ESTIMATING LINKAGE DISEQUILIBRIUM AND EFFECTIVE POPULATION SIZE FROM GENOMIC DATA IN SPANISH ATLANTIC SALMON POPULATIONS

Chtioui, A.[1,2], Villanueva*, B.[1], Morán, P.[3], Kent, M.P.[4], Saura, M.[1]

[1]INIA, Madrid, Spain; Instituto de Ciencia y Tecnología Animal. Universitat Politècnica de València, Spain; [3]Universidad de Vigo, Spain; [4]CIGENE, Norwegian University of Life Sciences, Norway

*villanueva.beatriz@inia.es

Introduction

A detailed knowledge of the magnitude and extent of linkage disequilibrium (LD) - the non-random association of alleles at different loci - is required for determining the power of genome-wide association studies and gene mapping, and the accuracy of genomic selection. LD patterns are also used to infer past population sizes (N_e) and to provide thus an insight into the evolutionary history of populations. With the advent of high throughput genotyping techniques and the current availability of dense SNP panels in most farm animal species, genomic estimates of LD and N_e have attracted increasing attention in recent years. Although, in general, genomic tools for aquaculture species are far behind those available for terrestrial farm species, important advances have been made in the last years, particularly for Atlantic salmon, a species of major economic, social and environmental importance. A high-quality reference genome assembly has been published recently (Lien et al., 2016), and at least there have been three independent initiatives for this species to generate SNP arrays of different densities.

The Spanish populations of Atlantic salmon represent the southern distribution limit of the species in Europe and have suffered a dramatic decline over the last century due to the hydroelectric development and dam construction, river pollution and overfishing (Saura et al., 2010). Supportive breeding programs using native salmon were established at the beginning of the 1990s in order to re-establish these populations. The objective of this study was to evaluate, using a high-density SNP chip, LD patterns and ancestral N_e of Spanish Atlantic salmon populations.

Material and Methods

Data used originated from192 samples from six Spanish rivers (Figure 1) that cover all the distribution range of the species in Spain: Miño ($n = 16$), Ulla ($n = 108$), Eo ($n = 16$), Sella ($n = 16$), Urumea ($n = 16$) and Bidasoa ($n = 16$). A custom 220K Affymetrix SNP genotyping array (Aquagen/CIGENE) was used to genotype the samples. An Atlantic salmon physical map was provided by CIGENE. Quality control (QC) was performed with the R package SNPolisher (Affymetrix). SNP that did not satisfy QC thresholds, unmapped SNP, SNP with MAF < 0.01, and those that deviated significantly from Hardy Weinberg equilibrium ($p < 10^{-5}$) were excluded from the analyses. The number of SNP and samples retained after filtering was 164,722 and 187, respectively.

The LD measure used was r^2, the squared correlation coefficient between SNP pairs (Hill and Robertson, 1968) and was corrected for sample size. Pairwise r^2 between syntenic SNP separated by up to a distance of 5 Mb were obtained using Plink v1.09 (Figure 2).

Estimates of N_e were obtained from $N_{e(t)} = (4d)^{-1}[(r_d^2 - N^{-1})^{-1} - \alpha]$, where $N_{e(t)}$ is the effective population size t generations ago and r_d^2 is the mean value of r^2 for markers d Morgans apart. Based on this equation a non-linear least squares approach was implemented to statistically model the observed r^2. In order to avoid dependence between LD and linkage distance estimated from the same data, estimates of recombination rates from a different Atlantic salmon population (Lien et al., 2011) were

used. Thus, for each SNP pair in LD separated by a particular physical distance (Mb), an equivalent linkage distance (Morgans) was calculated as the product of the recombination rate for a particular chromosome obtained from Lien et al. (2011) and the physical distance.

Results and Discussion

Figure 2 shows that LD was relatively high between closely linked markers (average r^2 = 0.52 at 5 kb). However, LD declined rapidly with increasing distance between SNP pairs. The average r^2 decreased by half over the first 0.02 Mb and by 90% over the first 0.34 Mb, and was reduced to non-syntenic levels (0.0076) at distances greater than 20 Mb. The high levels of LD at short distances and the fast decline as distance increases has been also observed by Gutierrez et al. (2015) in a Canadian population of Atlantic salmon when using a 6.5K SNP array. In contrast, these rates of decrease in r^2 with increasing distance are much higher than those found in terrestrial farm animals. This could be explained by the large family sizes of salmon since recombination would have more chances to act.

**Figure 1.** Geographic location of the six rivers analysed in this study.
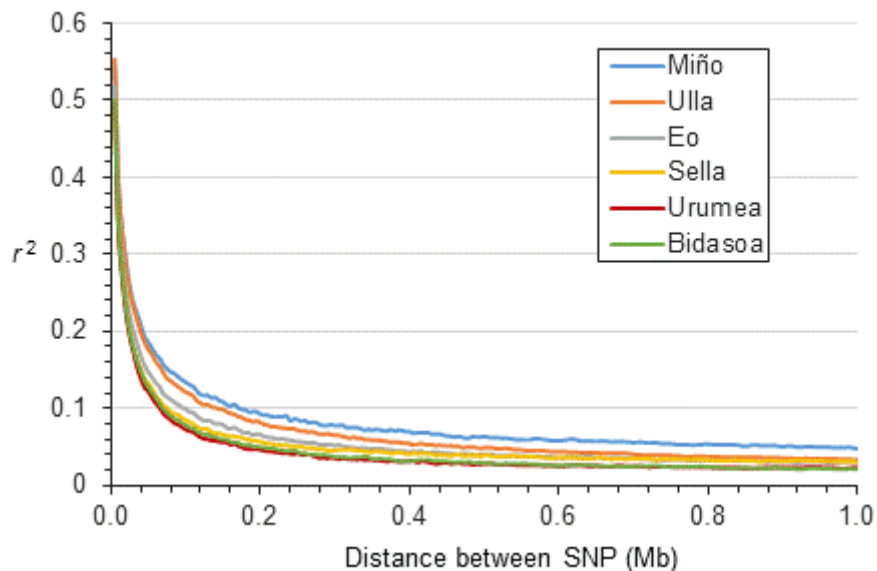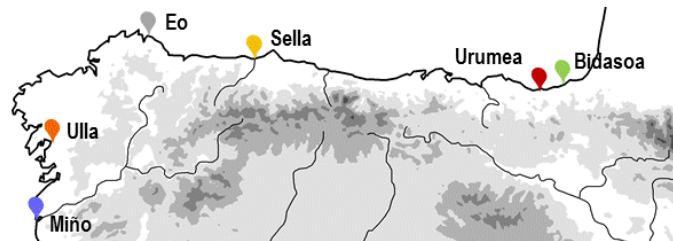
Figure 2. Average linkage disequilibrium measured as r^2, plotted against distance (Mb) for the different rivers.

Figure 3 shows that N_e six generations ago was 3-5 times lower than 50 generations ago (before the decrease in census size) in all rivers. Values for N_e ranged from about 500 (Miño) to 1200 (Bidasoa) individuals 50 generations ago and from 100 to 400 six generations ago. Average slopes of the regressions of N_e on time (generations in the past) before and after hydroelectric development (that occurred in the 1950s, about 25 generations ago) indicated that the rate of decrease in N_e was stronger after such development started ($b = 16$) compared with the previous period ($b = 12$).

It is interesting to observe that LD measures (and consequently N_e estimates) for the different rivers follow an isolation by distance pattern. This is expected given the homing behaviour of the species, with gene flow occurring among geographically close populations. Results for N_e are also in agreement with the demographic history of the populations. Atlantic populations have been maintained more isolated (and therefore more inbred) than populations draining into the Cantabric Sea.

Figure 3. Effective population size (N_e) based on linkage disequilibrium plotted against generations in the past for the different rivers.
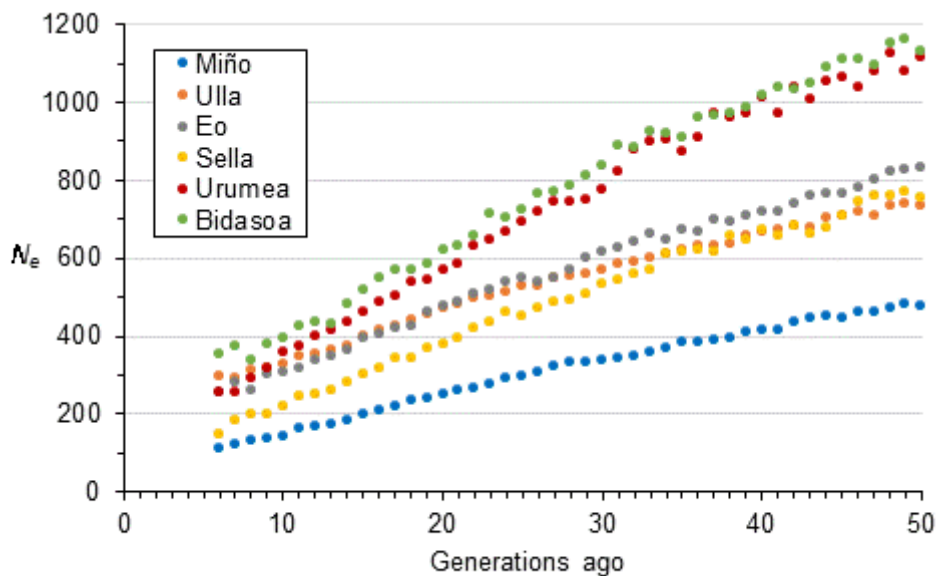
To our knowledge, this is the first study investigating LD patterns and the magnitude of N_e from genomic data in this particular metapopulation of Atlantic salmon which suffers the most extreme conditions of the distribution range of the species in the world.

References
Gutierrez AP, Yáñez JM, Fukui S et al. (2015). PLoS ONE 10(3):e0119730.
Hill WG, Robertson A (1968). Theoretical and Applied Genetics 38: 226-231.
Lien S, Gidskehaug L, Moen T et al. (2011). BMC Genomics 12:615.
Lien S, Koop BF, Sandve SR et al. (2016). Nature doi:10.1038/nature17164
Saura M, Morán P, Brotherstone S et al. (2010). Freshwater Biology 55: 923–930.