) and BY2008 (n = 224) with their corresponding parents with SPAGEDI version 3.1a (Hardy and Vekemans 2002). Pair-wise relationship coefficients were also estimated using only pedigree information with SAS Proc INBRED (SAS 2007). Accuracy of relationship estimates from marker data were assessed by the correlation coefficient between relationships estimated from pedigree and marker data with SAS Proc CORR (SAS 2007).

[*]National Center for Cool and Cold Water Aquaculture, USDA-ARS, 11861 Leetown Rd., Kearneysville, West Virginia, USA

Table 1: Accuracy of pair-wise relationship coefficients[1] estimated from microsatellite loci[2] typed in samples from NCCCWA brood-years 2007 and 2008

Brood-year	Source of information	N[3]	Mean	S.D.	Correlation coefficient[4]	*P*-value
BY2007	Marker genotype	24531	0.080	0.147	0.59	<.0001
	Pedigree	24531	0.054	0.150		
BY2008	Marker genotype	25200	0.074	0.126	0.43	<.0001
	Pedigree	25200	0.080	0.148		

[1]Pair-wise relationship coefficients estimated with SPAGEDI version 3.1a (Hardy and Vekemans 2002). [2]Eight microsatellite loci were typed in fish sampled from BY2007 and BY2008 with their corresponding parents. [3]*N* indicates the estimated number of pair-wise relationship coefficients. [4]Pearson's correlation coefficient estimated with SAS Proc CORR (SAS 2007).

Table 2: Estimates of population *F*-statistics[1] using pedigree information and microsatellite loci[2] typed on samples from NCCCWA brood-years 2007 and 2008

Source of information	Strain	F_{IS}	F_{IT}	F_{ST}
Pedigree	Odd-year	-0.0243		
	Even-year	-0.0377		
	Global	-0.0313	-0.0174	0.0135
Marker genotype	Global	-0.0349	0.0109	0.0443
	P-value[1]	0.0	0.1575	0.0

[1]*F*-statistics were estimated with SPAGEDI version 3.1a (Hardy and Vekemans 2002), and exact *P*-values for global *F*-statistics were estimated using 10,000 permutations. [2]Eight microsatellite loci were typed in fish sampled from BY2007 and BY2008 with their corresponding parents.

***F*-Statistics from pedigree and marker data by strain.**
Eight microsatellite loci were typed in individuals sampled from BY2007 (*n*= 177) with their parents (BY2005, *n*= 44); and BY2008 (*n*= 171) with their parents (BY2006, *n*= 53). The individuals were grouped by odd- and even-year strain, and *F*-statistics as defined by Weir and Cockerham (1984) were estimated with SPAGEDI version 3.1a (Hardy and Vekemans 2002). The *F*-statistics were also estimated using only pedigree information with ENDOG version 4.6 (Gutiérrez and Goyache 2005).

***F*-statistics from marker data by brood-year (BY).** Eight microsatellite loci were genotyped in individuals sampled from BY2005 (*n* = 143), BY2006 (*n* = 243), BY2007 (*n* = 177), and BY2008 (*n* = 171). The BY pair-wise and global *F*-statistics were estimated with SPAGEDI version 3.1a (Hardy and Vekemans 2002).

Population structure. Eight markers were genotyped in fish sampled from BY2005 and BY2006. The sampled individuals comprised 81 unrelated full-sib (FS) families, each with 1-20 siblings. We generated three sets of unrelated individuals (*n* = 81 per set) by random sampling a sib from each FS family. Bayesian cluster analysis was run with STRUCTURE version 2.3.1 (Pritchard et al. 2000) to infer the number of sub-populations (*K*) present in the dataset (*X*). The prior probability of data [*Ln Pr(X|K)*] and posterior probability of *K* clusters [*Ln Pr(K|X)*] were estimated from independent runs of Markov chain-Monte Carlo (MCMC)

simulations using admixture and independent allele frequency model. A run included $1.5x10^6$ MCMC simulations from which the first third was discarded as a burn-in period.

Table 3: Estimates of population *F*-statistics[1] microsatellite loci[2] genotyped on samples from NCCCWA brood-year 2005, 2006, 2007 and 2008

Brood-year	BY2005	BY2006	BY2007	F_{IS}	F_{IT}	F_{ST}
BY2006	0.0226					
BY2007	0.0091	0.0341				
BY2008	0.0362	0.0070	0.0452			
Global				0.0265	0.0012	0.0254
P-value[1]				0.0003	0.8588	0.0

[1]*F*-statistics were estimated with SPAGEDI version 3.1a (Hardy and Vekemans 2002), and exact *P*-values for global *F*-statistics were estimated using 10,000 permutations. [2]Eight microsatellite loci were genotyped in individuals sampled from BY2005 (n = 143), BY2006 (n = 243), BY2007 (n = 177), and BY2008 (n = 171).

Results and discussion

The correlation coefficient between pair-wise relationship coefficients estimated with marker data and pedigree information was 0.59 and 0.43 in BY2007 and BY2008, respectively (Table 1). The low accuracy for pair-wise relationship coefficients estimated with marker genotype data might be due to the use of few microsatellite loci. As part of this study, we plan to determine the number of markers needed for accurate prediction of relationship coefficients via computer simulations.

The global F_{ST} = 0.0443 ($P < 0.01$) with marker data indicates there is significant genetic divergence between odd- and even-year NCCCWA rainbow trout strains (Table 2). The pedigree information provided marginal support for genetic divergence between NCCCWA strains (F_{ST} = 0.0135). The negative global F_{IS} = -0.035 ($P < 0.01$) with marker genotype data indicates there is a significant excess of heterozygotes in the NCCCWA population which is supported by similar and negative F_{IS} estimates from pedigree data. These results are consistent with the notion that the NCCCWA population is subdivided into discrete breeding units (i.e., odd- and even-year strains). The total inbreeding coefficient F_{IT} from marker data was non-significant in the NCCCWA population.

The pair-wise F_{ST} estimates between even and odd BY samples (BYs) indicate there is significant genetic differentiation between even and odd BYs which increased with generations of selective breeding (Table 3). The pair-wise F_{ST} estimates between either odd-year BYs (BY2005 vs. BY2007) or even-year BYs (BY2006 vs. BY2008) were small because BYs from the same strain (even- or odd-year strain) were compared. The global F_{ST} = 0.0254 ($P < 0.01$) supports also a significant genetic differentiation in the NCCCWA population. The estimated F_{IS} = 0.0265 ($P < 0.01$) is enigmatic and not consistent with the notion that the studied population is subdivided into discrete breeding units (Long 1986). This F_{IS} estimate is biased due to the use of a sampling framework (BY structured) that does not reflect the population structure model.

The likelihood support of the data *X* [*Ln Pr(X|K)*] and the posterior probability of the inferred *K* [*Ln Pr(K|X)*] were the highest for *K* = 4 in each analyzed dataset which indicates there is population sub-structure within the NCCCWA broodstock. The proportion of membership assigned to each unrelated individual is consistent with the number of founder strains in the NCCCWA population (Figure 1), and founder effects might be responsible for the structuring of populations. Clusters with predominantly red bars denote membership to Shasta (2), blue bars to Troutlodge (3), yellow bars to University of Washington (4), and green bars to Arlee (5) strain. Individuals of mixed ancestry showed membership to more than one cluster.

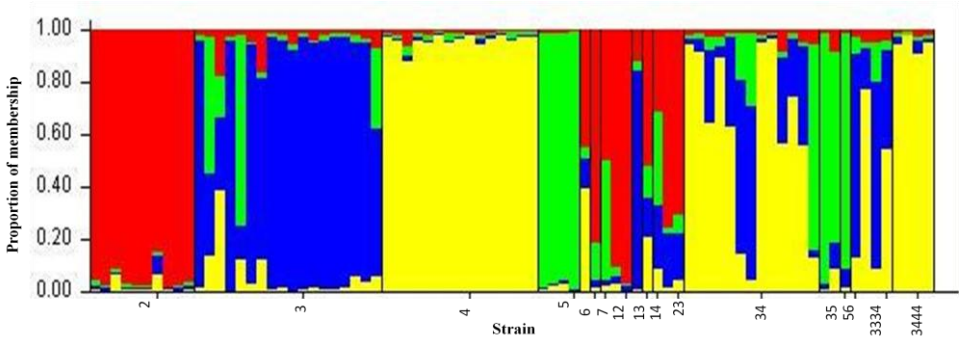

Figure 1. Proportion of membership assigned to inferred populations (*K* = 4) for each individual (*x*-axis) using unrelated individuals from BY2005 and BY2006 (*n* = 81) with STRUCTURE version 2.3.1 (Pritchard et al. 2000). Individuals are grouped by NCCCWA founder strains: 1= House Creek, 2= Shasta , 3= Troutlodge , 4= University of Washington, 5= Arlee, 6= Kamloop, 7= Hayspur.

Conclusion

The F_{ST} estimates and cluster analysis indicate there is a significant genetic differentiation and population sub-structure in the NCCCWA broodstock. Non-significant estimates of total inbreeding coefficient F_{IT} and negative F_{IS} values indicate there is an excess of heterozygotes in NCCCWA population. Pair-wise relationships from marker data had low accuracy.

References

Gutiérrez, J. P., and Goyache F. (2005). *J. Anim. Breed. Genet.*, 122:172-176.

Hardy, O. J., and Vekemans X. (2002). *Molec. Ecol. Notes*, 2: 618-620.

Johnson, N. A., Rexroad III, C. E., , Hallerman, E., *et al.* (2007). *Aquaculture*, 266:53-62.

Long, J. C. (1986). *Genetics*, 112:629-647.

SAS (2007). *SAS 9.1.3 Help and Documentation*, SAS Institute Inc., Cary, NC, USA.

Silverstein, J. T., Rexroad III, C. E., and King, T. L. (2004). *Aquaculture Res*., 35:40-48.

Weir, B. S., and Cockerham, C. C. (1984). *Evolution*, 38:1358-1370.