VERSION 2

JUL 11, 2019

OPEN ACCESS

DOI:
dx.doi.org/10.17504/protocols.io.3xqgpmw

External link:
<https://doi.org/10.1093/gbe/evz128>

Document Citation: Changsheng Nie 2019. Genome-wide SNP data unveil admixture of Chinese indigenous chicken breeds with commercial breeds. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.3xqgpmw>

MANUSCRIPT CITATION:
Nie C, Almeida P, Jia Y, Bao H, Ning Z, Qu L, Genome-Wide Single-Nucleotide Polymorphism Data Unveil Admixture of Chinese Indigenous Chicken Breeds with Commercial Breeds. *Genome Biology and Evolution* 11(7). doi: [10.1093/gbe/evz128](https://doi.org/10.1093/gbe/evz128)

License: This is an open access document distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Genome-wide SNP data unveil admixture of Chinese indigenous chicken breeds with commercial breeds V.2

Genome Biology and Evolution

Changsheng Nie¹

¹State Key Laboratory of Animal Nutrition, Department of Animal Genetics and Breeding, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China.

Changsheng Nie

ABSTRACT

In this study, we aimed to evaluate the genetic diversity and population structure of eight chicken breeds (including conserved chicken breeds) from different geographic origin and to identify admixture within these breeds using a 600-K SNP (single nucleotide polymorphism) panel for genotyping.

Using the genotype of 580 961 single nucleotide polymorphism markers scored in 1 200 animals, we evaluated the genetic diversity (heterozygosity and proportion of polymorphic markers), linkage disequilibrium (LD) decay, population structure (principal component analysis and neighbor-joining tree), genetic differentiation (FST and genetic distance), and migration events (TreeMix and f-statistics) of the eight domesticated chicken breeds.

Created: Jun 09, 2019

Last Modified: Jul 11, 2019

DOCUMENT integer ID:
24272

Documentation

#Pairwise LD measures for multiple SNPs (genome-wide)

#source: <http://zzz.bwh.harvard.edu/plink/ld.shtml>

plink --file mydata --r2

#running ADMIXTURE in multithreaded mode

#source: <http://software.genetics.ucla.edu/admixture/admixture-manual.pdf>

admixture ~Data/dataset.bed 3 -j4

for K in 1 2 3 4 5; \

do admixture --cv dataset.bed \$K | tee log\${K}.out; done

#quickly view the CV errors

grep -h CV log*.out

#TreeMix: Build the ML tree

#source: https://bitbucket.org/nygcresearch/treemix/downloads/treemix_manual_10_1_2012.pdf

treemix -i inputfile.gz -o outstem

#Build the ML graph with migration (-m)

treemix -i inputfile.gz -m 2 -o outstem

#Graph visualization(R code)

source("src/plottingfuncs.R")

plot_tree("outstem")

plot_resid("outstem", "poporder")

#fstatistics

#source: https://bitbucket.org/nygcresearch/treemix/downloads/treemix_manual_10_1_2012.pdf

#threepop

threepop -i input.gz -k 500

#fourpop

fourpop -i input.gz -k 500

#calculate Fst between two populations (R code).

#Setting path

setwd("F:\\summary_files")

#Import data: two population summary files

af1<- read.table("summary_population1.csv",header = T, sep = "")

```

af2<- read.table("summary-population2.csv",header = T, sep = "")
#summary file generated by AxAS software: contain SNP site information and genotype frequency.
(https://assets.thermofisher.com/TFS-Assets/LSG/manuals/axiom\_genotyping\_solution\_analysis\_guide.pdf)
#calculate Fst between two populations
af1<- transform(af1,
Fst_population1_vs_population2= ((1-(((2*af1$AA+af1$AB+2*af2$AA+af2$AB)/(2*
(af1$AA+af1$AB+af1$BB)+2*(af2$AA+af2$AB+af2$BB)))**2+
((2*af1$BB+af1$AB+2*af2$BB+af2$AB)/(2*(af1$AA+af1$AB+af1$BB)+2*
(af2$AA+af2$AB+af2$BB)))**2))-(((1-(((af1$AA*2+af1$AB)/((af1$AA+af1$AB+af1$BB)*2)))**2+
((af1$BB*2+af1$AB)/((af1$AA+af1$AB+af1$BB)*2)))**2)*(af1$AA+af1$AB+af1$BB)+(1-
(((af2$AA*2+af2$AB)/((af2$AA+af2$AB+af2$BB)*2)))**2+
((af2$BB*2+af2$AB)/((af2$AA+af2$AB+af2$BB)*2)))**2)*
(af2$AA+af2$AB+af2$BB))/(af1$AA+af1$AB+af1$BB+af2$AA+af2$AB+af2$BB)))/(1-
(((2*af1$AA+af1$AB+2*af2$AA+af2$AB)/(2*(af1$AA+af1$AB+af1$BB)+2*
(af2$AA+af2$AB+af2$BB)))**2+((2*af1$BB+af1$AB+2*af2$BB+af2$AB)/(2*
(af1$AA+af1$AB+af1$BB)+2*(af2$AA+af2$AB+af2$BB)))**2)))
#n_AA, n_AB, n_BB: three different genotypes frequency
#Save results
write.csv(af1,"Fst_population1_vs_population2.csv",row.names = F, col.names = F, quote = F)

```