

BIOL 502 Population Genetics Spring 2017

Week 14 Human Population Genomics

Arun Sethuraman

California State University San Marcos

Peoples and Cultures

Cavalli Sforza 1997 PNAS

Table 1. Genetic distances among major continents or continental areas, based on 120 classical polymorphisms Africa Oceania East Asia Europe Oceania 24.7

10.0

9.7

8.9

9.5

13.5 14.6 Information for this table was adapted from refs. 1 and 2.

20.6

16.6

22.6

East Asia

Europe

America

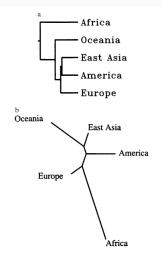


Fig. 1. (a) Tree derived from data in Table 1 by unweighted pair-group method with arithmetic mean. (b) Tree from same data by neighbor joining. Note the difference in the location of the branch leading to Europe.

Cavalli Sforza 1997 PNAS

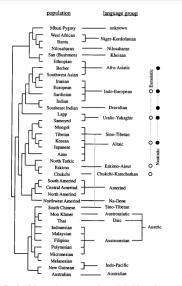
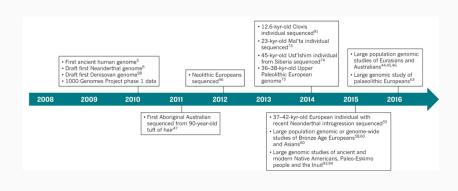


Fig. 3. Coherence between a genetic tree derived from 42 populations with 120 classical polymorphisms (*Left*) and what is known of the linguistic tree (*Right*), including two recently reconstructed superfamilies (shown at the extreme right). (From ref. 23.)

The Human Genome Project - 2001

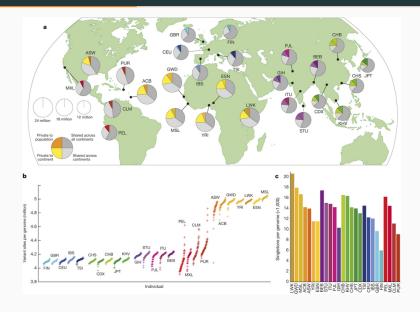


A timeline - Nielsen et al. 2017 DOI:10.1038/nature21347

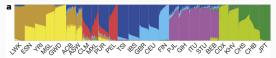


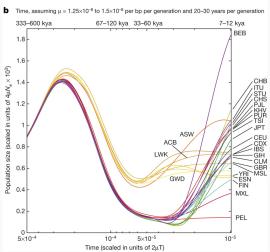
Human Genomic Variation

The 1000 Genome Project - 2015 DOI:10.1038/nature15393

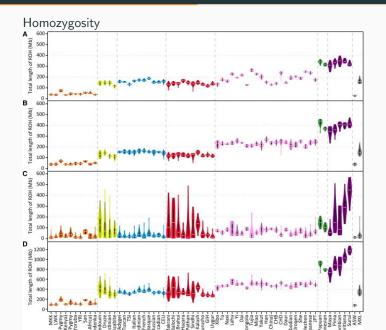


The 1000 Genome Project - 2015

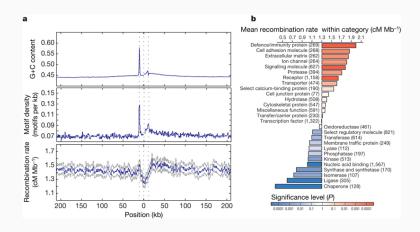




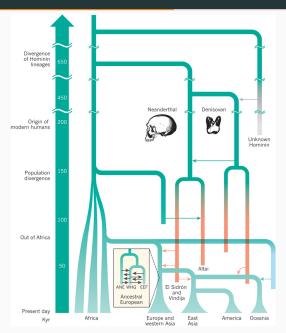
Pemberton et al. 2012 DOI:10.1016/j.ajhg.2012.06.014



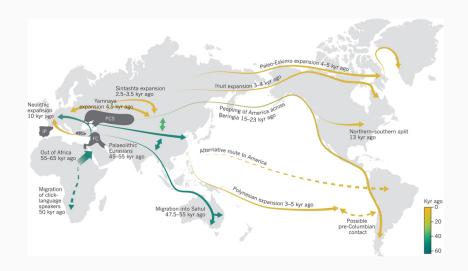
Frazer et al 2007 DOI: 10.1038/nature06258



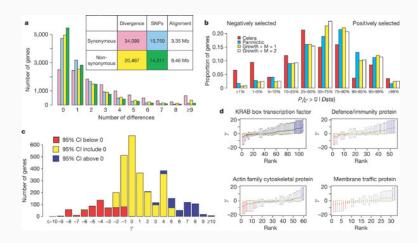
Nielsen et al. 2017 DOI: 10.1038/nature21347



Nielsen et al. 2017 DOI:10.1038/nature21347



Bustamante et al. 2005 DOI:10.1038/nature04240



Where to access data:

- 1000 Genome Project http://www.internationalgenome.org/
- UK10K Project https://www.uk10k.org/
- DECODE Project Genome of Iceland https://www.decode.com/
- Complete Genomics Data http://www.completegenomics.com/public-data/69-genomes/
- Private databases 23andMe, ancestry.com, etc.

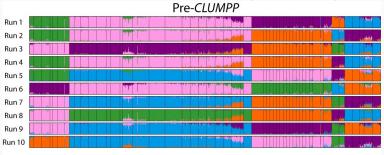
Statistical Issues

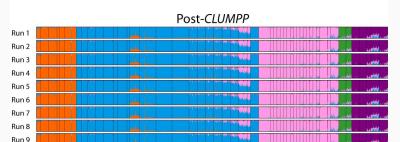
Statistical Issues

- Sample size
- Multiple testing and corrections
- Missing data/unsampled variation
- Genotyping/sequencing errors
- Quality control
- Bioinformatics errors
- Identifiability
- Confounding signals
- Label switching

e.x. Label switching

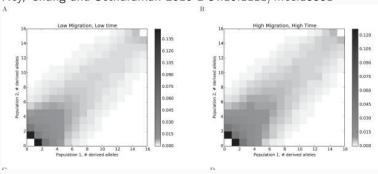
Jakobsson and Rosenberg 2007 DOI:10.1093/bioinformatics/btm233





e.x. Identifiability





e.x. Confounding effects

Falush et al. 2017 DOI: http://biorxiv.org/content/biorxiv/early/2016/07/28/066431.full.pdf

