BEAGLE

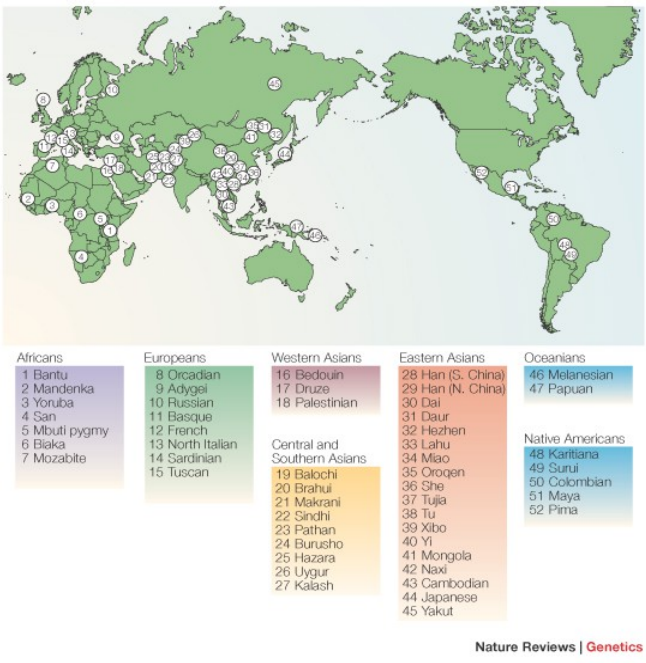
Science Highlight

IMPUTING GENOME SEQUENCE VARIATION IN THE HUMAN GENOME DIVERSITY PROJECT PANEL

D. Witonsky, C. Jeong, A. Di Rienzo

To enable studies of human genetic adaptations, we have imputed genotypes for the entire genome of the 938 individuals in the Human Genome Diversity Project panel, which includes samples from 52 populations from 6 continents genotyped at a set of 637,707 autosomal single nucleotide polymorphisms. As a reference panel, we used whole genome sequence data for the 2,504 individuals in the 1000 Genomes Project. After applying filters for quality control to the imputed genotype data, we obtained high-confidence imputed genotype data at 9,162,006 autosomal SNPs, thus achieving a greater than 14-fold increase in the amount of available sequence variation data for the Human Genome Diversity Project panel. The imputed genotype data will be used for investigating models of human adaptations to different local environments. In addition, the data set will be made publicly available to investigators interested in the study of human evolution...

Our laboratory has been particularly interested in detecting human genetic adaptations to different local environments. Since their migration out of Africa 50–100 thousand years ago (kya), anatomically modern humans have colonized a wide range of environments in a relatively short time period. For example, archaeological studies provide evidence of human habitation in environments extremely divergent from those of sub-Saharan Africa, such as cold climates in arctic Siberia or high-altitude environments in the Tibetan plateau, as early as 27 and 30 kya, respectively. In addition to differences in the physical environment, cultural transitions such as the introduction of agriculture and pastoralism also contributed to divergence of human environments. Understanding the genetic basis of heritable beneficial traits is a major goal of human genetics...



Training:

Intro to Python
///date to be announced///

Topics will include:

Python on Beagle2

Scripts/modules/paths

Functions

Strings

Numerics with Python

Plotting with Python

Monte Carlo methods for the steady state diffusion equation

Parallel scripting lectures
///date to be announced///

Topics will include:

Why we need to pack jobs on Beagle: data location hierarchies and types of parallel problems

How to write pbs/bash scripts that run many applications sequentially

How to write swift scripts that run many applications sequentially and in parallel

Hands on user problem solution (if there are any)

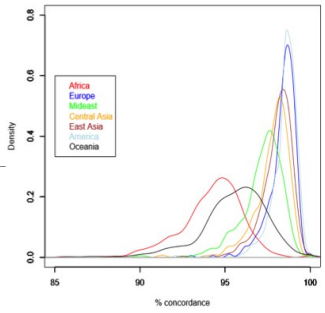
Beagle2 Events To learn more about Beagle2 trainings.

Request for publications made using Beagle2

Please send us your most recent publications made possible by Beagle2. Make sure that your papers acknowledge Beagle and contain the grant number: 1S10OD018495-01

The plot below shows the distribution of the concordance rates for SNP genotypes imputed in different groups of populations. In general, the concordance rate is very high, i.e. >90%, for all populations. However, there are significant differences across ethnic groups. For example, the distributions for African and Oceanian populations are clearly shifted to the left, reflecting lower accuracy in imputations in these groups. Because genotype imputation of unassayed SNPs is based on patterns of linkage disequilibrium among SNPs, which in turn is a property of the history of the population, the lower concordance rate in African individuals is likely due to the well-documented lower extent of LD in these populations. This explanation is unlikely to account for the same finding in Oceanian populations, which are known to be heavily drifted and therefore harbor high levels of LD. The lack of representation of Oceanian populations in the 1KGP reference panel is more likely to underlie the lower imputation accuracy. Both factors are well known to affect imputation accuracy, thus caution should be taken in interpreting the results for these populations.

We will now use this imputed data set that we generated using Beagle to investigate polygenic adaptations at the genome-wide level. In addition, the full imputed data set, with information about estimates of the uncertainty for each imputed genotype, will be made publicly available through the Di Rienzo lab web site (http://genapps.uchicago.edu/newlabweb/index.html). We anticipate that this resource will be widely used by investigators interested in similar questions as well as in projects focusing on the evolutionary history of our species.



To read full article go to: **Science on Beagle2**

Beagle2 Related Publications

A.M. Fluitt, J.J. de Pablo
An analysis of biomolecular force fields for simulations of polyglutamine in solution
(in press) Biophysical Journal 2015

W. Guo, Y. Ji, D. Catenacci
A Subgroup Cluster Based Bayesian Adaptive Design for Precision Medicine
(submitted to JASA)

B. Dhakshnamoorthy, L. Blachowicz, H. Rui, B. Roux
Structural and functional characterization of a calcium-activated cation channel from Tsukamurella paurometabola
(in preparation & to be submitted)

S.M. Islam, R. Cheng, A. R. Stein, at all., M. C. Maduke
Monitoring H+-driven structural changes in a CLC chloride-proton antiporter using computationally-guided double electron-electron resonance spectroscopy
(in preparation to be submitted)

M. Fajer, Y. Meng, B. Roux
The Complete Activating Transition in c-Src Kinases
(in preparation to be submitted)

Searle Chemistry Laboratory
5735 South Ellis Avenue
Chicago, IL 60637

Help E-Mail:
beagle-support@ci.uchicago.edu
Beage2 CI Website:
beagle.ci.uchicago.edu