GeLaTo: Genes and Languages Together

The GeLaTo dataset is a worldwide diversity panel of available population genetic samples matched with databases of linguistic, cultural and environmental diversity. Population genetic samples are assigned to existing GlottoCodes, following ethnolinguistic criteria: the data is filtered following the indication of geneticists, linguists, cultural anthropologists and historians. The choice of genetic data corresponds to essential guidelines: maximum compatibility and standardization, modern high quality data, avoidance of ascertainment bias, availability for different regions of the world, and finally high resolution to capture recent events. The dataset provides elaborated summary statistics such as genetic diversity within a population, genetic proximity between pairs of populations, sharing of identical motifs, and demographic history reconstructions. The genetic samples are directly linked to Glottolog and D-Place databases, and to the original publication. The current version hosts summary statistics from the genetic diversity panel of autosomal STR from Pemberton et al. 2013 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3656735/. It will be expanded to include mtDNA genomes, Y chromosome STRs, and autosomal SNPs.

GeLaTo goals:

* Allowing geneticists to properly characterize the human history behind the molecular data, and give an accessible reference dataset for regional or worldwide comparisons.
* Allowing linguists, historians and cultural anthropologists to integrate information on genealogical relatedness and demography, which can be robustly extrapolated from the genetic data.
* Allowing scholars of various disciplines to approach questions of major relevance on human diversity in a true multidisciplinary perspective, and develop a more realistic understanding of the complex mechanisms behind human migration, contact and cultural transmission.

data – page

measures for the genetic samples.

ExpectedHeterozygosity:

Expected heterozygosity calculated with the function Hs from package package {adegenet}

Hs is commonly used for measuring within population genetic diversity (and as such, it still has sense when computed from haploid data).

Matrix of genetic distances. the matrix of genetic distance between populations calculated with the function Dist.genpop (Nei 1972) from package adegenet.

AverageDistance

Mean of the genetic distance between the given genetic sample and the other samples of the dataset. See Matrix of genetic distances for all the pairwise distances.

AverageDistanceRegional

Mean of the genetic distance between the given genetic sample and the other samples of the same geographic region.

See Matrix of genetic distances for all the pairwise distances.