

Large-scale genome-wide SNP analysis reveals the rugged (and ragged) landscape of global ancestry, phylogeny, and demographic history in chicken breeds

Natalia V. DEMENTIEVA¹, Yuri S. SHCHERBAKOV¹, Olga I. STANISHEVSKAYA¹, Anatoly B. VAKHRAMEEV¹, Tatiana A. LARKINA¹, Artem P. DYSIN¹, Olga A. NIKOLAEVA¹, Anna E. RYABOVA¹, Anastasiia I. AZOVTSEVA¹, Olga V. MITROFANOVA¹, Grigoriy K. PEGLIVANYAN¹, Natalia R. REINBACH¹, Darren K. GRIFFIN², Michael N. ROMANOV^{2,3}

¹Russian Research Institute of Farm Animal Genetics and Breeding – Branch of the L. K. Ernst Federal Research Centre for Animal Husbandry, Pushkin, St. Petersburg, 196601, Russia

²School of Biosciences, University of Kent, Canterbury, CT2 7NJ, UK

³L. K. Ernst Federal Research Center for Animal Husbandry, Dubrovitsy, Podolsk, Moscow Oblast, 142132, Russia

Table S2 Calculation formulae for main population genetic parameters in this study

Parameter	Formula	Reference	Software
coefficient of inbreeding (F_{IS})	$F_{IS} = \frac{1-H_o}{H_e},$ where H_o is observed heterozygosity, H_e is expected heterozygosity	Purcell et al. (2007)	PLINK
observed heterozygosity (H_o)	$H_o = \frac{N(NM) - O(HOM)}{N(NM)},$ where $N(NM)$ is number of non-missing genotypes, $O(HOM)$ is observed number of homozygotes	Purcell et al. (2007)	PLINK
linkage disequilibrium (LD)	$r^2 = \frac{D^2}{p_{A_1}(1-p_{A_1})p_{B_1}(1-p_{B_1})},$ where r^2 is correlation coefficient of 1/0 (all or none) indicator variables corresponding to the presence of A_1 and B_1 , D^2 is difference between the frequency of gametes having the pair of alleles A_1 and B_1	Purcell et al. (2007)	PLINK
mean length of runs of homozygosity (ROH)	$Mean_{ROH} = \frac{\sum L_{ROH}}{Count_{ROH}},$ where L_{ROH} is homozygous region length, $Count_{ROH}$ is number of homozygous regions	Biscarini et al. (2019)	detectRuns
index of genomic inbreeding (F_{ROH})	$F_{ROH} = \frac{\sum L_{ROH}}{L_{genome}},$ where L_{ROH} is homozygous region length, L_{genome} is genome length	Biscarini et al. (2019)	detectRuns
effective population size (N_e)	$N_e = (4f(c_t))^{-1} (E[r^2_{adj} c_t]^{-1} - \alpha),$ where c_t is the recombination rate t generations ago in the past, r^2_{adj} is linkage disequilibrium estimation adjusted for sampling bias, α is a constant	Barbato et al. (2015)	SneP

References

- Purcell S, Neale B, Todd-Brown K, et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3):559-575. <https://doi.org/10.1086/519795>
- Biscarini F, Cozzi P, Gaspa G, et al., 2019. detectRUNS: detect runs of homozygosity and runs of heterozygosity in diploid genomes. Version 0.9.6. The Comprehensive R Archive Network.
- Barbato M, Orozco-terWengel P, Tapio M, et al., 2015. *SNeP*: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front Genet*, 6:109. <https://doi.org/10.3389/fgene.2015.00109>