

Evolutionary history of Native Americans drawn by deep learning approach

O. Dolgova*, I. Maceda, O. Lao

Population Genomics Team, Centre Nacional d'Anàlisi Genòmica, Barcelona, Spain

* e-mail: olga.dolgova@cnag.crg.eu

Key words: human population genomics, deep learning, Approximate Bayesian Computation, Native Americans

Motivation and Aim: Population history of present-day Native Americans has been a hot debated topic in last decades. The consensus view on the peopling of the Americas is that ancestors of modern Native Americans entered the Americas from Siberia via the Bering Land Bridge and this occurred at least ~14.6 thousand years ago (ka). However the number and timing of migrations into the Americas remain controversial, with conflicting interpretations based on anatomical and genetic evidence [1, 2]. Our aim is to detect the demographic processes along the evolutionary history of Native American populations, including multiple introgression and admixture events among ancient and even unknown archaic populations.

Methods and Algorithms: Quantifying the amount of admixture and the relationship of (sometimes) unknown ancestral populations is a complex task in human population genomics [3]. Here we present a novel approach based on coupling of Deep Learning with Approximate Bayesian Computation (ABC), which overcomes the common problems related with summary statistics (SS) redundancy. Six admixture models including introgression events from unknown populations were developed for subsequent simulation testing (100,000 simulations per model). Unfolded multidimensional site frequency spectrum (SFS) of 29,899 callable intergenic regions ~49 kb long on average and free from CpG islands were computed and used as ABC SS. We trained 100 Artificial Neural Networks per each of nine Native American populations forming Deep Learning system of four layers in order to infer the informative SS and decrease the number of needed simulations. SFS estimation, model parameter computation and Deep Learning development were implemented in Java environment, while ABC was conducted using R scripts.

Results and Conclusions: Approximate Bayesian Computation coupled with Deep Learning constitutes a highly reliable approach for proposing and contrasting population genetics hypotheses. Positive correlations between simulated values of population parameters and predicted by DL were strong and significant in most of the cases as an evidence of robustness of Neural Network results. Estimation of posterior probabilities from six models elucidated the footprints of Neanderthal-Denisovan unknown ancestor introgression at least in three Native American populations. The posterior distributions of mean values for 51 parameters were also estimated using ABC-DL algorithm and full evolutionary history was reconstructed.

References

1. Raghavan M. et al. (2015) POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*. 349(6250):aab3884.
2. Soglund P. et al. (2015) Genetic evidence for two founding populations of the Americas. *Nature*. 525:104-108.
3. Slatkin M. (2016) Statistical methods for analyzing ancient DNA from hominins. *Curr. Opin. Genet. Dev.* 41:72-76.