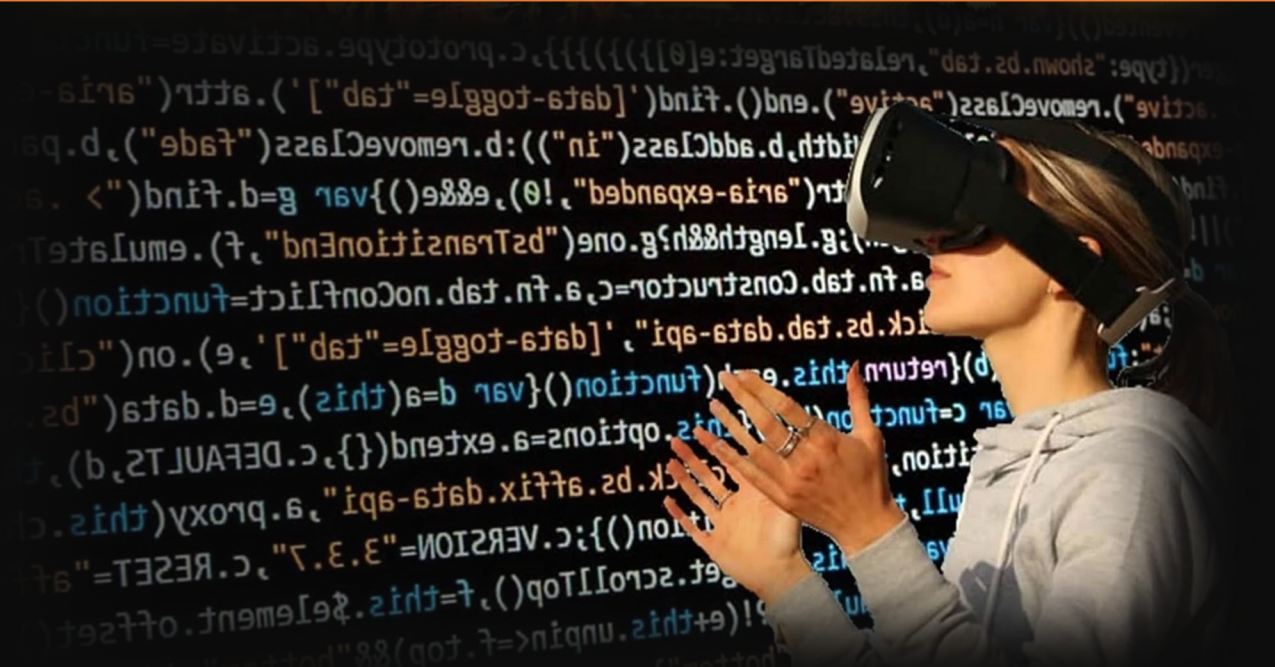
From prehistoric time to the futuristic age of AI

A human genetic perspective

Linh Tran

University of Arizona



Presenter Financial Disclosure

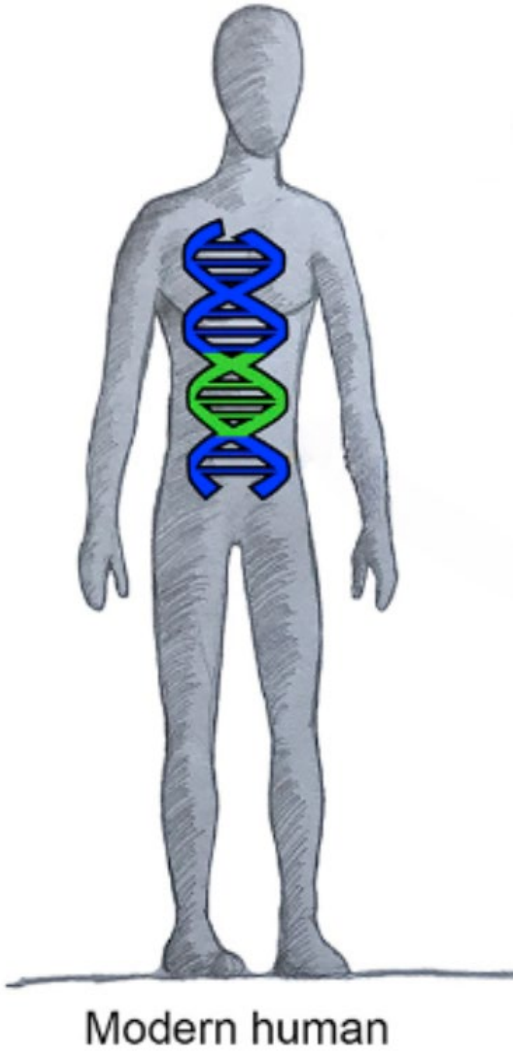
I do not have any relationships to report within the last 24 months with ACCME defined ineligible companies.

Unlabeled/Investigational Uses

I will not be discussing unlabeled/investigational uses of medical devices or pharmaceuticals during this presentation.

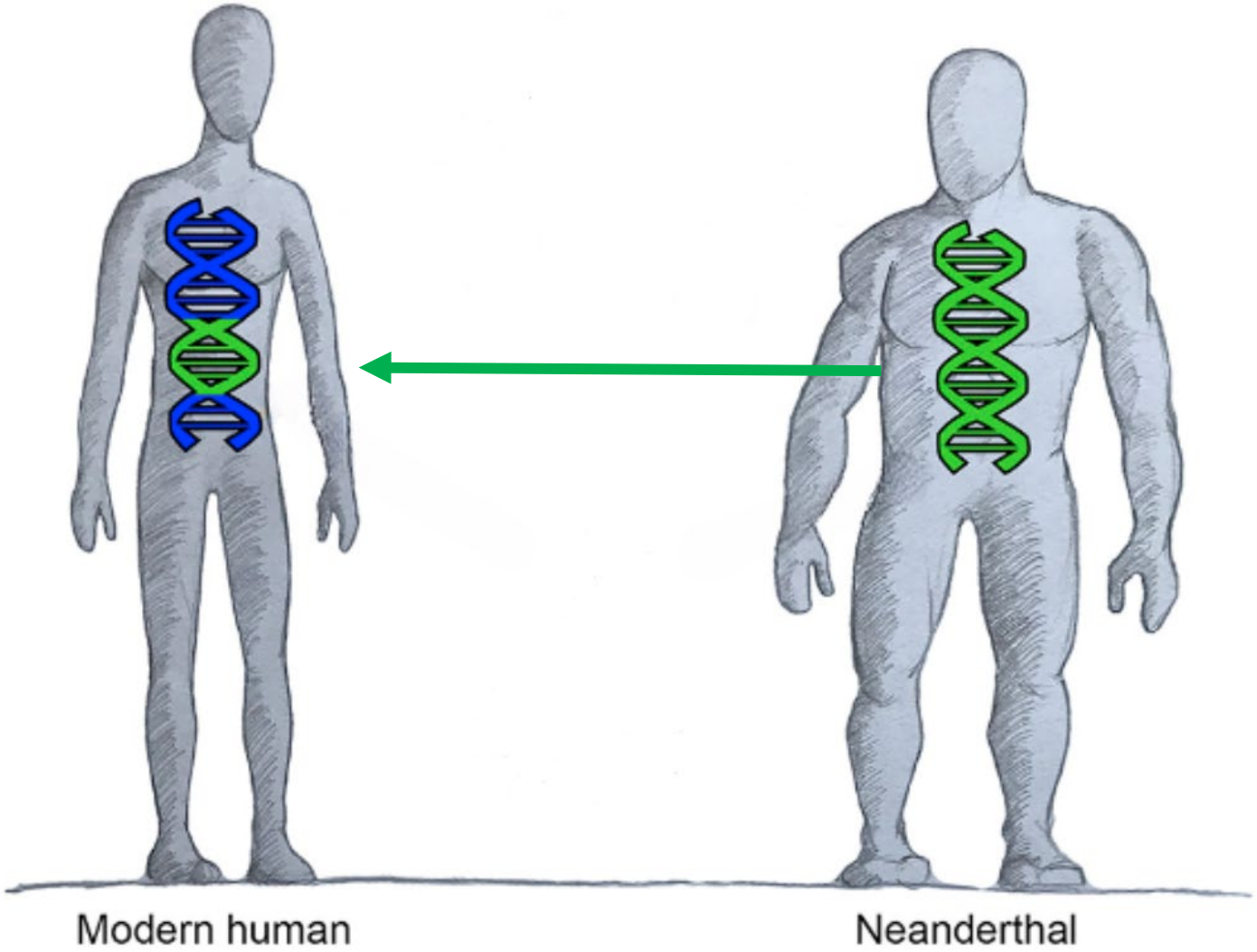
Study the past
if you would define
the future.

Confucius



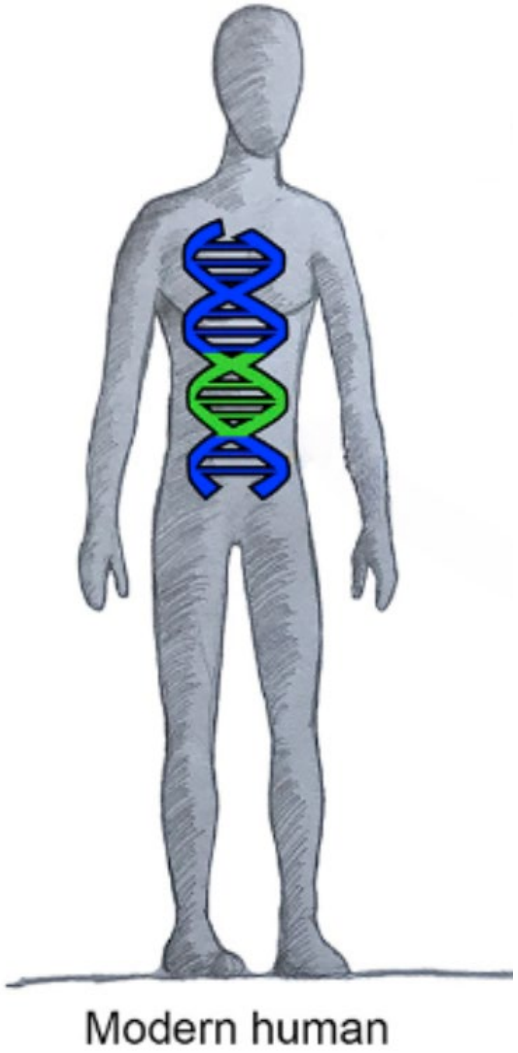
Adapted from Enard & Petrov (2018) Cell

Up to 4% of Eurasian descent genomes are of Neanderthals origin

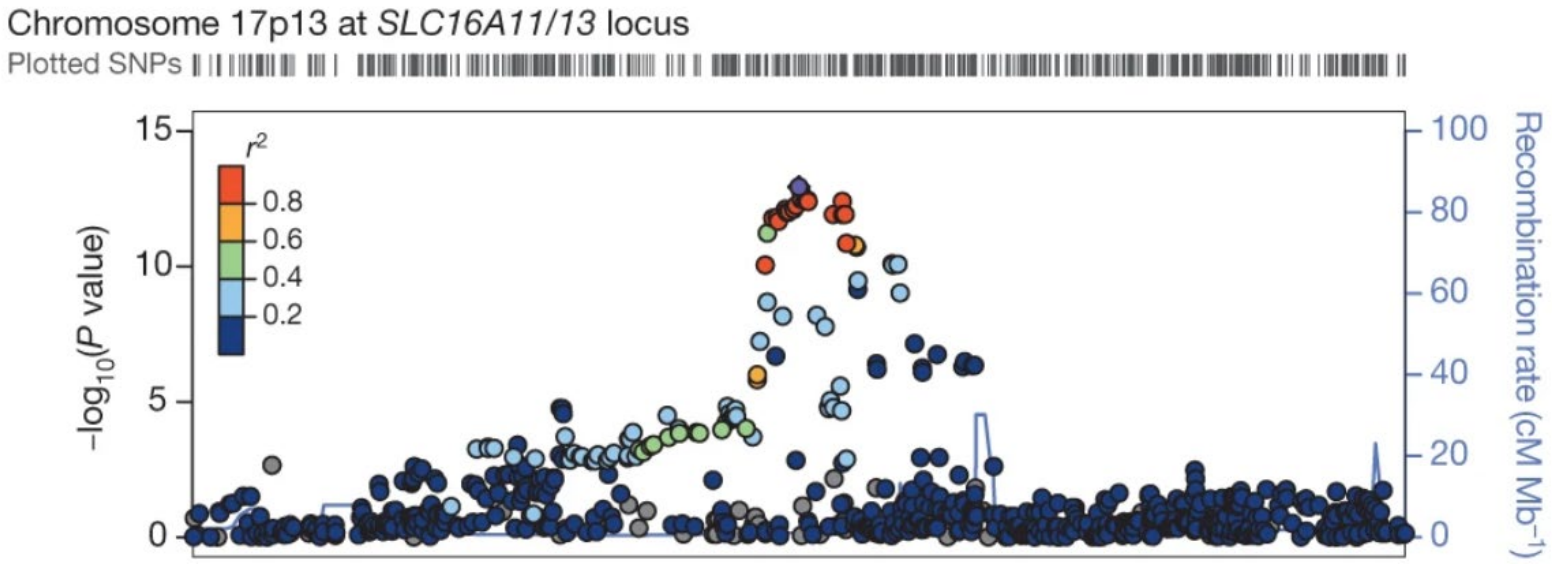


Adapted from Enard & Petrov (2018) Cell

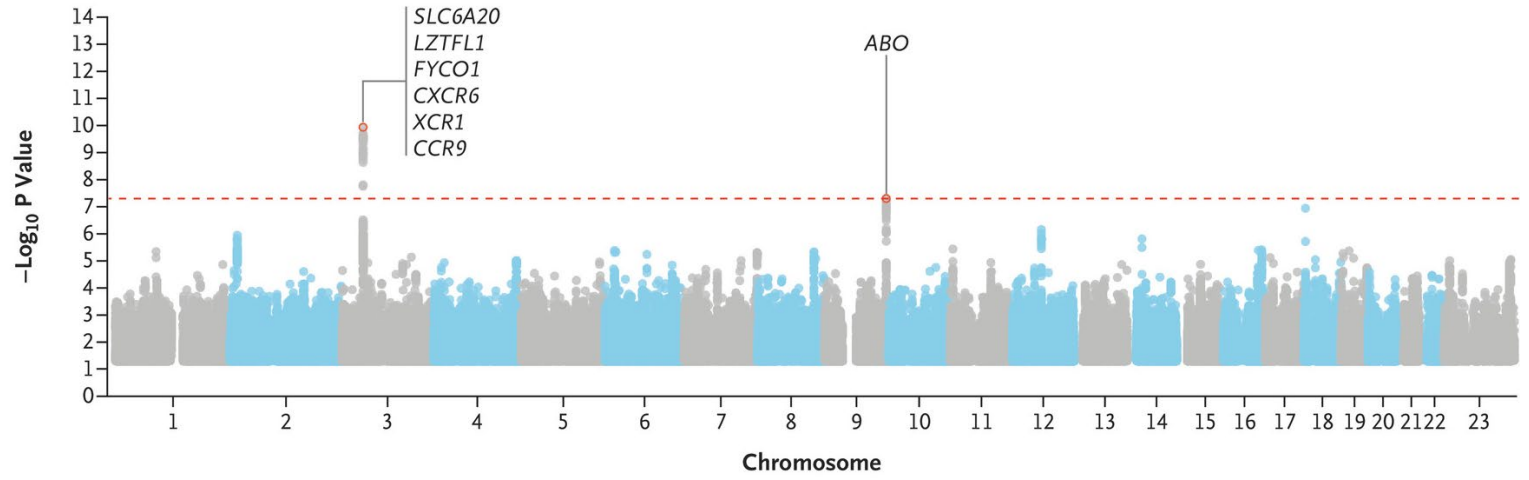
Some Neanderthals-originated variants are associated with disease risk



Adapted from Enard & Petrov (2018) Cell

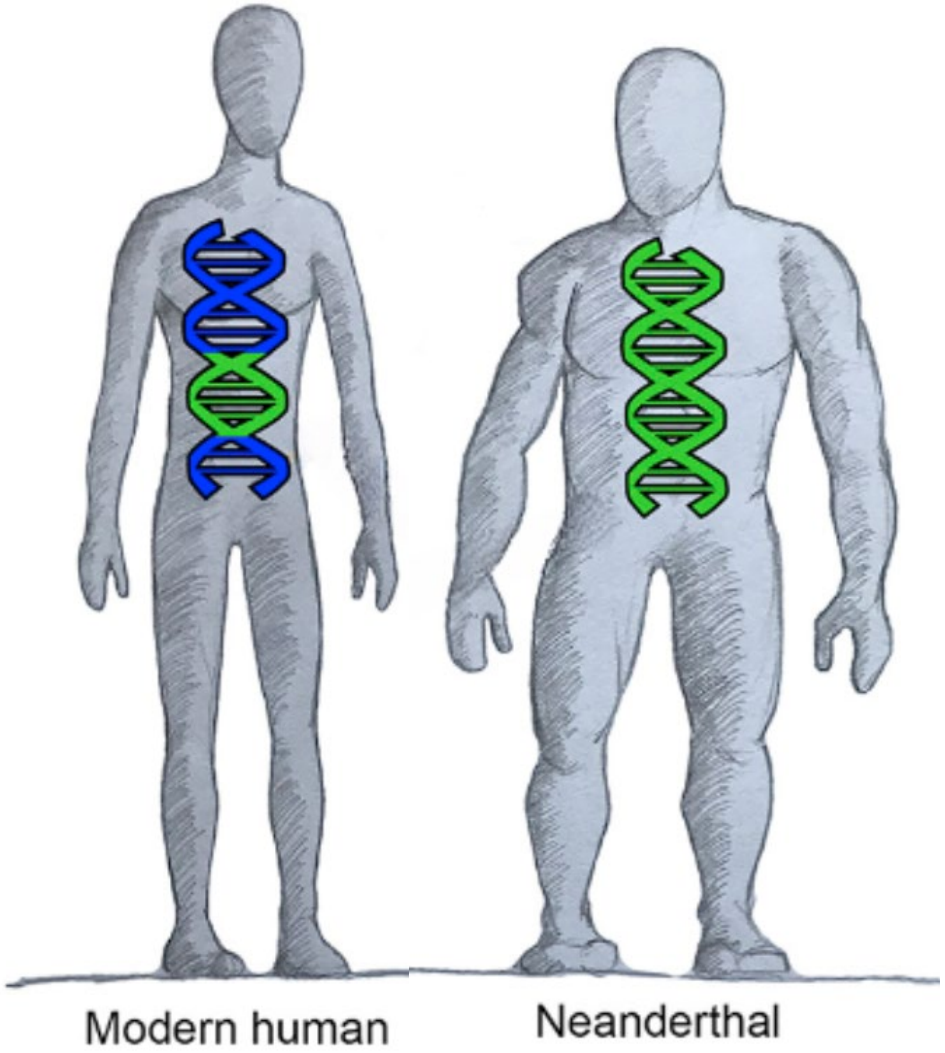


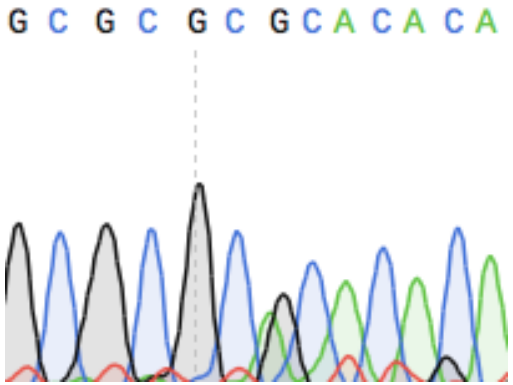
The SIGMA Type 2 Diabetes Consortium (2014) Nature



Severe Covid-19 GWAS group (2020) New England J Med

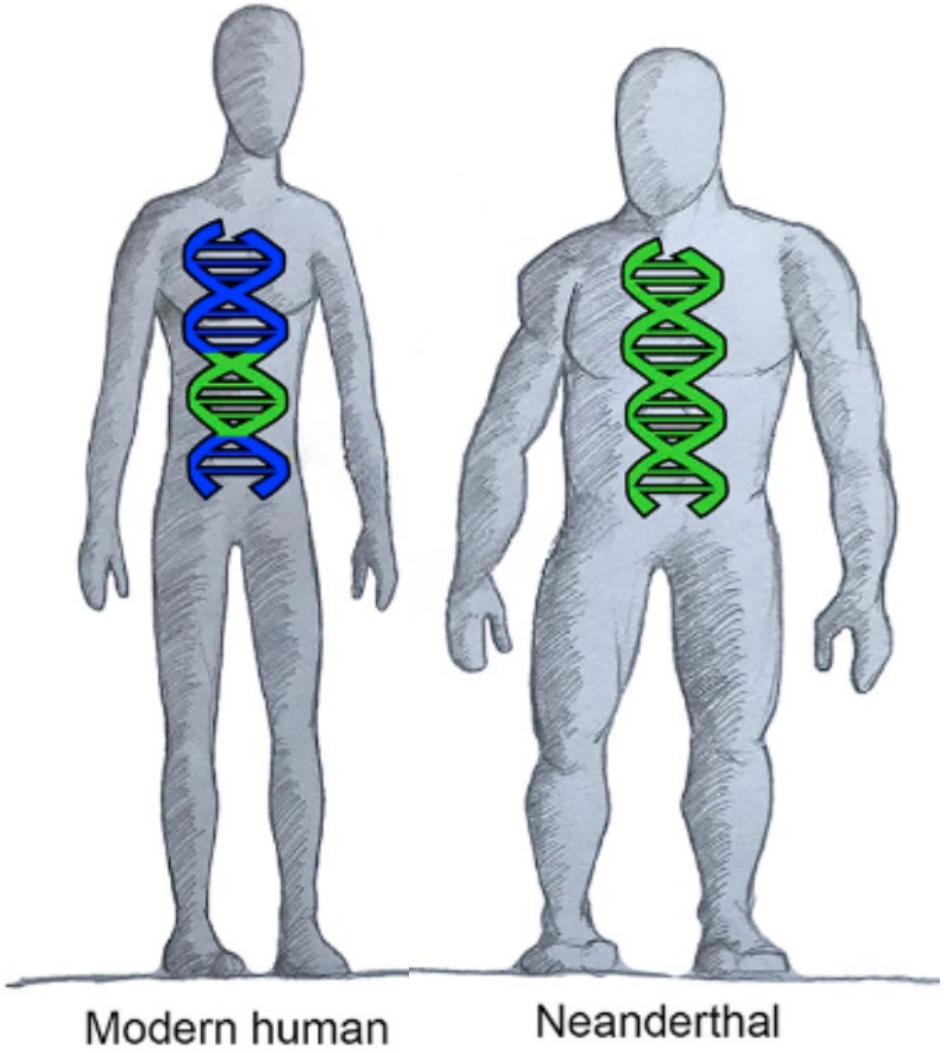
The past, present, and future within our genomes

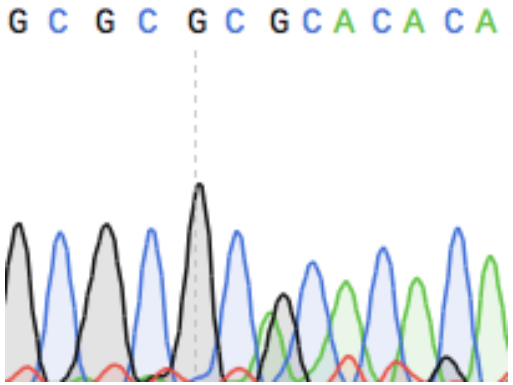


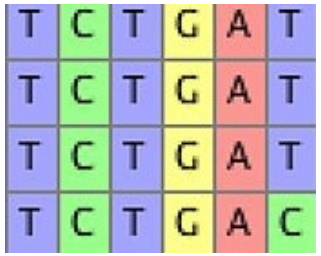


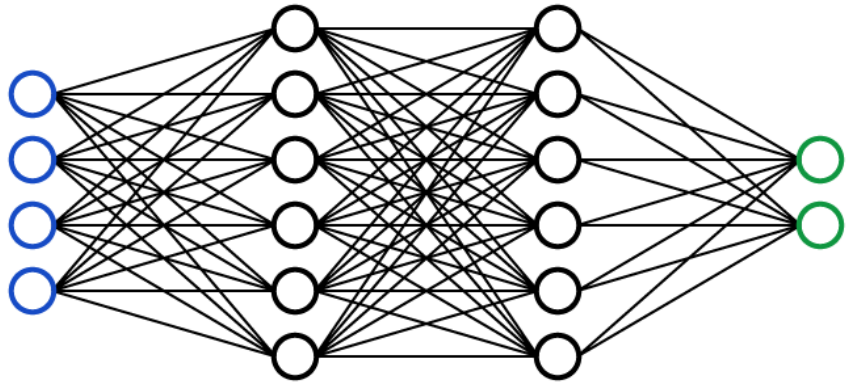
Adapted from Enard & Petrov (2018) Cell

The past, present, and future within our genomes

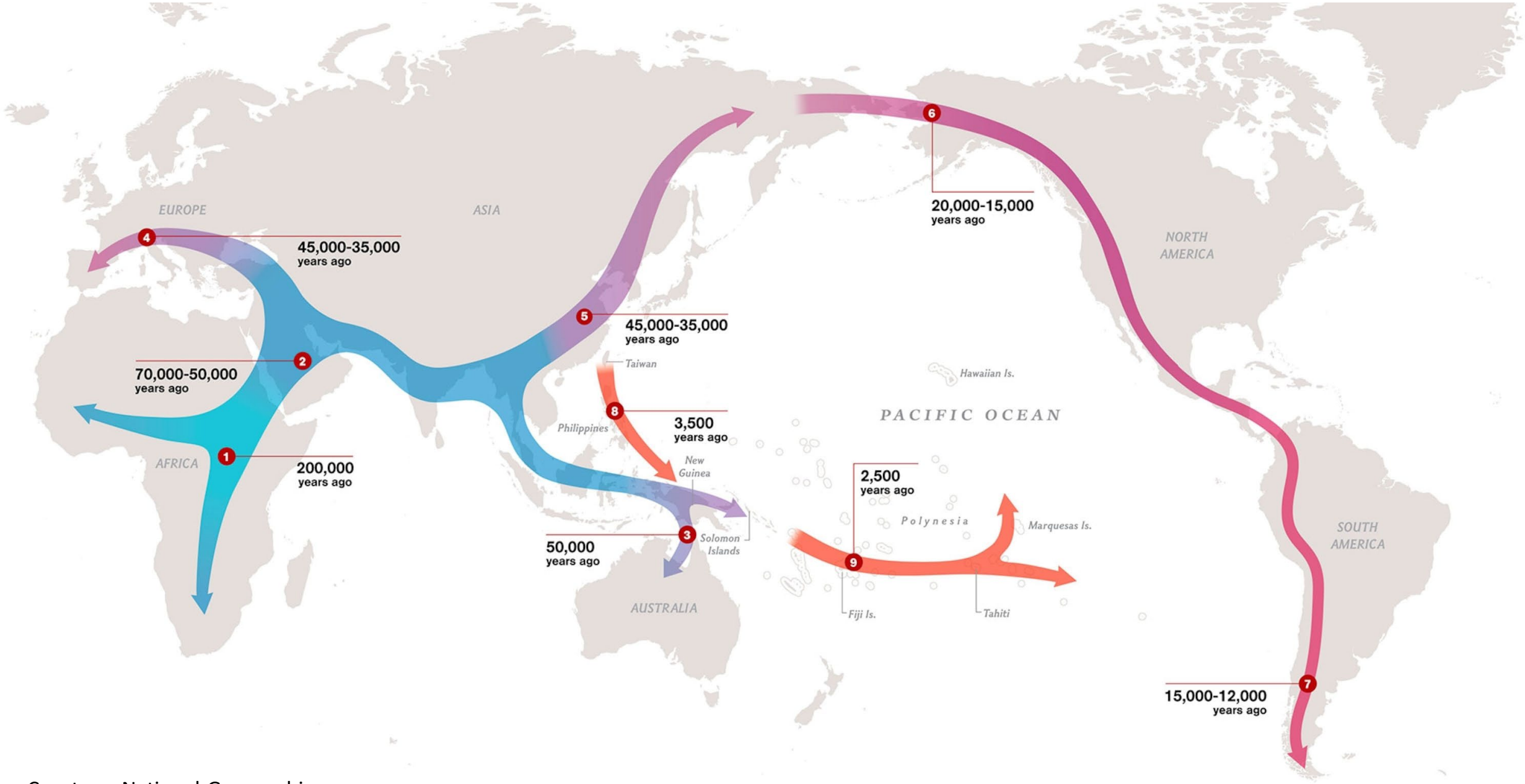








Adapted from Enard & Petrov (2018) Cell



Courtesy: National Geographic

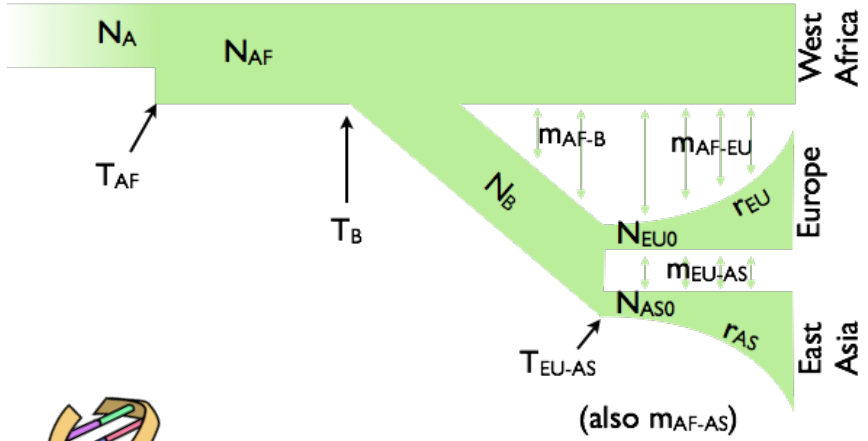
PLOS GENETICS

1,538 Citation | 59,629 View

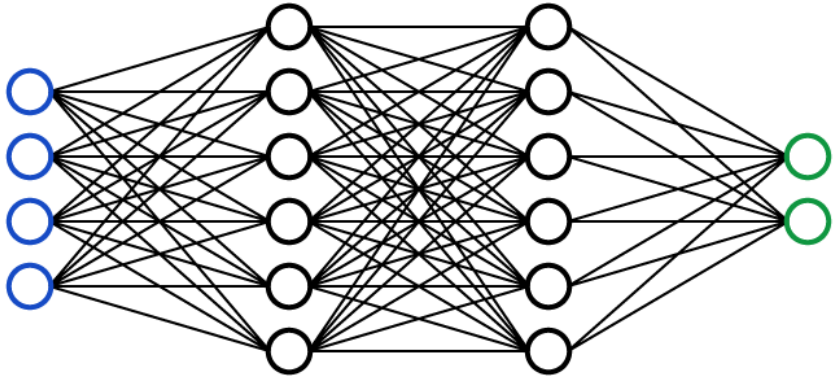
Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data

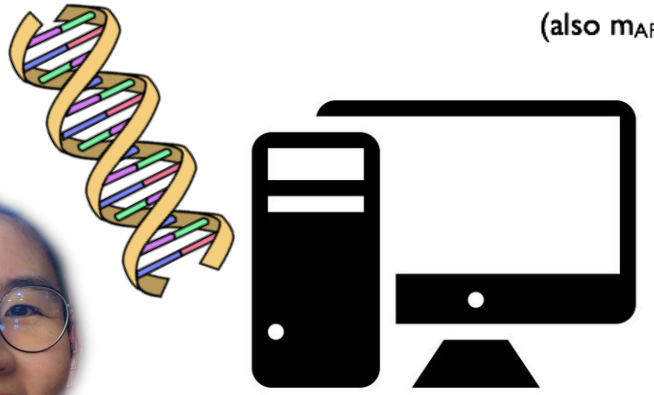
Ryan N. Gutenkunst, Ryan D. Hernandez, Scott H. Williamson, Carlos D. Bustamante

Published: October 23, 2009 • https://doi.org/10.1371/journal.pgen.1000695

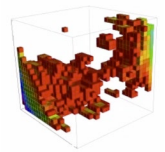


2009





2020



dadi: diffusion approximation for demographic inference

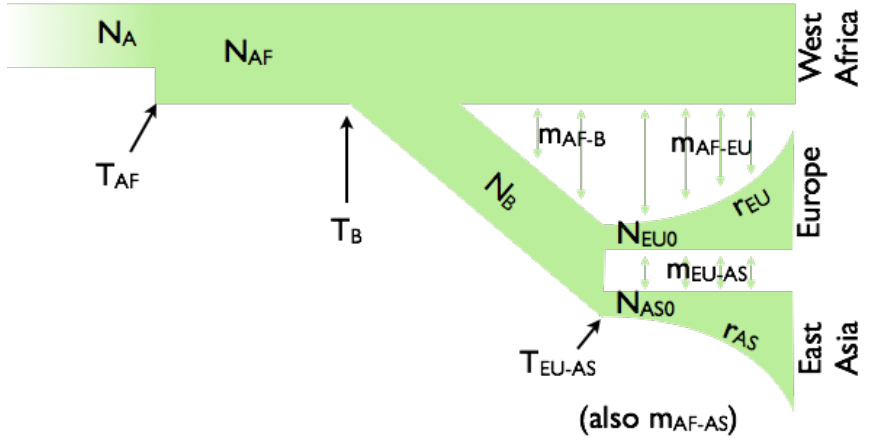
PLOS GENETICS

1,538 Citation | 59,629 View

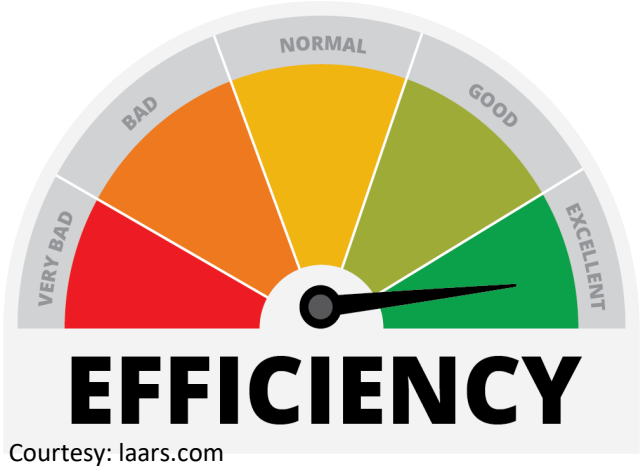
Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data

Ryan N. Gutenkunst, Ryan D. Hernandez, Scott H. Williamson, Carlos D. Bustamante

Published: October 23, 2009 • https://doi.org/10.1371/journal.pgen.1000695

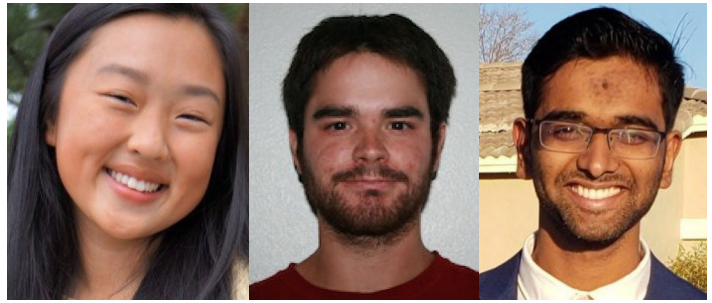


2009

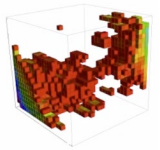


Courtesy: laars.com

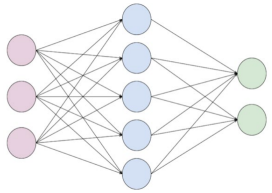
Connie Sun, Travis Struck, Mathews Sajan



2020



dadi: diffusion approximation for demographic inference



donni: demography optimization via neural network inference
Tran et al. (BioRxiv 2023)

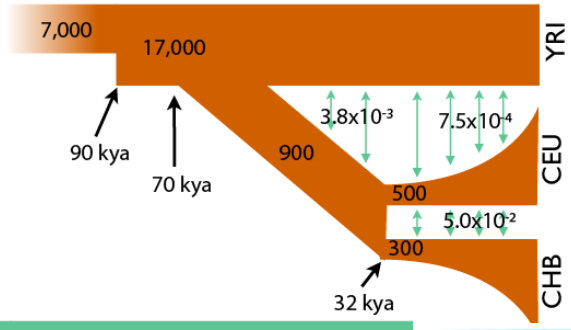
Data (SNP)

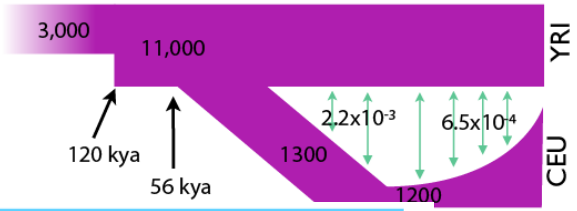


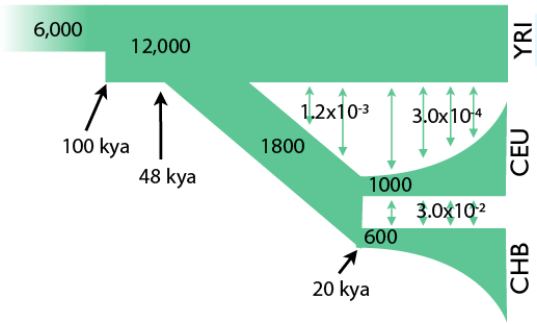
TGGTCACTCTTAT
TGGTCACTCTTAT
TGGTCACTCTTAA
TGCTCATTCTTAT
TGCTCATTCTTAT
TGCTCACTCTTAA

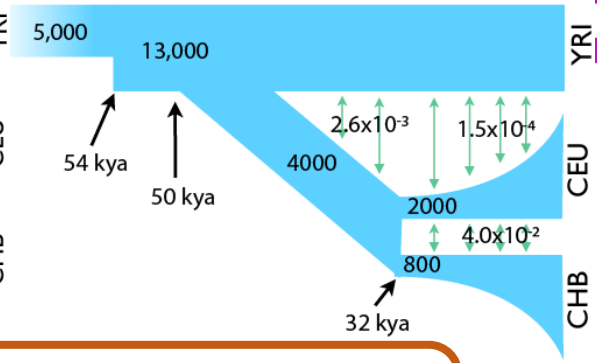
Likelihood Optimization

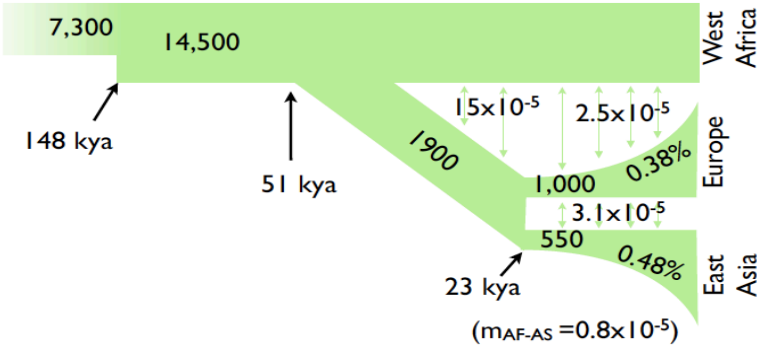
Expensive

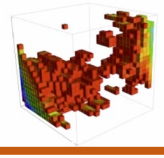




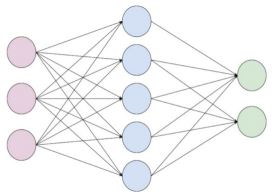








dadi: diffusion approximation for demographic inference



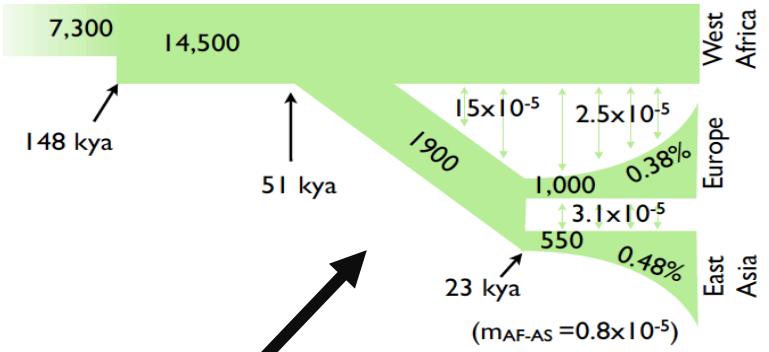
donni: demography optimization via neural network inference
Tran et al. (BioRxiv 2023)

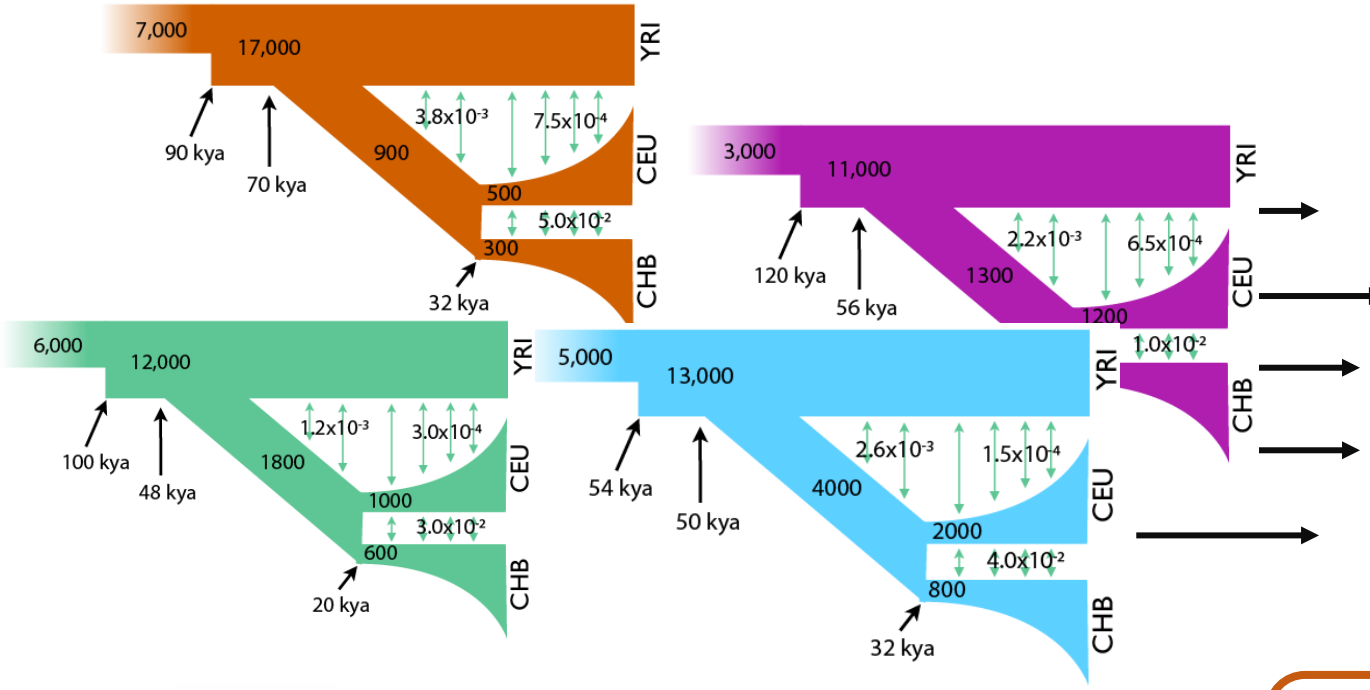
Data (SNP)

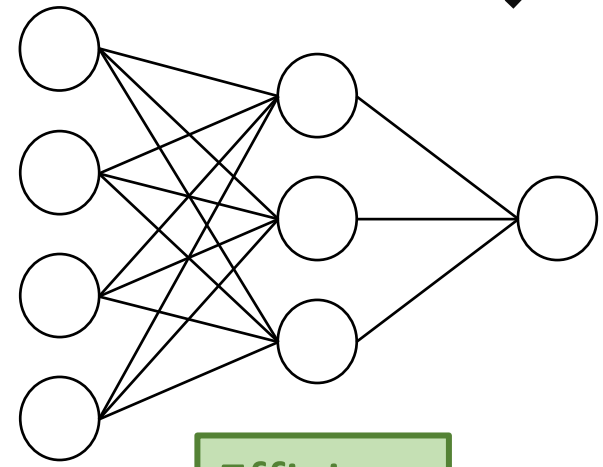


TGGTCACTCTTAT
TGGTCACTCTTAT
TGGTCACTCTTAA
TGCTCATTCTTAT
TGCTCATTCTTAT
TGCTCACTCTTAA

Likelihood Optimization

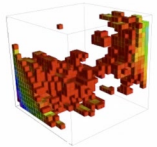




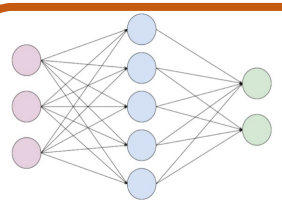


Multilayer Perceptron Regressor

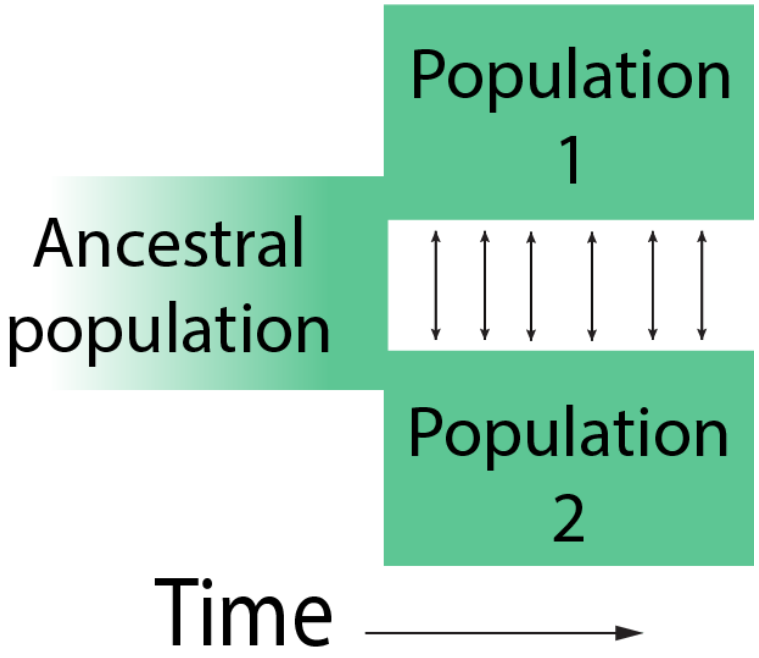
Efficient



dadi: diffusion approximation for demographic inference

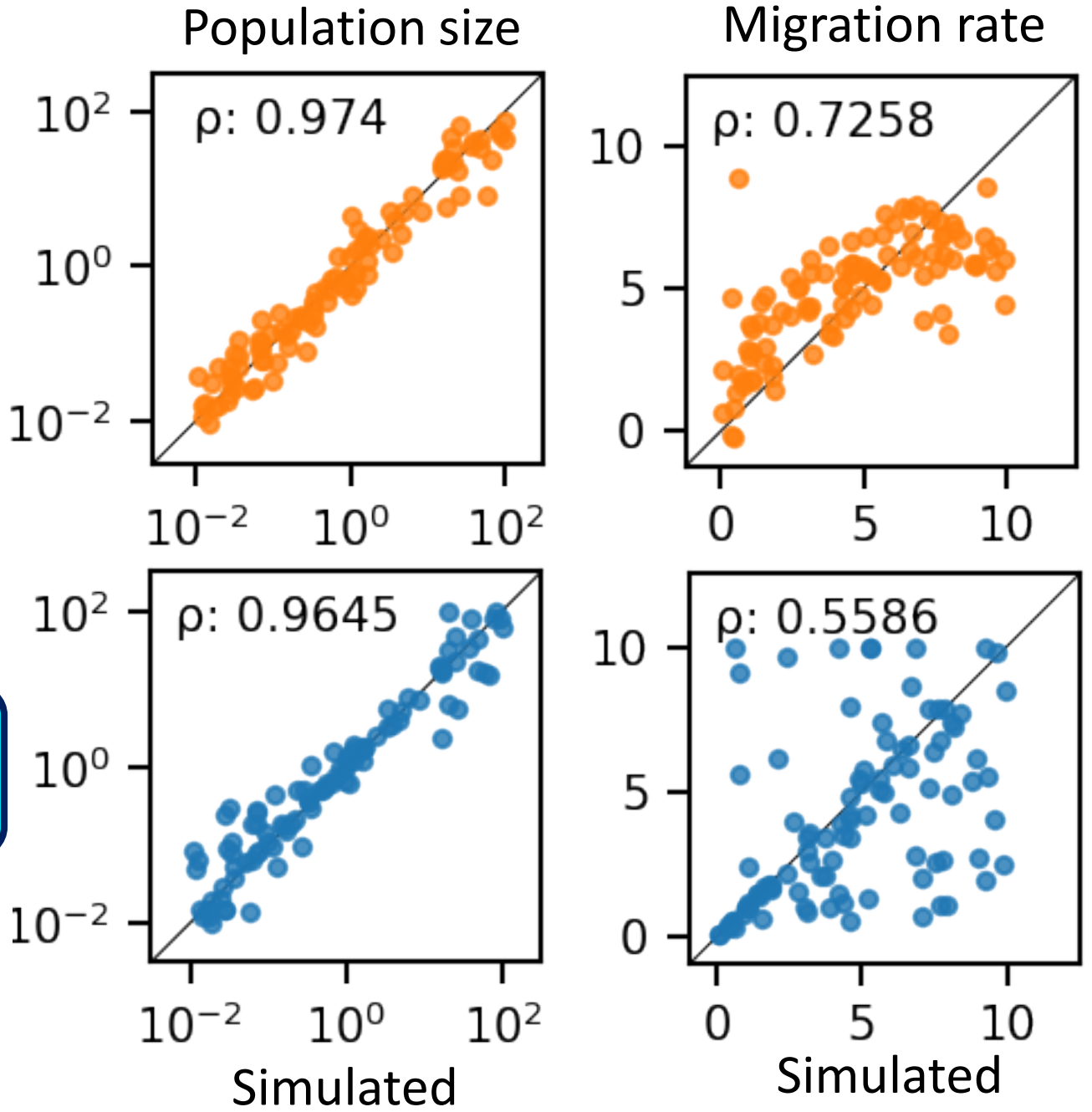


donni: demography optimization via neural network inference
Tran et al. (BioRxiv 2023)



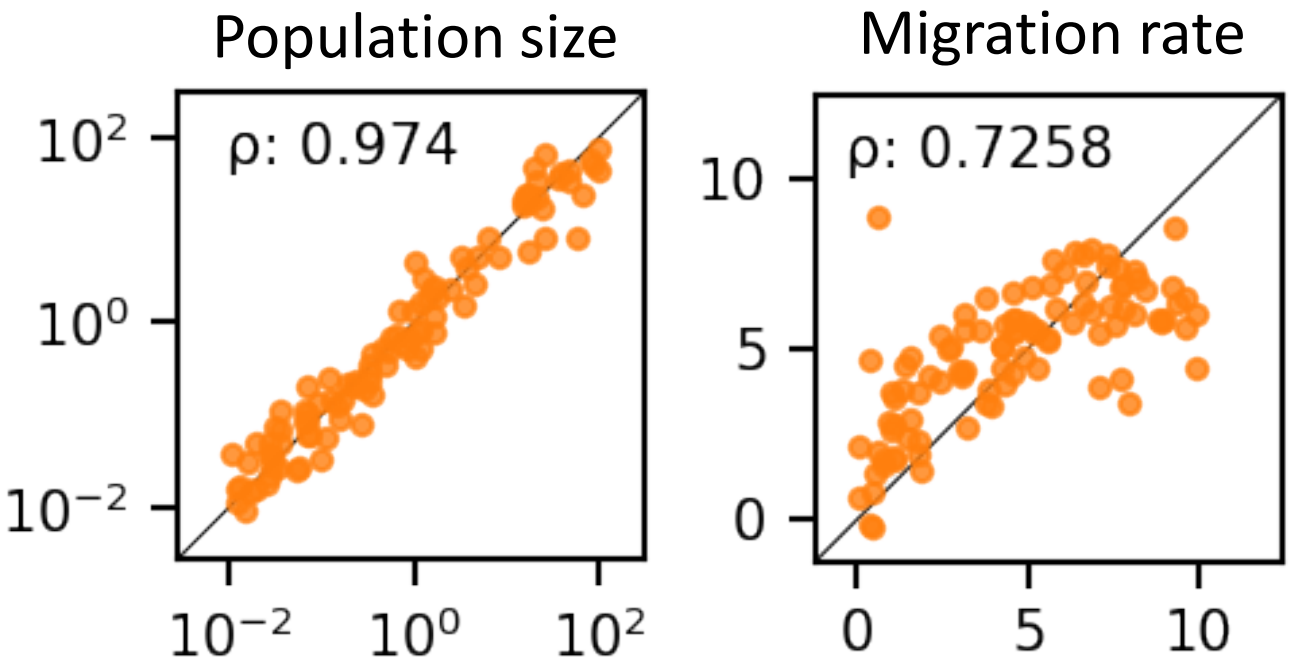
Donni inferred

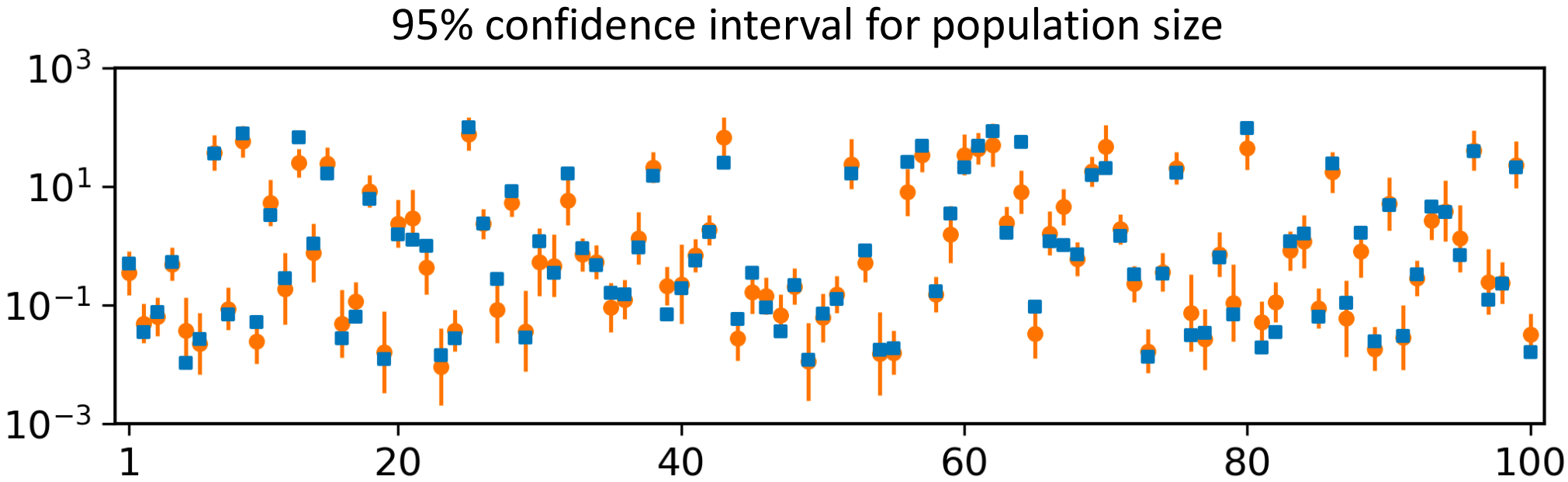
Dadi inferred



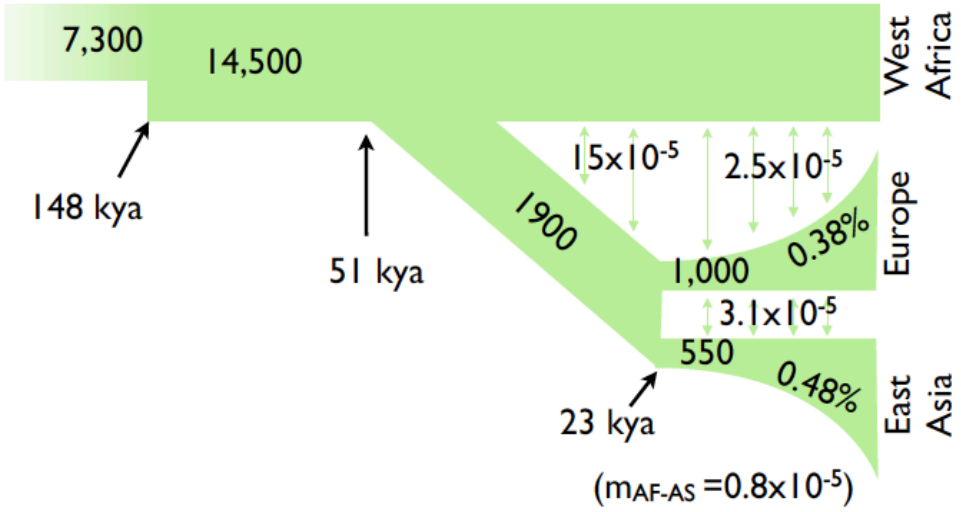
Tran et al. (2023) BioRxiv, in revision

Tensorflow Keras
Mean-Variance Estimation Net



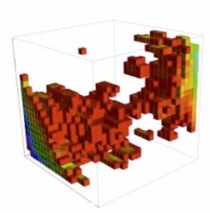


Tran et al. (2023) BioRxiv, in revision



Distribution of fitness effects (DFE)

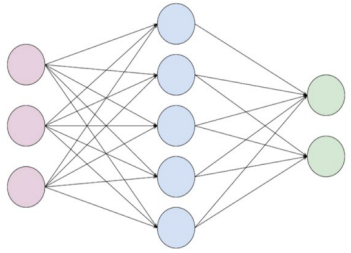
2009 2017 2020

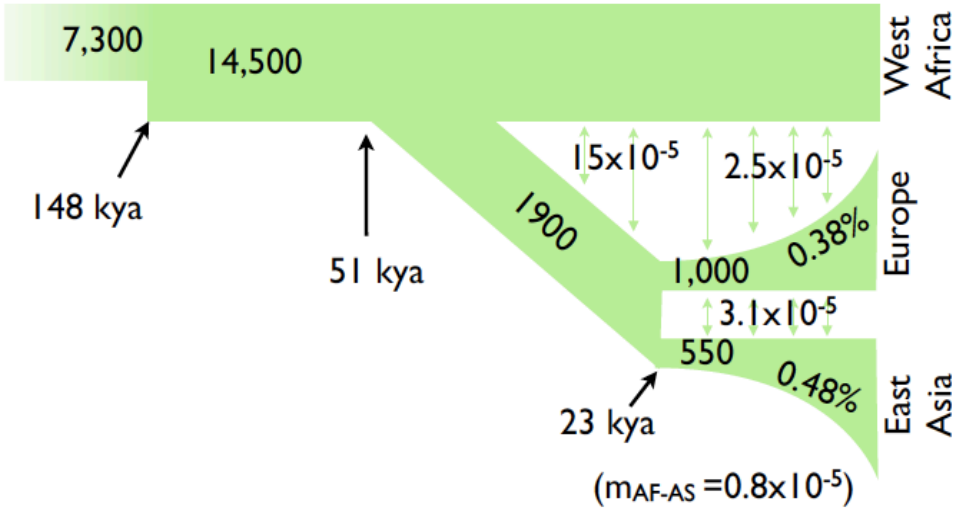


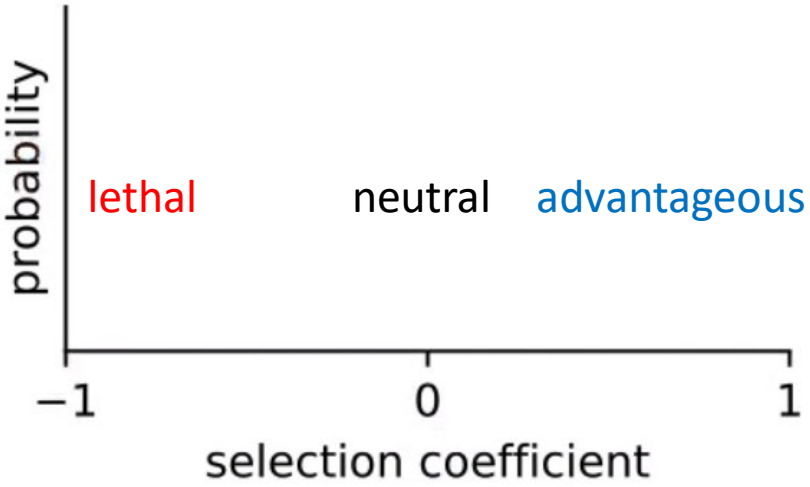
dadi

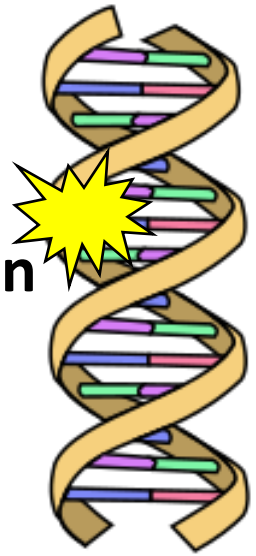
fit-dadi
Kim et al.
(Genetics)

donni



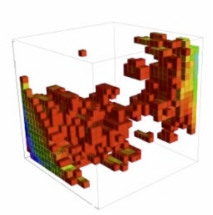






Distribution of fitness effects (DFE)

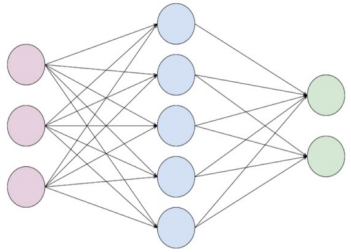
2009 2017 2020

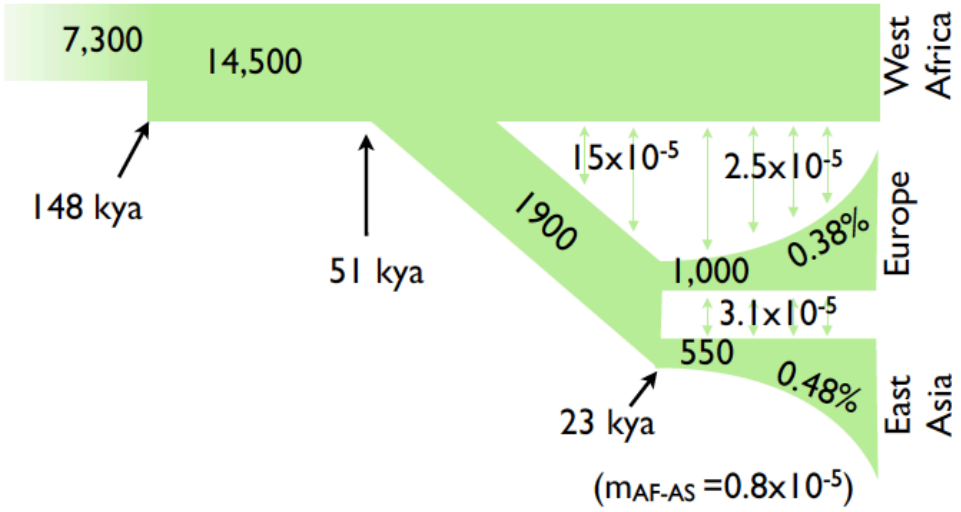


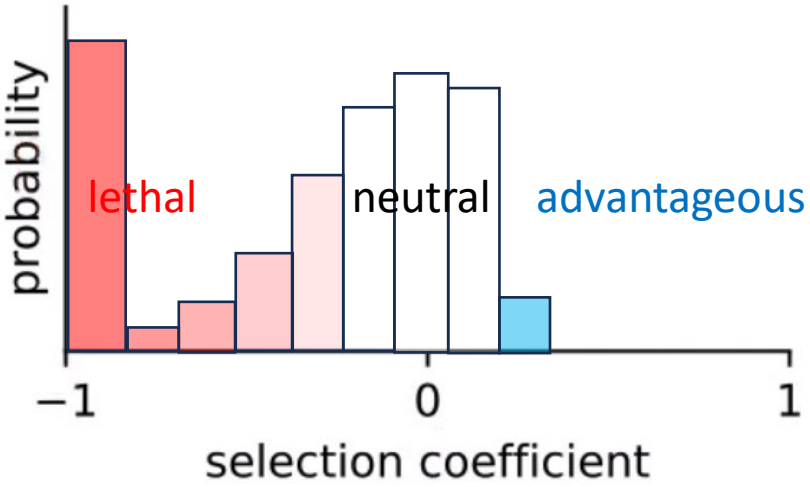
dadi

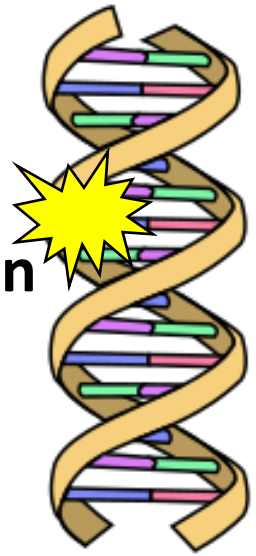
fit-dadi
Kim et al.
(Genetics)

donni



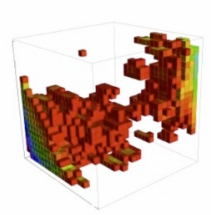






Distribution of fitness effects (DFE)

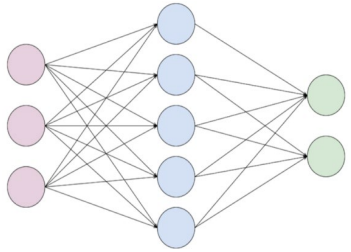
2009 2017 2020



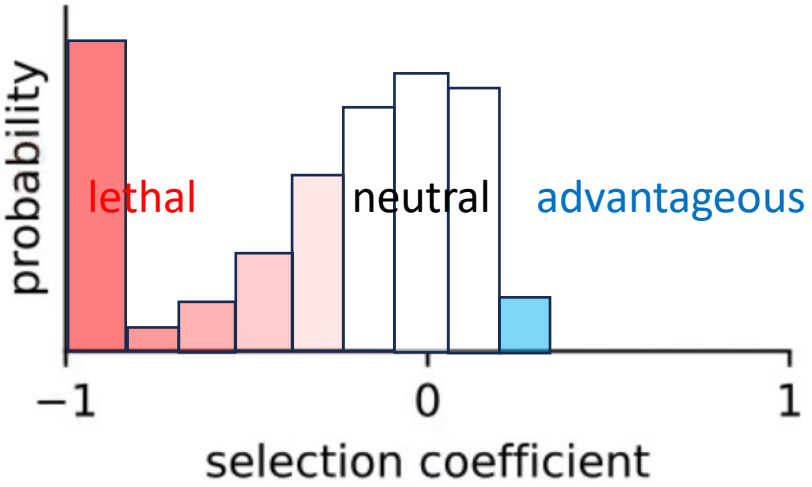
dadi

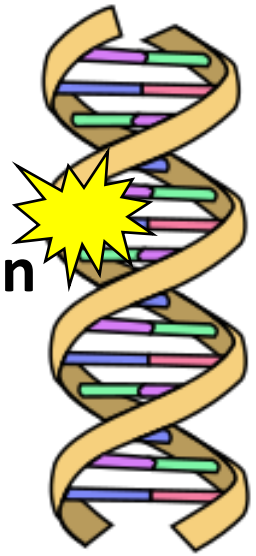
fit-dadi
Kim et al.
(Genetics)

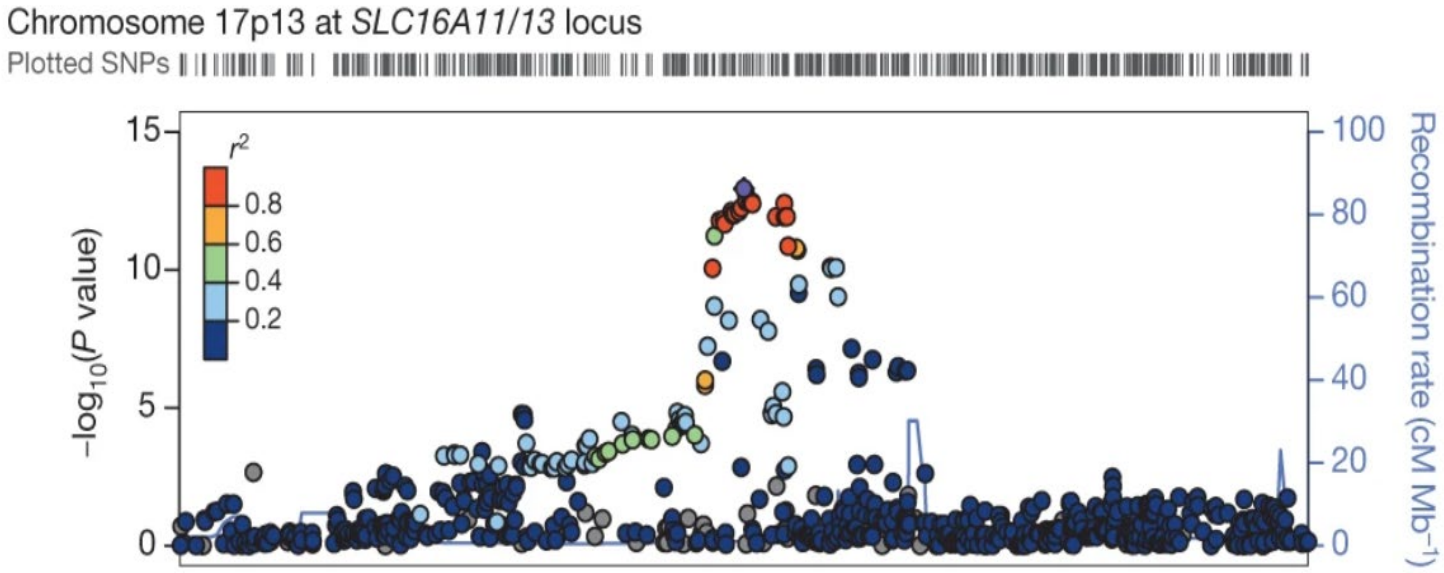
donni



- Mutations associated with diseases are likely deleterious
- The DFE quantifies the portion of deleterious mutations along with their average affect
- Significant implications for the design and interpretation of GWAS







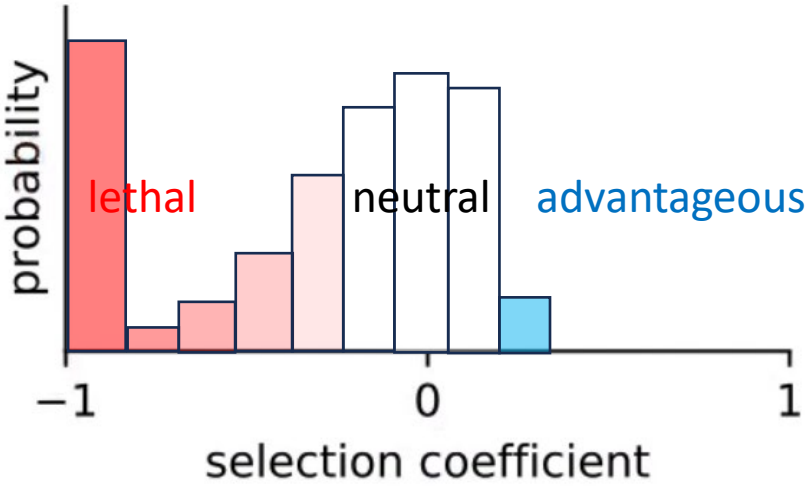
The SIGMA Type 2 Diabetes Consortium (2014) Nature

Data (SNP)

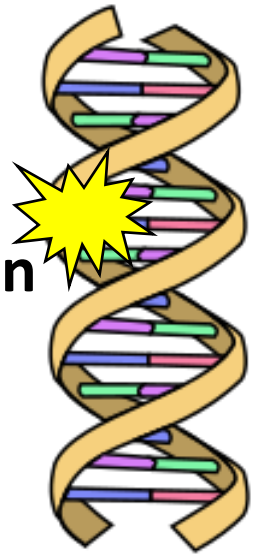

TGGTCACTCTTAT
TGGTCACTCTTAT


TGGTCACTCTTAA
TGCTCATTCTTAT


TGCTCATTCTTAT
TGCTCACTCTTAA



Mutation



Distribution of
fitness effects (DFE)

2009 2017 2020 2023

dadi
Gutenkunst et al.
(PLOS Genetics)

fit-dadi
Kim et al.
(Genetics)

donni
Tran et al.
(BioRxiv)

DFEnn

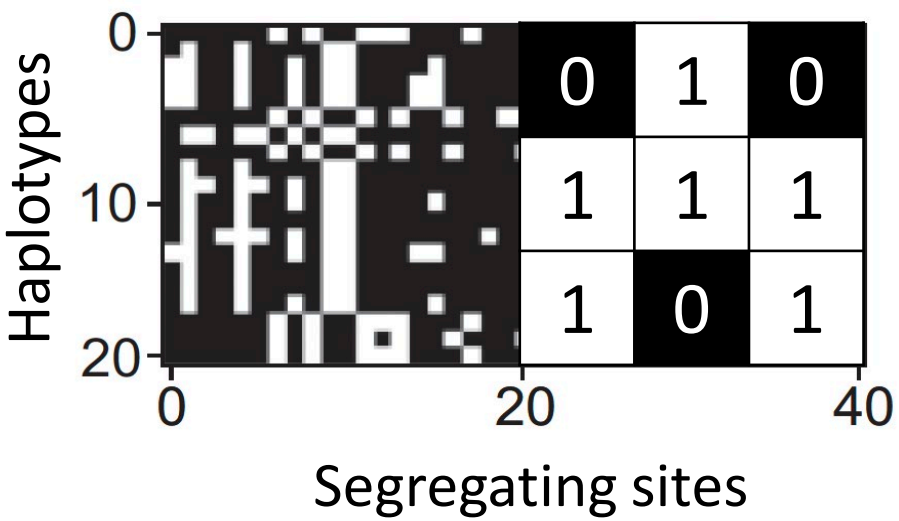
Data (SNP)


TGGTCACTCTTAT
TGGTCACTCTTAT


TGGTCACTCTTAA
TGCTCATTCTTAT


TGCTCATTCTTAT
TGCTCACTCTTAA

Flagel *et al.* MBE (2018)



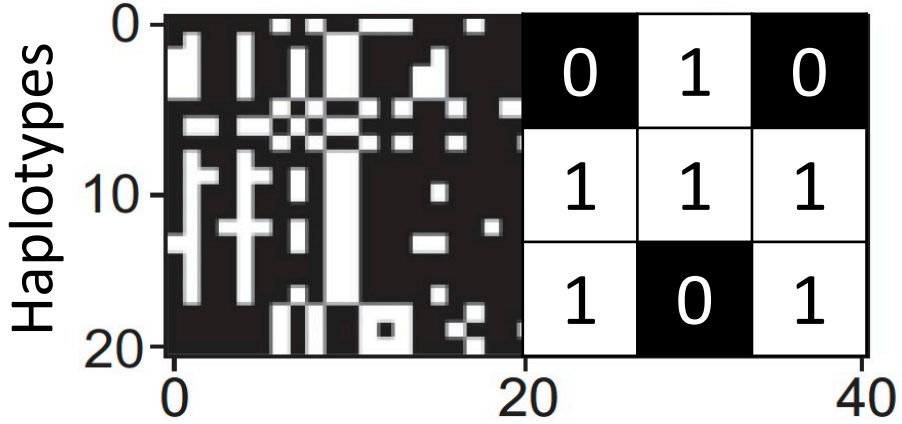
Data (SNP)

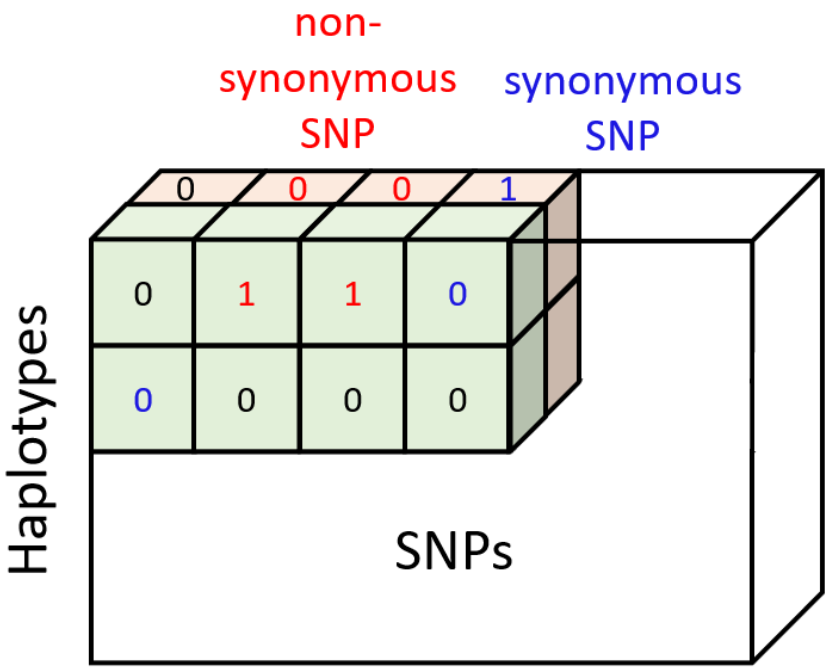

TGGTCACTCTTAT
TGGTCACTCTTAT

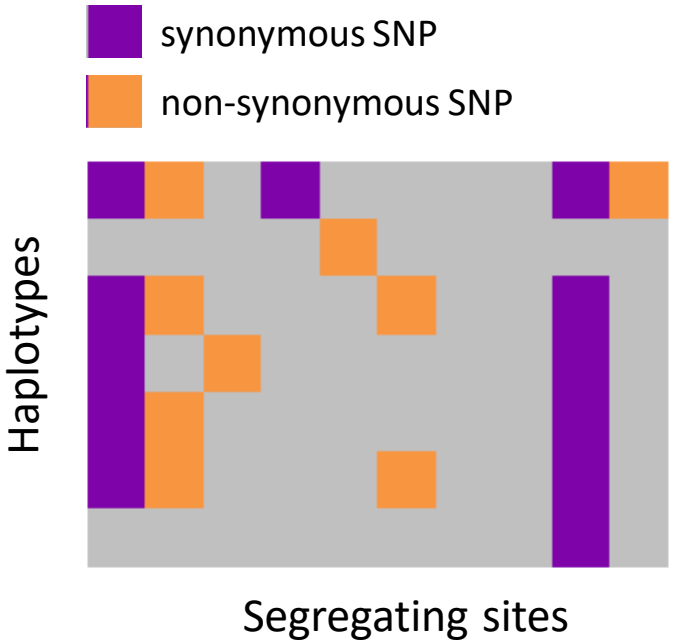

TGGTCACTCTTAA
TGCTCATTCTTAT


TGCTCATTCTTAT
TGCTCACTCTTAA

Flagel *et al.* MBE (2018)







Courtesy: The Wizard of Oz (1939)

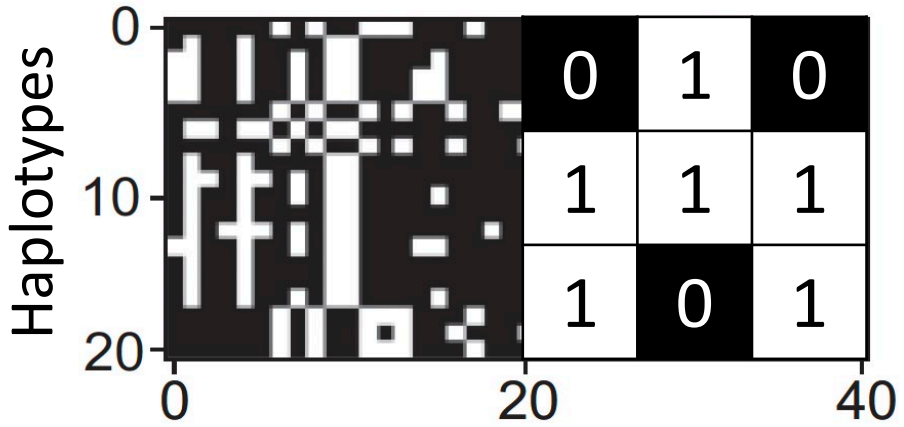
Data (SNP)

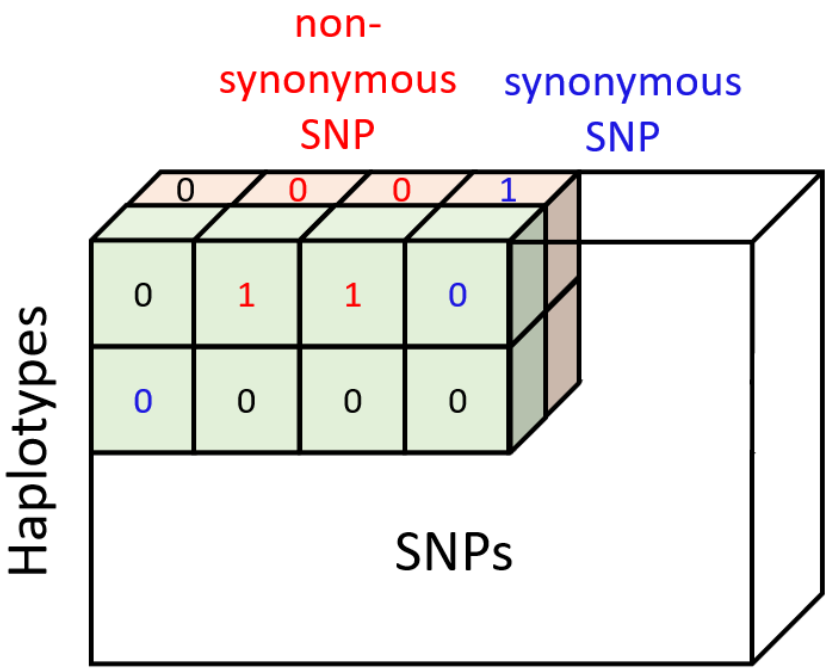

TGGTCACTCTTAT
TGGTCACTCTTAT


TGGTCACTCTTAA
TGCTCATTCTTAT


TGCTCATTCTTAT
TGCTCACTCTTAA

Flagel *et al.* MBE (2018)

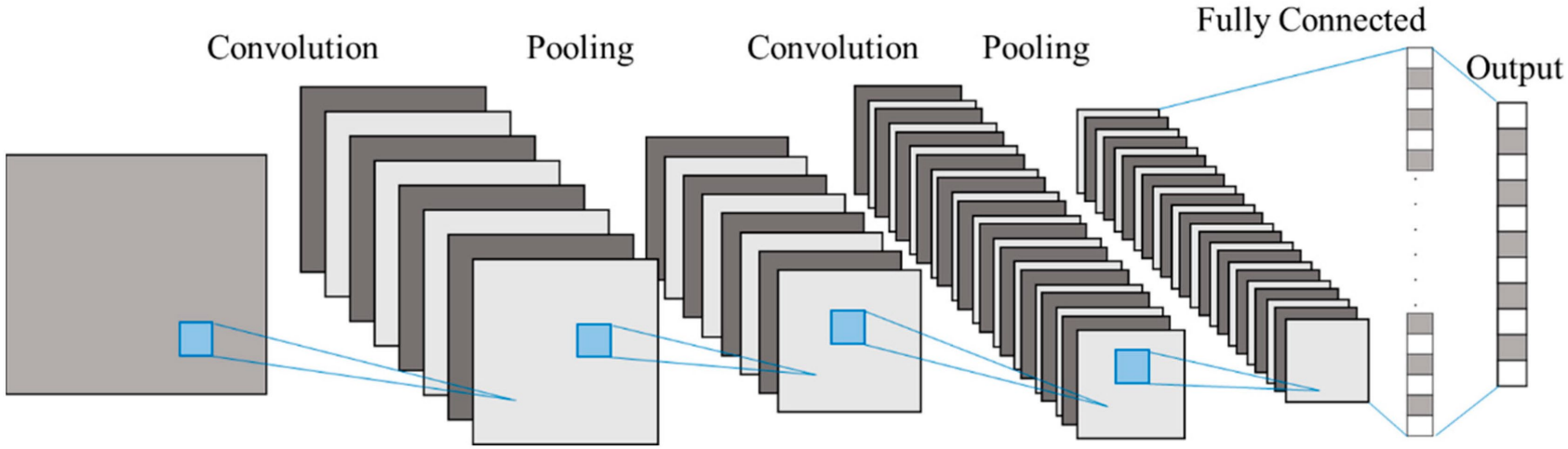




DFEnn

synonymous SNP

non-synonymous SNP



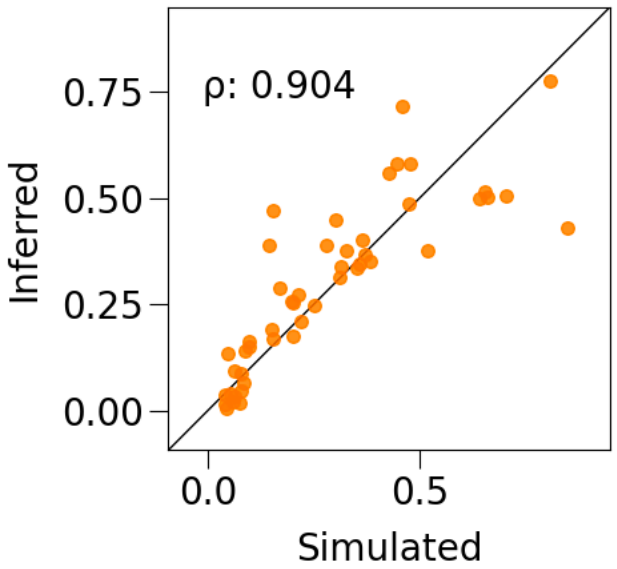
DFEnn

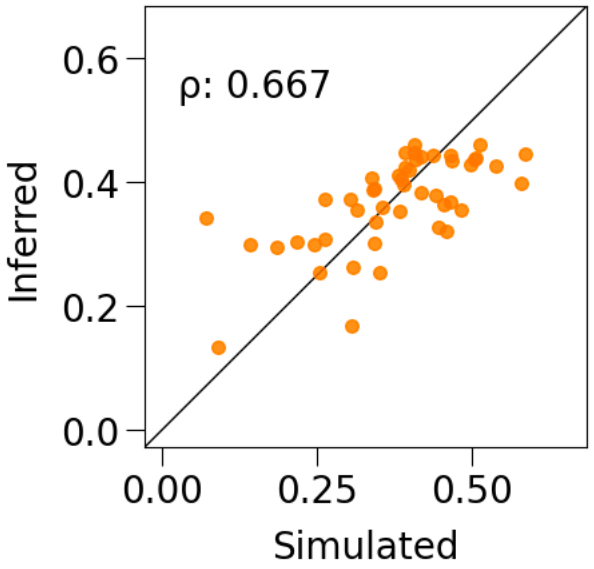
dadi

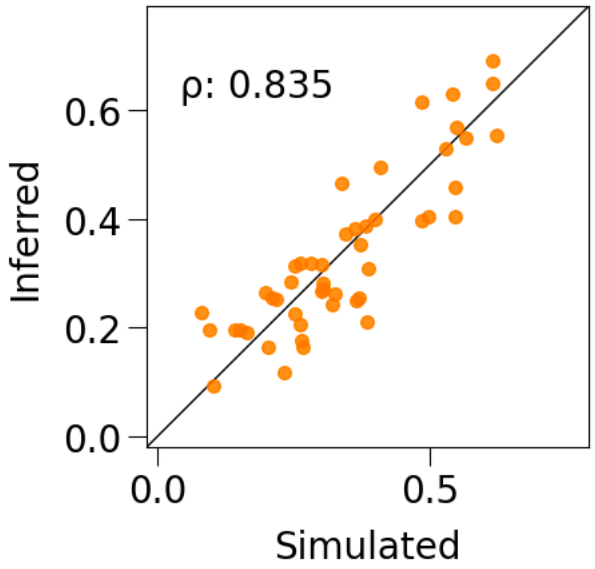
Weakly deleterious

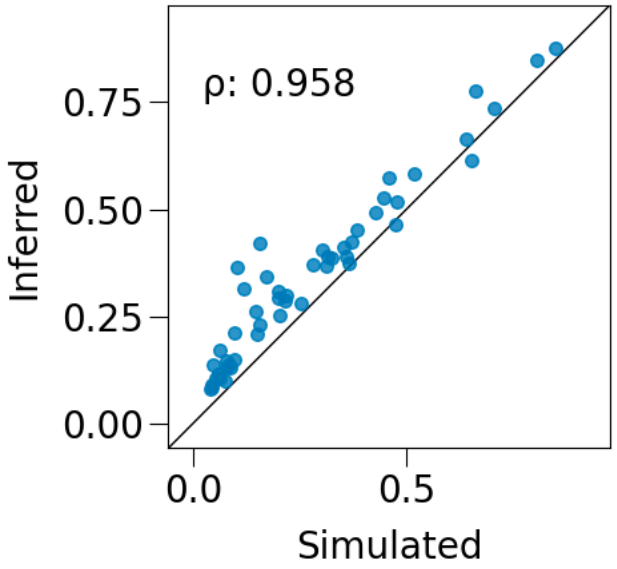
Moderately deleterious

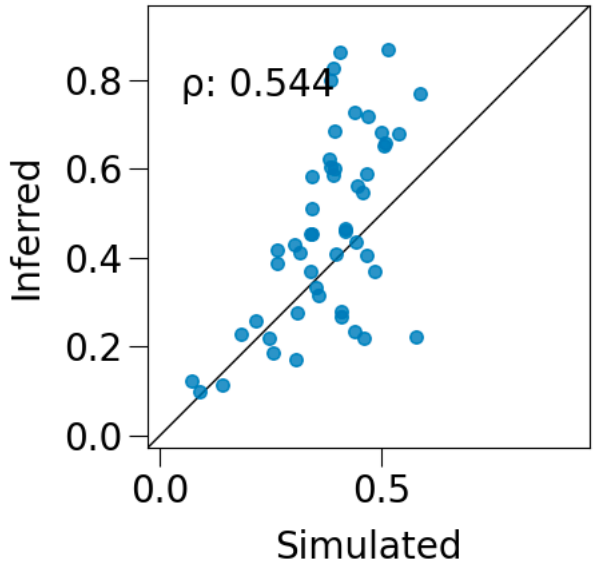
Strongly deleterious

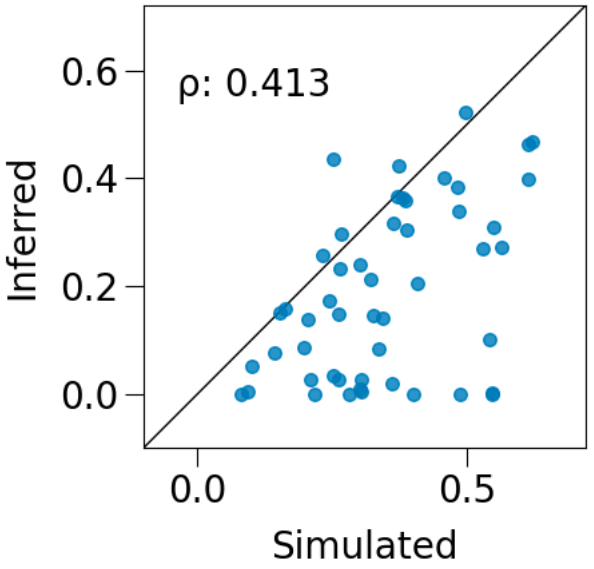






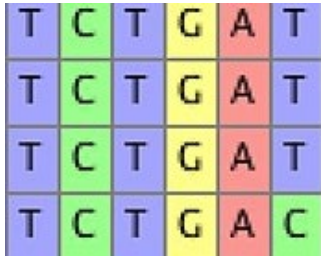


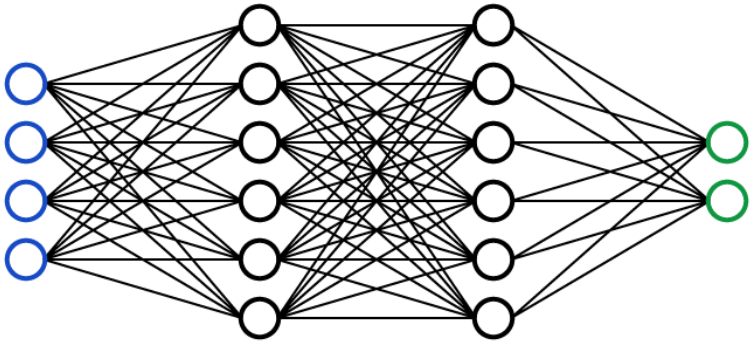




Preliminary results

Summary





- Genomic data contain a wealth of information about our genetic past and disease risk
- Deep neural networks are powerful and effective tools for extracting insights from genomic data
- donni: computationally efficient inference of population history with uncertainty quantification
- DFEnn: informative genomic data representation and processing with convolutional neural network

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

lnt@arizona.edu
linhngohoangtran

NIH NIGMS
THE UNIVERSITY OF ARIZONA