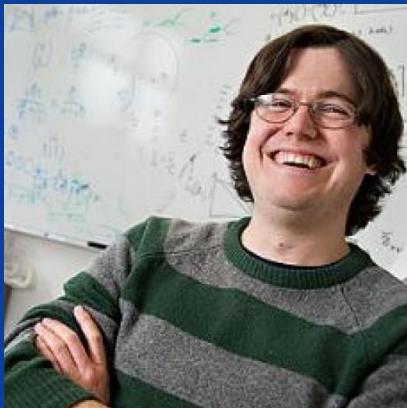




University of
Pittsburgh®

School of
Public Health



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

The contribution of gene flow, selection, and genetic drift to five thousand years of human allele frequency change

Alexis Simon^{1,2,*} and Graham Coop^{1,2}

¹Center for Population Biology, University of California, Davis, CA 95616

²Department of Evolution and Ecology, University of California, Davis, CA 95616

Posted to bioRxiv: January 11, 2024

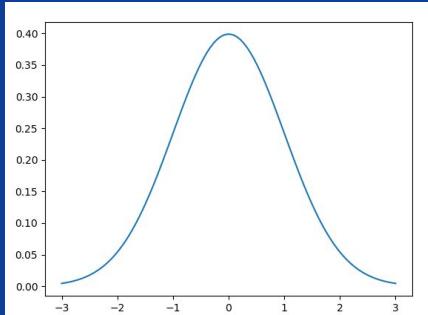
Alexis Simon (top)
Graham Coop (bottom)

Presentation by
Dylan Maher
HUGEN 2028
January 26, 2024



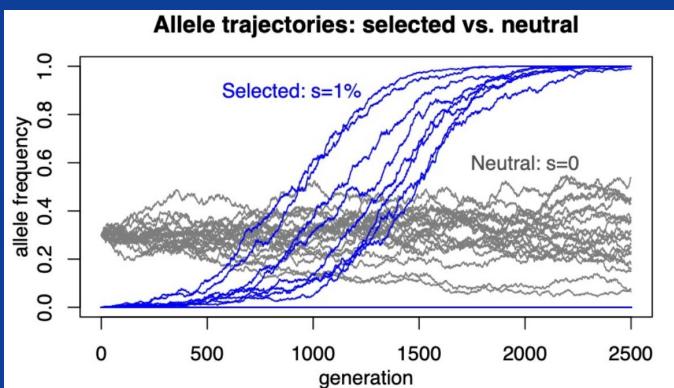
A Bit of Jargon

“Gaussian” Stabilizing Selection



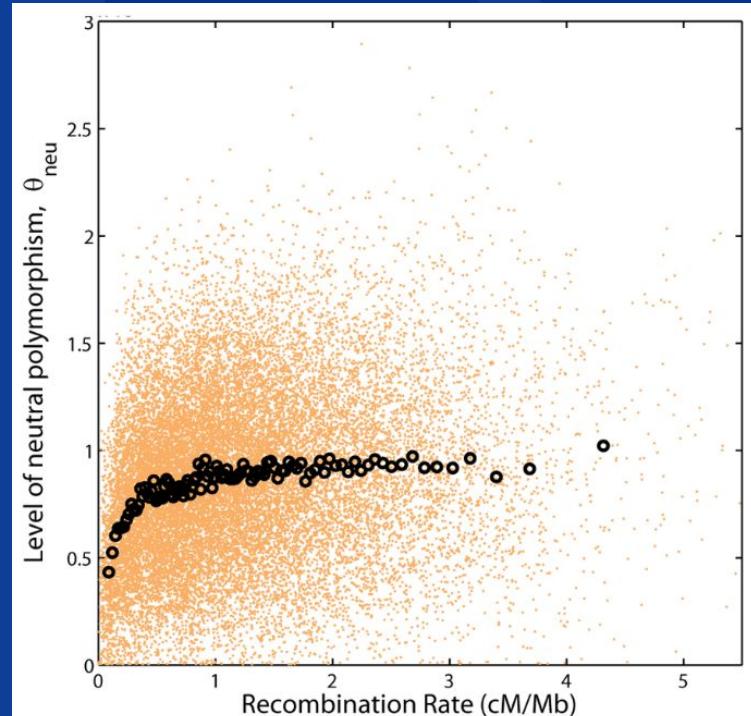
- Linked Selection
- Background selection
 - Hitchhiking

“Effective”
Population Size



Genetic
Drift/Neutral
theory

Pritchard
(2023)



Cai et al. (2009)



University of
Pittsburgh®

School of
Public Health

“Long-standing debate about the role of genetic drift vs. selection in evolutionary change...”

The Neutral Theory in Light of Natural Selection

Andrew D. Kern^{*1} and Matthew W. Hahn²

¹Department of Genetics, Rutgers University, Piscataway, NJ

²Department of Biology and Department of Computer Science, Indiana University Bloomington, IN

The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018

Jeffrey D. Jensen,^{1,2} Bret A. Payseur,³ Wolfgang Stephan,⁴ Charles F. Aquadro,⁵ Michael Lynch,⁶

Deborah Charlesworth,⁷ and Brian Charlesworth⁷

The neutral theory is dead. Long live the neutral theory

Martin Kreitman

Background

Widespread signatures of natural selection across human complex traits and functional genomic categories

Jian Zeng¹✉, Angli Xue¹, Longda Jiang¹, Luke R. Lloyd-Jones¹, Yang Wu¹, Huanwei Wang¹, Zhili Zheng¹, Loic Yengo¹, Kathryn E. Kemper¹, Michael E. Goddard^{2,3}, Naomi R. Wray^{1,4}, Peter M. Visscher¹ & Jian Yang¹ 1,5,6

Classic Selective Sweeps Were Rare in Recent Human Evolution

Ryan D. Hernandez,^{1*} Joanna L. Kelley,¹ Eyal Elyashiv,² S. Cord Melton,¹ Adam Auton,³ Gilean McVean,^{3,4} 1000 Genomes Project, Guy Sella,^{2,†} Molly Przeworski^{1,5,6,‡}

The evolution of the G matrix: selection or drift?

DEREK ROFF
Department of Biology, McGill University, 1205 Dr Penfield Ave., Montreal, Quebec, Canada H3A 1B1

Pervasive selection or is it...? why are F_{ST} outliers sometimes so frequent?

NICOLAS BIERNE,^{*} DENIS ROZE^{‡§} and JOHN J. WELCH[¶]



Model

Basic Idea:

Decompose variance
in allele frequency
change into
contributions from
different evolutionary
forces

$$\begin{aligned}\text{Var}(p_T - p_0) = & \underbrace{\sum_{i=0}^{T-1} \text{Var}(\Delta_D p_i)}_{\text{Drift}} + \underbrace{\sum_{i=0}^{T-1} \text{Var}(\Delta_S p_i)}_{\text{Selection}} + \underbrace{\sum_{i=0}^{T-1} \text{Var}(\Delta_A p_i)}_{\text{Admixture}} \\ & + \underbrace{\sum_{i \neq j}^{T-1} \text{Cov}(\Delta_S p_i, \Delta_S p_j)}_{\text{Selection}} + \underbrace{\sum_{i \neq j}^{T-1} \text{Cov}(\Delta_A p_i, \Delta_A p_j)}_{\text{Admixture}}\end{aligned}$$

Note: Above model formulation omits interaction between drift in one time period that admixture in later time periods erases. (Derivation including correction for this in Appendix C)



Theoretical Development

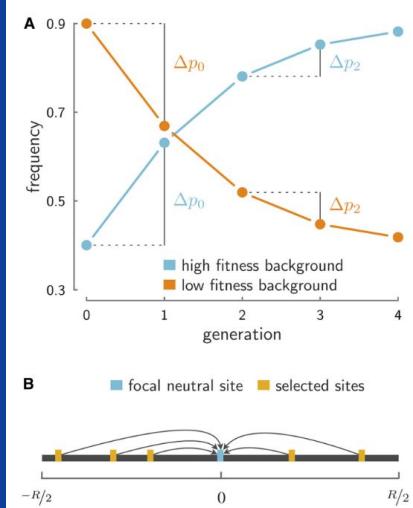
The Linked Selection Signature of Rapid Adaptation in Temporal Genomic Data

Vince Buffalo^{a,b,1} and Graham Coop^b

^aPopulation Biology Graduate Group and ^bCenter for Population Biology, Department of Evolution and Ecology, University of California, Davis, California 95616

ORCID IDs: 0000-0003-4510-1609 (V.B.); 0000-0001-8431-0302 (G.C.)

Buffalo and
Coop (2019)



Background

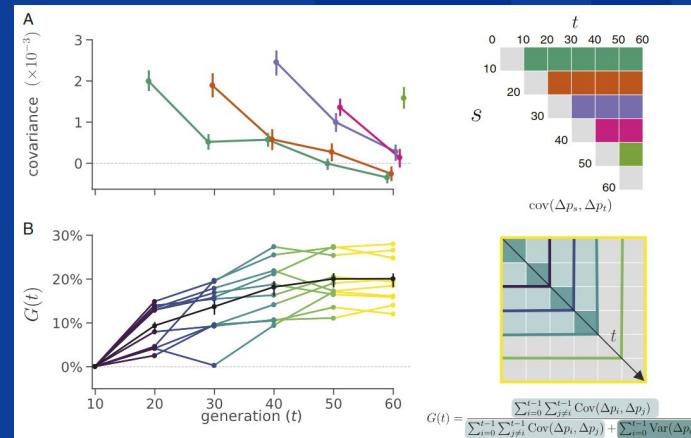
Application with Empirical Data (selection experiment)

Estimating the genome-wide contribution of selection to temporal allele frequency change

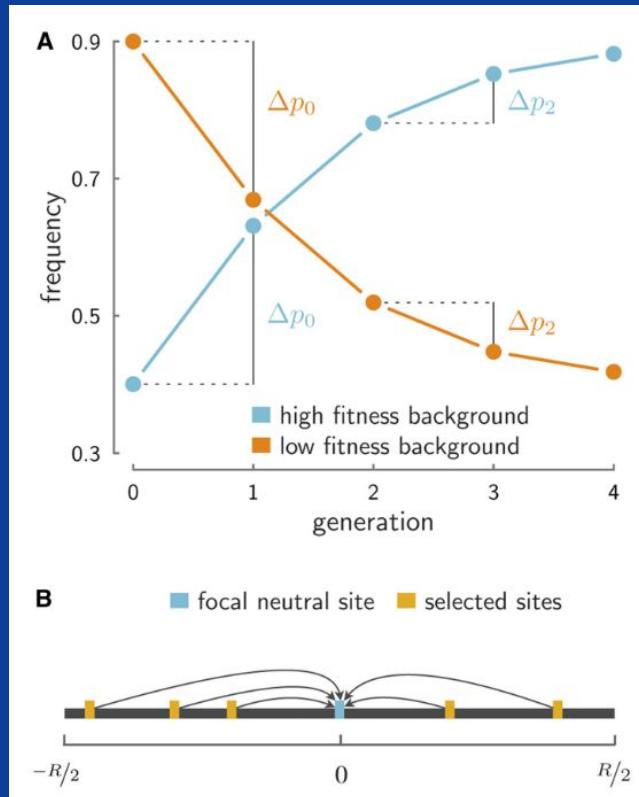
Vince Buffalo^{a,b,1} and Graham Coop^b

^aPopulation Biology Graduate Group, University of California, Davis, CA 95616; and ^bCenter for Population Biology, Department of Evolution and Ecology, University of California, Davis, CA 95616

Edited by Montgomery Slatkin, University of California, Berkeley, CA, and approved July 13, 2020 (received for review October 31, 2019)



Buffalo
and
Coop
(2020)



Basic Idea Behind Buffalo & Coop's Model

Drift: no effect on temporal covariance

Selection: generates covariance

So, track allele frequency data across time—if correlated, this is evidence for selection.

Problem(?): assumes panmictic (“closed”) populations



Polygenic Adaptation has Impacted Multiple Anthropometric Traits

Jeremy J. Berg^{1,3}, Xinjun Zhang², Graham Coop¹

Author's Note on Failure to Replicate

After this preprint was posted, the UK Biobank dataset was released, providing a new and open GWAS resource. When attempting to replicate the height selection results from this preprint using GWAS data from the UK Biobank, we discovered that we could not. In subsequent analyses, we determined that both the GIANT consortium height GWAS data, as well as another dataset that was used for replication, were impacted by stratification issues that created or at a minimum substantially inflated the height selection signals reported here. The results of this second investigation, written

Admixture has obscured signals of historical hard sweeps in humans

Received: 21 November 2021

Accepted: 16 September 2022

Published online: 31 October 2022

Check for updates

Yassine Souilmi , Raymond Tobler , Angad Johar , Matthew Williams¹, Shane T. Grey^{4,5}, Joshua Schmidt , João C. Teixeira , Adam Rohrlach , Jonathan Tuke , Olivia Johnson¹, Graham Gower , Chris Turnley , Murray Cox¹⁰, Alan Cooper^{11,12,14} and Christian D. Huber

Assumption of panmictic (“closed”) populations: problem?

As it turns out, yes.

Reduced signal for polygenic adaptation of height in UK Biobank

Jeremy J Berg^{1,2*}, Arbel Harpak^{1,2†}, Nasa Sinnott-Armstrong^{3†}, Anja Moltke Joergensen⁴, Hakhamanesh Mostafavi¹, Yair Field³, Evan August Boyle³, Xinjun Zhang⁵, Fernando Racimo⁴, Jonathan K Pritchard^{2,6*}, Graham Coop^{7,8*}

Abstract Several recent papers have reported strong signals of selection on European polygenic height scores. These analyses used height effect estimates from the GIANT consortium and replication studies. Here, we describe a new analysis based on the the UK Biobank (UKB), a large, independent dataset. We find that the signals of selection using UKB effect estimates are strongly attenuated or absent. We also provide evidence that previous analyses were confounded by population stratification. Therefore, the conclusion of strong polygenic adaptation now lacks

historic hard sweeps in modern European genomes. Our results imply that this canonical mode of selection has probably been underappreciated in the evolutionary history of humans and suggest that our current understanding of the tempo and mode of selection in natural populations may be inaccurate.



$$\Delta p_t = \underbrace{\Delta_N p_t + \Delta_M p_t}_{\text{drift}} + \underbrace{\Delta_H p_t}_{\text{selection}}$$

Buffalo & Coop (2019)

Simon & Coop (2024)

$$\begin{aligned} \text{Var}(p_T - p_0) &= \sum_{i=0}^{T-1} \underbrace{\text{Var}(\Delta_D p_i)}_{\text{Drift}} + \sum_{i=0}^{T-1} \underbrace{\text{Var}(\Delta_S p_i)}_{\text{Selection}} + \sum_{i=0}^{T-1} \underbrace{\text{Var}(\Delta_A p_i)}_{\text{Admixture}} \\ &\quad + \sum_{i \neq j}^{T-1} \underbrace{\text{Cov}(\Delta_S p_i, \Delta_S p_j)}_{\text{Selection}} + \sum_{i \neq j}^{T-1} \underbrace{\text{Cov}(\Delta_A p_i, \Delta_A p_j)}_{\text{Admixture}} \end{aligned}$$

Basic Idea Behind Simon & Coop's Extension

Problem: Buffalo and Coop's model assumes panmictic population

This has been shown to confound inference

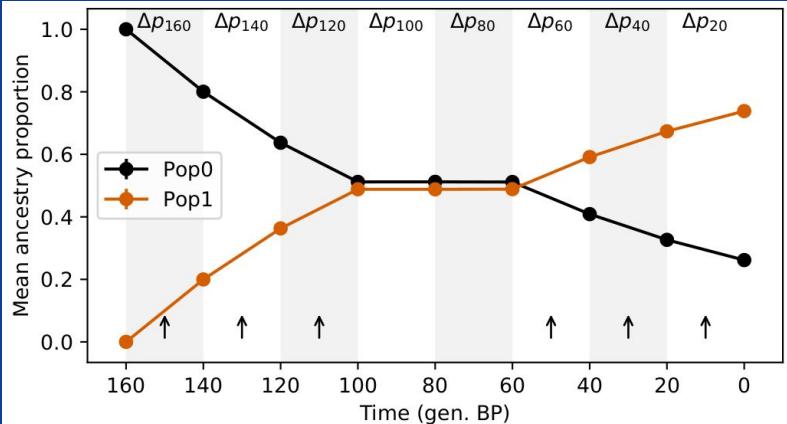
Solution: “Correct” for known gene flow events

The proportion of total variance in allele frequency change between 0 and t due to linked selection, $G(t)$, is defined as the ratio of the total covariance due to linked selection over the total variance (Buffalo & Coop, 2019). Under our model, $G(t)$ can be estimated by correcting the empirical covariance by the estimated covariance term due to admixture:

$$\begin{aligned} G(t) &\equiv \frac{\sum_{i \neq j}^{t-1} \text{Cov}(\Delta_S p_i, \Delta_S p_j)}{\text{Var}(p_t - p_0)} \\ &= \frac{\sum_{i \neq j}^{t-1} \text{Cov}(\Delta p_i, \Delta p_j) - \text{Cov}(\Delta_A p_i, \Delta_A p_j)}{\text{Var}(p_t - p_0)} \end{aligned} \tag{4}$$

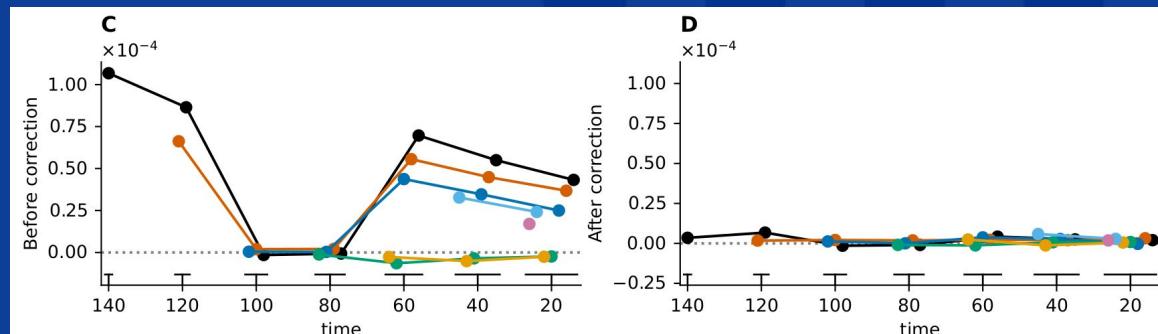
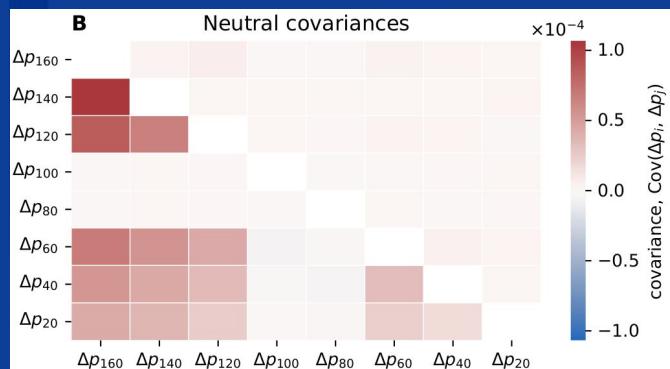


Simulation Work



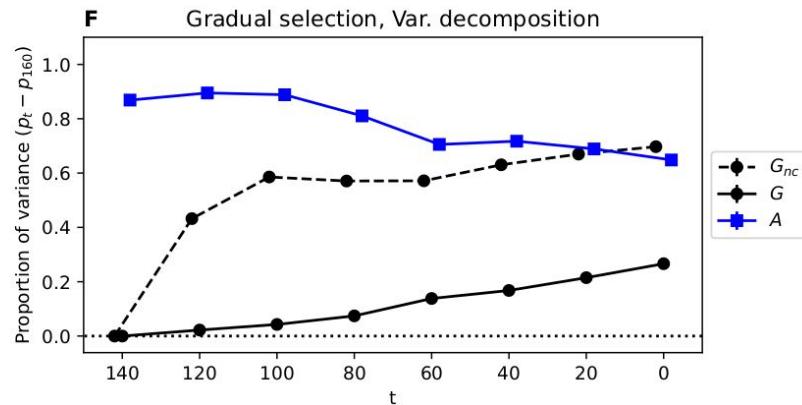
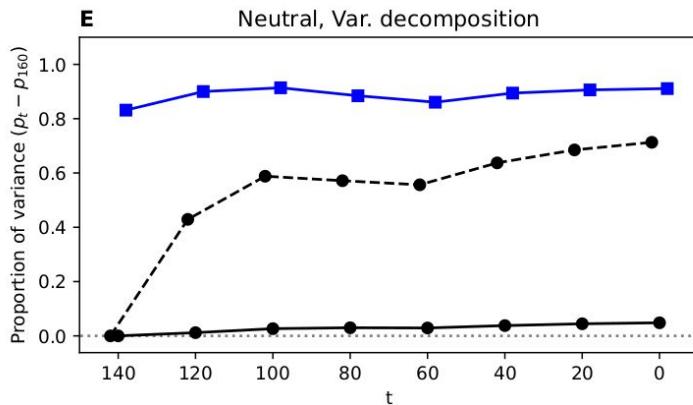
Admixture causes positive covariance (lower diagonal) and C, corrected for, covariance disappears (upper diagonal) and D

Simulated population receives pulses of admixture (arrows)





Simulation Work



Not accounting
for admixture
generates
spurious signal
of selection

Two simulation scenarios: neutral (drift) [left] and gradual (positive selection) [right]

Variance attributable to admixture (up to time t)

$$A(t) \equiv \frac{\sum_{i=0}^{t-1} \text{Var}(\Delta_A p_i) + \sum_{i \neq j}^{t-1} \text{Cov}(\Delta_A p_i, \Delta_A p_j)}{\text{Var}(p_t - p_0)}$$



(Very) Simplified Speedrun Through Human History

~6 mya: humans split from chimps

~2 mya: first human (species homo)

~300 kya: oldest divergence for modern populations

~50 kya: Out-of-Africa event

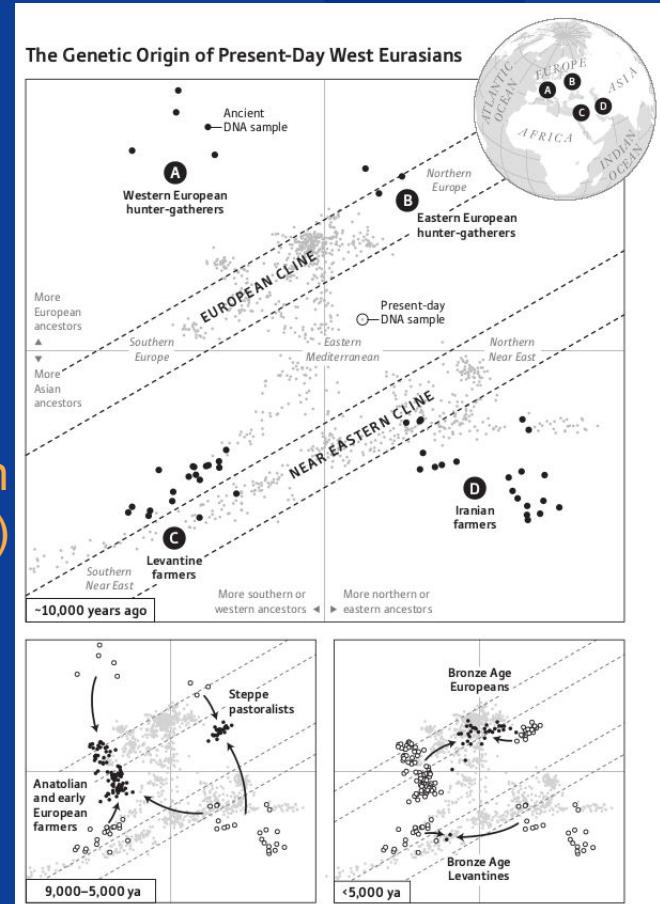
~14 kya: Europe primarily consists of hunter-gatherers (WHG)

~9 kya: farmers from Anatolia move into Europe (EEF)

~5kya: Yamnaya arrive in Europe (and Asia) and spread (Steppe-Like)

Historical Background

Reich
(2018)

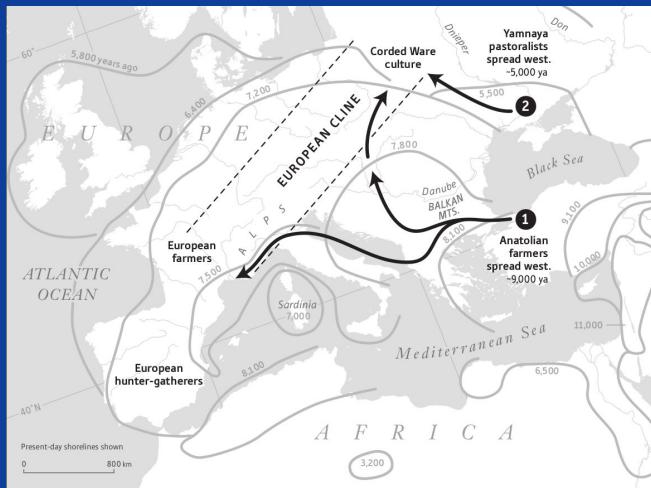




Historical Background

Table S9 from Papac et al. (2021). In concordance with the literature on European Human demographic history during the last 5000 years, we consider the simplest three-way admixture between populations genetically most similar to European early farmers (EEF-like, early migrants from Anatolia), Western hunter-gatherers (WHG-like) and individuals associated to the Steppe pastoralists Yamnaya culture (Steppe-like).

Assumed
demographic
makeup of model



Reich (2018)

(Very) simplified version of
Europe since Behavioral
Modernity (50kya)

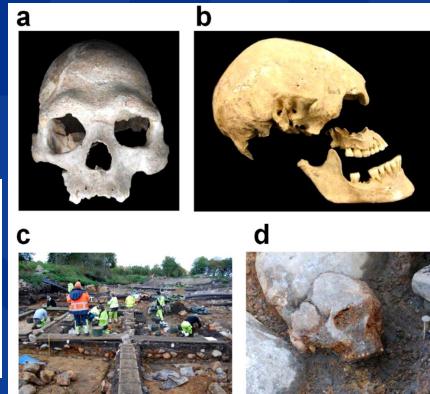
Lazaridis (2014)

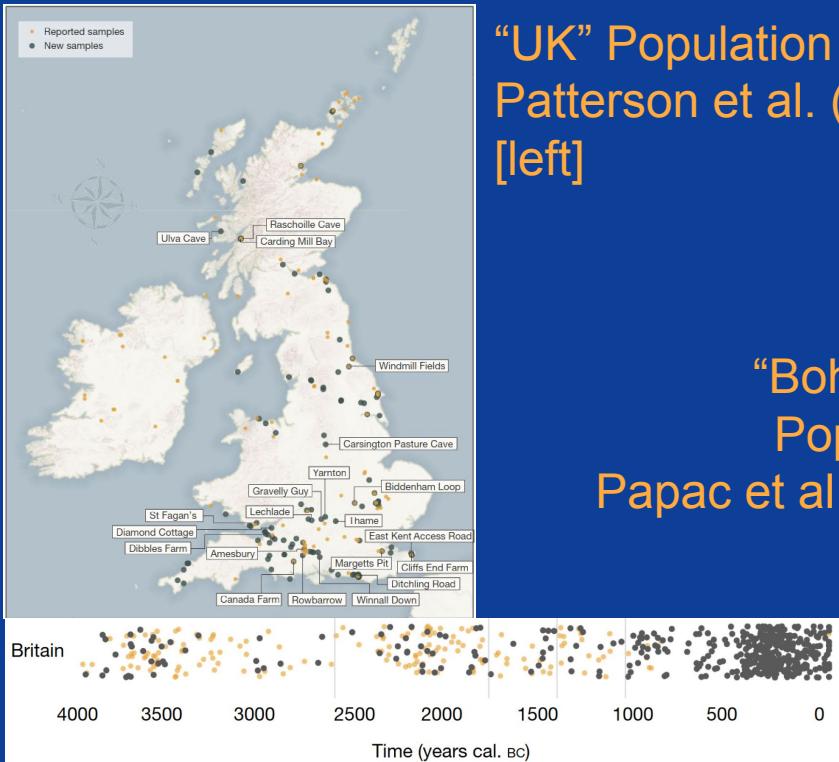
LETTER

doi:10.1038/nature13673

Ancient human genomes suggest three ancestral populations for present-day Europeans

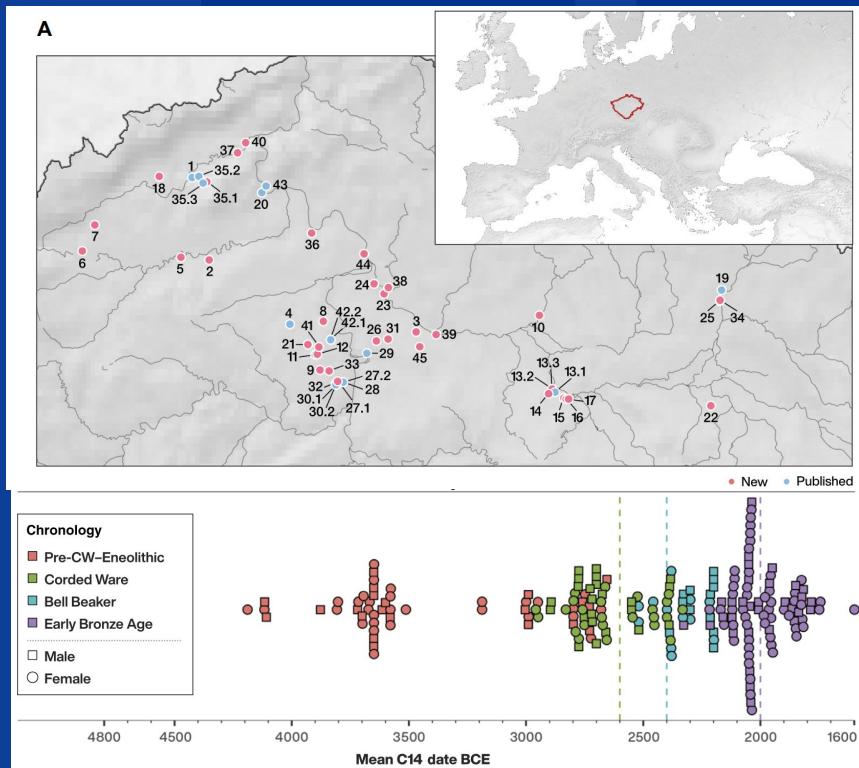
A list of authors and their affiliations appears at the end of the paper





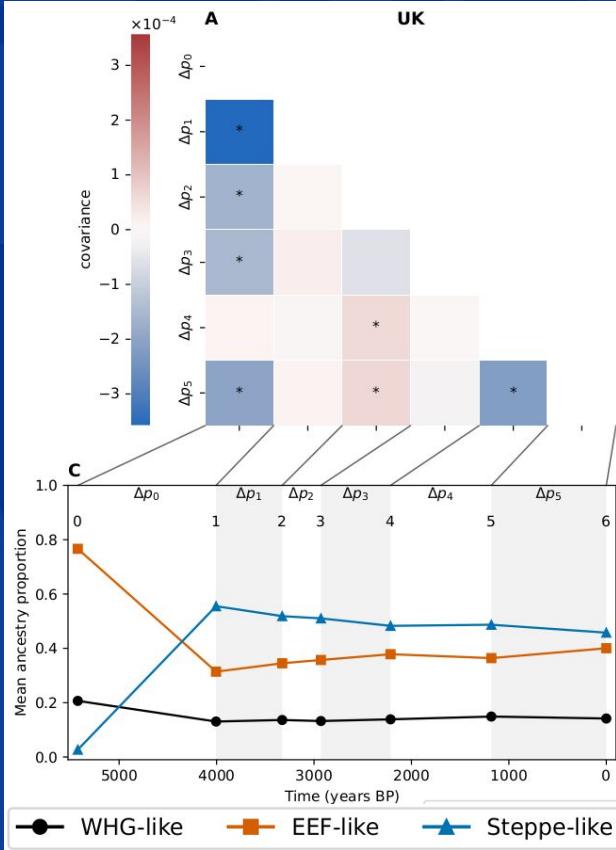
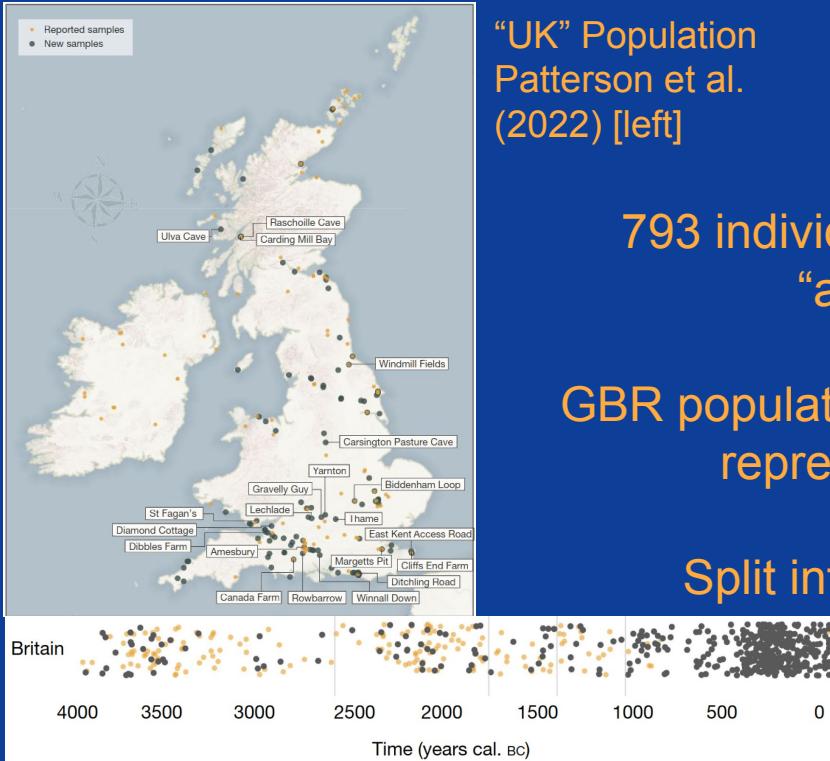
“UK” Population
Patterson et al. (2022)
[left]

“Bohemian”
Population
Papac et al. (2021)
[right]



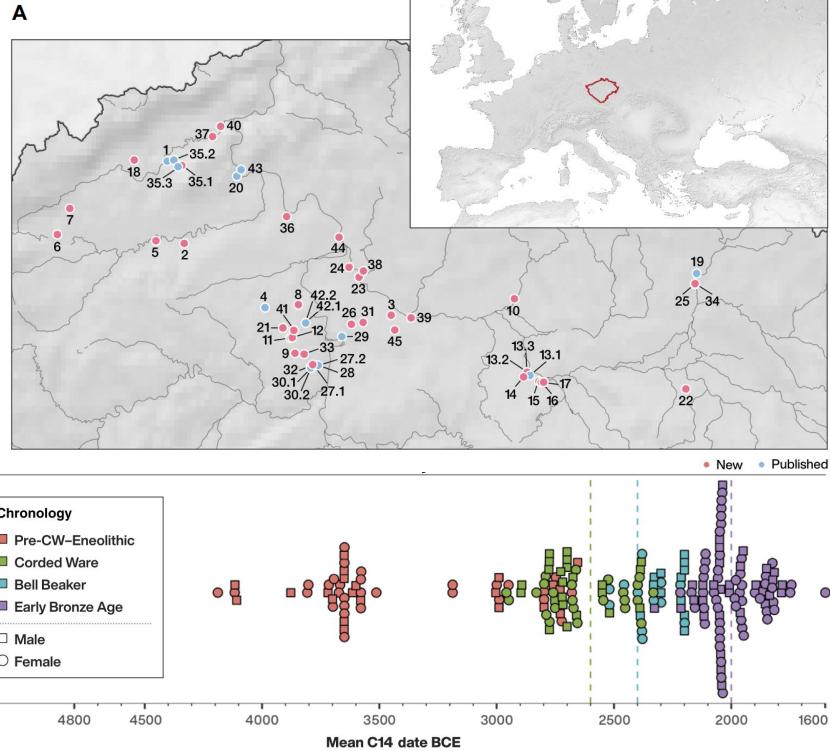


“UK” Data

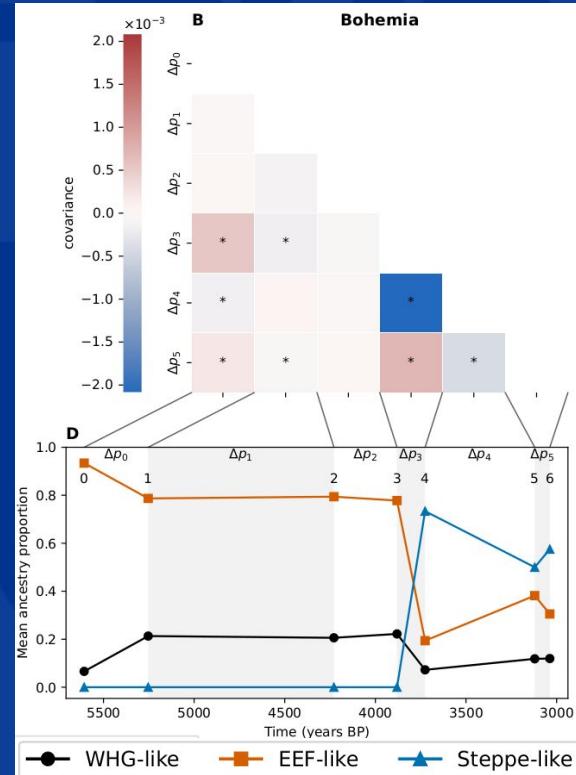


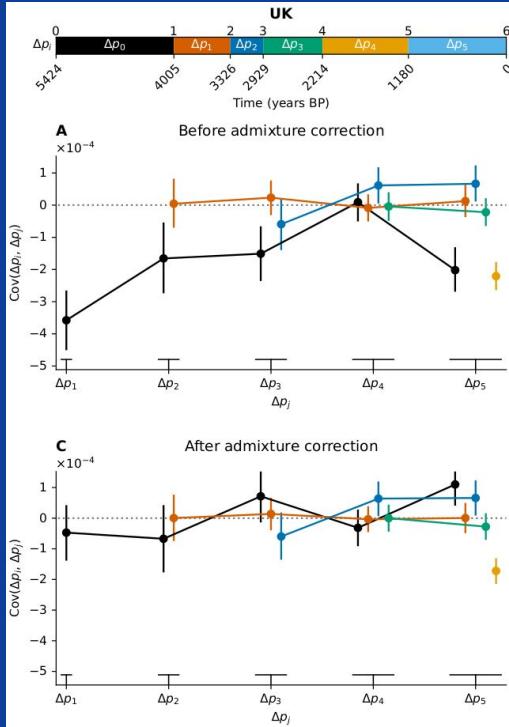


“Bohemia” Data



“Bohemian” Population
Papac et al. (2021)

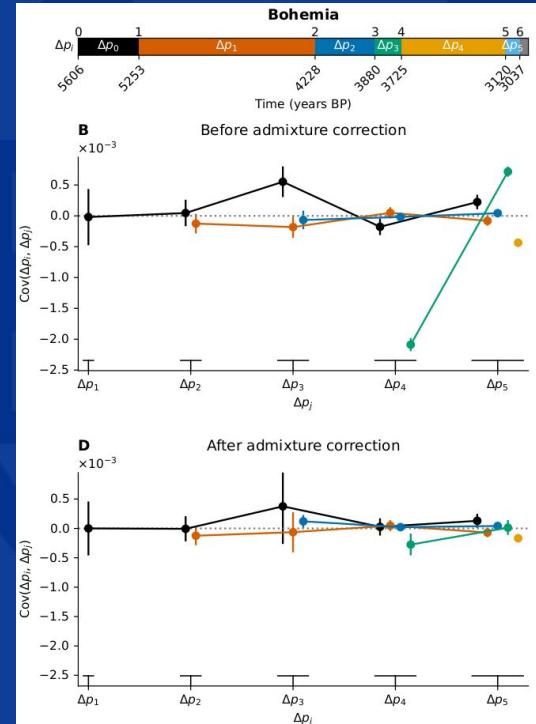




Data Analysis Results

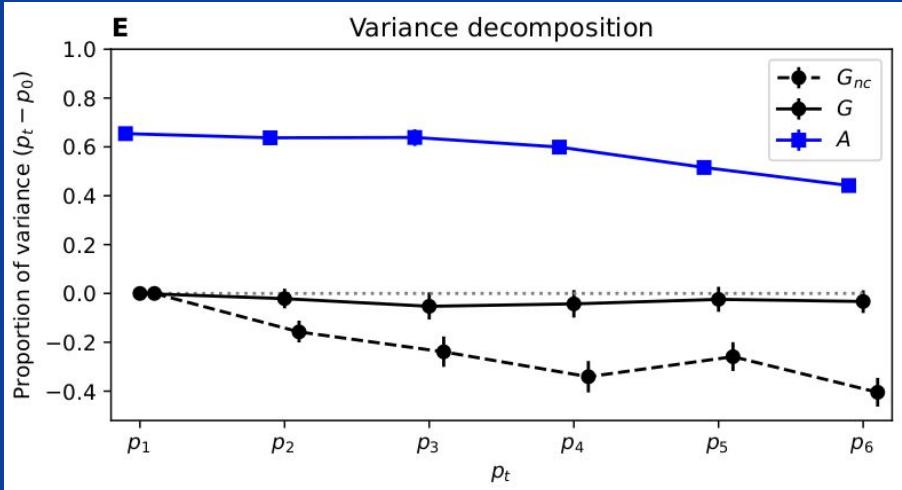
Bohemia Data

Both datasets
show spurious
signals of
selection before
correction for
admixture



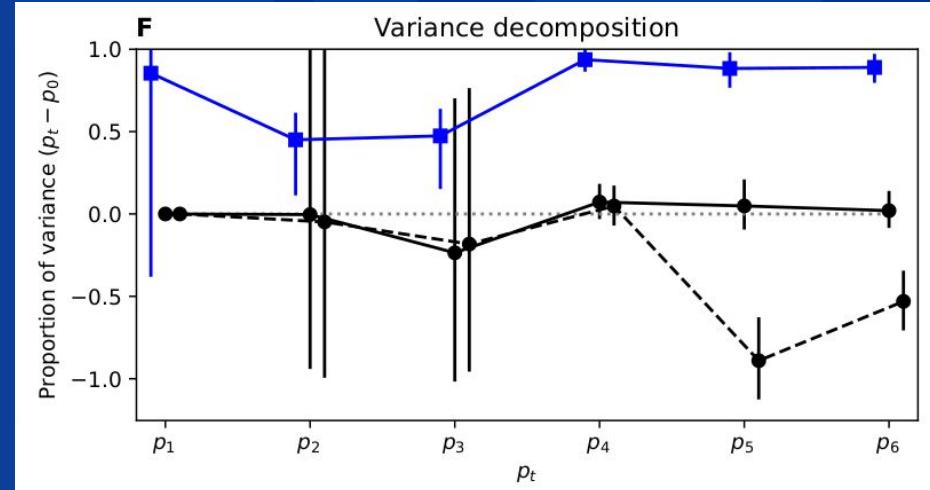


UK Data

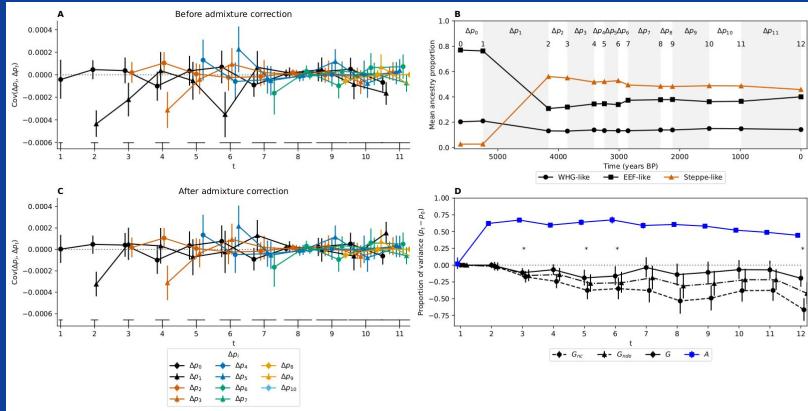


Data Analysis Results

Bohemia Data



After correction, variance decompositions show negligible evidence for selection—majority of covariance generated by drift and gene flow

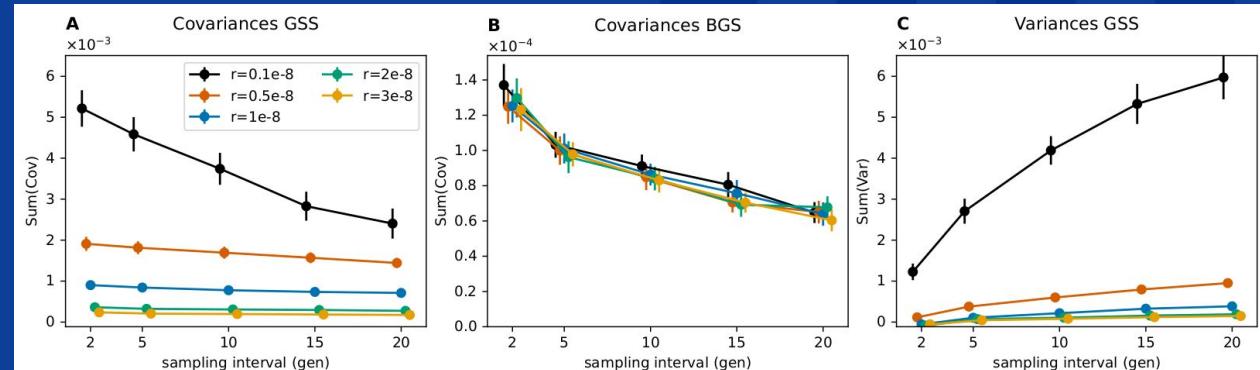


Different selection schemes
GSS = gaussian stabilizing selection (A & C)
BGS = background selection (B)

Data Analysis (Robustness Checks)

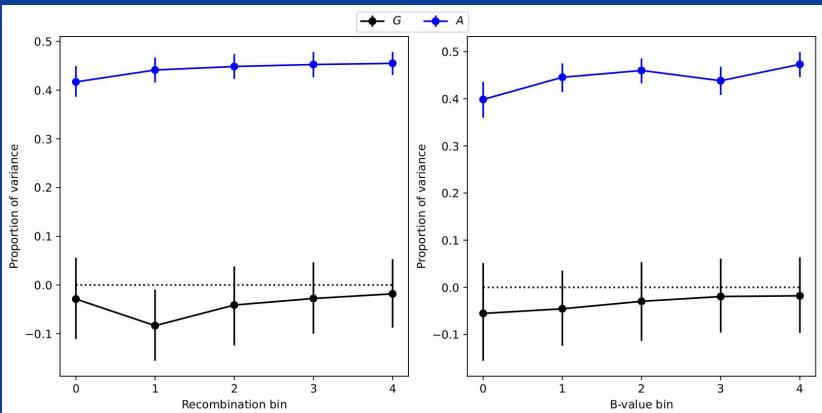
To examine effects of time transects, further split each transect in two (left)

Takeaway: sum of covariances increases but only appreciably when recombination rate is low



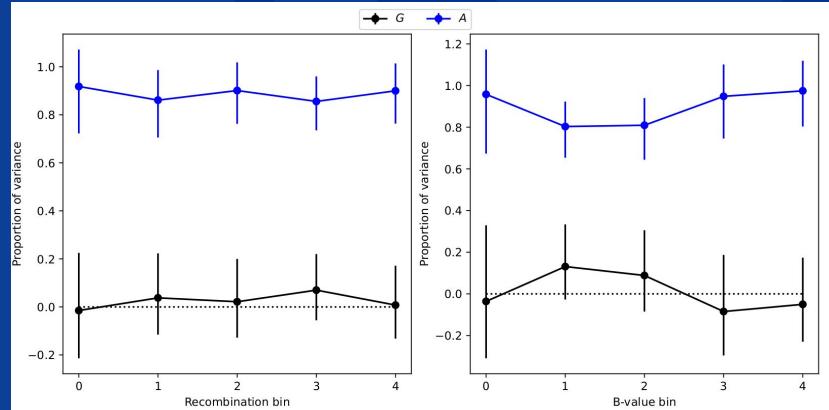


UK Data



Temporal covariances decrease with longer time intervals, but linked selection makes greater contribution to variance change *within* time intervals. Strategy to assess: bin SNPs by local recombination rate and B-value (**carries information about recombination rate and density of coding sites**).

Data Analysis (Robustness Checks) Bohemia Data

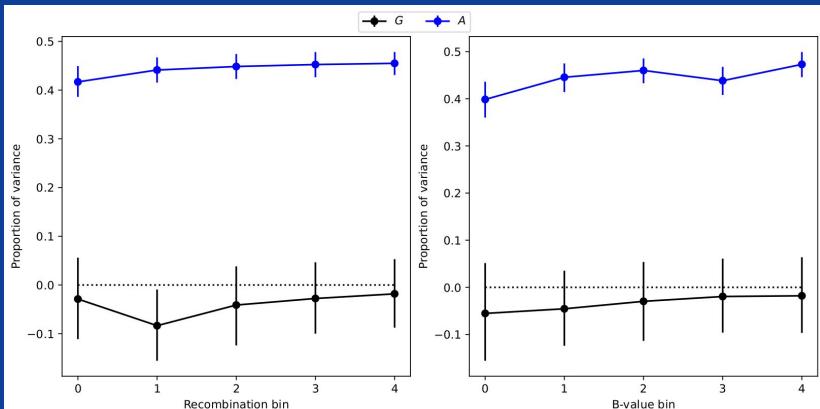


$$B = B^{(v)} = \frac{\pi_e}{\pi_0} = \frac{N_e}{N_0} \approx \exp \left[- \sum_x \int_0^1 \frac{u_x f_x(t) dt}{t(1 + (1-t)r_{x,v}/t)^2} \right]$$

B-value, McVicker et al. (2009)



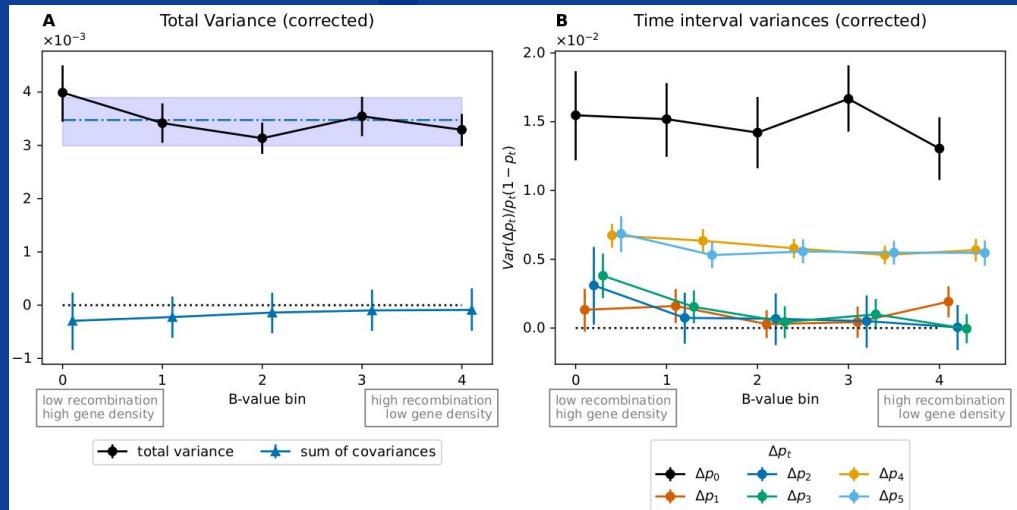
UK Data



$$B = B^{(v)} = \frac{\pi_e}{\pi_0} = \frac{N_e}{N_0} \approx \exp \left[- \sum_x \int_0^1 \frac{u_x f_x(t) dt}{t(1 + (1-t)r_{x,v}/t)^2} \right]$$

B-value, McVicker et al. (2009)

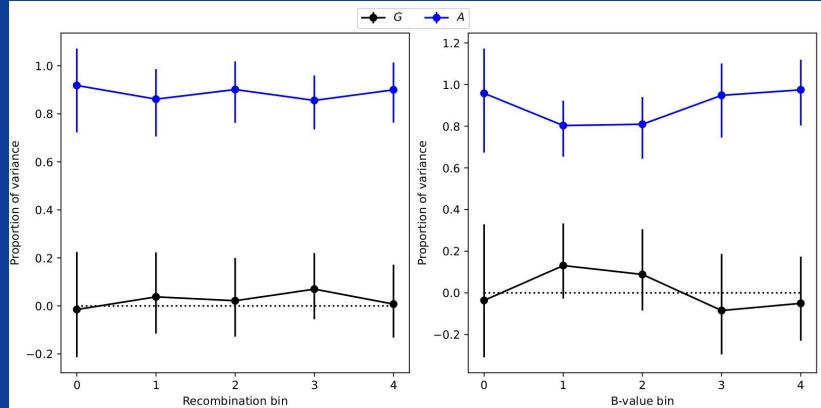
Data Analysis (Robustness Checks)



figs. S10 and S14). However, in the UK transect, we do see a significant increase in the total variance in allele frequency change in the lowest B-value bin (corresponding to the largest decrease in effective population size due to background selection, fig. 5A, first bin mean is above the genome-wide 95 % CI) and in the variances of change within some of the time intervals (fig. 5B). Noise in the smaller Bohemia time transect precludes seeing



Bohemia Data

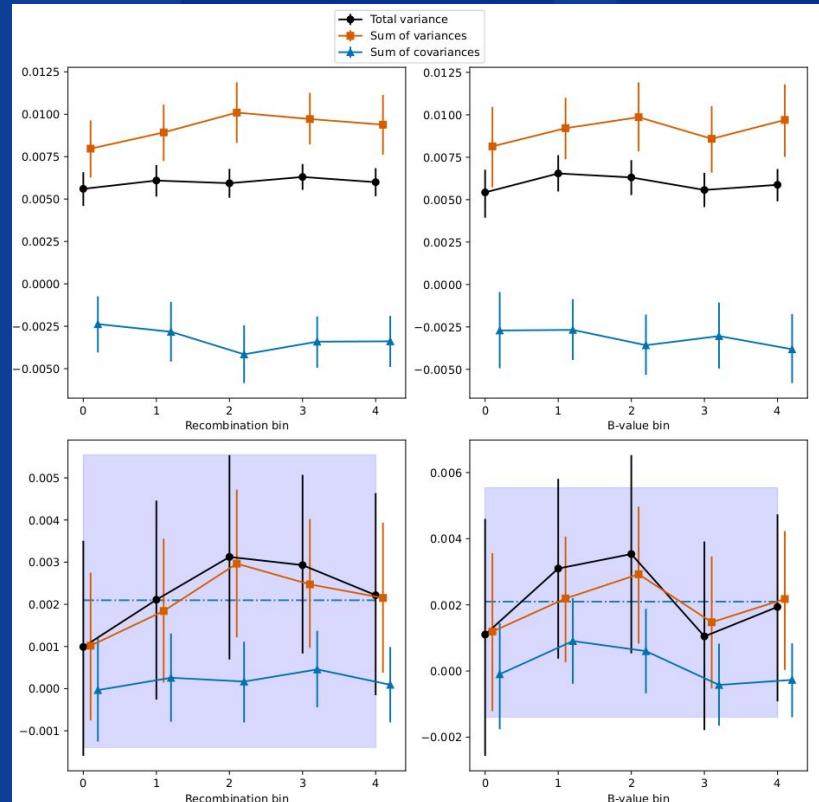


$$B = B^{(v)} = \frac{\pi_e}{\pi_0} = \frac{N_e}{N_0} \approx \exp \left[- \sum_x \int_0^1 \frac{u_x f_x(t) dt}{t(1 + (1-t)r_{x,v}/t)^2} \right]$$

B-value, McVicker et al. (2009)

Attempted to do the same for Bohemian data
but too much noise

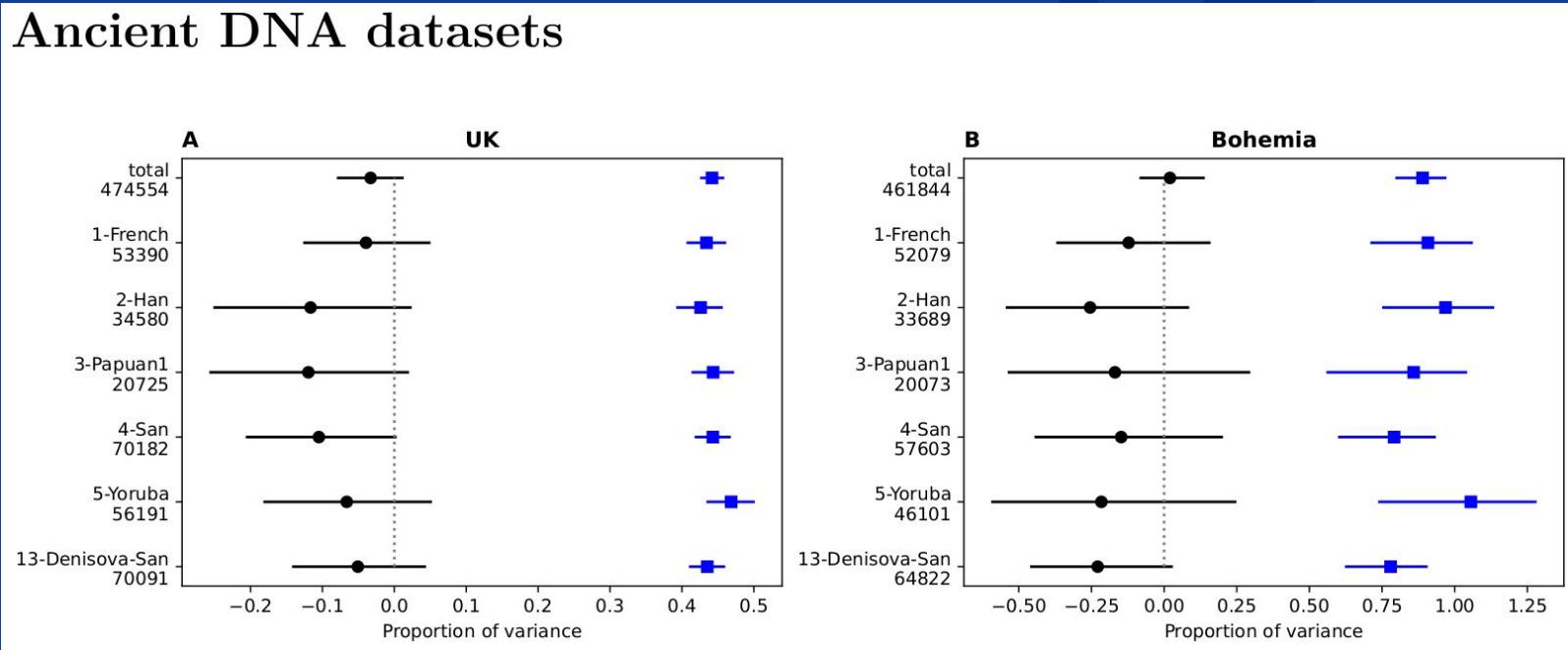
Data Analysis (Robustness Checks)





Check for Ascertainment Bias

Ancient DNA datasets





Caveats

Inferences only valid for *these* populations during *this* time period

Gene flow modeled as pulses (i.e. not continuous)

Assumed that allele frequencies in source populations are constant (and that estimates are accurate)

Estimates of variance in allele frequency change attributable to gene flow might be underestimated

Estimates of variance in allele frequency change attributable to linked selection are lower bounds

Negative selection may be working too fast to be detected

Estimates are genome-wide and so may not pick up small fraction of loci subject to strong selection

Sample sizes relatively small (such as aDNA...)

Additional Caveat: THIS IS NOT THE PRESENTER'S AREA OF EXPERTISE



Model larger (and more diverse)
populations, over longer time scales

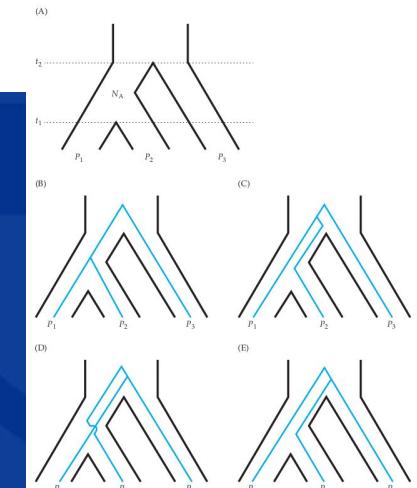
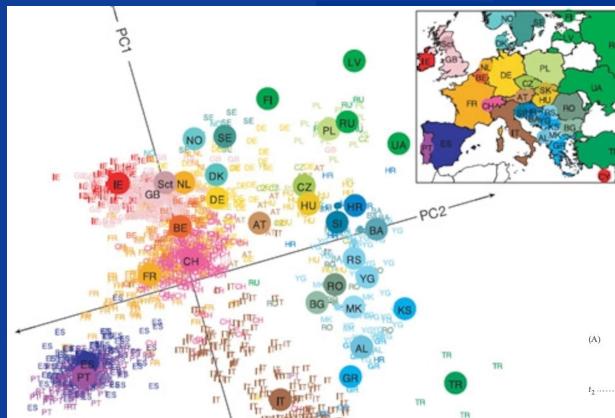
Applications to non-human organisms

Interesting suggestions by authors:

our focal time series. One future extension might be to use principal components analysis to learn about major axes of population structure involving samples in a time series and then to regress these PCAs out of our genotypes to account for variation in ancestry composition in a more model-free manner.

frequencies might not fully account for the impact of migration. More detailed modeling with admixture graphs and tree sequences could help better resolve the sources of gene flow in time series (e.g. Allentoft et al., 2022; Irving-Pease et al., 2022; Pearson and

Future Extensions





Direct detection of natural selection in Bronze Age Britain

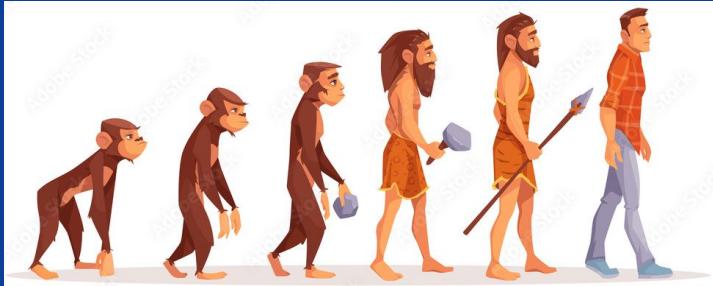
Iain Mathieson¹ and Jonathan Terhorst²

¹Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA;

²Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA

We developed a novel method for efficiently estimating time-varying selection coefficients from genome-wide ancient DNA data. In simulations, our method accurately recovers selective trajectories and is robust to misspecification of population size. We applied it to a large data set of ancient and present-day human genomes from Britain and identified seven loci with genome-wide significant evidence of selection in the past 4500 yr. Almost all of them can be related to increased vitamin D or calcium levels, suggesting strong selective pressure on these or related phenotypes. However, the strength of selection on individual loci varied substantially over time, suggesting that cultural or environmental factors moderated the genetic response. Of 28 complex anthropometric and metabolic traits, skin pigmentation was the only one with significant evidence of polygenic selection, further underscoring the importance of phenotypes related to vitamin D. Our approach illustrates the power of ancient DNA to characterize selection in human populations and illuminates the recent evolutionary history of Britain.

Mathieson & Terhorst (2022)



Le et al
(2023)

Future Extensions

More broadly: reconciliation with other recent results?

1,000 ancient genomes uncover 10,000 years of natural selection in Europe

Megan K. Le¹, Olivia S. Smith², Ali Akbari^{3,4,7}, Arbel Harpak^{2,5+}, David Reich^{3,4,6,7+} & Vagheesh M. Narasimhan^{2,8+}

Ancient DNA has revolutionized our understanding of human population history. However, its potential to examine how rapid cultural evolution to new lifestyles may have driven biological adaptation has not been met, largely due to limited sample sizes. We assembled genome-wide data from 1,291 individuals from Europe over 10,000 years, providing a dataset that is large enough to resolve the timing of selection into the Neolithic, Bronze Age, and Historical periods. We identified 25 genetic loci with rapid changes in frequency during these periods, a majority of which were previously undetected. Signals specific to the Neolithic transition are associated with body weight, diet, and lipid metabolism-related phenotypes. They also include immune phenotypes, most notably a locus that confers immunity to *Salmonella* infection at a time when ancient *Salmonella* genomes have been shown to adapt to human hosts, thus providing a possible example of human-pathogen co-evolution. In the Bronze Age, selection signals are enriched near genes involved in pigmentation and immune-related traits, including at a key human protein interactor of SARS-CoV-2. Only in the Historical period do the selection candidates we detect largely mirror previously-reported signals, highlighting how the statistical power of previous studies was limited to the last few millennia. The Historical period also has multiple signals associated with vitamin D binding, providing evidence that lactase persistence may have been part of an oligogenic adaptation for efficient calcium uptake and challenging the theory that its adaptive value lies only in facilitating caloric supplementation during times of scarcity. Finally, we detect selection on complex traits in all three periods, including selection favoring variants that reduce body weight in the Neolithic. In the Historical period, we detect selection favoring variants that increase risk for cardiovascular disease plausibly reflecting selection for a more active inflammatory response that would have been adaptive in the face of increased infectious disease exposure. Our results provide an evolutionary rationale for the high prevalence of these deadly diseases in modern societies today and highlight the unique power of ancient DNA in elucidating biological change that accompanied the profound cultural transformations of recent human history.



University of
Pittsburgh®

School of
Public Health

FIN.