UNIVERSITY OF TARTU
1632

Digging genomes to understand our past - modern tools for old puzzles

Mait Metspalu

INSTITUTE OF GENOMICS

29.06.2021

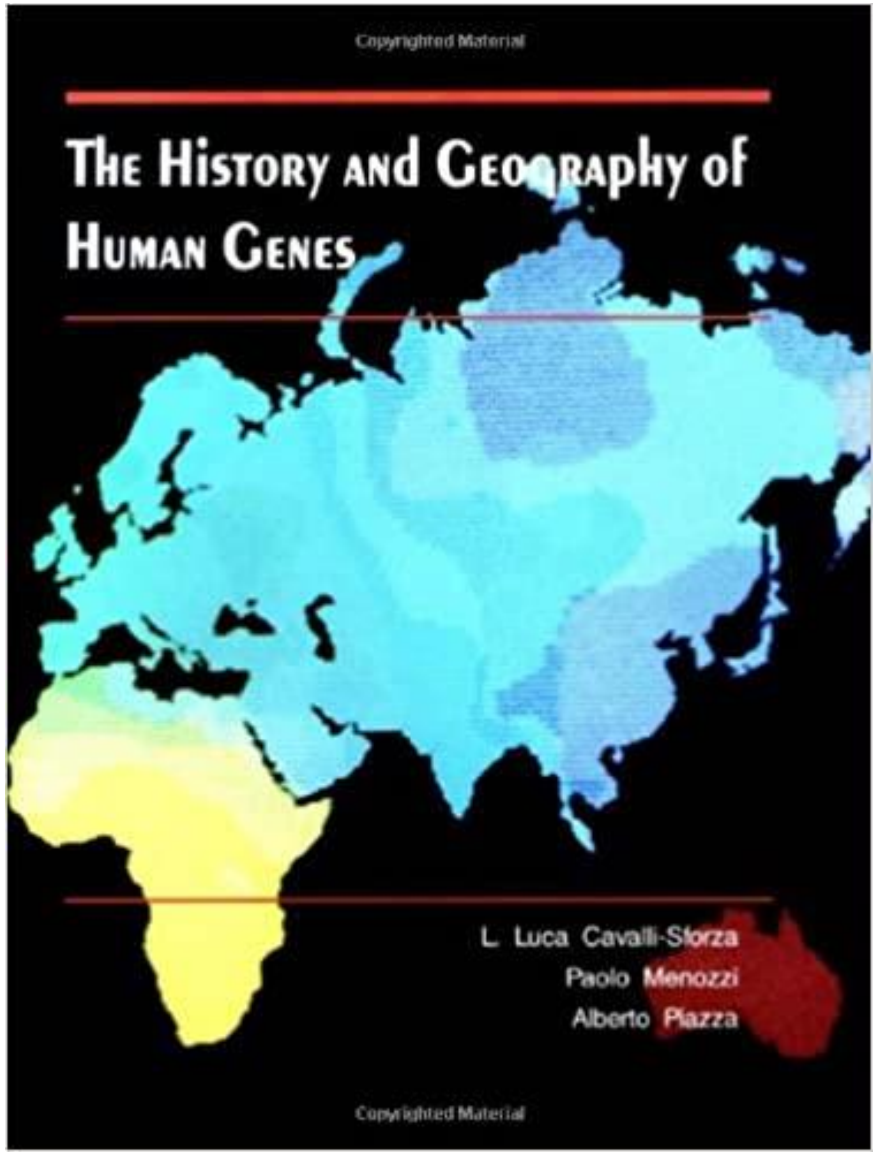
(Human) genetic diversity is shaped by

- demographic (evolutionary) history (genetic drift)
  - Time depth of most recent common ancestor
  - Migratory patterns
  - Admixture patterns
  - …
- natural selection
- culture

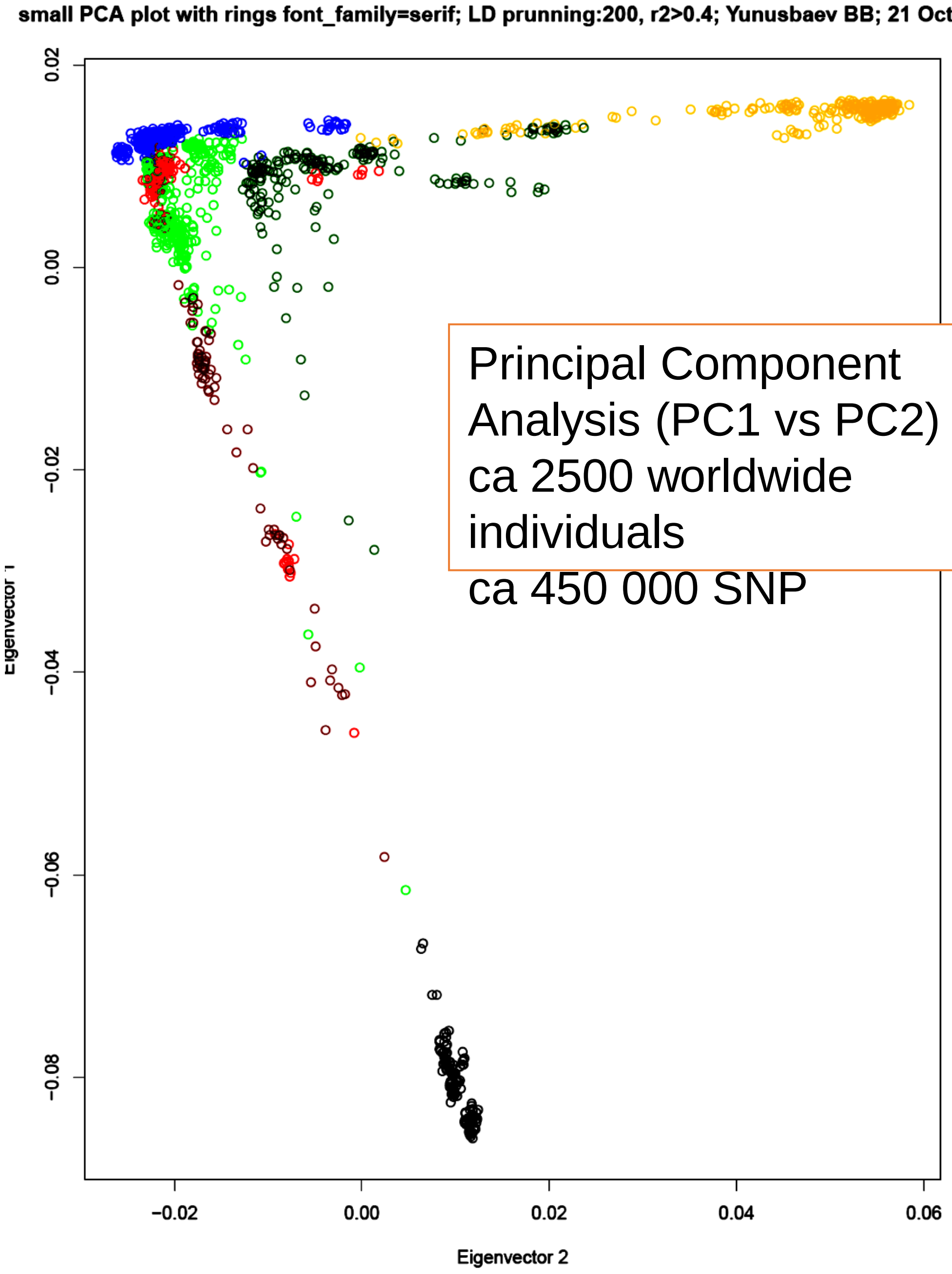
Hence we can learn about our demographic history, natural selection and even culture by studying (the genesis of) genetic diversity

Principal components analysis

- Invented by Karl Pearson in 1901
- Reduces complexity in the dataset. Creates synthetic characters, which do not correlate with each other and orders them by the amount of variation explained*.
- Goes by many names
- Made mainstream in population genetics in the “bible” published in 1994
- PCA describes variation in the sample being completely oblivious to the processes which create the variation
- Most of the genome evolves over time through random genetic drift (and not natural selection). The more time has passed since the split between two populations, the more genetically different they become. In most cases this separation translates also to geographical distance. Hence PCA on genetic data often mirrors geography of the populations.
- Most population genetics papers include a PCA plot



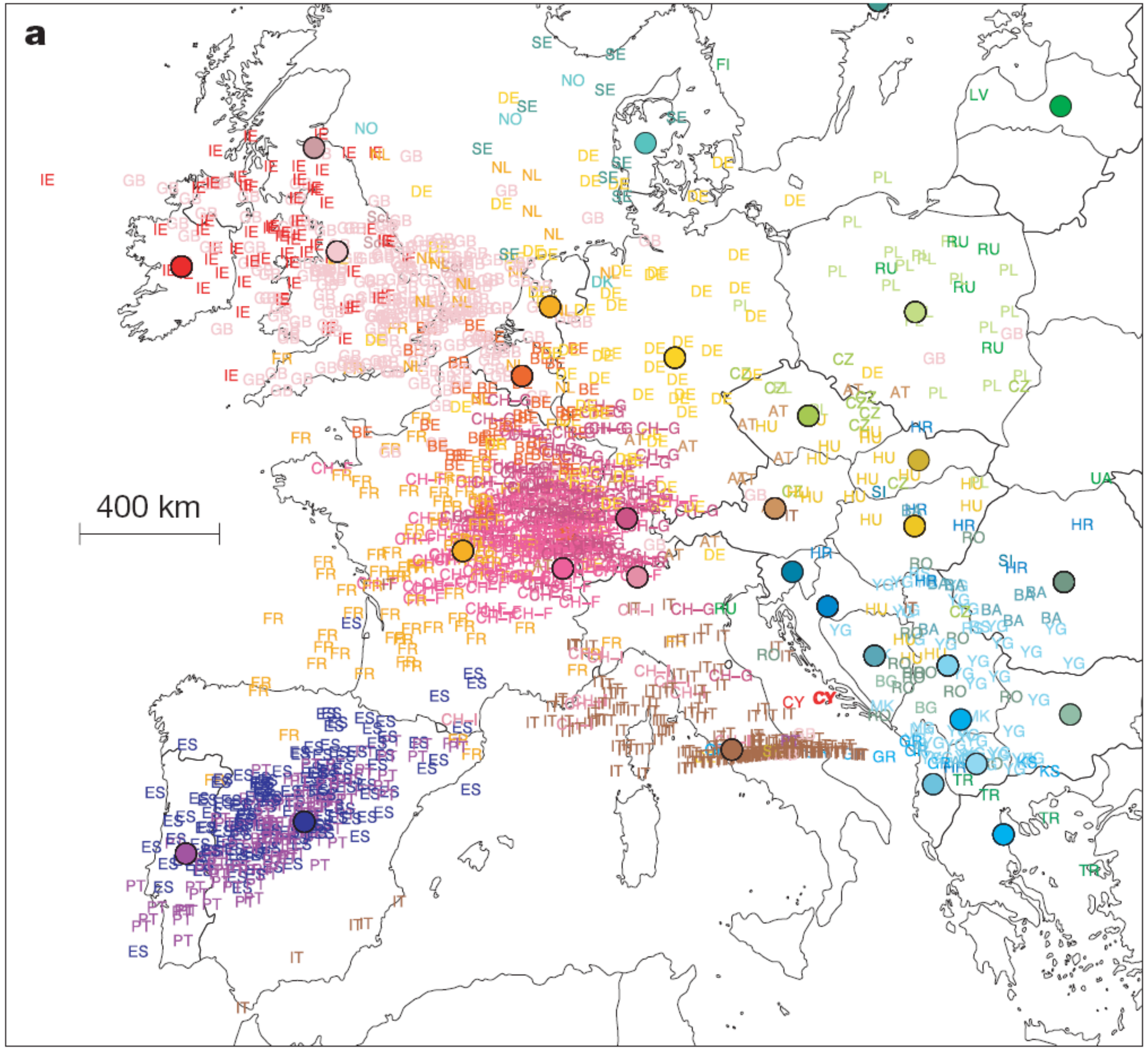
the “bible”



Principal Component Analysis (PC1 vs PC2)
ca 2500 worldwide individuals
ca 450 000 SNP

Genes mirror geography within Europe

John Novembre[1,2], Toby Johnson[4,5,6], Katarzyna Bryc[7], Zoltán Kutalik[4,6], Adam R. Boyko[7], Adam Auton[7],
Amit Indap[7], Karen S. King[8], Sven Bergmann[4,6], Matthew R. Nelson[8], Matthew Stephens[2,3] & Carlos D. Bustamante[7]



INSIGHT
Exoplanets

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

MELTING POT

THREE ANCESTRAL POPULATIONS FOR MODERN EUROPEANS

PAGE 409

INSIDE THIS ISSUE
Inclusive science is better science
A NATURE AND SCIENTIFIC AMERICAN SPECIAL ISSUE ON DIVERSITY

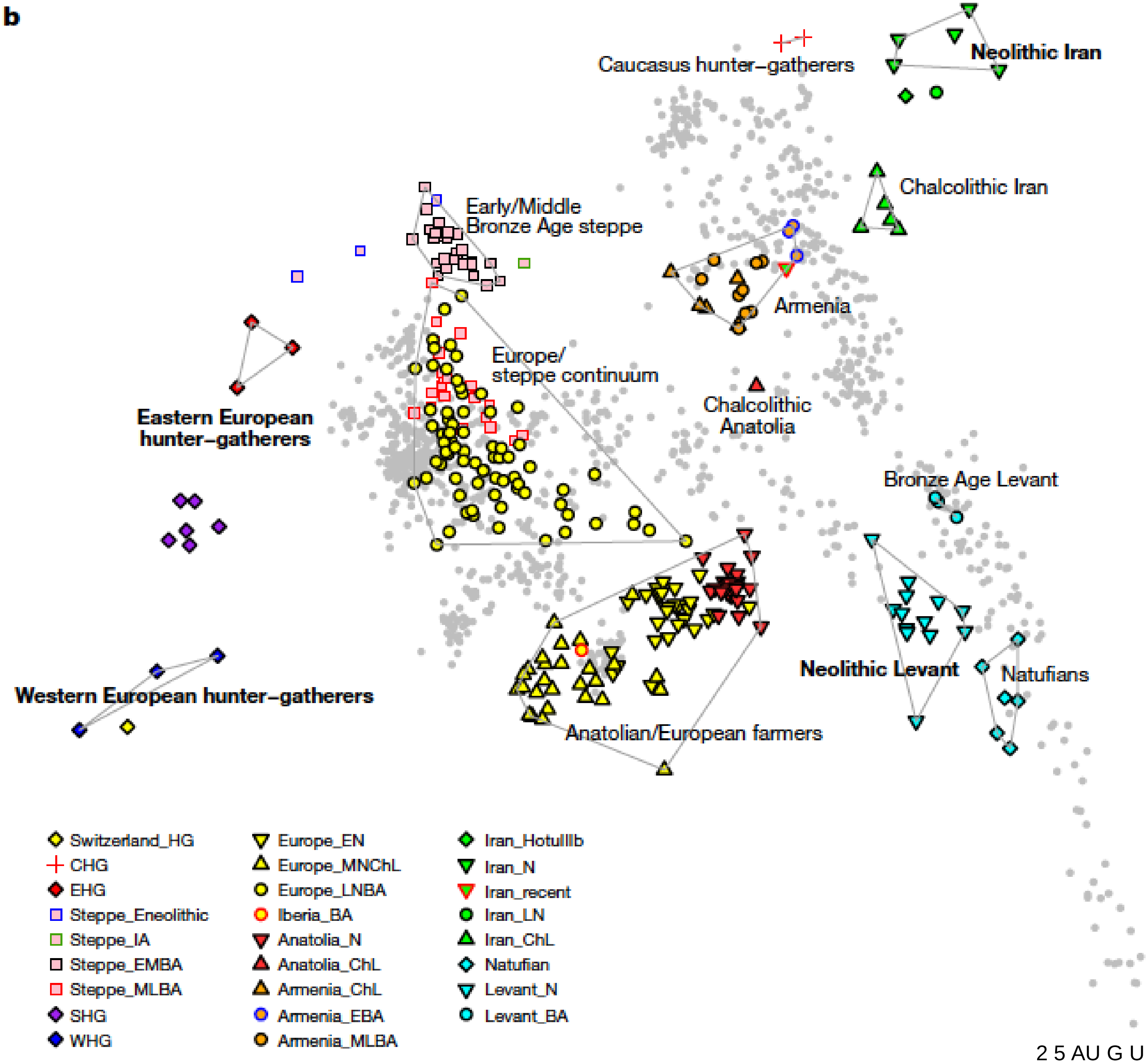
REVIEWS
AUTUMN BOOKS SPECIAL
E. O. Wilson, Steven Pinker, Naomi Klein and more
PAGE 306

ANIMAL BEHAVIOUR
BORN TO GET WILD
Chimpanzees don't need our help to be agressive
PAGES 321 & 414

GEODYNAMICS
ORIGIN OF PLATE TECTONICS
Did continental spread kick-start continental drift?
PAGE 405

NATURE.COM/NATURE
18 September 2014 £10
Vol. 513, No. 7518

38
9 770028 083095


b
Caucasus hunter–gatherers
Neolithic Iran
Chalcolithic Iran
Early/Middle Bronze Age steppe
Armenia
Europe/ steppe continuum
Chalcolithic Anatolia
Eastern European hunter–gatherers
Bronze Age Levant
Neolithic Levant
Natufians
Western European hunter–gatherers
Anatolian/European farmers
Switzerland_HG
CHG
EHG
Steppe_Eneolithic
Steppe_IA
Steppe_EMBA
Steppe_MLBA
SHG
WHG
Europe_EN
Europe_MNChL
Europe_LNBA
Iberia_BA
Anatolia_N
Anatolia_ChL
Armenia_ChL
Armenia_EBA
Armenia_MLBA
Iran_HotuIIIb
Iran_N
Iran_recent
Iran_LN
Iran_ChL
Natufian
Levant_N
Levant_BA

Current Biology
Report
CellPress

Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe

Lehti Saag,[1,2,7,*] Liivi Varul,[3] Christiana Lyn Scheib,[4] Jesper Stenderup,[5] Morten E. Allentoft,[5] Lauri Saag,[2] Luca Pagani,[2] Maere Reidla,[1,2] Kristiina Tambets,[2] Ene Metspalu,[1,2] Aivar Kriiska,[6] Eske Willerslev,[5] Toomas Kivisild,[1,2,4] and Mait Metspalu[2,*]
[1]Department of Evolutionary Biology, Institute of Cell and Molecular Biology, University of Tartu, Tartu 51010, Estonia
[2]Estonian Biocentre, Tartu 51010, Estonia
[3]School of Humanities, Tallinn University, Tallinn 10120, Estonia
[4]Department of Archaeology and Anthropology, University of Cambridge, Cambridge CB2 3QG, UK
[5]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen 1350, Denmark
[6]Department of Archaeology, Institute of History and Archaeology, University of Tartu, Tartu 51014, Estonia
[7]Lead Contact
*Correspondence: lehtis@ut.ee (L.S.), mait@ebc.ee (M.M.)

http://dx.doi.org/10.1016/j.cub.2017.06.022

The Arrival of Siberian Ancestry Connecting the Eastern Baltic to Uralic Speakers further East

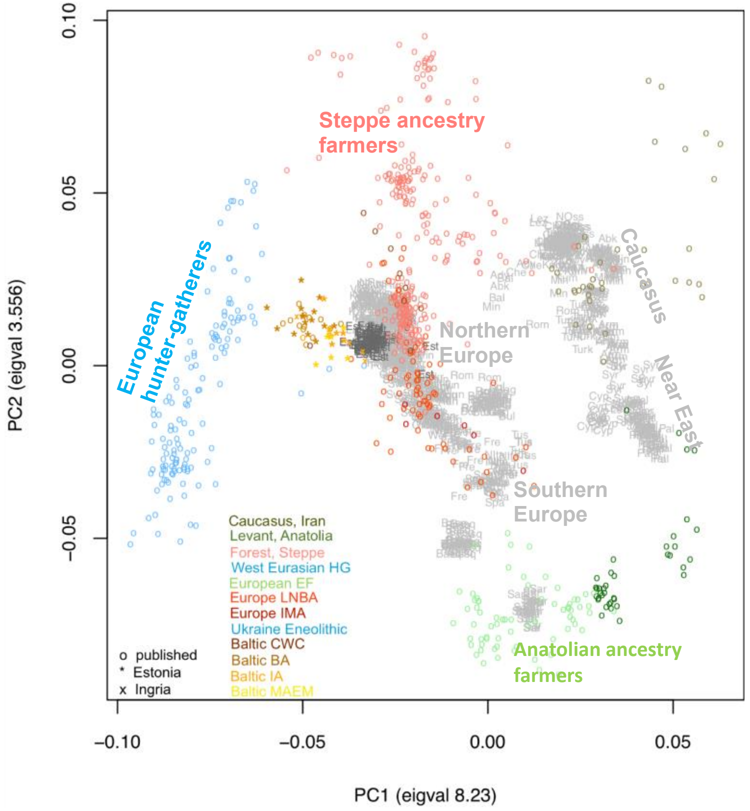
Lehti Saag,[1,2,16,*] Margot Laneman,[3] Liivi Varul,[4] Martin Malve,[3] Heiki Valk,[3] Maria A. Razzak,[5] Ivan G. Shirobokov,[6] Valeri I. Khartanovich,[6] Elena R. Mikhaylova,[7] Alena Kushniarevich,[1] Christiana Lyn Scheib,[1] Anu Solnik,[1] Tuuli Reisberg,[1] Jüri Parik,[1,2] Lauri Saag,[1] Ene Metspalu,[1] Siiri Rootsi,[1] Francesco Montinaro,[1] Maido Remm,[8] Reedik Mägi,[9] Eugenia D'Atanasio,[10] Enrico Ryunosuke Crema,[11] David Díez-del-Molino,[12,13] Mark G. Thomas,[14,15] Aivar Kriiska,[3] Toomas Kivisild,[1,2,10] Richard Villems,[1,2] Valter Lang,[3] Mait Metspalu,[1] and Kristiina Tambets[1,*]

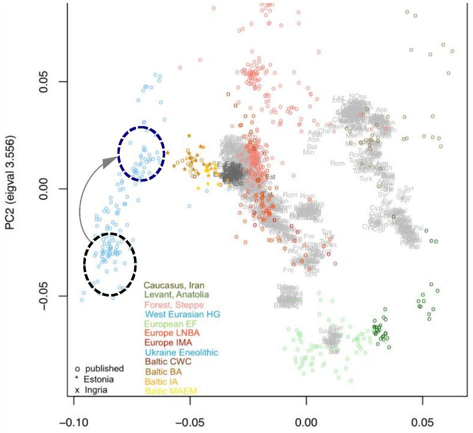
11,000 ya | 6,000 ya | 4,900 ya | 3,800 ya | 2,600 ya | 700 ya

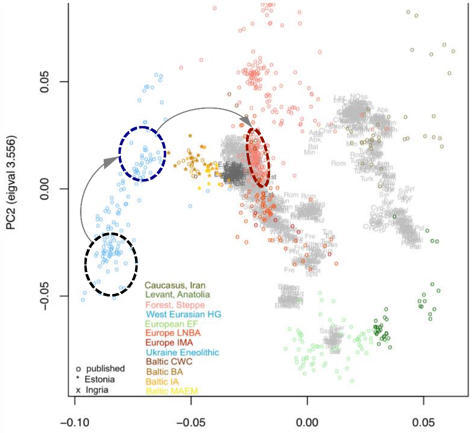
Mesolithic | Neolithic | Bronze Age | Iron Age | Middle Ages

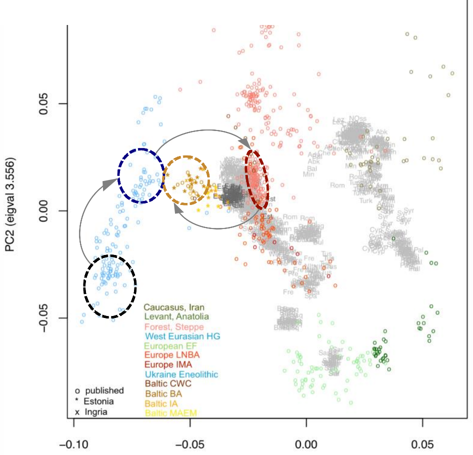
Kunda | Narva | Comb Ceramics | Corded Ware

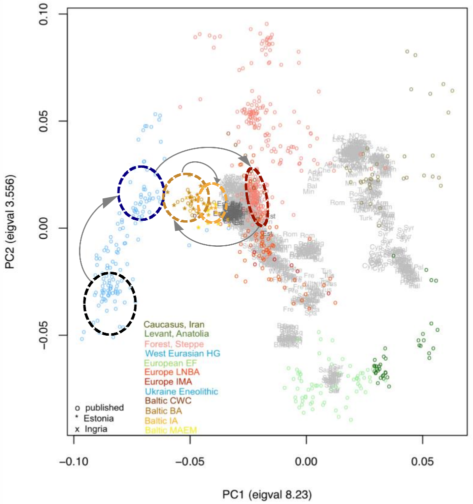
hunting-gathering | farming

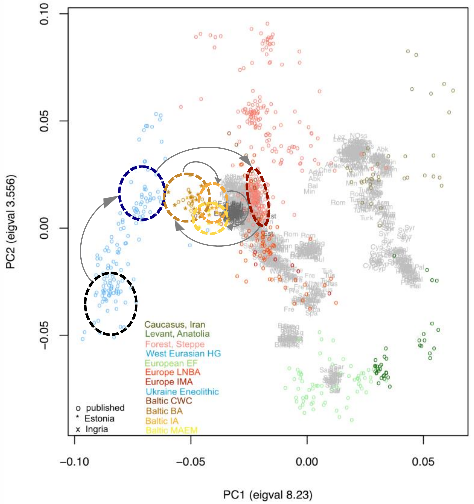


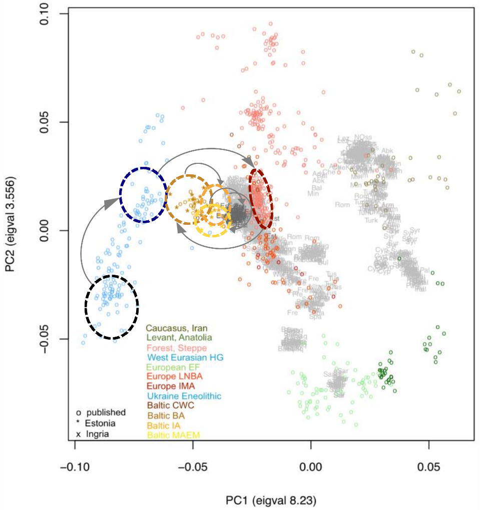


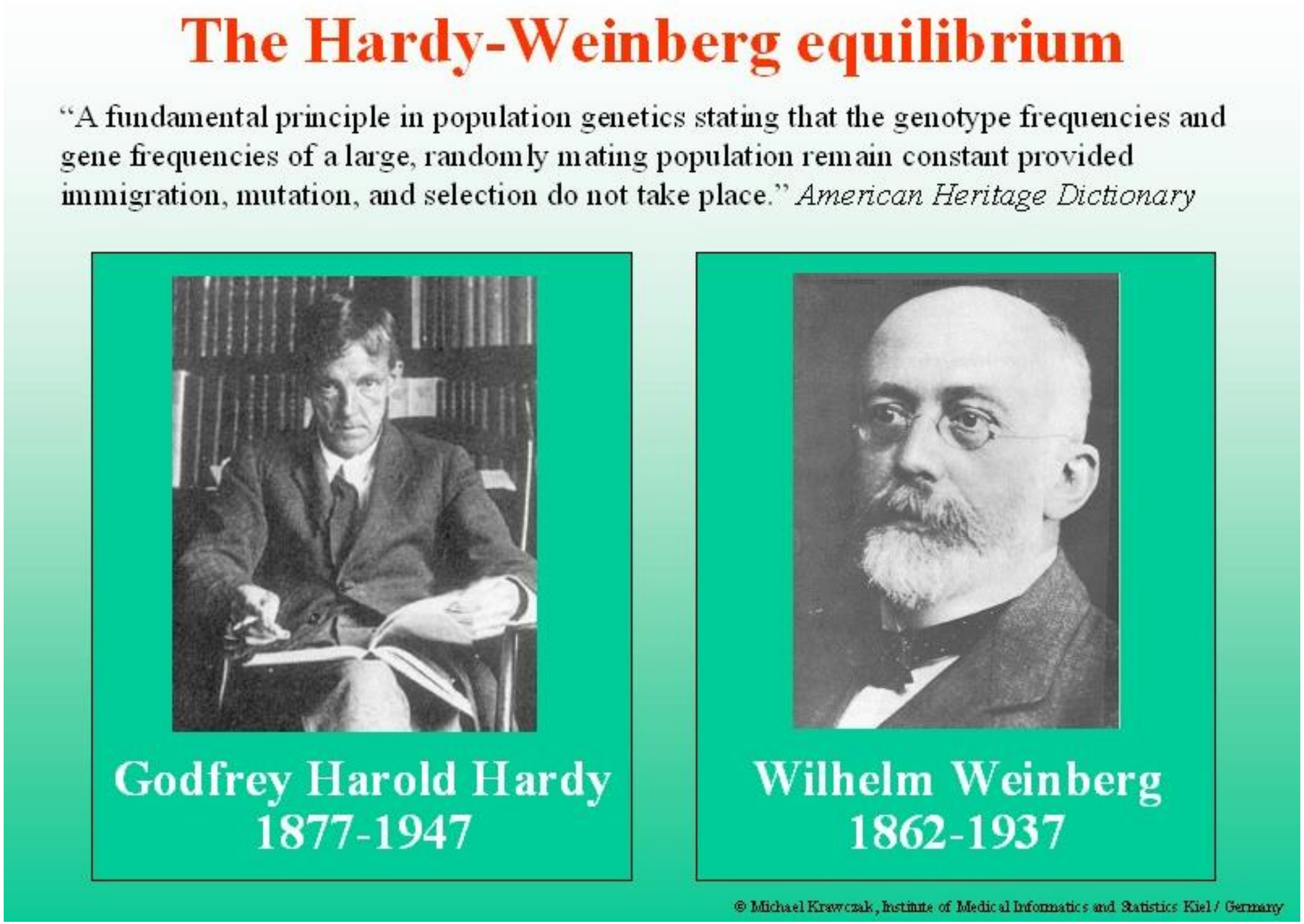












The Hardy-Weinberg equilibrium

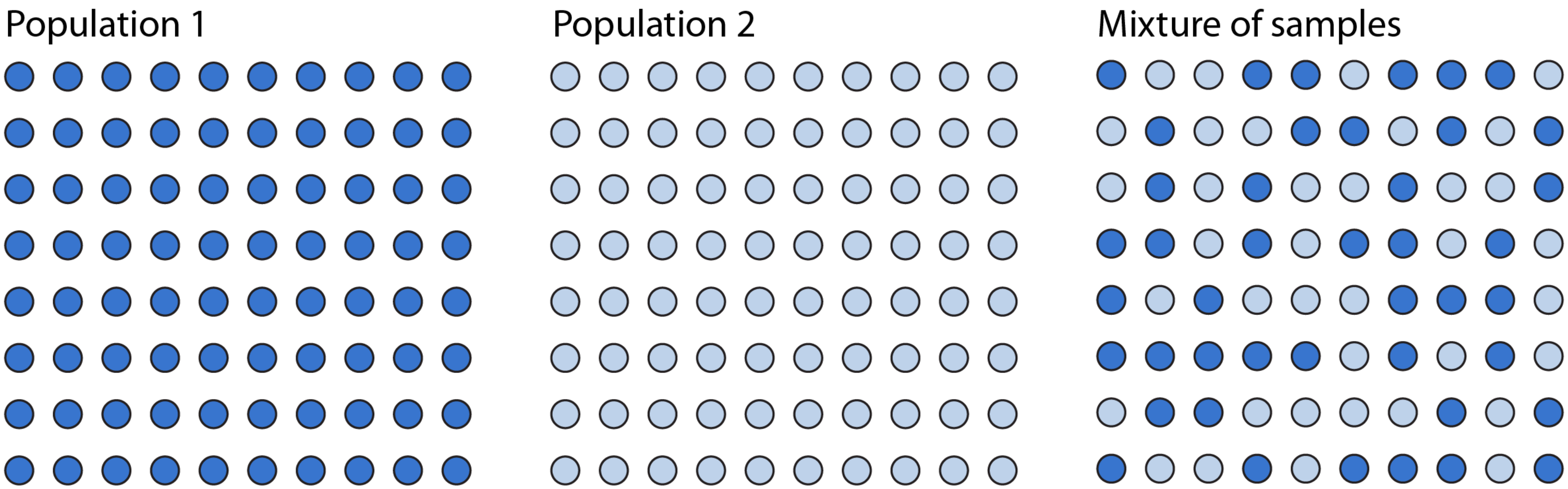
"A fundamental principle in population genetics stating that the genotype frequencies and gene frequencies of a large, randomly mating population remain constant provided immigration, mutation, and selection do not take place." *American Heritage Dictionary*

Godfrey Harold Hardy
1877-1947

Wilhelm Weinberg
1862-1937

© Michael Krawczak, Institute of Medical Informatics and Statistics Kiel / Germany

Gene variants = alleles (A_1, $A_{2;}$ frequencies p and q)
In diploid organisms (such as us) two alleles of a gene are present = genotypes (A_1A_1, A_1A, A_2A_2 ; frequencies P,Q,R)
In case of random combination of alleles, genotype frequencies are given by Hardy-Weinberg proportions:
$P=p^2$; $Q=2pq$; $R=q^2$

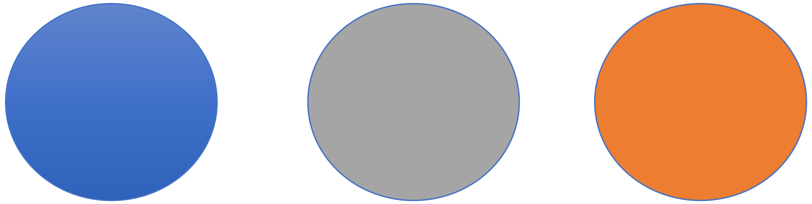


If we manage to sort the samples correctly into populations, we will recover HW equilibrium!

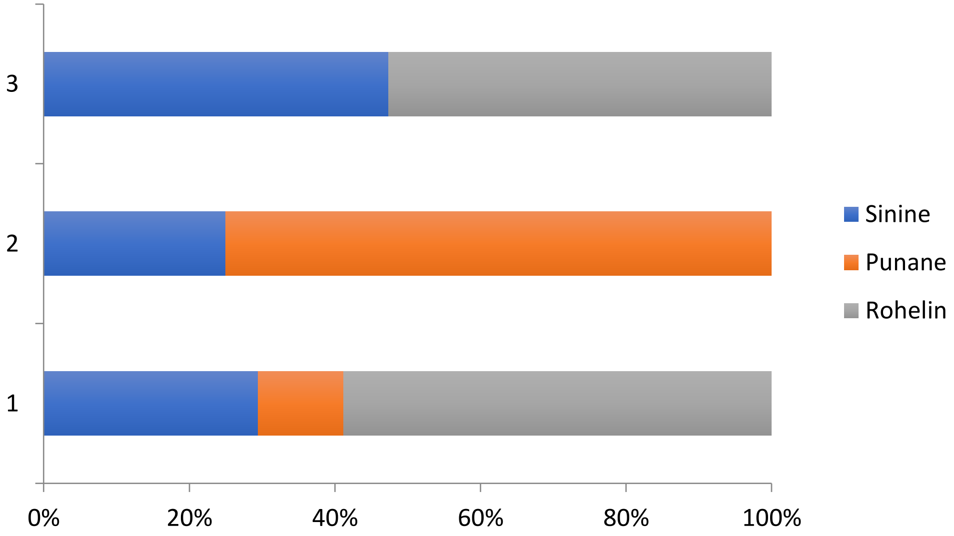
Random mating (combination of allelels) within population.
Hardy-Weinberg equilibrium holds: genotype frequencies are predicted from allele frequences

Nonrandom mating (combination of allelels).
Hardy-Weinberg equilibrium does not hold: genotype frequencies canot be predicted from allele frequences

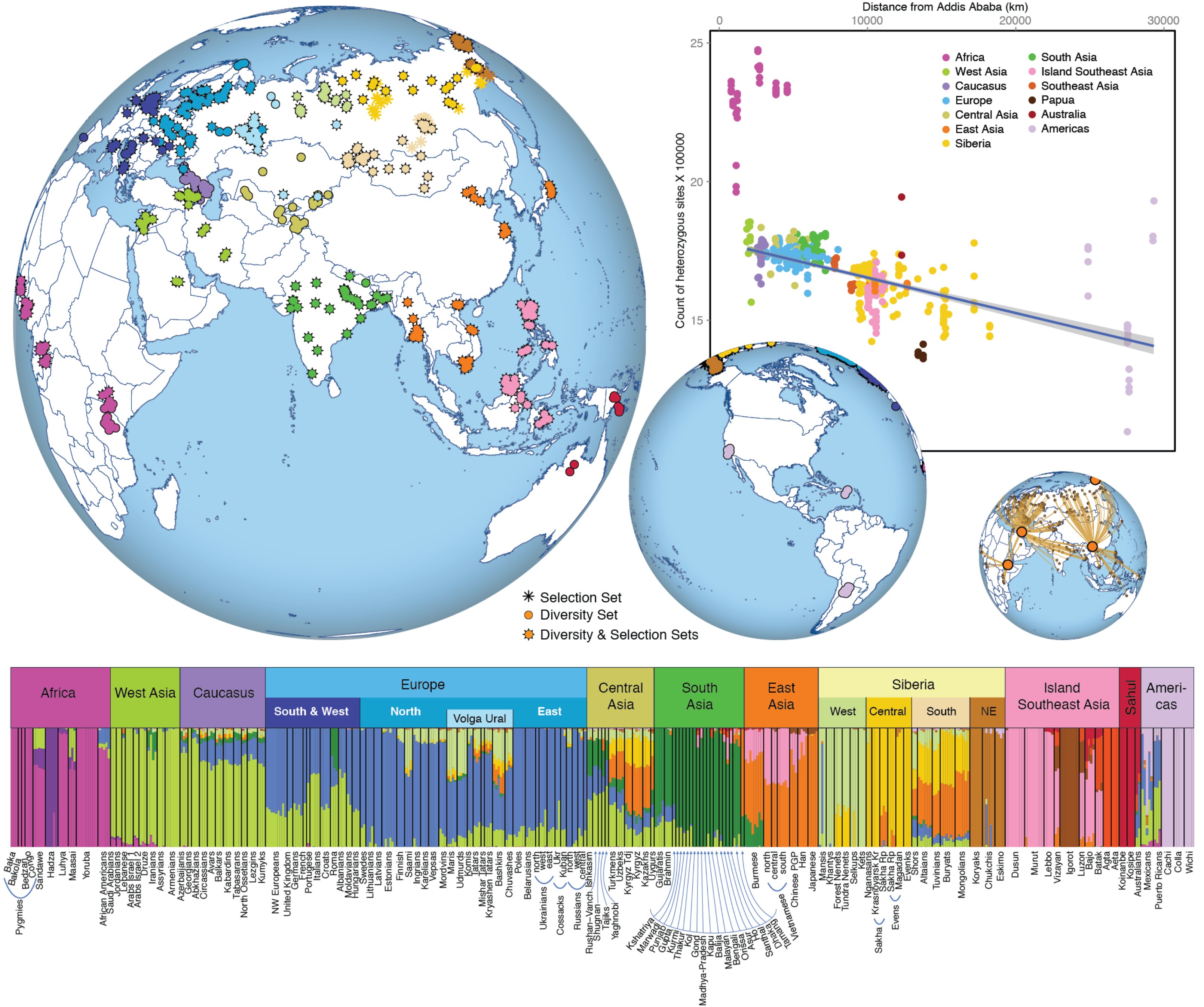
Proportions of “ancestral” components in a genetic profile of an individual



K3 - 3 “ancestral populations”

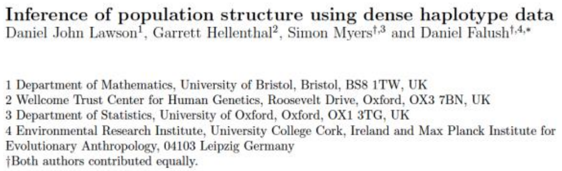


“Ancestry” proportions for individuals 1 to 3



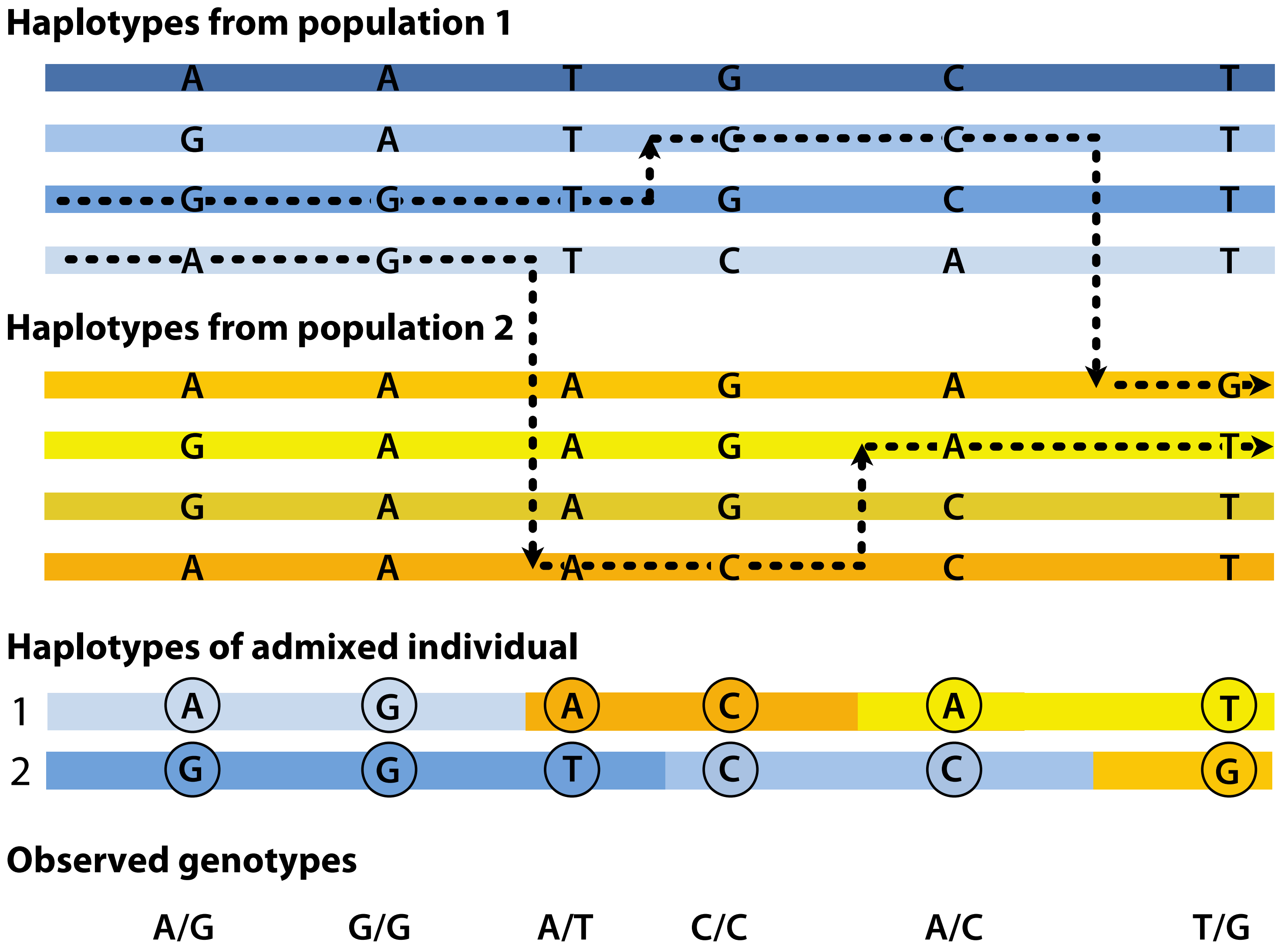
447 High Coverage genomes From 148 worldwide human populations

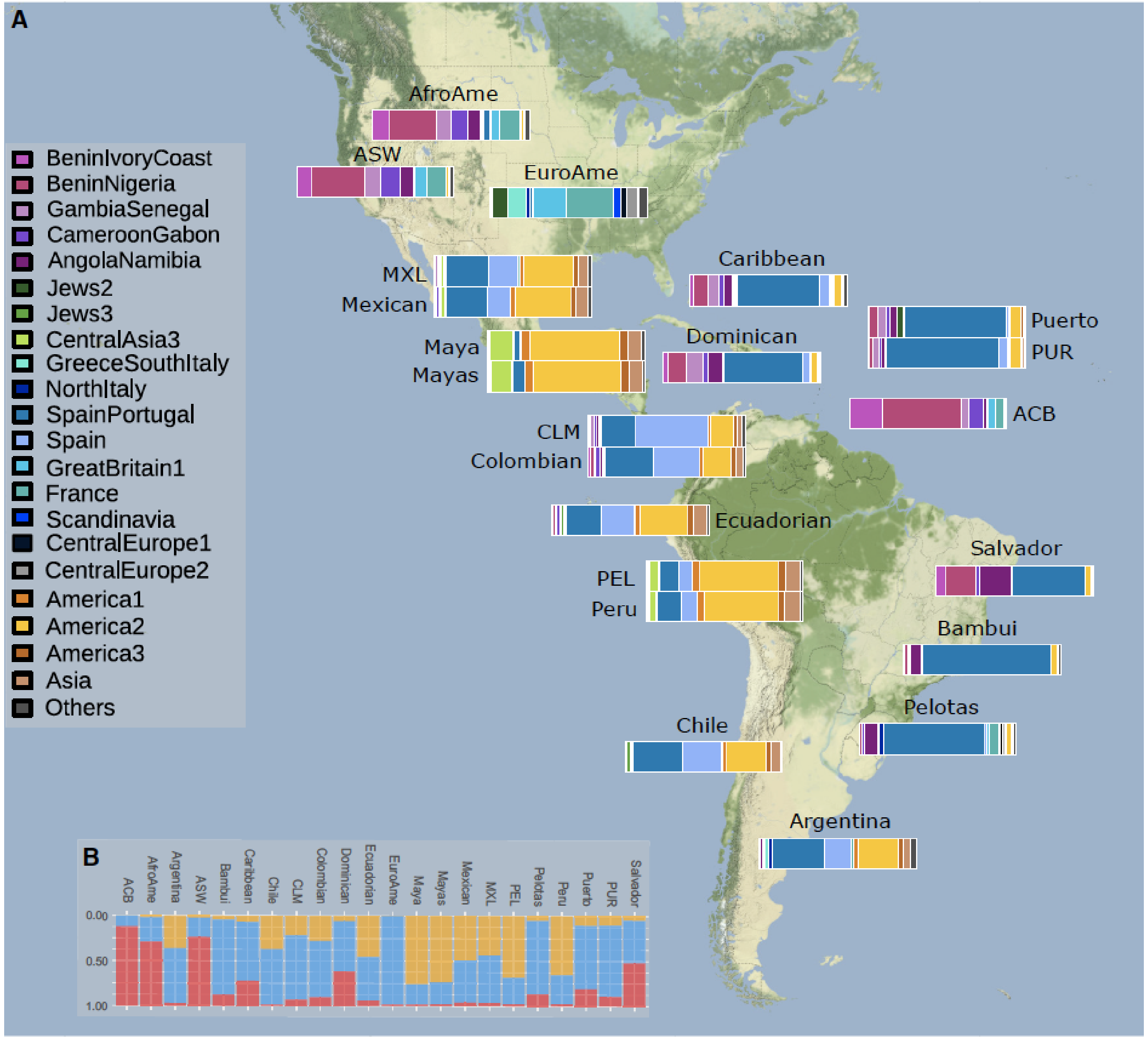
Genetic variation is organized in haplotypes – arrays of variants along the chromosome that are shared by multiple genomes. One can ”paint” a target chromosome by copying arrays of variants from a set of reference genomes.



Inference of population structure using dense haplotype data
Daniel John Lawson[1], Garrett Hellenthal[2], Simon Myers[†,3] and Daniel Falush[†,4,*]

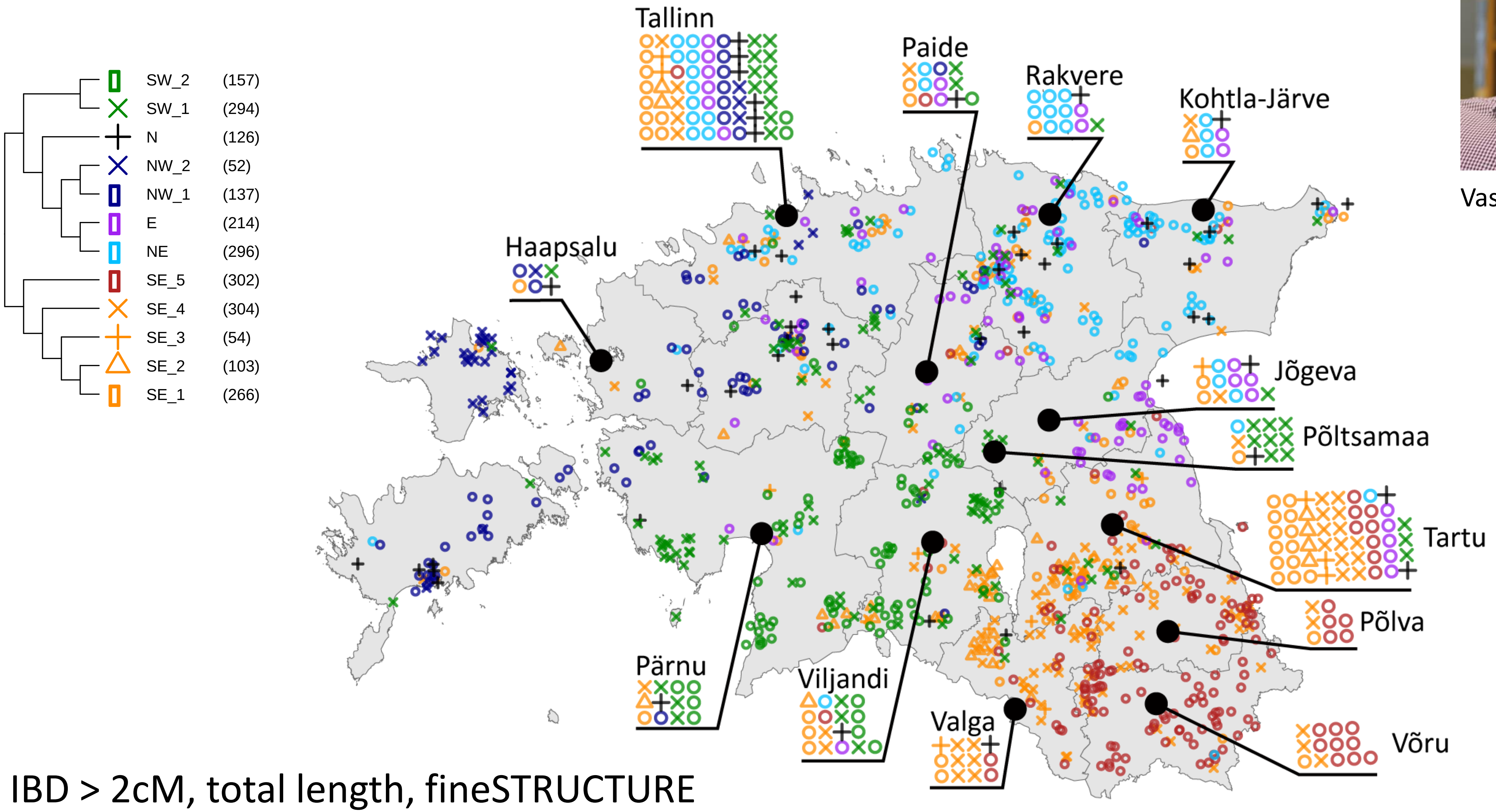
1 Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK
2 Wellcome Trust Center for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK
3 Department of Statistics, University of Oxford, Oxford, OX1 3TG, UK
4 Environmental Research Institute, University College Cork, Ireland and Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig Germany
†Both authors contributed equally.





Ongaro et al., The Genomic Impact of European Colonization of the Americas, Current Biology (2019)

Genetic structure of Estonians based on ~2300 genomes



Vasili Pankratov

IBD > 2cM, total length, fineSTRUCTURE

(Human) genetic diversity is shaped by

- demographic (evolutionary) history (genetic drift)
- natural selection
- culture

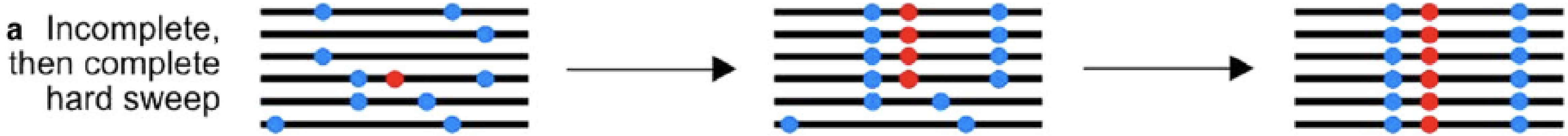
From the genomic perspective natural selection is change in allele frequency over generations FASTER than would be expected by random genetic drift.

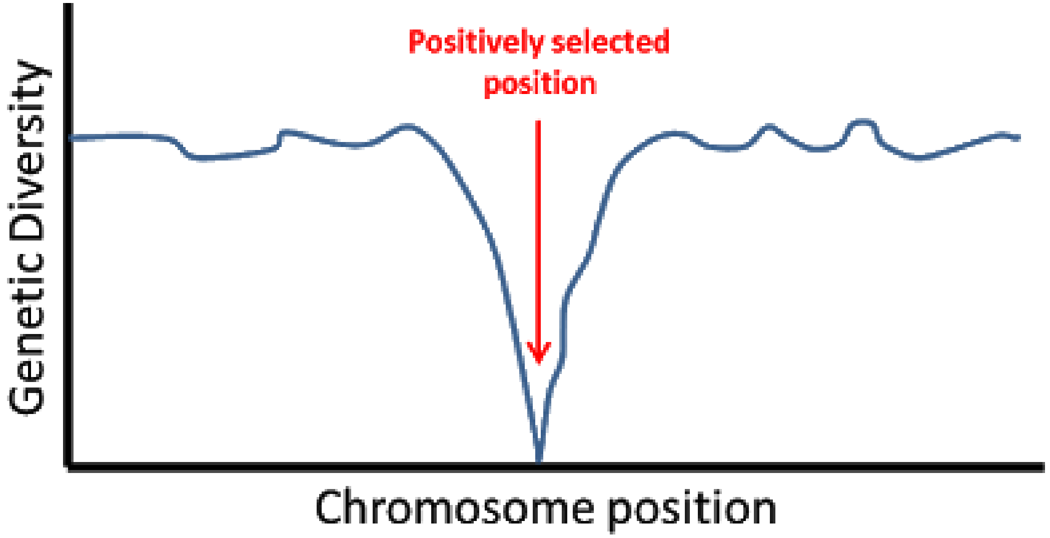
How to detect?

- Detect genetic sweeps
- Observe “directly” using ancient DNA

Selective sweep

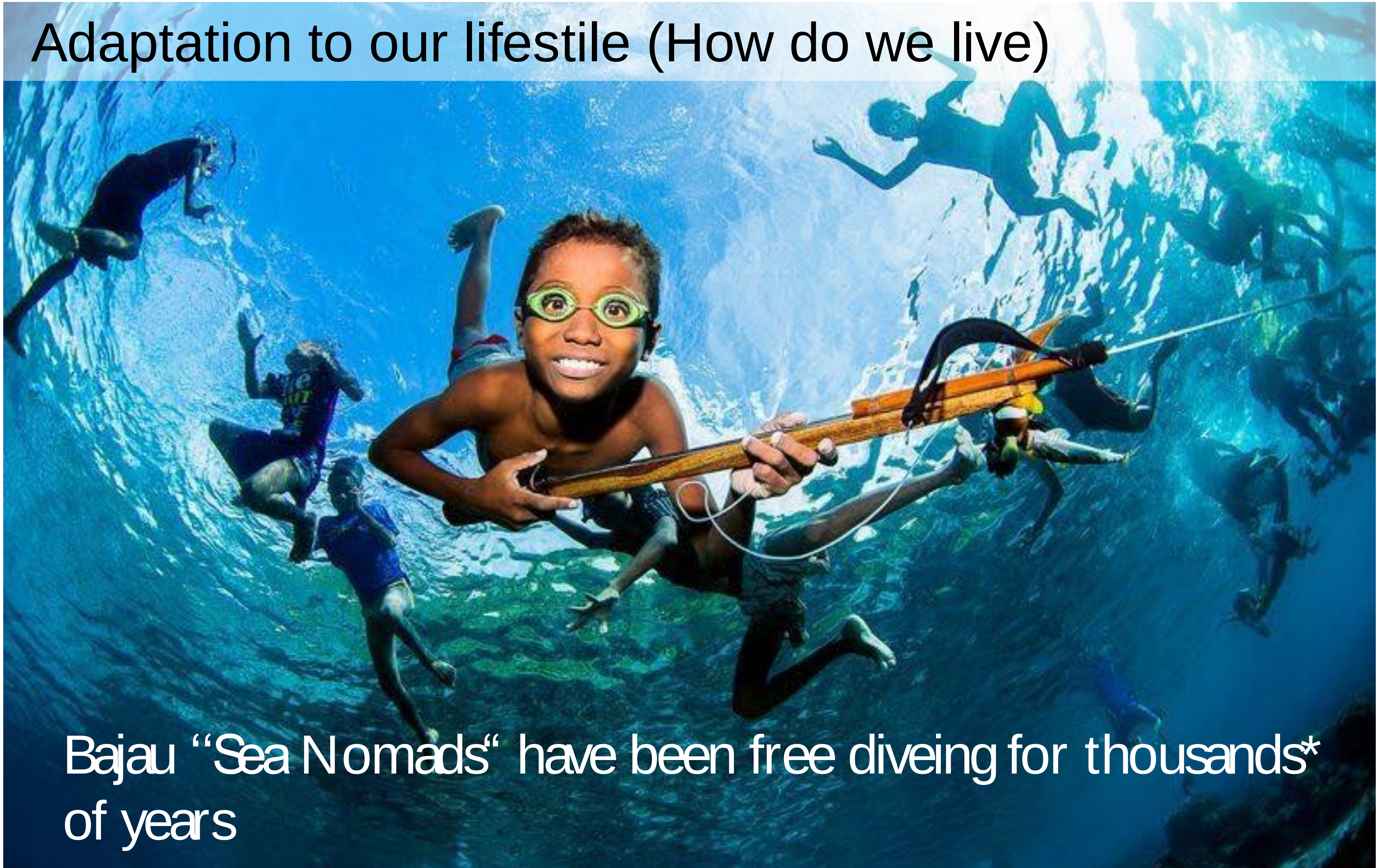
Booker et al. BMC Biology (2017) 15:98





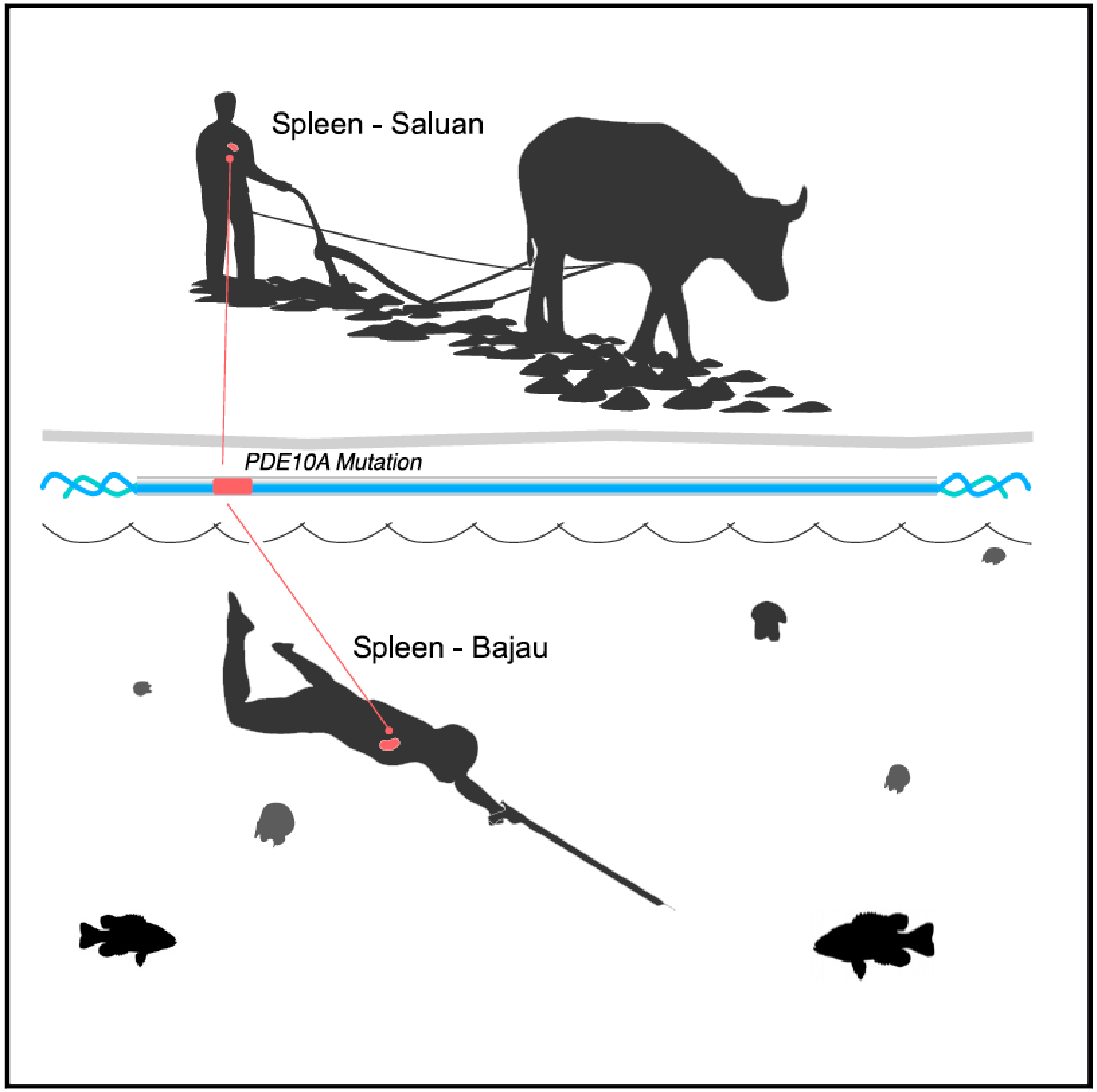
https://wp.unil.ch/genomeeee/2011/12/16/hard-selective-sweeps-do-not-seem-to-be-the-rule-in-human-evolution/

Adaptation to our lifestile (How do we live)



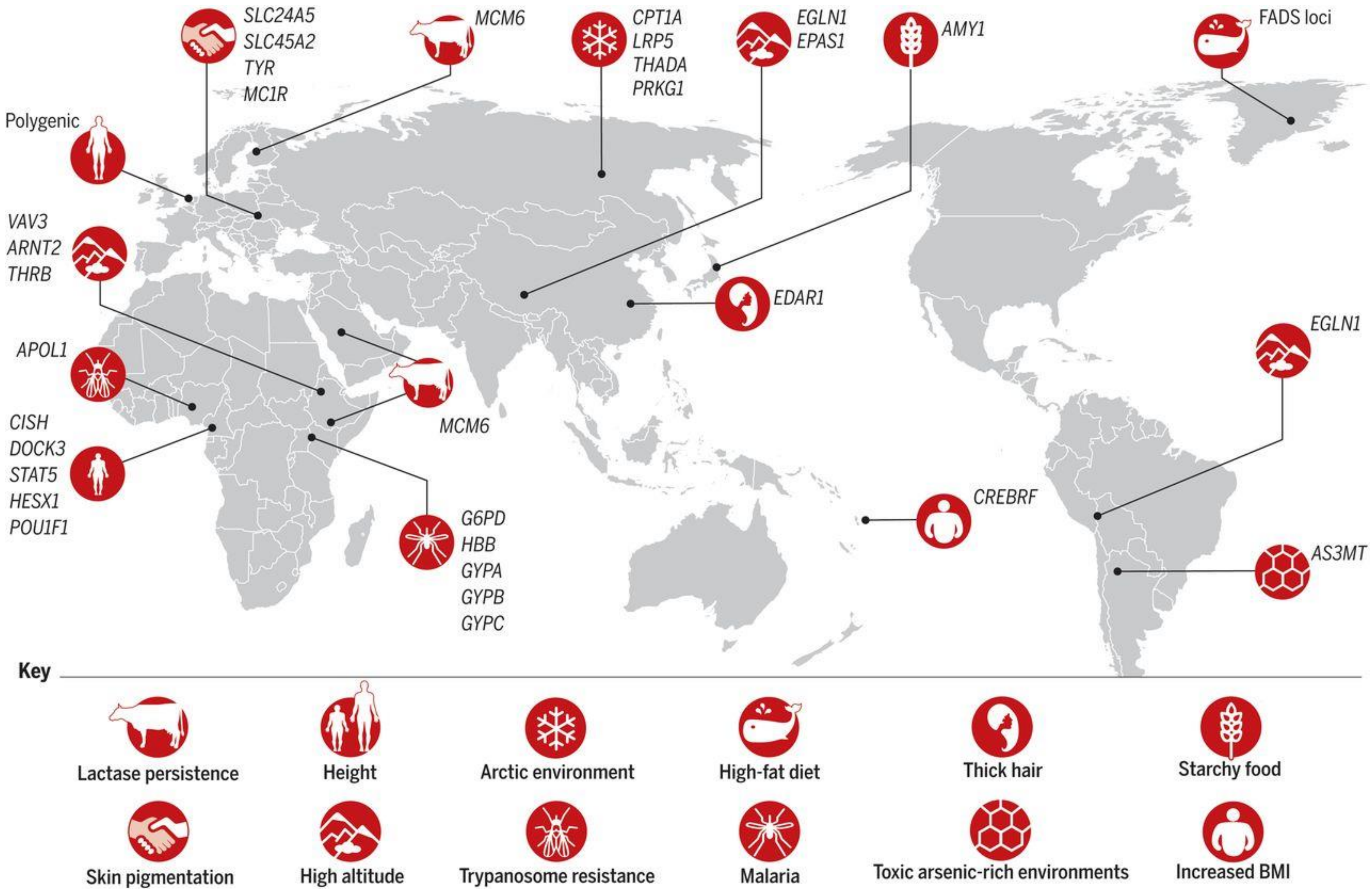
Bajau ‘‘Sea Nomads‘‘ have been free diveing for thousands* of years

The spleens on the Bajau have increased in size throught lnatural selection (quite like in sea otters). Spleen holds a reservoir of oxigen-rich blood. Quite helpful if diving for livelihood



Cell 173, 569–580, April 19, 2018 ª 2018 Elsevier Inc.

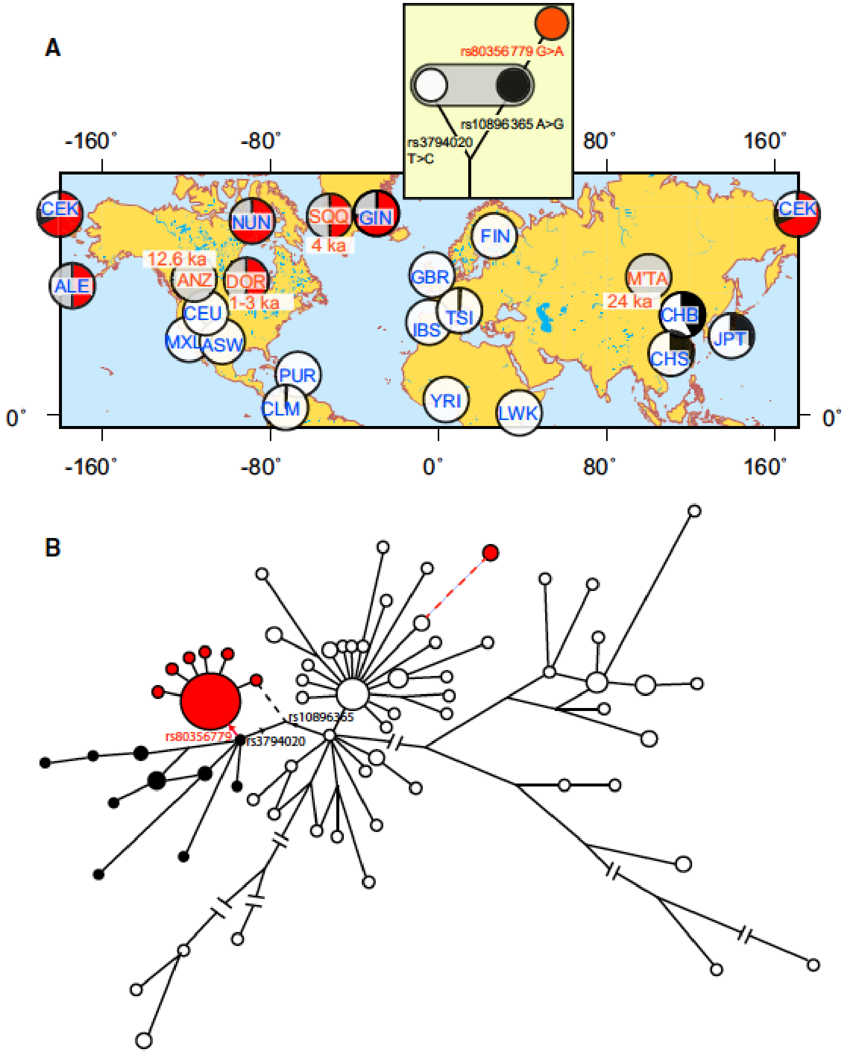
Examples of regional natural selection in humans



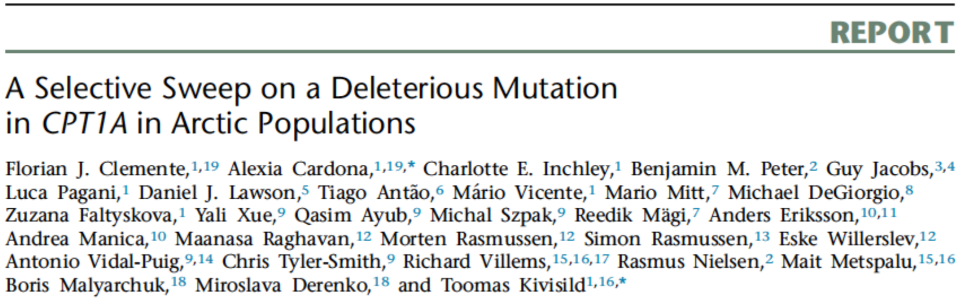
Shaohua Fan et al. Science 2016;354:54-59

Copyright © 2016, American Association for the Advancement of Science

Natural selection is not perfect - „side effects“



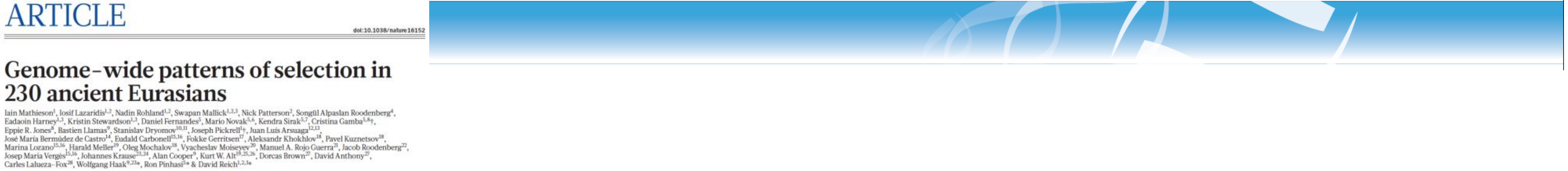
Carnitine palmitoyltransferase I


REPORT

A Selective Sweep on a Deleterious Mutation in *CPT1A* in Arctic Populations

Florian J. Clemente,[1,19] Alexia Cardona,[1,19,*] Charlotte E. Inchley,[1] Benjamin M. Peter,[2] Guy Jacobs,[3,4] Luca Pagani,[1] Daniel J. Lawson,[5] Tiago Antão,[6] Mário Vicente,[1] Mario Mitt,[7] Michael DeGiorgio,[8] Zuzana Faltyskova,[1] Yali Xue,[9] Qasim Ayub,[9] Michal Szpak,[9] Reedik Mägi,[7] Anders Eriksson,[10,11] Andrea Manica,[10] Maanasa Raghavan,[12] Morten Rasmussen,[12] Simon Rasmussen,[13] Eske Willerslev,[12] Antonio Vidal-Puig,[9,14] Chris Tyler-Smith,[9] Richard Villems,[15,16,17] Rasmus Nielsen,[2] Mait Metspalu,[15,16] Boris Malyarchuk,[18] Miroslava Derenko,[18] and Toomas Kivisild[1,16,*]

- CPT1A is a key regulator of mitochondrial long-chain fatty-acid oxidation
- the derived allele is associated with hypoketotic hypoglycemia and high infant mortality
- occurs at high frequency in Canadian and Greenland Inuits and was also found at 68% frequency in our Northeast Siberian sample.
- One of the strongest selective sweeps reported in humans despite associated deleterious consequences, possibly as a result of the selective advantage it originally provided to either a high-fat diet or a cold environm

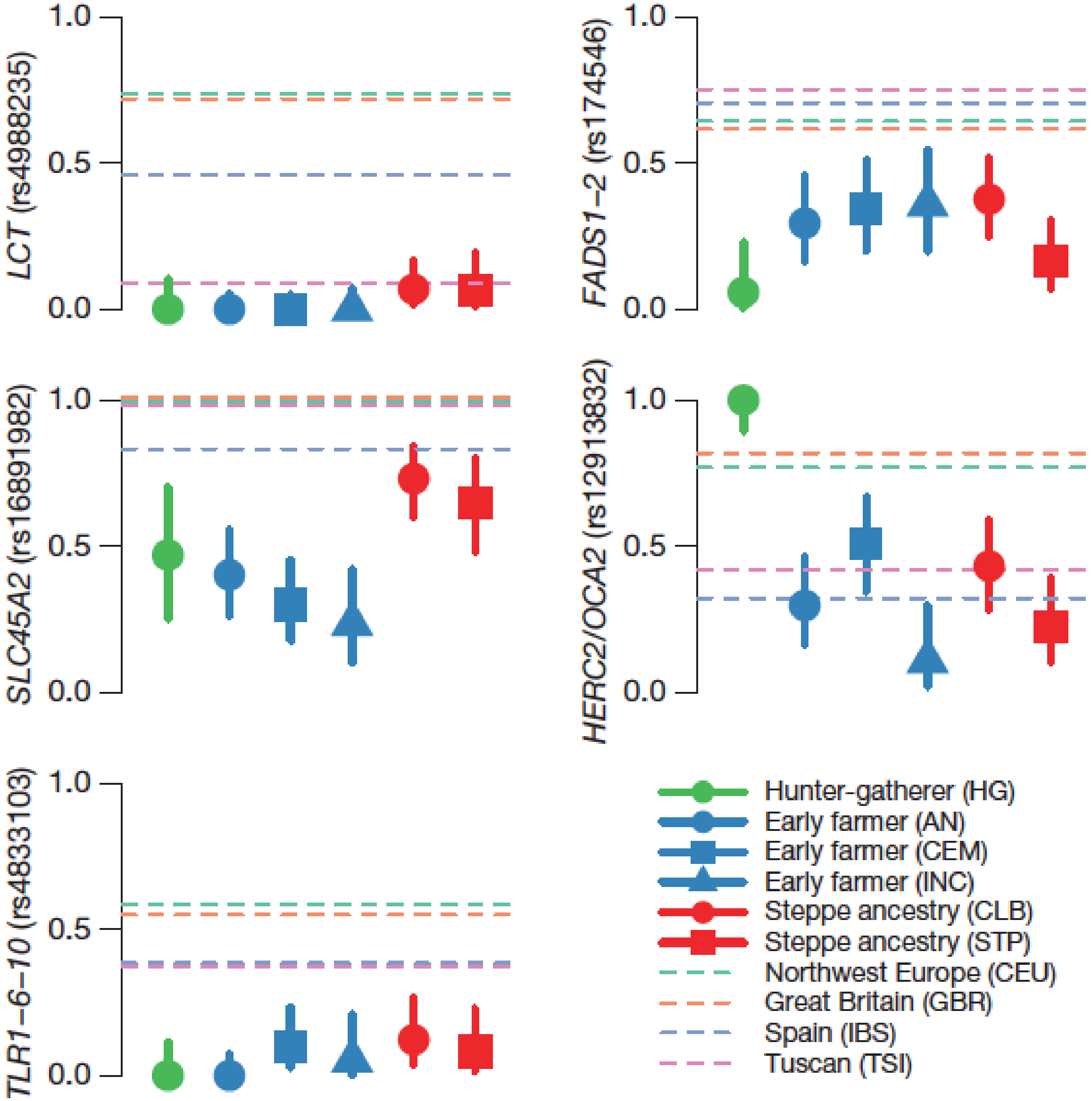


ARTICLE

doi:10.1038/nature16152

Genome-wide patterns of selection in 230 ancient Eurasians

Iain Mathieson[1], Iosif Lazaridis[1,2], Nadin Rohland[1,2], Swapan Mallick[1,2,3], Nick Patterson[2], Songül Alpaslan Roodenberg[4], Eadaoin Harney[1,3], Kristin Stewardson[1,3], Daniel Fernandes[5], Mario Novak[5,6], Kendra Sirak[5,7], Cristina Gamba[5,8]†, Eppie R. Jones[8], Bastien Llamas[9], Stanislav Dryomov[10,11], Joseph Pickrell[1]†, Juan Luís Arsuaga[12,13], José María Bermúdez de Castro[14], Eudald Carbonell[15,16], Fokke Gerritsen[17], Aleksandr Khokhlov[18], Pavel Kuznetsov[18], Marina Lozano[15,16], Harald Meller[19], Oleg Mochalov[18], Vyacheslav Moiseyev[20], Manuel A. Rojo Guerra[21], Jacob Roodenberg[22], Josep Maria Vergès[15,16], Johannes Krause[23,24], Alan Cooper[9], Kurt W. Alt[19,25,26], Dorcas Brown[27], David Anthony[27], Carles Lalueza-Fox[28], Wolfgang Haak[9,23]*, Ron Pinhasi[5]* & David Reich[1,2,3]*



Lactose tolerance

Fatty acid metabolism

Light skin

Light (blue) eye

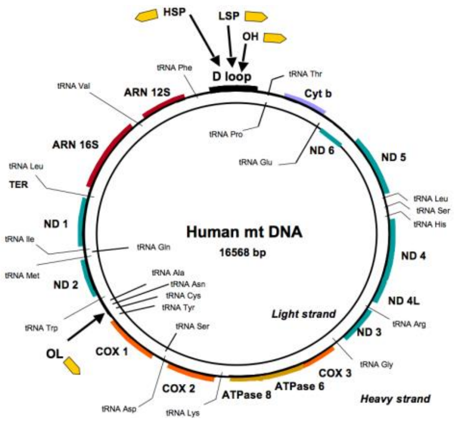
Resistance to (some) diseases

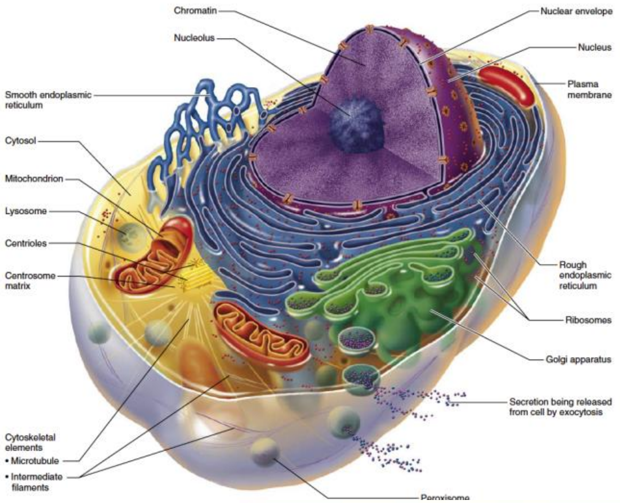
Nature 528, 499–503 (2015)

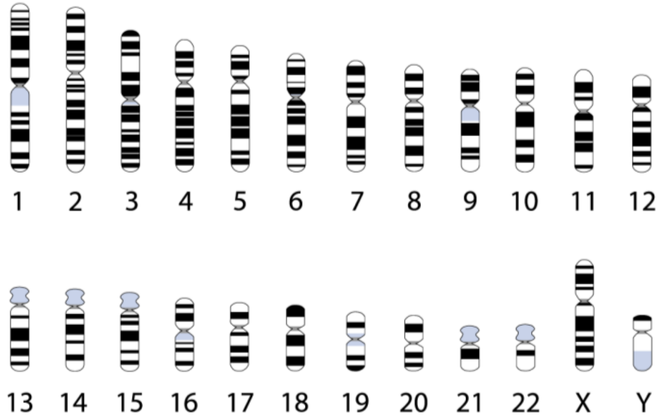
European hunter gatherers had a rather dark skin and blue eyes

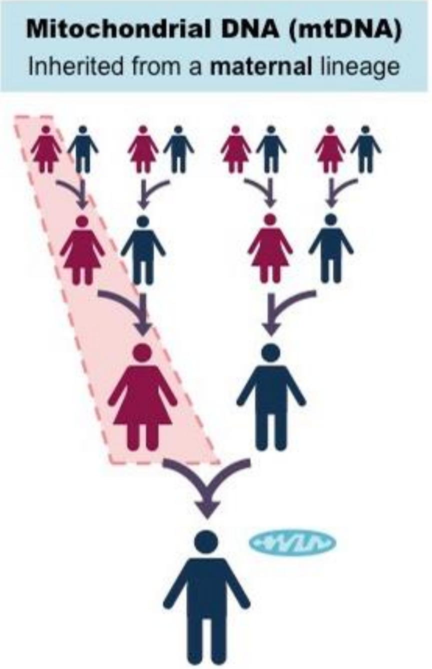
(Human) genetic diversity is shaped by

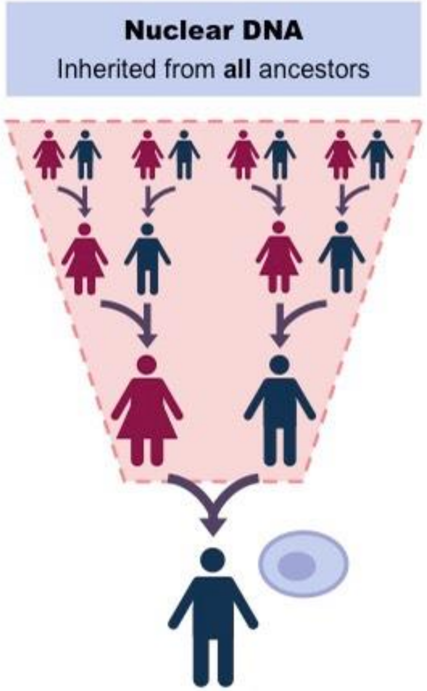
- demographic (evolutionary) history (genetic drift)
- natural selection
- culture



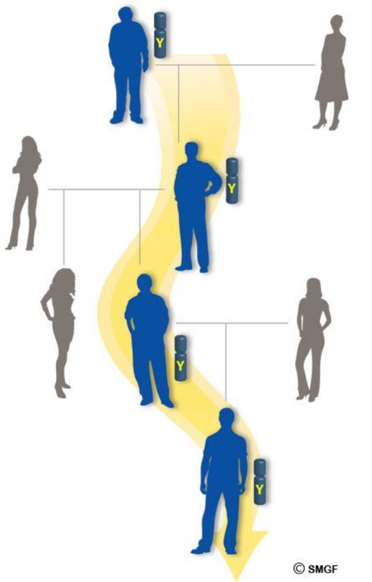




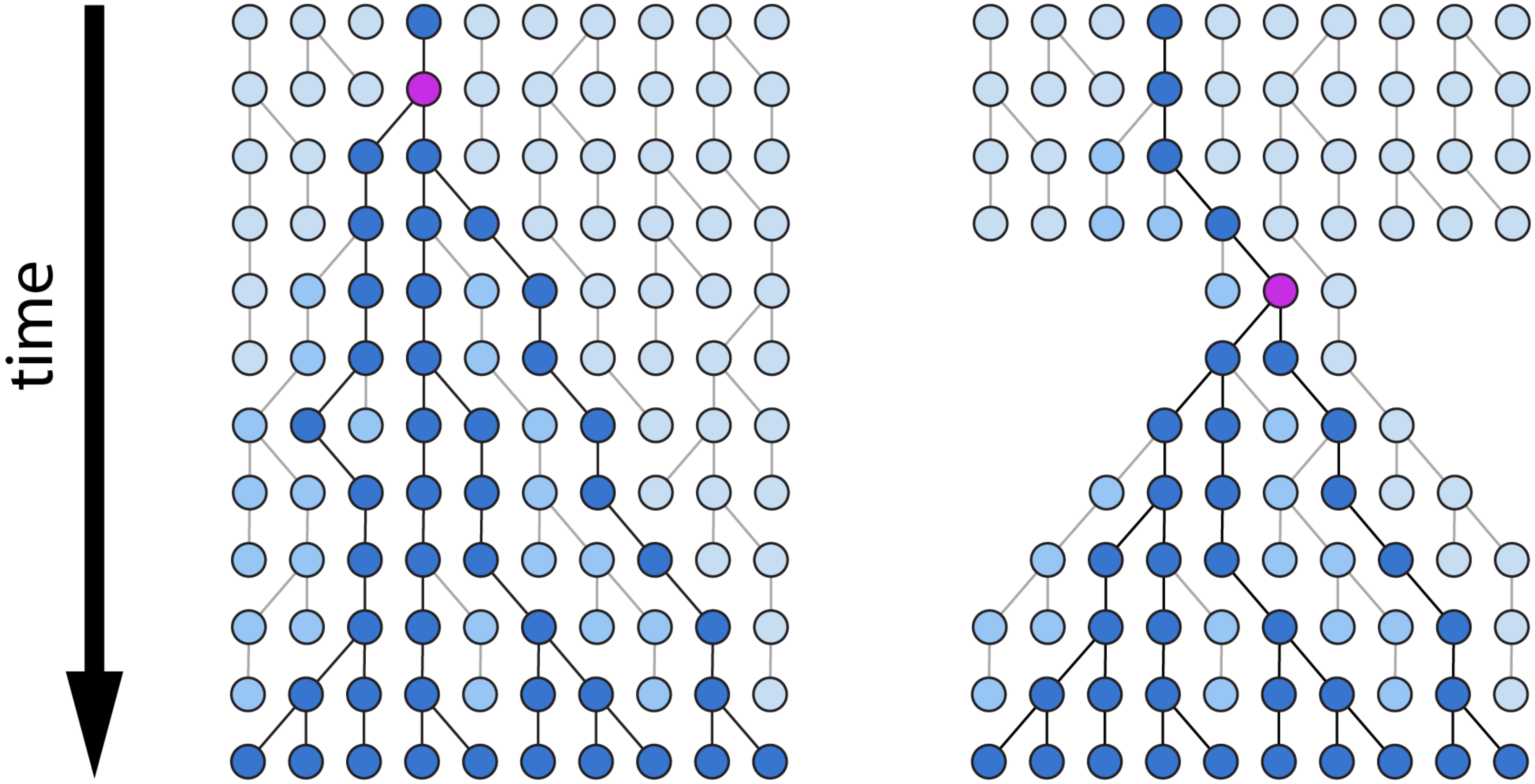




Y-chromosome inheritance



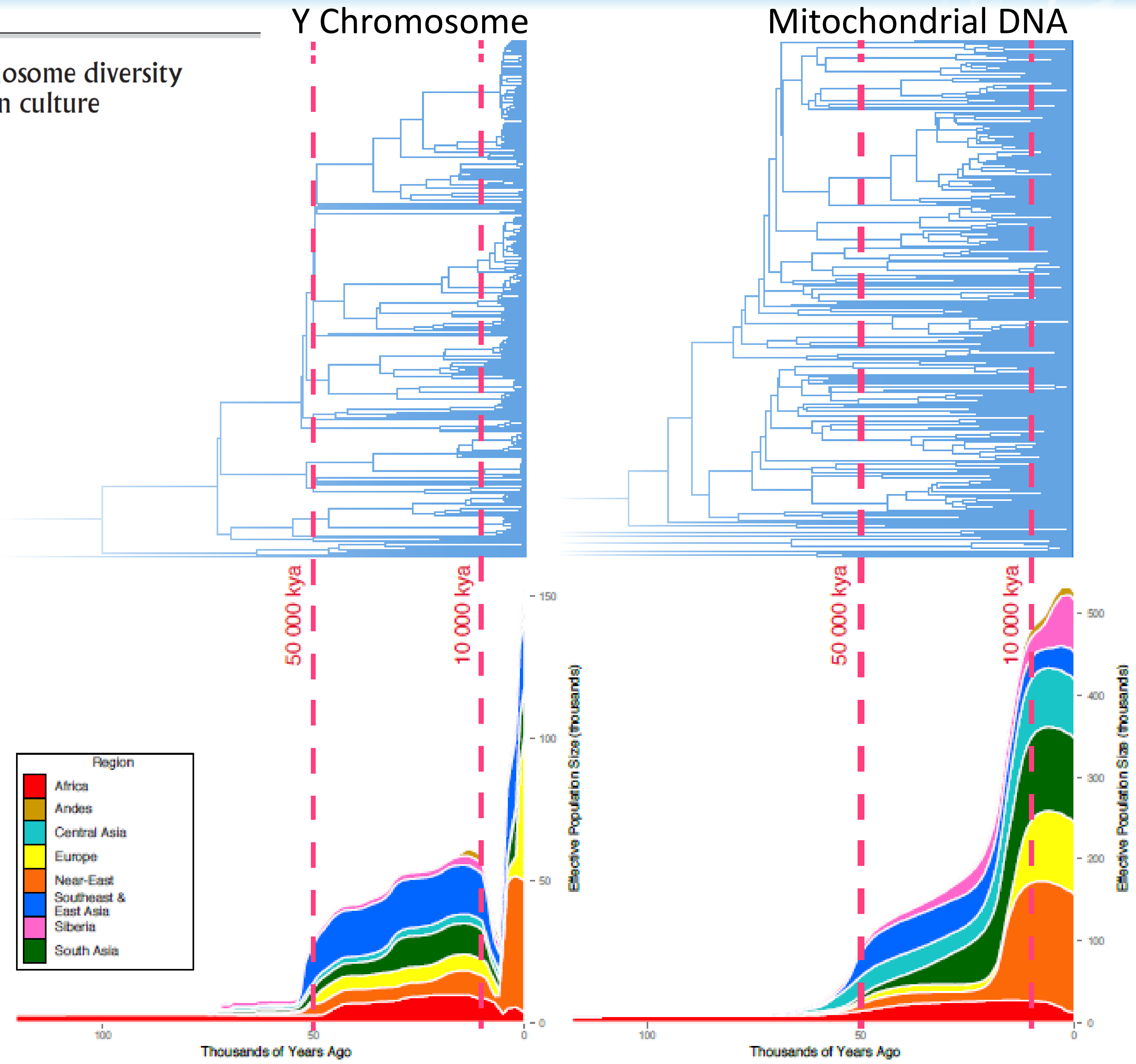
http://nutritionreview.org/wp-content/uploads/2013/09/MitoFig1.png; https://cdn-images-1.medium.com/max/1600/1*_b5RMmV5XT7MM22OgokTjg.png; http://ib.bioninja.com.au/_Media/mtdna_med.jpeg



● Most Recent Common Ancestor

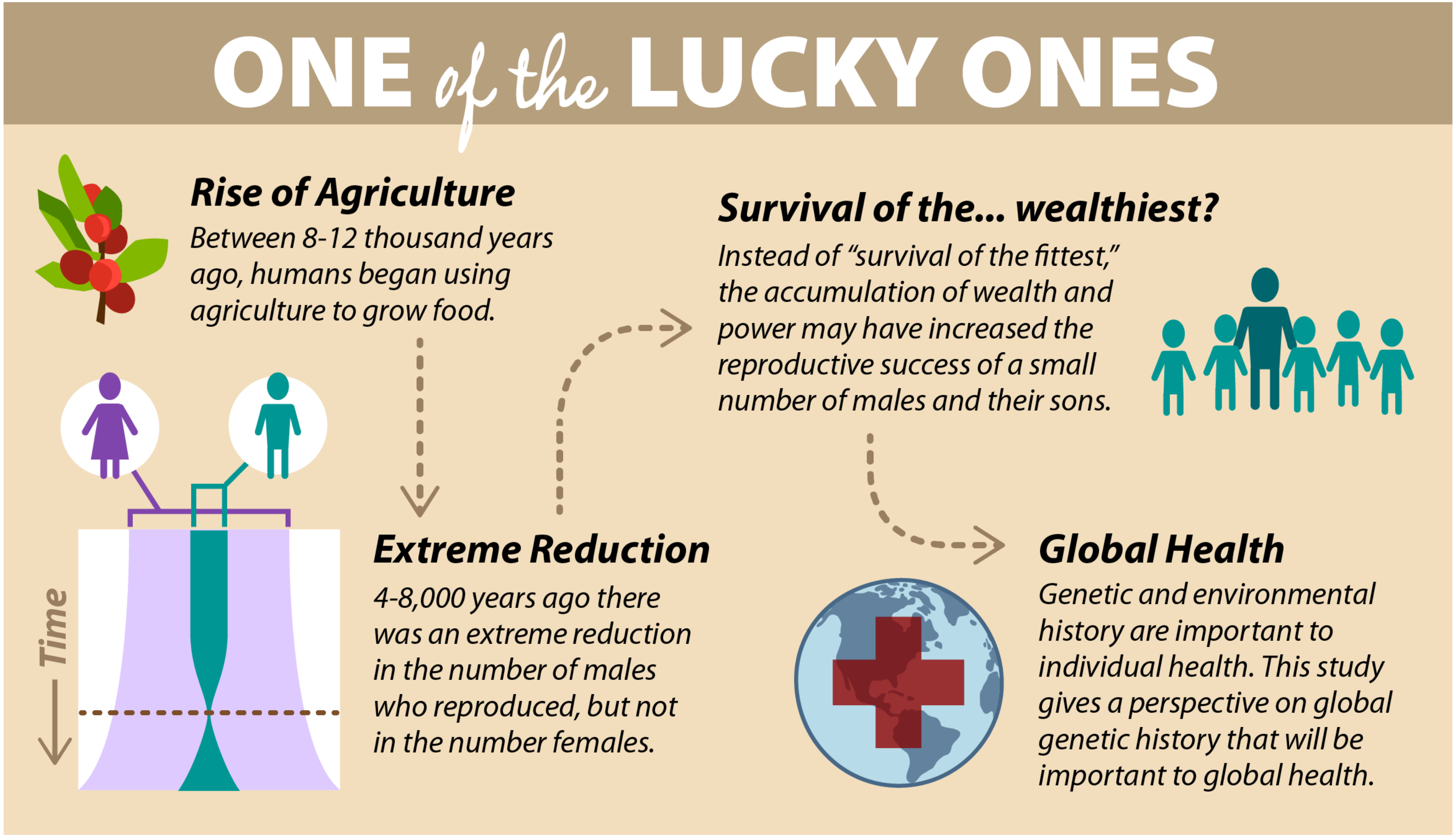
Research

A recent bottleneck of Y chromosome diversity coincides with a global change in culture


Y Chromosome

Mitochondrial DNA

Analysis run by:
Lauri Saag, Mario Vicente etc



ONE *of the* LUCKY ONES

Rise of Agriculture

Between 8-12 thousand years ago, humans began using agriculture to grow food.

Survival of the... wealthiest?

Instead of "survival of the fittest," the accumulation of wealth and power may have increased the reproductive success of a small number of males and their sons.

Extreme Reduction

4-8,000 years ago there was an extreme reduction in the number of males who reproduced, but not in the number females.

Global Health

Genetic and environmental history are important to individual health. This study gives a perspective on global genetic history that will be important to global health.

Infographics by Arizona State University

UNIVERSITY OF TARTU
Institute of Genomics

UNIVERSITY OF TARTU

1632

Thank you!

www.genomics.ut.ee

unitartu

tartuuniversity

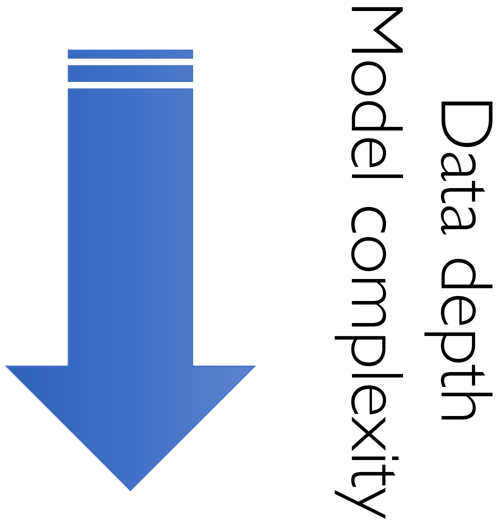
Archaeogenetics - application of the techniques of molecular population genetics to the study of the human past.

Why?

- We want to know our past.
- Evolutionary history shapes genetic diversity/susceptibility to disease.
- Personalized medicine in admixed populations/individuals.

How?

- Allele frequency patterns
- Coalescence patterns
- Haplotype sharing patterns

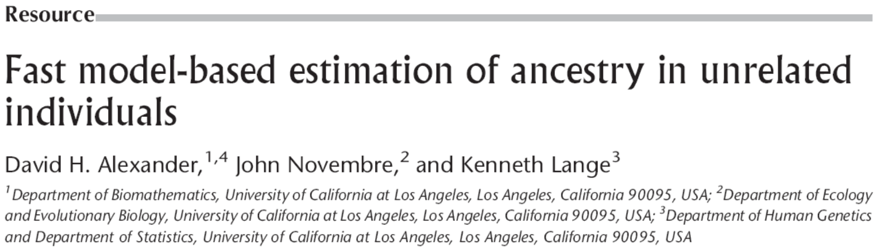


structure-like analyses

The name comes from a program STRUCTURE. **This method is a model based approach to reveal genetic structure in the sample.**

The principle of the method is to find a genetic structure that would maximize Hardy-Weinberg equilibrium and minimize linkage disequlibrium. The premise is that within (sub)populations there is HW but when you have two populations in your sample there is no HW.

The approach constructs K number of “ancestral” populations (defined by allele frequencies at each locus) and proportions for each sample in the study with which it belongs to each of the ”ancestral” populations (sums up to 1).



Resource

Fast model-based estimation of ancestry in unrelated individuals

David H. Alexander,[1,4] John Novembre,[2] and Kenneth Lange[3]

[1]Department of Biomathematics, University of California at Los Angeles, Los Angeles, California 90095, USA; [2]Department of Ecology and Evolutionary Biology, University of California at Los Angeles, Los Angeles, California 90095, USA; [3]Department of Human Genetics and Department of Statistics, University of California at Los Angeles, Los Angeles, California 90095, USA