Lessons from the amylase locus

Andrea Guarracino
Postdoctoral Scholar

@AndresGuarahino

HPRC Annual Meeting 2024
Seymour Marine Discovery Center, Santa Cruz, USA
2024/09/05

Panteam










- Human Technopole, Italy
- UTHSC, Tennessee
- UC Berkeley, California
- GeneDX and others, California

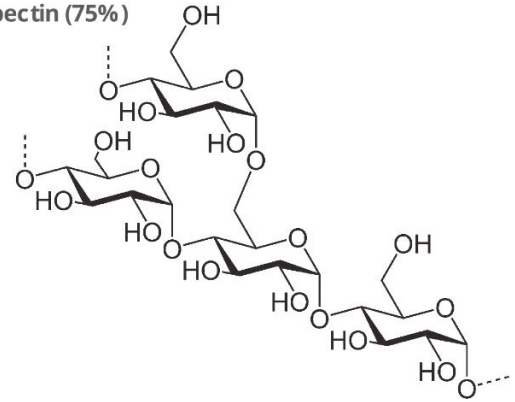
Outline

- Introduction
- Method
- Result

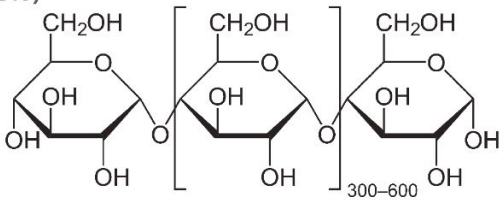
Amylase breaks down starches to sugars

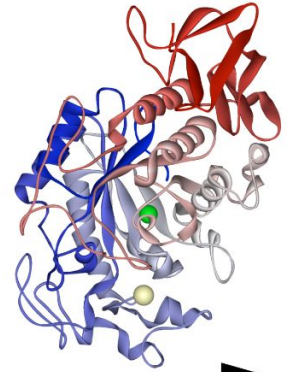
Starch

amylopectin (75%)



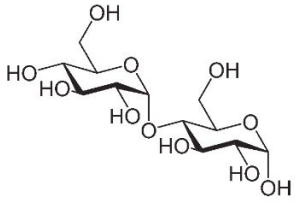
amylose (25%)



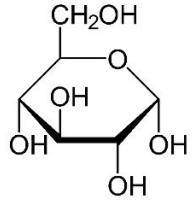


human salivary
α amylase

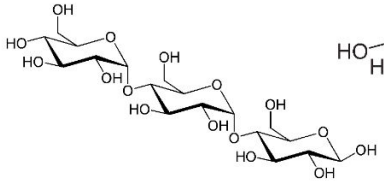
Sugar



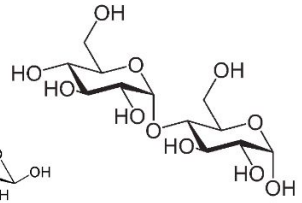
Maltose



Glucose



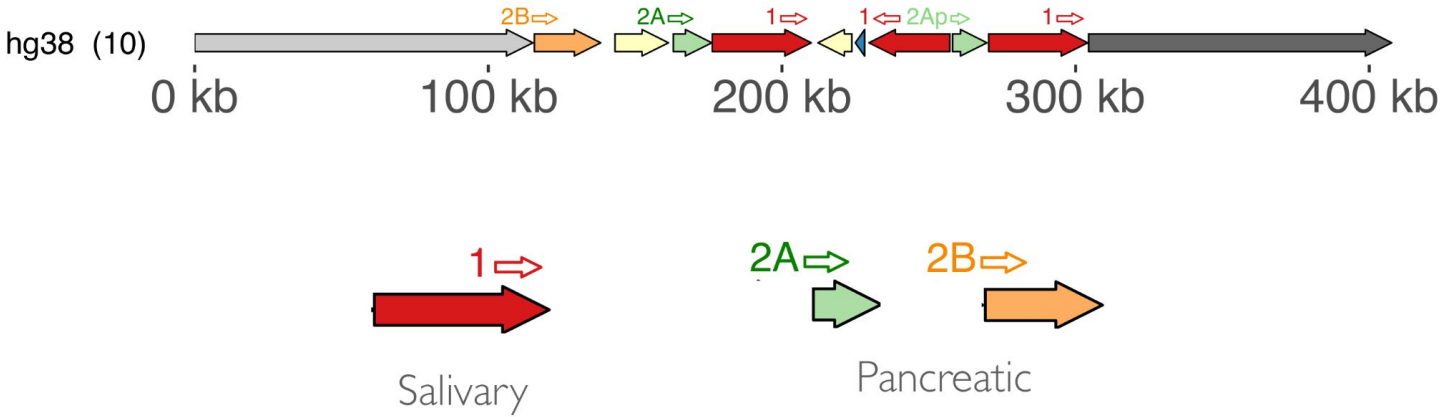
Maltotriose



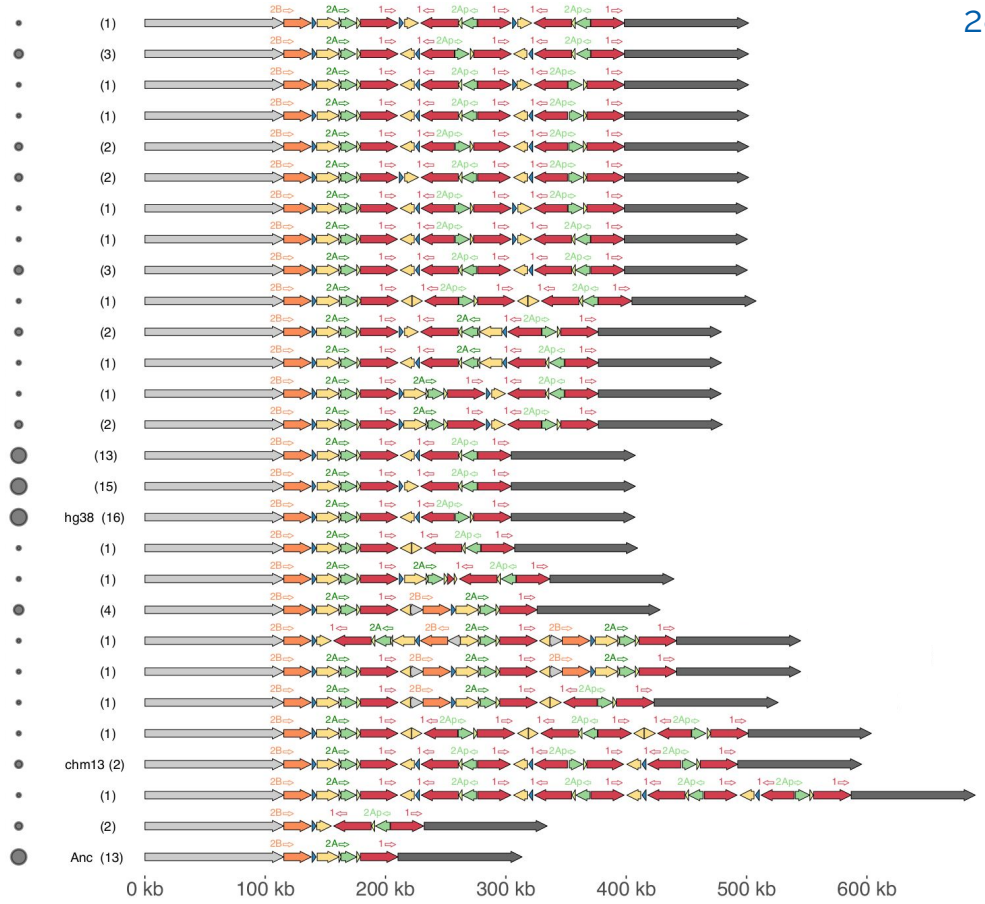
Maltose

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER.

Amylase is a multi-copy gene family

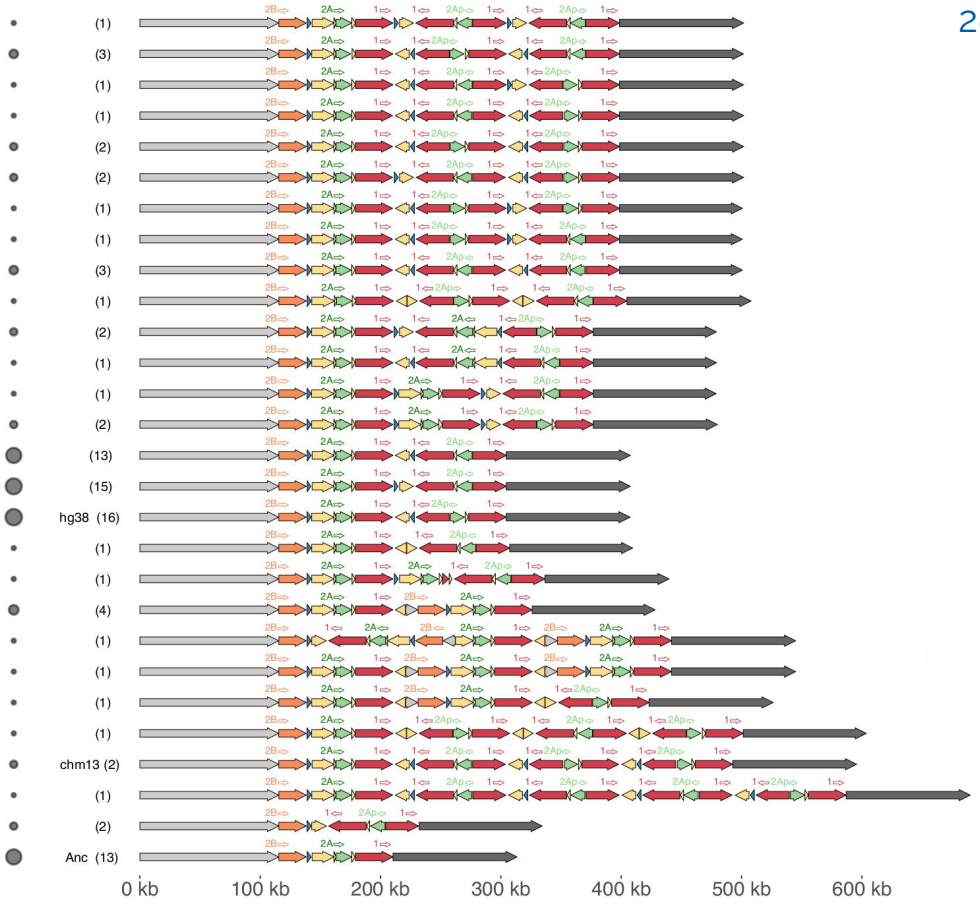


Human amylase copy number diversity

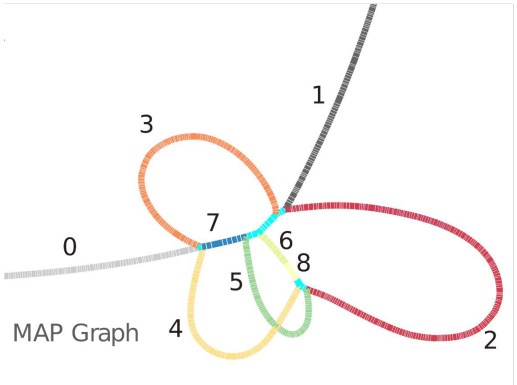


28 amylase structural haplotypes

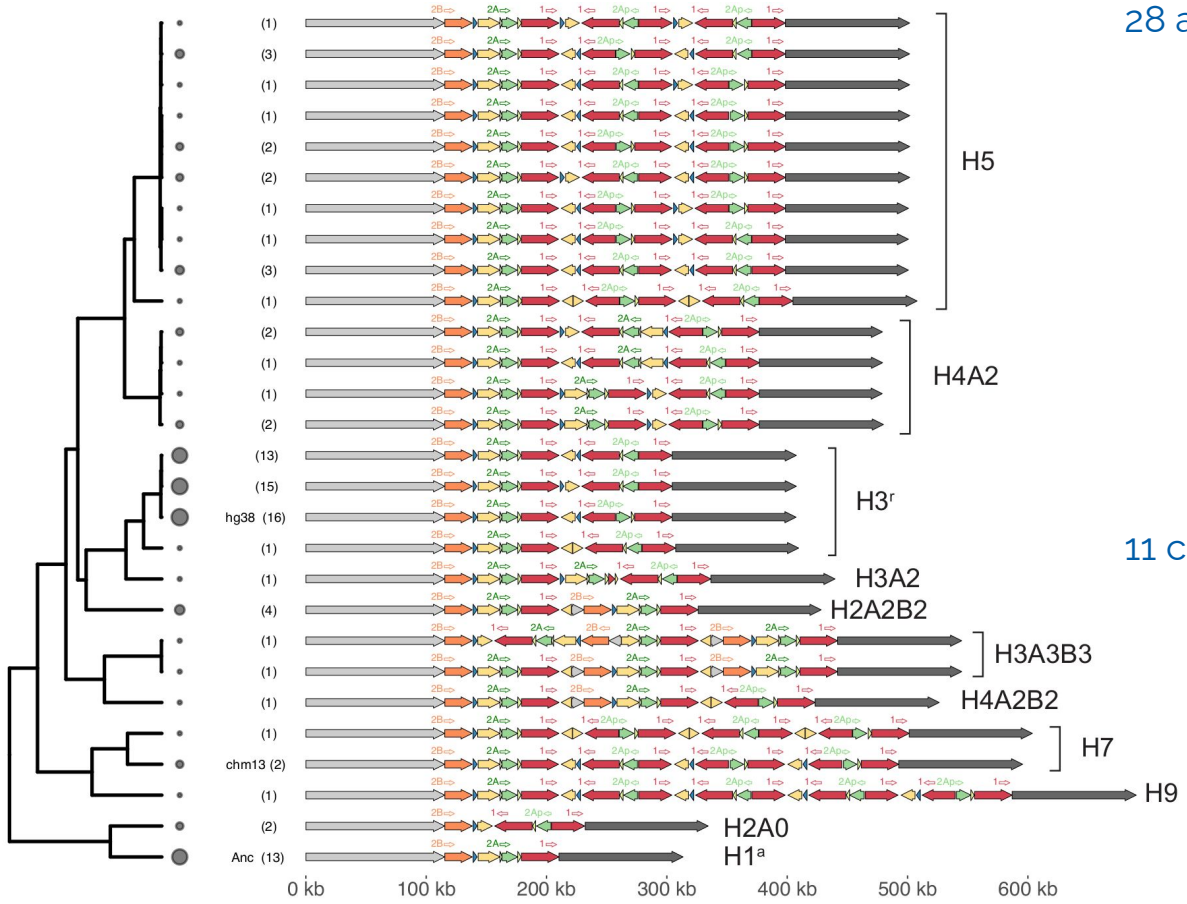
Human amylase copy number diversity



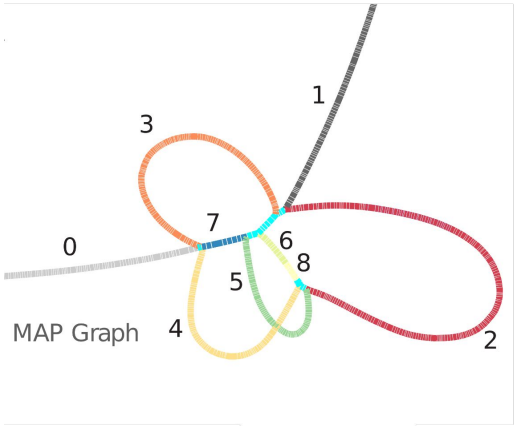
28 amylase structural haplotypes



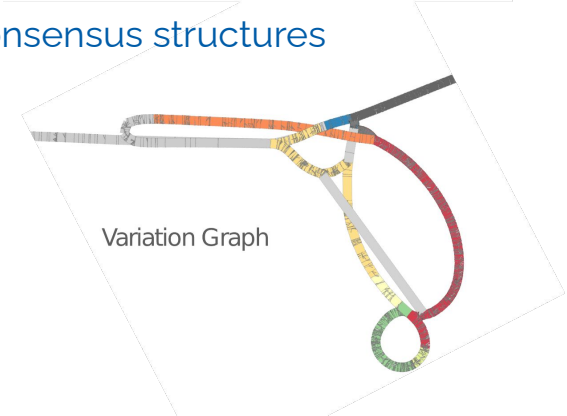
Human amylase copy number diversity



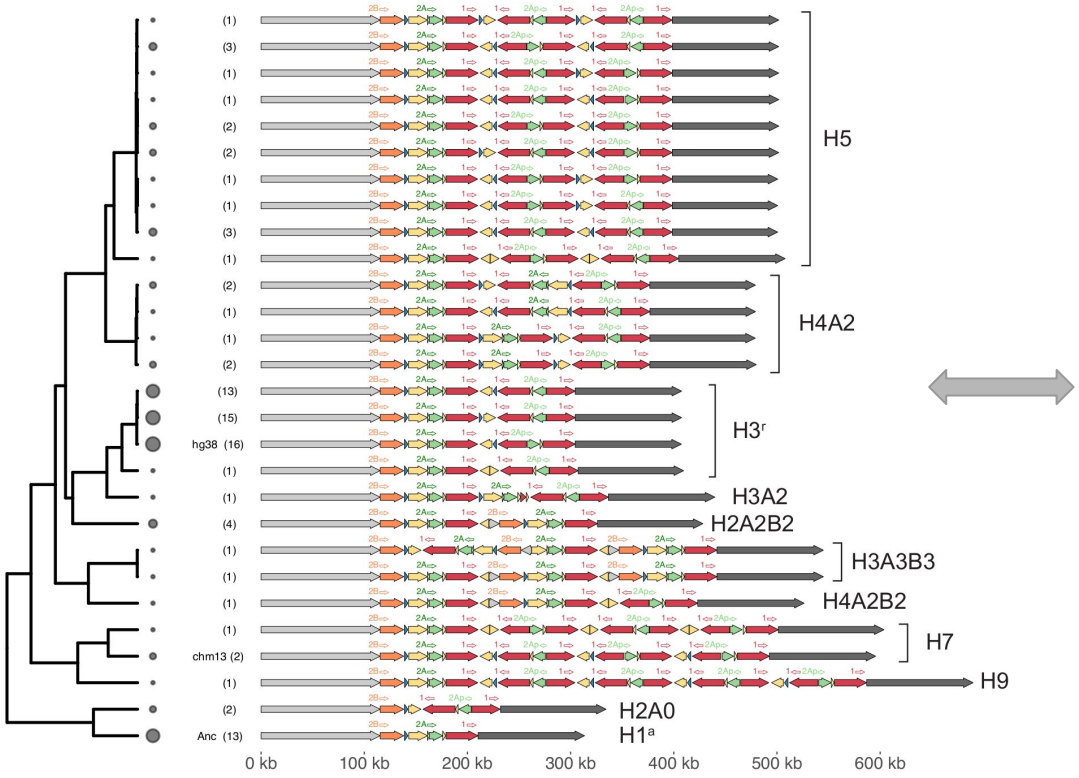
28 amylase structural haplotypes

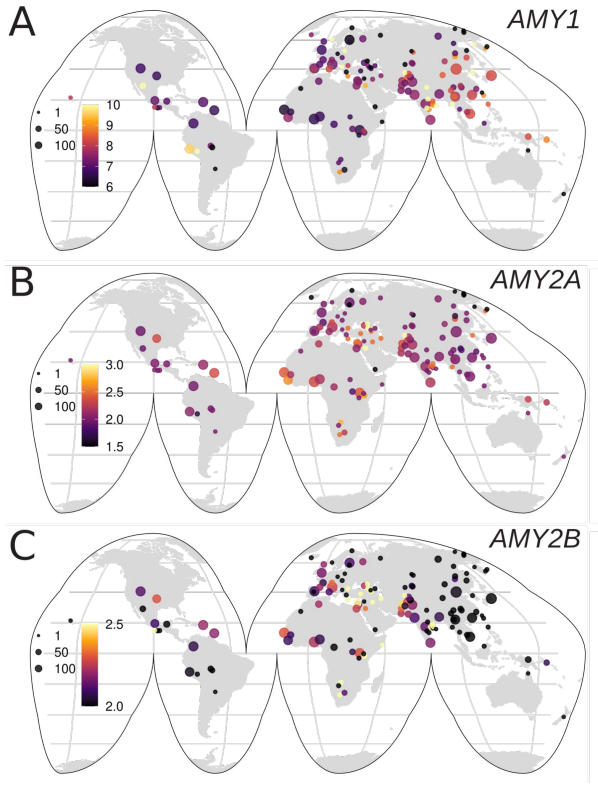


11 consensus structures



We want more





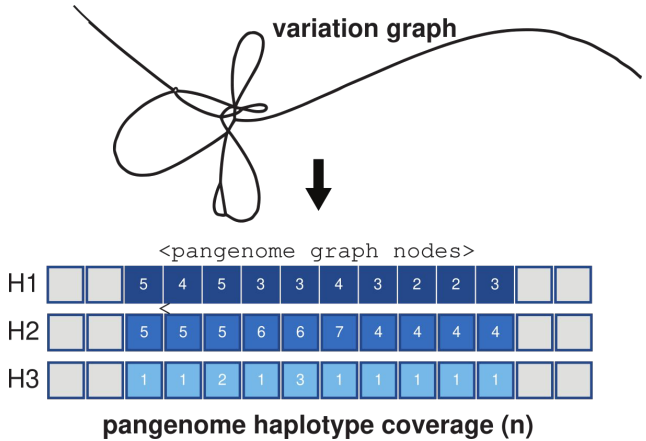
94 long read haplotypes

~5600 short read haplotypes

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Outline

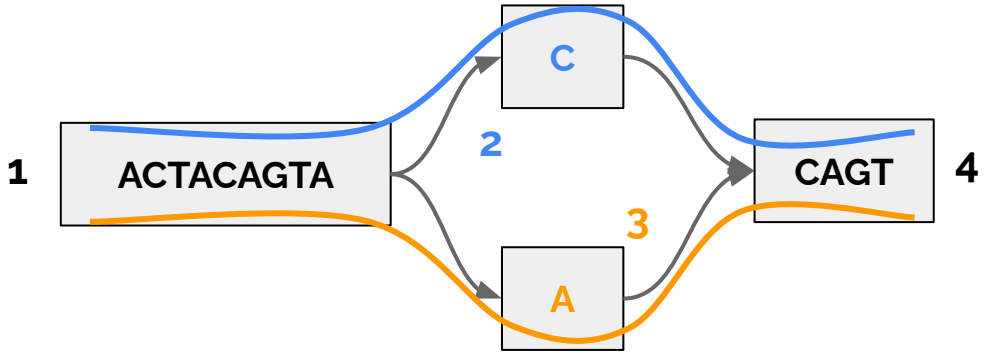
- Introduction
- Method
- Result



Node coverage matrix

Genome 1: **ACTACAGTACCAGT**

Genome 2: **ACTACAGTAACAGT**

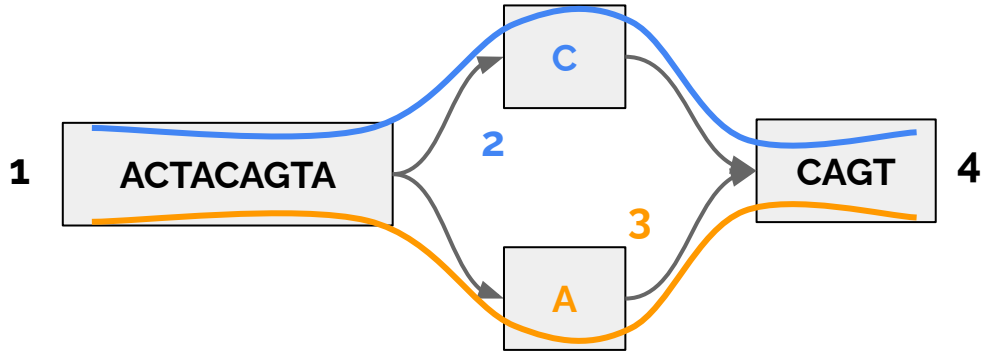


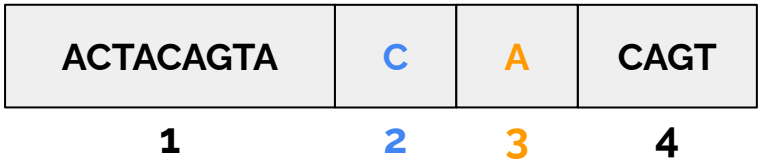
ACTACAGTA	C	A	CAGT
1	2	3	4

Node coverage matrix

Genome 1: **ACTACAGTACCAGT**

Genome 2: **ACTACAGTAACAGT**





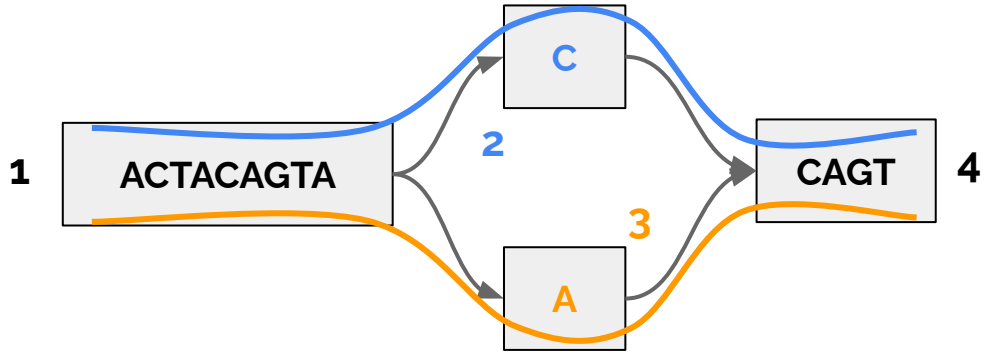
Path coverage matrix

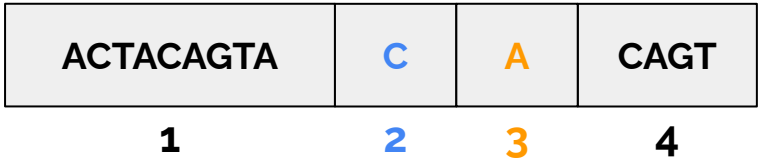
	1	2	3	4
Genome 1	1	1	0	1
Genome 2	1	0	1	1

Node coverage matrix

Genome 1: **ACTACAGTACCAGT**

Genome 2: **ACTACAGTAACAGT**





Path coverage matrix

	1	2	3	4
Genome 1	1	1	0	1
Genome 2	1	0	1	1

Sample coverage matrix

	1	2	3	4
Sample 1	29	31	1	27
Sample 2	28	0	30	28

Graph-based genotyping with COSIGT

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

https://github.com/davidebolo1993/cosigt

COSIGT

COSINE SIMILARITY-BASED GENOTYPER

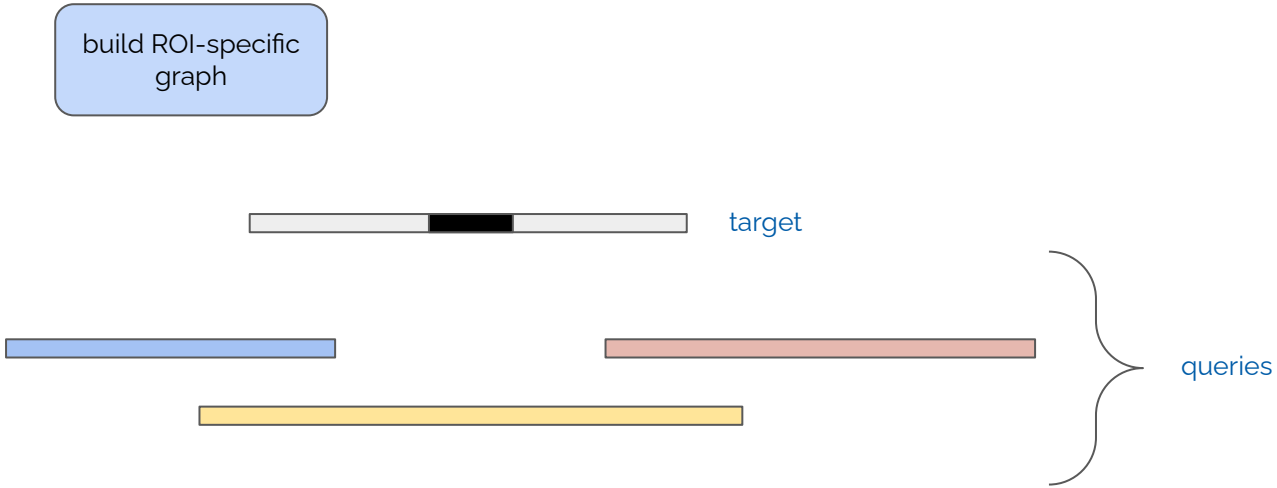
Graph-based genotyping with COSIGT

build ROI-specific graph

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

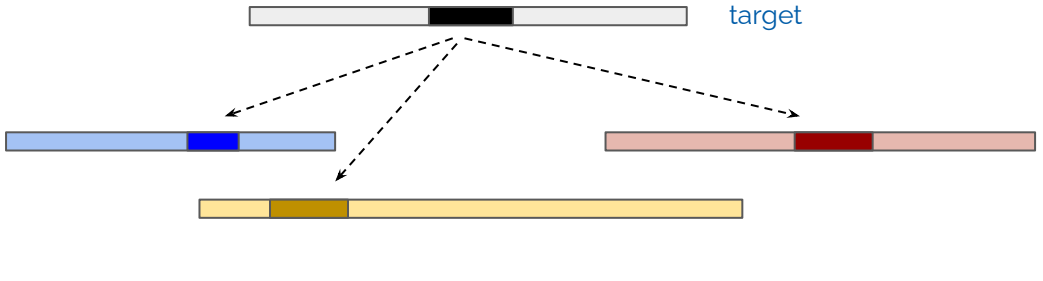


https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

build ROI-specific graph



IMPG

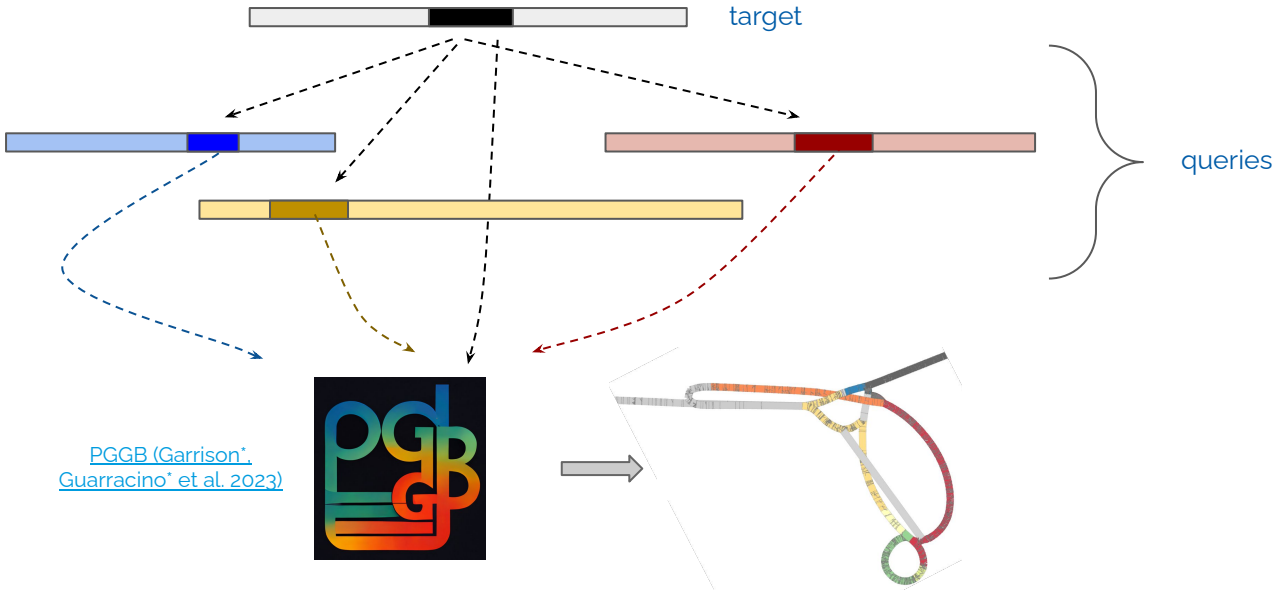
impg

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

build ROI-specific graph



IMPG

impg

https://github.com/davidebolo1993/cosigt

COSIGT

COSINE SIMILARITY-BASED GENOTYPER

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

Graph-based genotyping with COSIGT

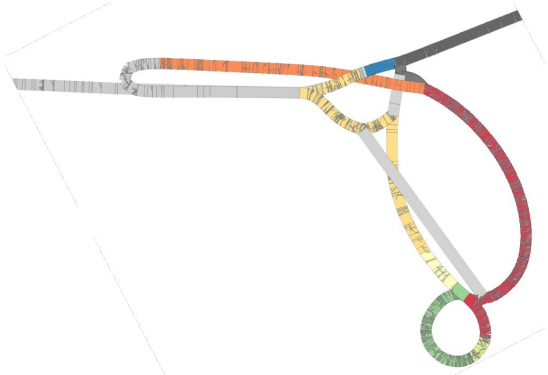
THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

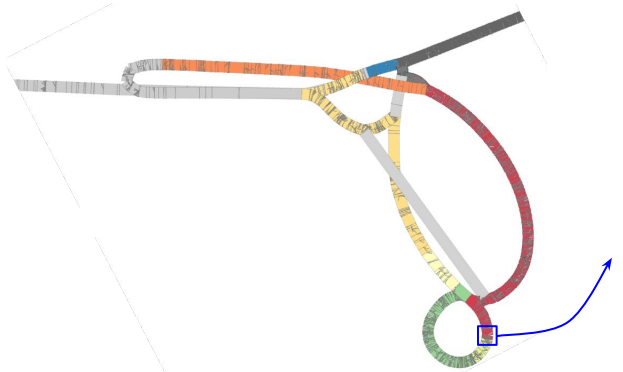
Graph-based genotyping with COSIGT

build ROI-specific graph → compute path coverage matrix

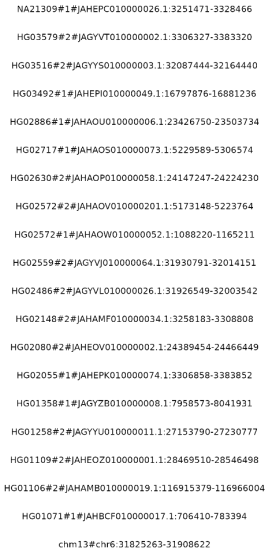


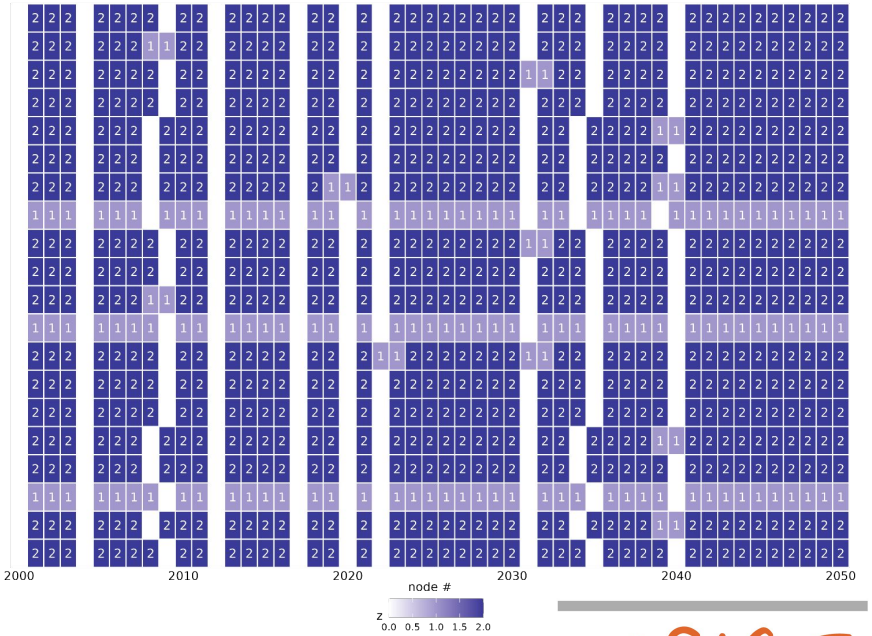
Graph-based genotyping with COSIGT

build ROI-specific graph → compute path coverage matrix



ODGI (Guarracino, Heumos et al., 2022)



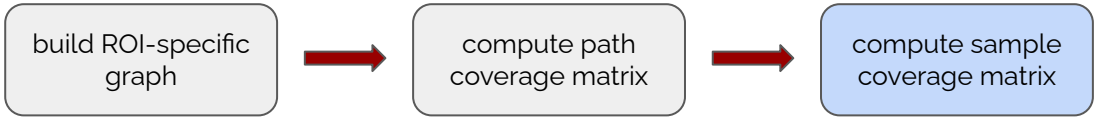


THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER.

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

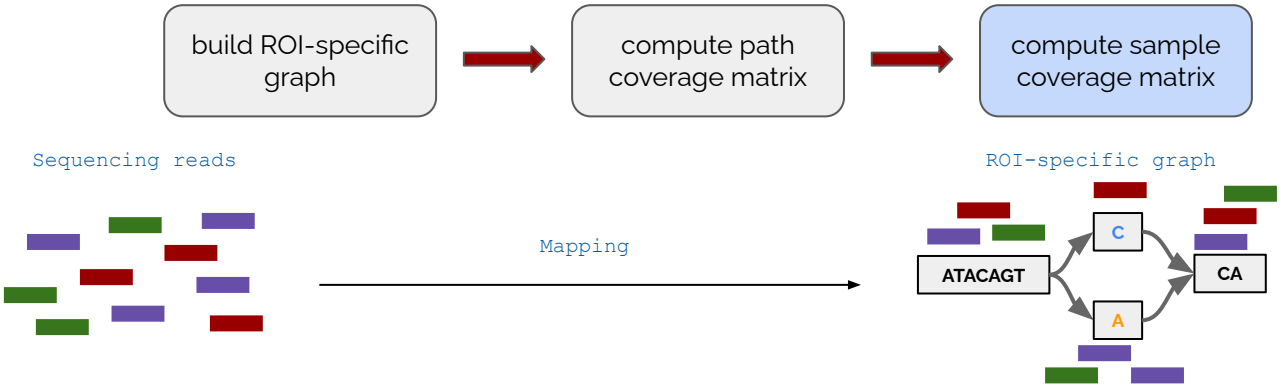


THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

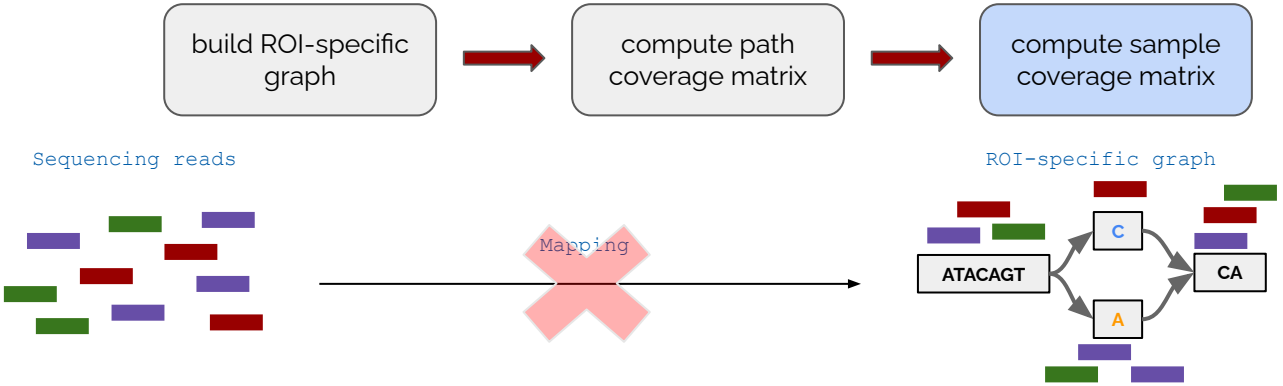


THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

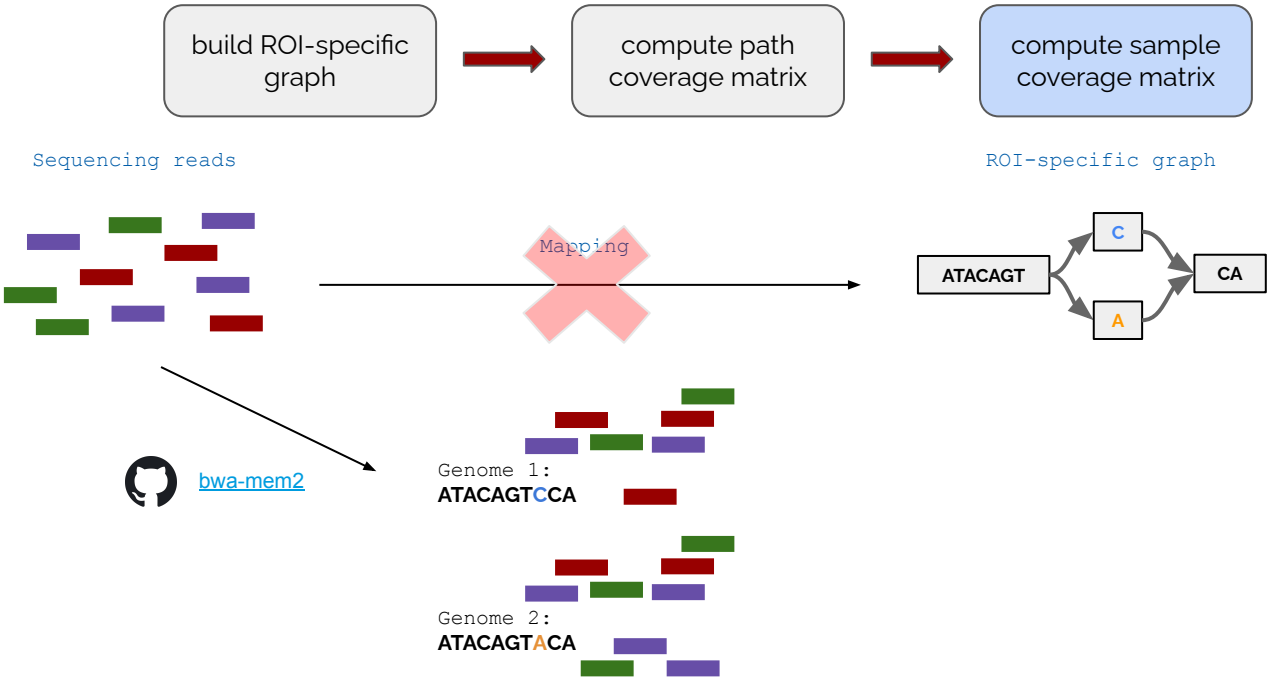


THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

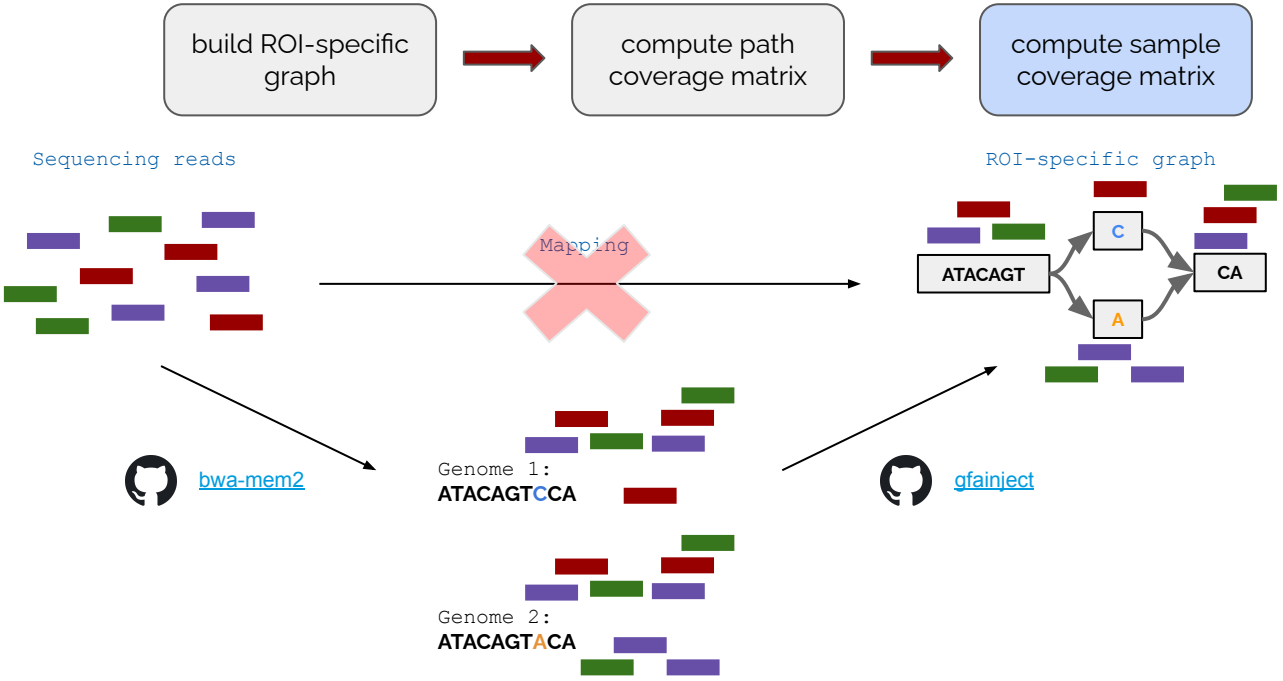
https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

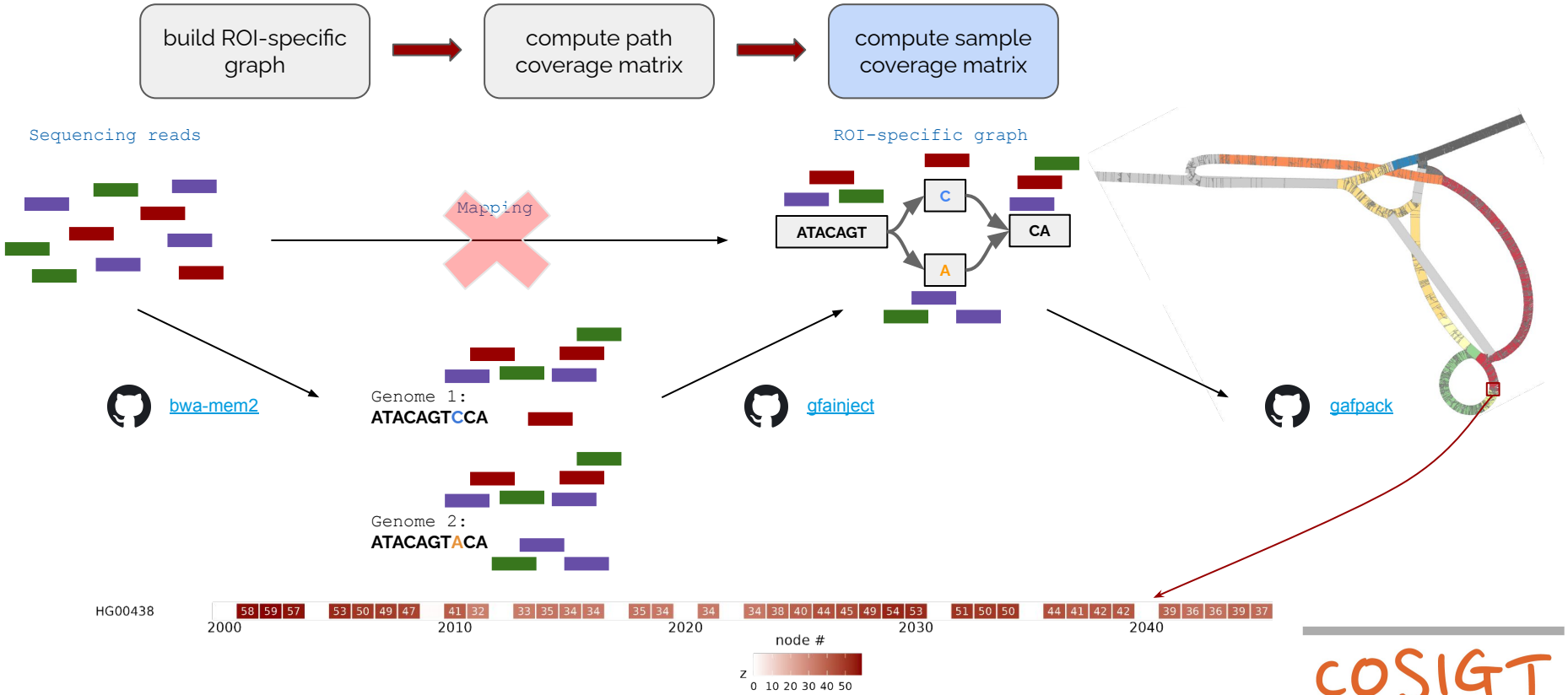
Graph-based genotyping with COSIGT



Graph-based genotyping with COSIGT



Graph-based genotyping with COSIGT



https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

haplotype deconvolution: find the combination of *N* path coverage vectors best representing the trend observed in the sample coverage vector

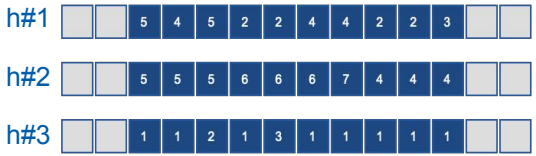
https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

haplotype deconvolution: find the combination of *N* path coverage vectors best representing the trend observed in the sample coverage vector

path coverage matrix



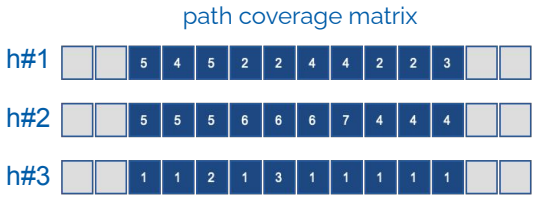
THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER.

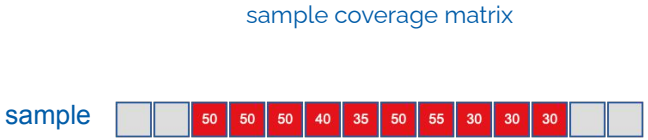
https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

haplotype deconvolution: find the combination of *N* path coverage vectors best representing the trend observed in the sample coverage vector





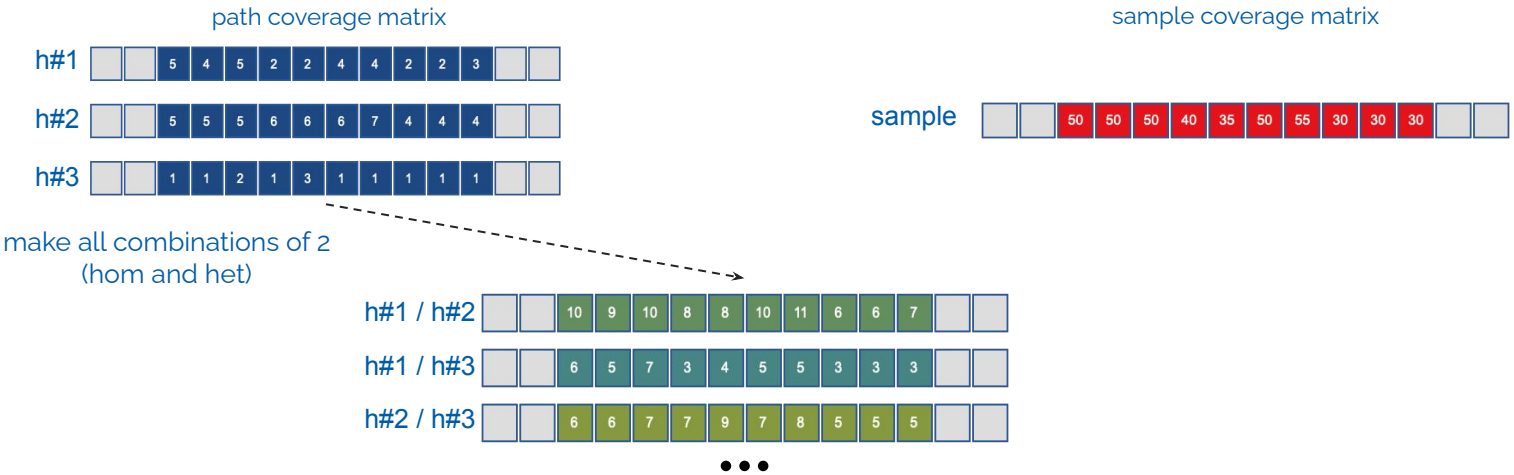
THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER.

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

haplotype deconvolution: find the combination of *N* path coverage vectors best representing the trend observed in the sample coverage vector



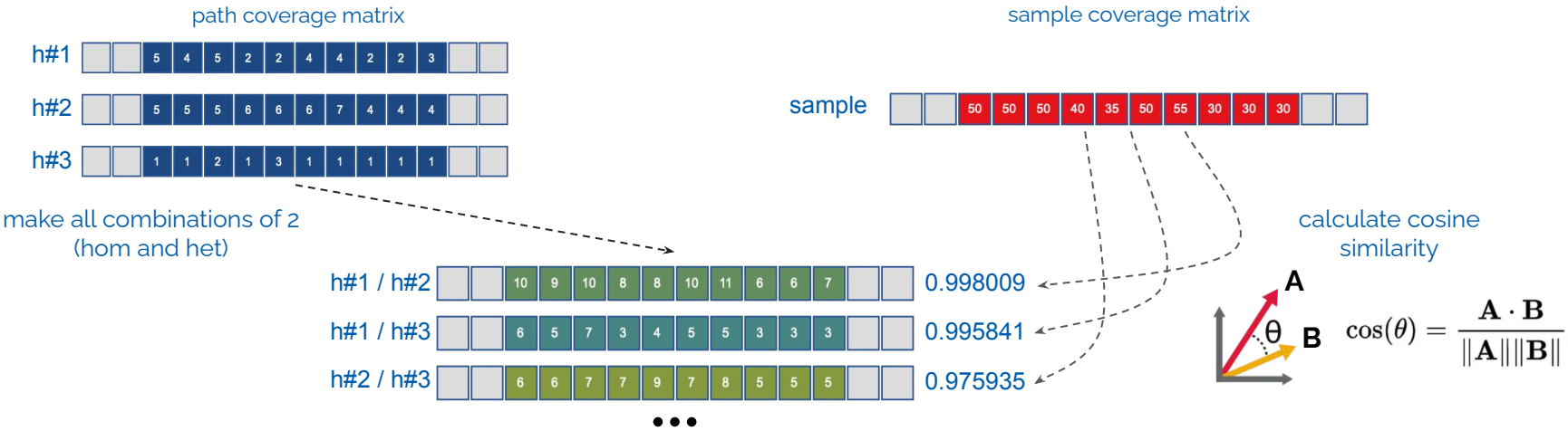
THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

haplotype deconvolution: find the combination of *N* path coverage vectors best representing the trend observed in the sample coverage vector



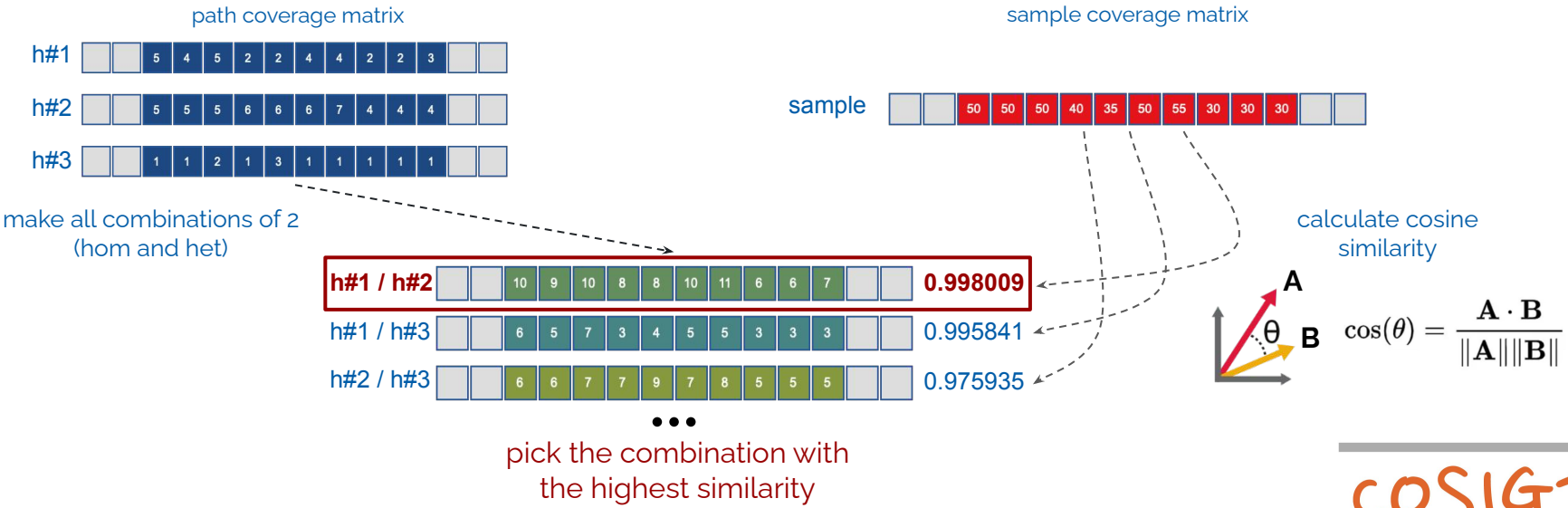
THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Graph-based genotyping with COSIGT

haplotype deconvolution: find the combination of *N* path coverage vectors best representing the trend observed in the sample coverage vector



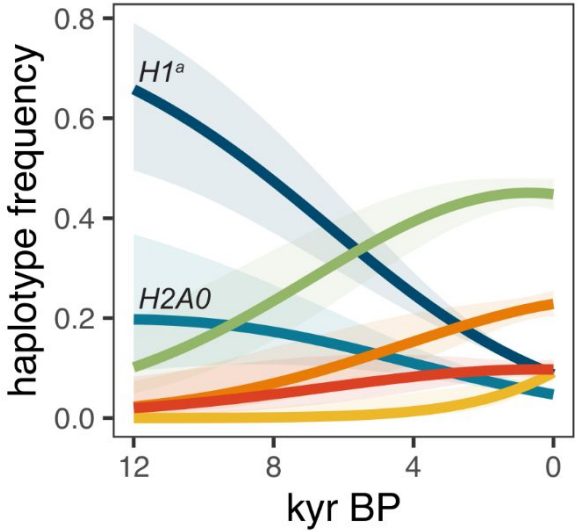
https://github.com/davidebolo1993/cosigt

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

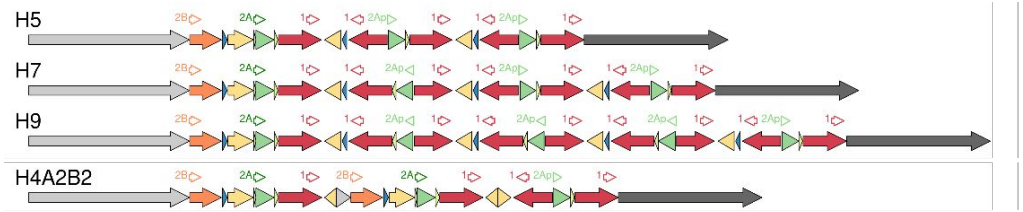
Outline

- Introduction
- Method
- Result

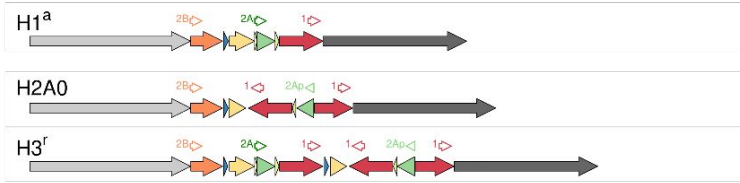


Allele frequencies of complex structural haplotypes

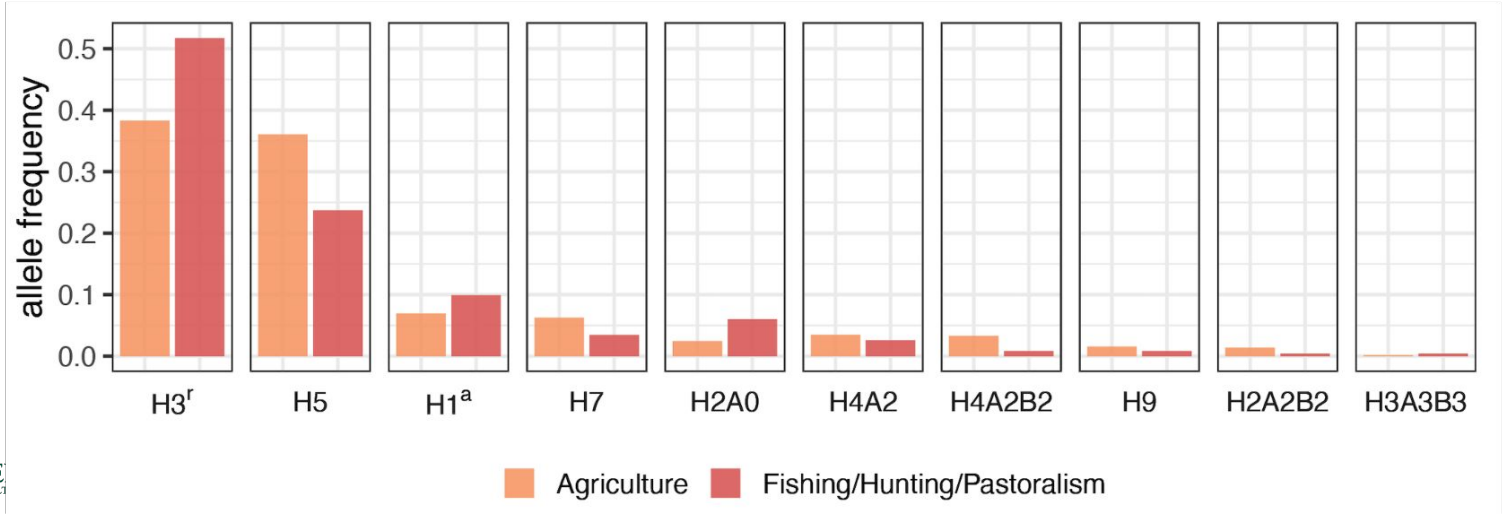
Pangenome-based haplotype deconvolution for ~5600 short read haplotypes



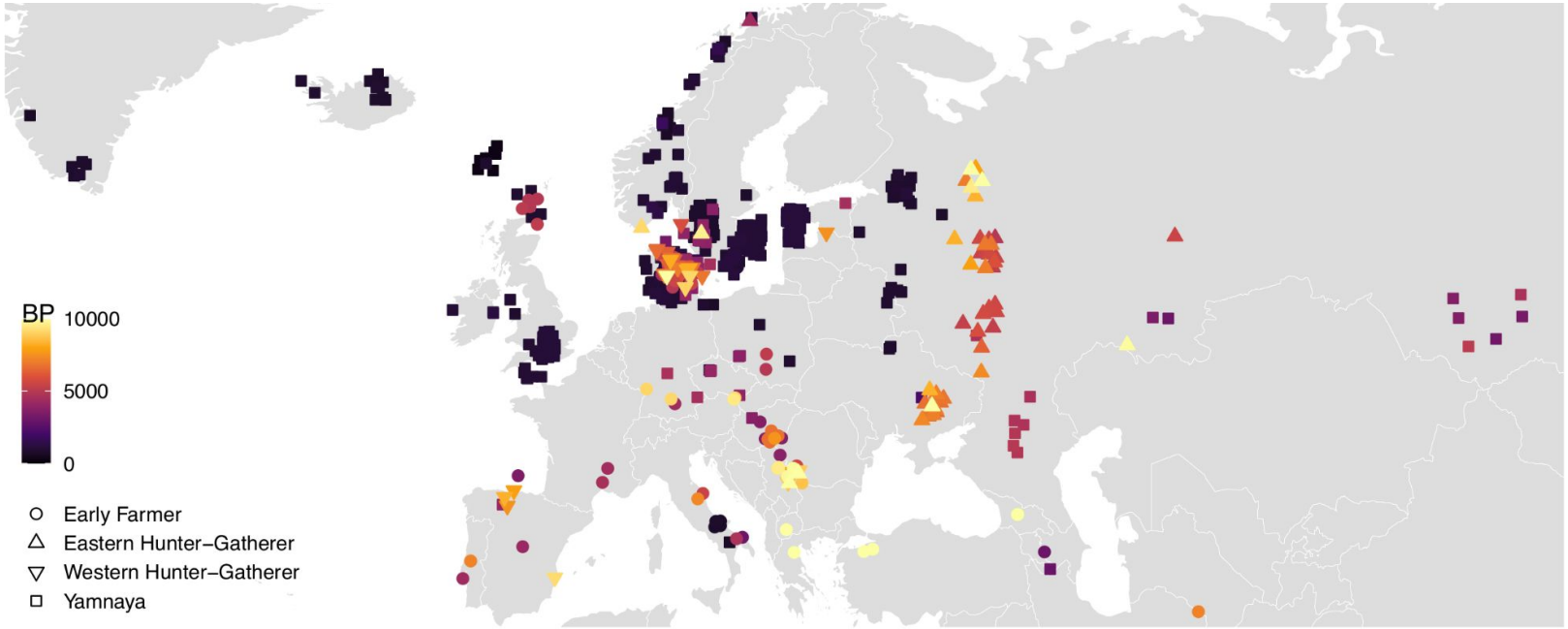
Agriculture enriched (LONG!)



Fishing/Hunting/Pastoralism Enriched (SHORT!)

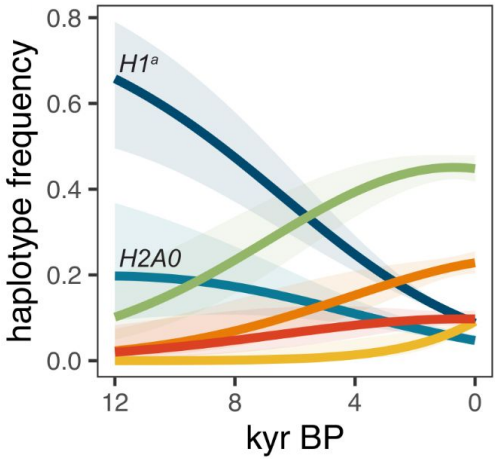


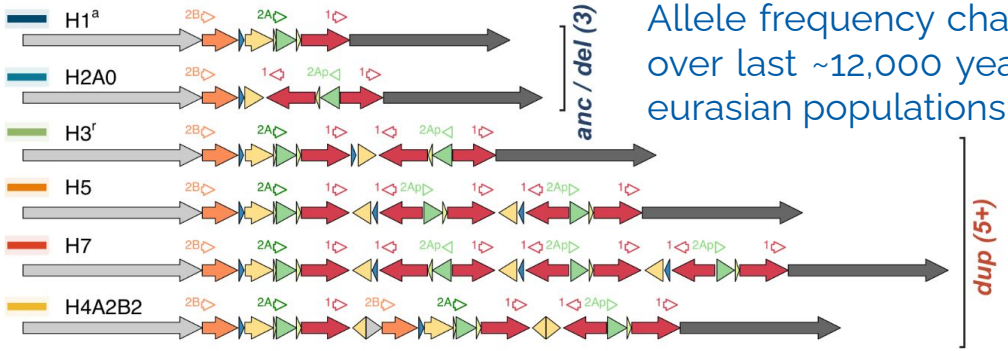
Recent evolution from 533 ancient European genomes



Data from Allentoft et al. 2024, Nature

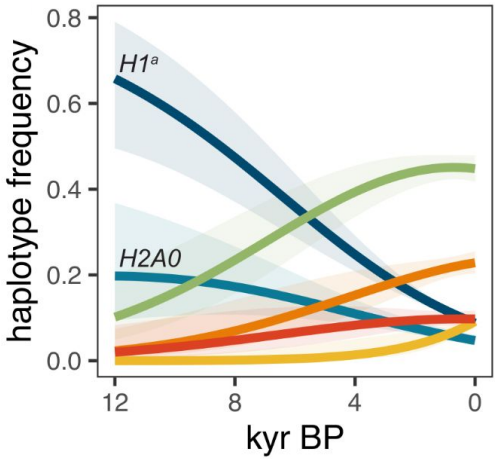
Evidence for selection of high-copy amylase haplotypes

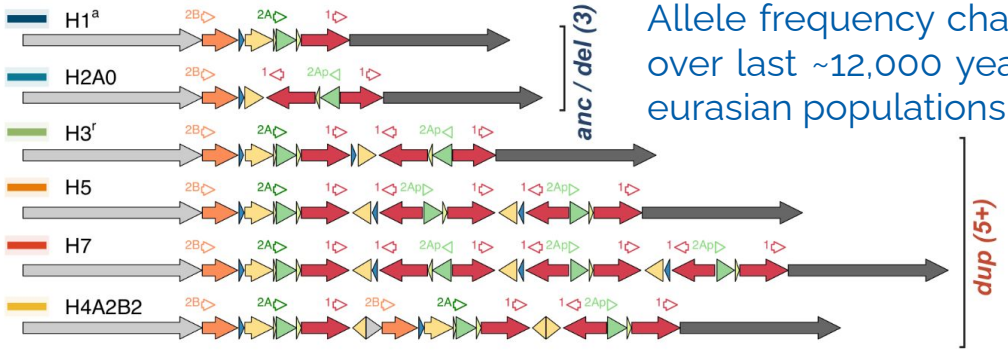




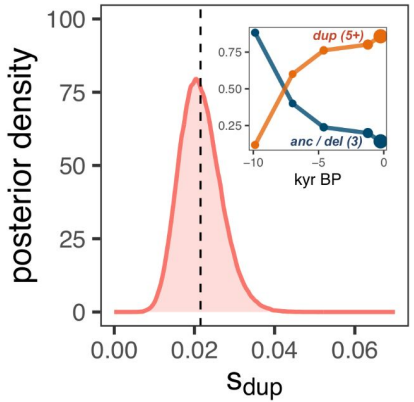
Allele frequency changes over last ~12,000 years in eurasian populations.

Evidence for selection of high-copy amylase haplotypes





Allele frequency changes over last ~12,000 years in eurasian populations.



Selection coefficient of 0.02 is equivalent to selection at lactase (Mathieson, 2020).

Panpublication

Article

Recurrent evolution and selection shape structural diversity at the amylase locus

https://doi.org/10.1038/s41586-024-07911-1

Received: 29 November 2023

Accepted: 6 August 2024

Published online: 04 September 2024

Davide Bolognini[1,10], Alma Halgren[2,10], Runyang Nicolas Lou[2,10], Alessandro Raveane[1,10], Joana L. Rocha[2,10], Andrea Guarracino[3], Nicole Soranzo[1,4,5,6,7], Chen-Shan Chin[8], Erik Garrison[3] ✉ & Peter H. Sudmant[2,9] ✉

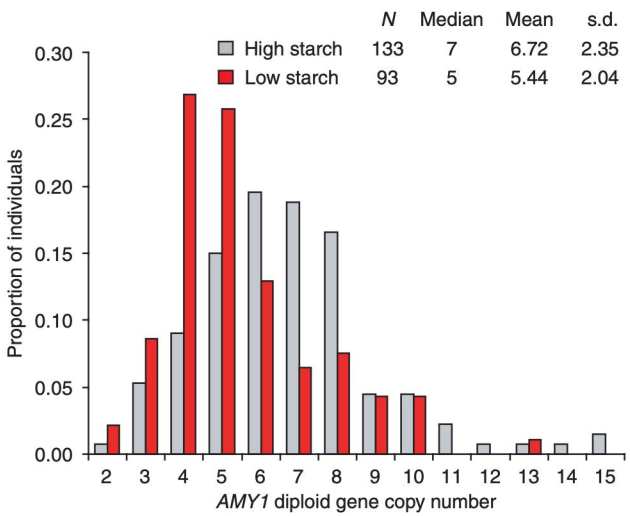
Integrative Biology

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER.

HUMAN TECHNOPOLE

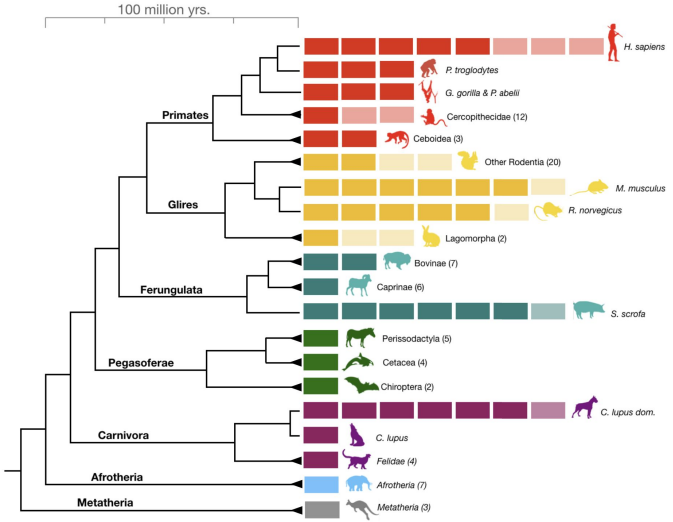
THE VALLEE FOUNDATION

Diet correlates with amylase copy number



Across humans

Perry et al., 2007

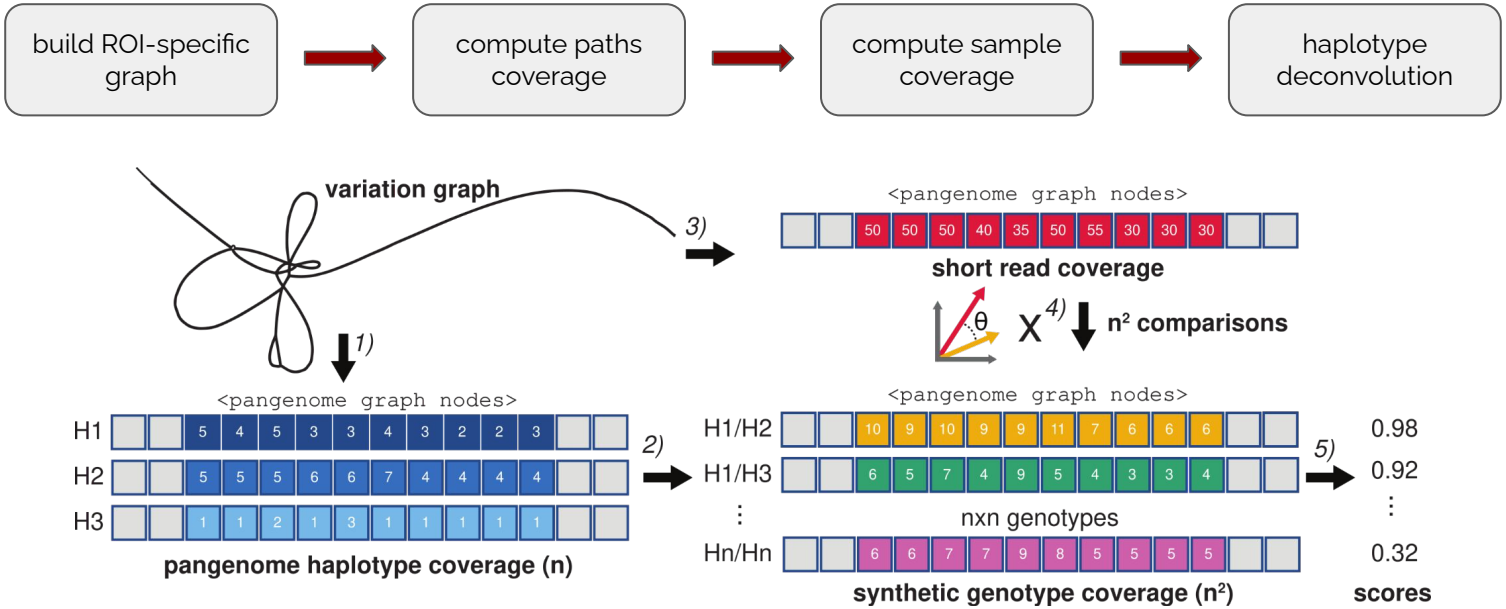


Across mammals

Pajic et al., 2019

No evidence of recent selection

Graph-based genotyping with COSIGT works!



- 100% accurate when mapping individuals with both long reads and short reads (n=35)
- ~95% accurate by mendelian inheritance patterns (561 short read trios)
- 95-99% accurate by copy number prediction compared to read depth (n=3102)

THE UNIVERSITY OF TENNESSEE HEALTH SCIENCE CENTER

https://github.com/davidebolo1993/cosigt

COSIGT
COSINE SIMILARITY-BASED GENOTYPER

Allele frequencies of complex structural haplotypes

Pangenome-based haplotype deconvolution for ~8000 short read haplotypes