

# Population Genomics – Where are we going? (in 60 minutes....)



Andrew Clark  
Cornell University

EMBO short course  
Procida - 30 Mar 2011

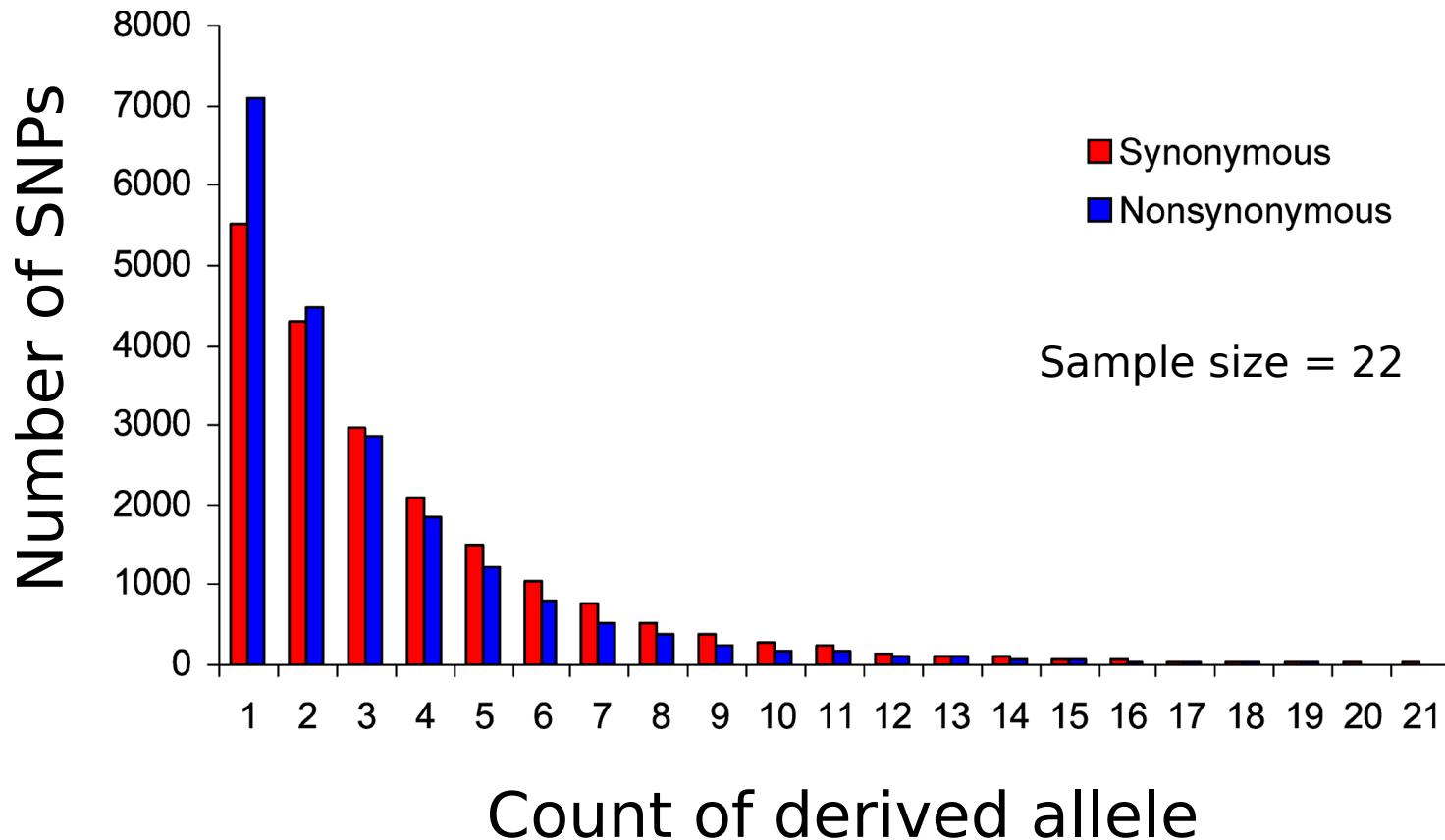
# Outline

- Demographic inference
- Population structure and history
- Admixture / Introgression
- Random genetic drift
- Natural selection
- Mutation spectrum
- Disease association
- Genome functional variation

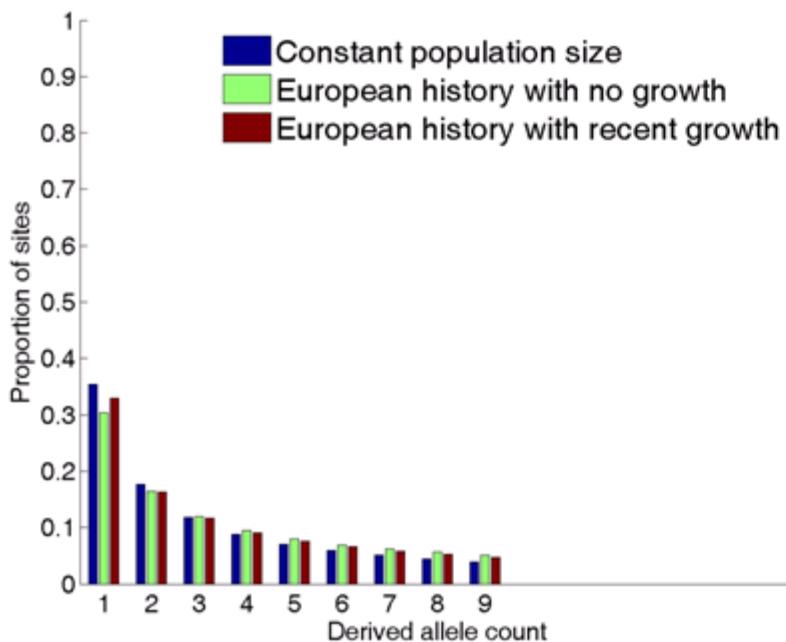
# **DEMOGRAPHY**

How can we infer past changes in population size, bottlenecks, etc.?

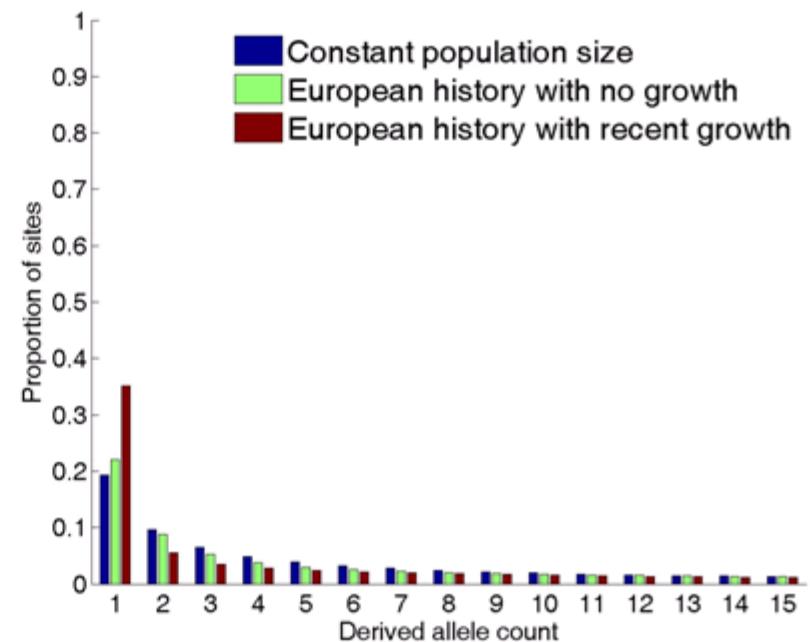
# The Site Frequency Spectrum



# Large samples are needed to see extent of skew to SFS



$n = 10$

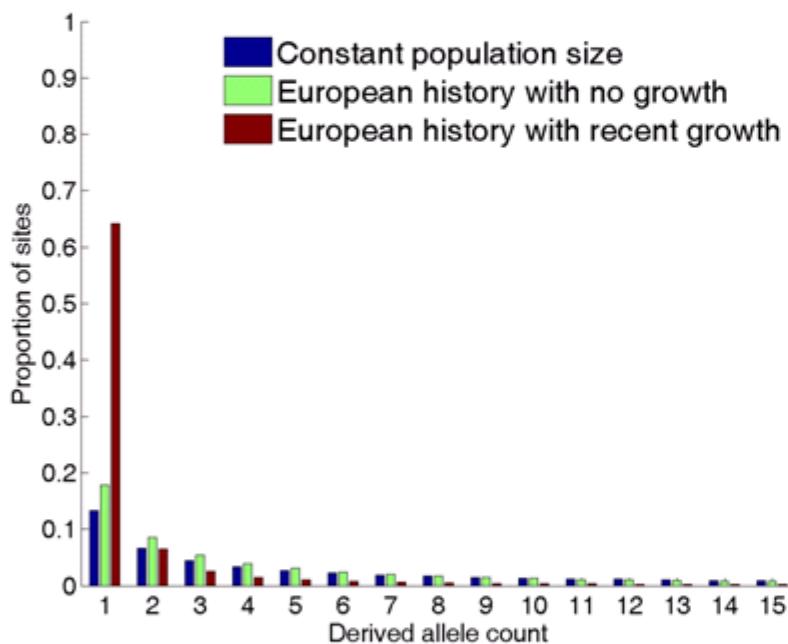


$n = 100$

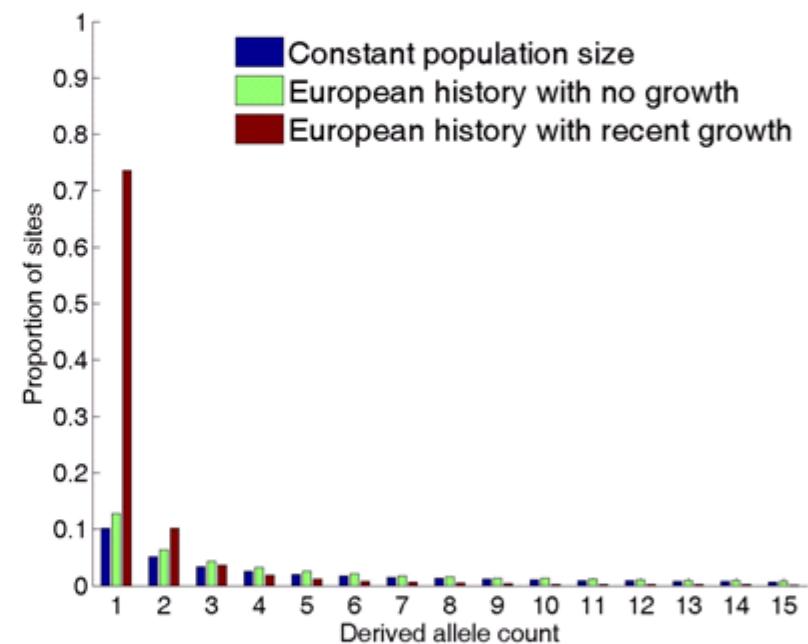
Simulation results

Keinan and Clark 2012

# Large samples are needed to see extent of skew to SFS

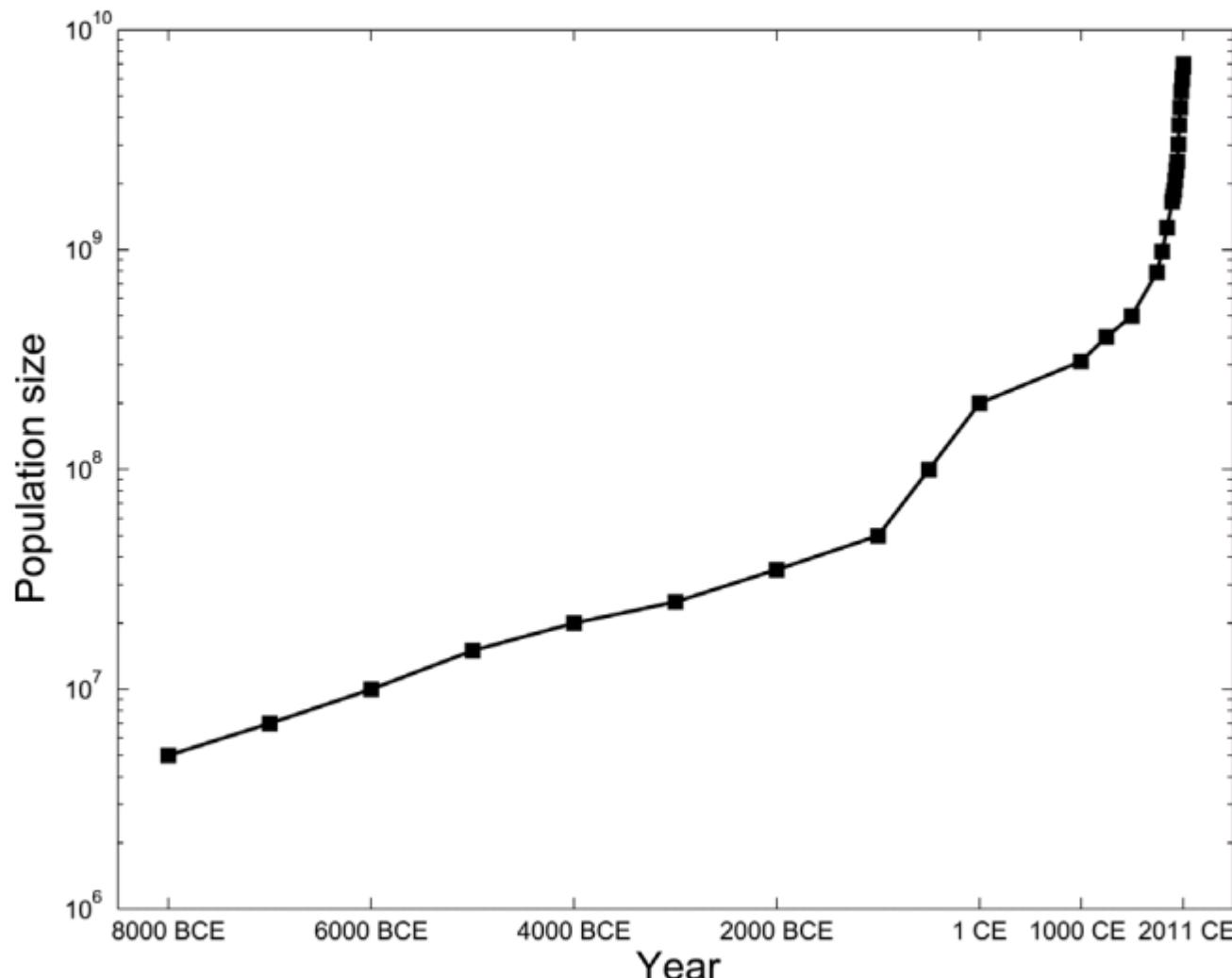


$n = 1000$



$n = 10,000$

# The human population has grown super-exponential



# An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People

Matthew R. Nelson,<sup>1,\*†</sup> Daniel Wegmann,<sup>2,\*</sup> Margaret G. Ehm,<sup>1</sup> Darren Kessner,<sup>2</sup>  
Pamela St. Jean,<sup>1</sup> Claudio Verzilli,<sup>3</sup> Judong Shen,<sup>1</sup> Zhengzheng Tang,<sup>4</sup> Silviu-Alin Bacanu,<sup>1</sup>  
Dana Fraser,<sup>1</sup> Liling Warren,<sup>1</sup> Jennifer Aponte,<sup>1</sup> Matthew Zawistowski,<sup>5</sup> Xiao Liu,<sup>6</sup> Hao Zhang,<sup>6</sup>  
Yong Zhang,<sup>6</sup> Jun Li,<sup>7</sup> Yun Li,<sup>4</sup> Li Li,<sup>1</sup> Peter Woollard,<sup>3</sup> Simon Topp,<sup>3</sup> Matthew D. Hall,<sup>3</sup>  
Keith Nangle,<sup>1</sup> Jun Wang,<sup>6,8</sup> Gonçalo Abecasis,<sup>5</sup> Lon R. Cardon,<sup>9</sup> Sebastian Zöllner,<sup>5,10</sup>  
John C. Whittaker,<sup>3</sup> Stephanie L. Chissoe,<sup>1</sup> John Novembre,<sup>2,†‡</sup> Vincent Mooser<sup>9‡</sup>

# An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People

Matthew R. Nelson,<sup>1,\*†</sup> Daniel Wegmann,<sup>2,\*</sup> Margaret G. Ehm,<sup>1</sup> Darren Kessner,<sup>2</sup> Pamela St. Jean,<sup>1</sup> Claudio Verzilli,<sup>3</sup> Judong Shen,<sup>1</sup> Zhengzheng Tang,<sup>4</sup> Silviu-Alin Bacanu,<sup>1</sup> Dana Fraser,<sup>1</sup> Liling Warren,<sup>1</sup> Jennifer Aponte,<sup>1</sup> Matthew Zawistowski,<sup>5</sup> Xiao Liu,<sup>6</sup> Hao Zhang,<sup>6</sup> Yong Zhang,<sup>6</sup> Jun Li,<sup>7</sup> Yun Li,<sup>4</sup> Li Li,<sup>1</sup> Peter Woollard,<sup>3</sup> Simon Topp,<sup>3</sup> Matthew D. Hall,<sup>3</sup> Keith Nangle,<sup>1</sup> Jun Wang,<sup>6,8</sup> Gonçalo Abecasis,<sup>5</sup> Lon R. Cardon,<sup>9</sup> Sebastian Zöllner,<sup>5,10</sup> John C. Whittaker,<sup>3</sup> Stephanie L. Chissoe,<sup>1</sup> John Novembre,<sup>2,†‡</sup> Vincent Mooser<sup>9</sup>‡

# Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes

Jacob A. Tennessen,<sup>1,\*</sup> Abigail W. Bigham,<sup>2,\*†</sup> Timothy D. O'Connor,<sup>1,\*</sup> Wenqing Fu,<sup>1</sup> Eimear E. Kenny,<sup>3</sup> Simon Gravel,<sup>3</sup> Sean McGee,<sup>1</sup> Ron Do,<sup>4,5</sup> Xiaoming Liu,<sup>6</sup> Goo Jun,<sup>7</sup> Hyun Min Kang,<sup>7</sup> Daniel Jordan,<sup>8</sup> Suzanne M. Leal,<sup>9</sup> Stacey Gabriel,<sup>4</sup> Mark J. Rieder,<sup>1</sup> Gonçalo Abecasis,<sup>7</sup> David Altshuler,<sup>4</sup> Deborah A. Nickerson,<sup>1</sup> Eric Boerwinkle,<sup>6,10</sup> Shamil Sunyaev,<sup>4,8</sup> Carlos D. Bustamante,<sup>3</sup> Michael J. Bamshad,<sup>1,2,†‡</sup> Joshua M. Akey,<sup>1,‡</sup> Broad GO, Seattle GO, on behalf of the NHLBI Exome Sequencing Project

# An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People

Matthew R. Nelson,<sup>1,\*†</sup> Daniel Wegmann,<sup>2\*</sup> ...  
Pamela St. Jean,<sup>1</sup> Claudio Verzilli,<sup>3</sup> ...  
Dana Fraser,<sup>1</sup> Liling Warren,<sup>1</sup> ...  
Yong Zhang,<sup>6</sup> Jun Li,<sup>7</sup> Yun Li,<sup>8</sup> ...  
Keith Nangle,<sup>1</sup> Jun Wang,<sup>9</sup> ...  
John C. Whittaker,<sup>3</sup> ...

International weekly journal of science

Sci nature LETTER

## Evolution of rare coding variants in most human protein-coding exomes

Jacob A. Tennessen,<sup>1,\*</sup> Abigail S. Bigham,<sup>1</sup> ... Timothy D. O'Connor,<sup>1,\*</sup> Wenqing Fu,<sup>1</sup> Eimear E. Kenny,<sup>3</sup> Simon J. Evans,<sup>4</sup> ... Michael J. Bamshad,<sup>1,4</sup> ... Broad GO, Seattle GO, on behalf of the NHLBI Exome Sequencing Project

Wenqing Fu,<sup>1</sup> Timothy D. O'Connor,<sup>1</sup> Goo Jun,<sup>2</sup> Hyun Min Kang,<sup>2</sup> Goncalo Abecasis,<sup>2</sup> Suzanne M. Leal,<sup>3</sup> Stacey Gabriel,<sup>4</sup> David Altshuler,<sup>4</sup> Jay Shendure,<sup>1</sup> Deborah A. Nickerson,<sup>1</sup> Michael J. Bamshad,<sup>1,4</sup> ... NHLBI Exome Sequencing Project\*

doi:10.1038/nature11690

These studies all show massive excesses of rare variation.

# Implications of recent growth for complex disease studies

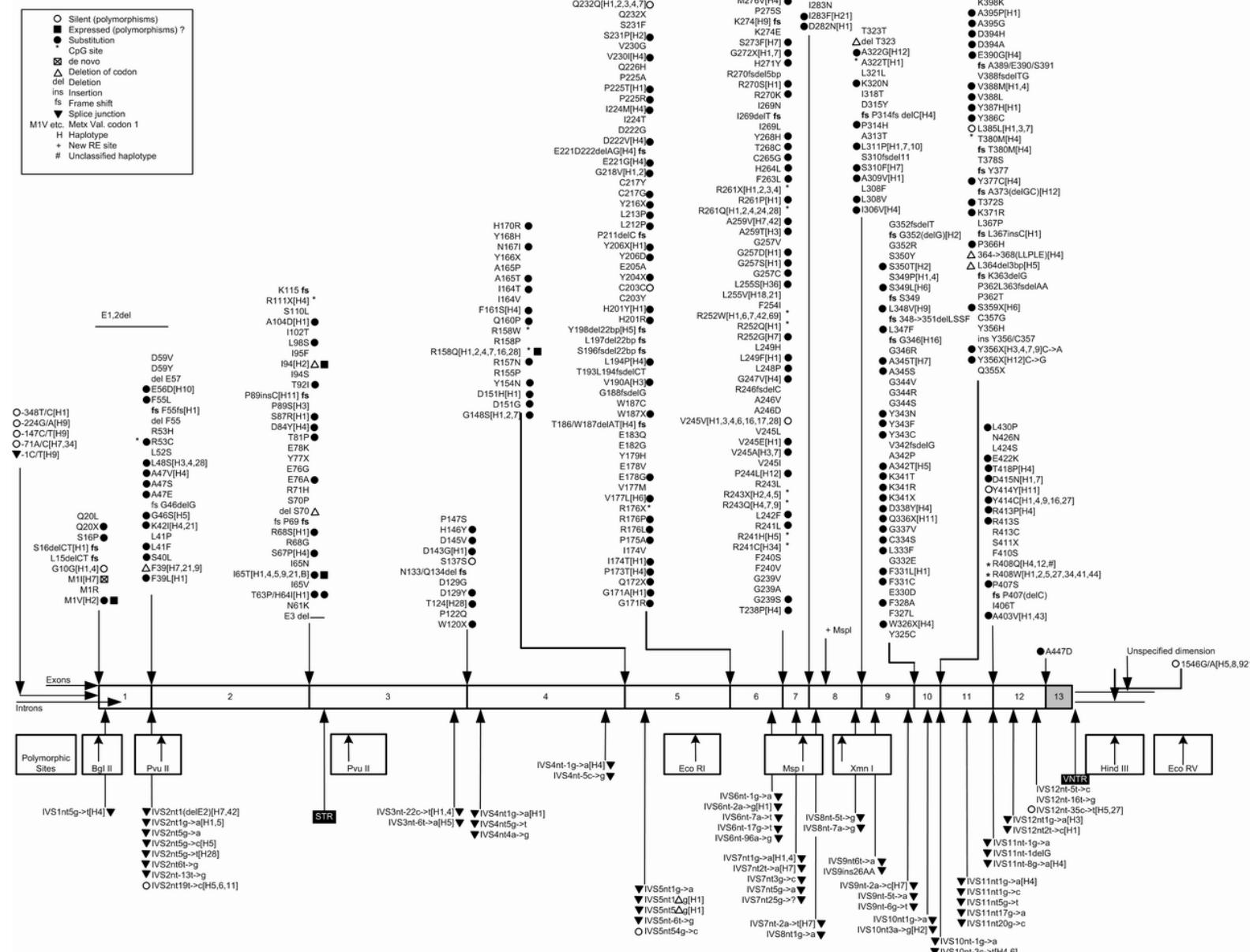
- Many more variants of all classes.
- Massive excess of rare variants.
- Rare variants are more likely to have deleterious effects.

# Implications of recent growth for complex disease studies

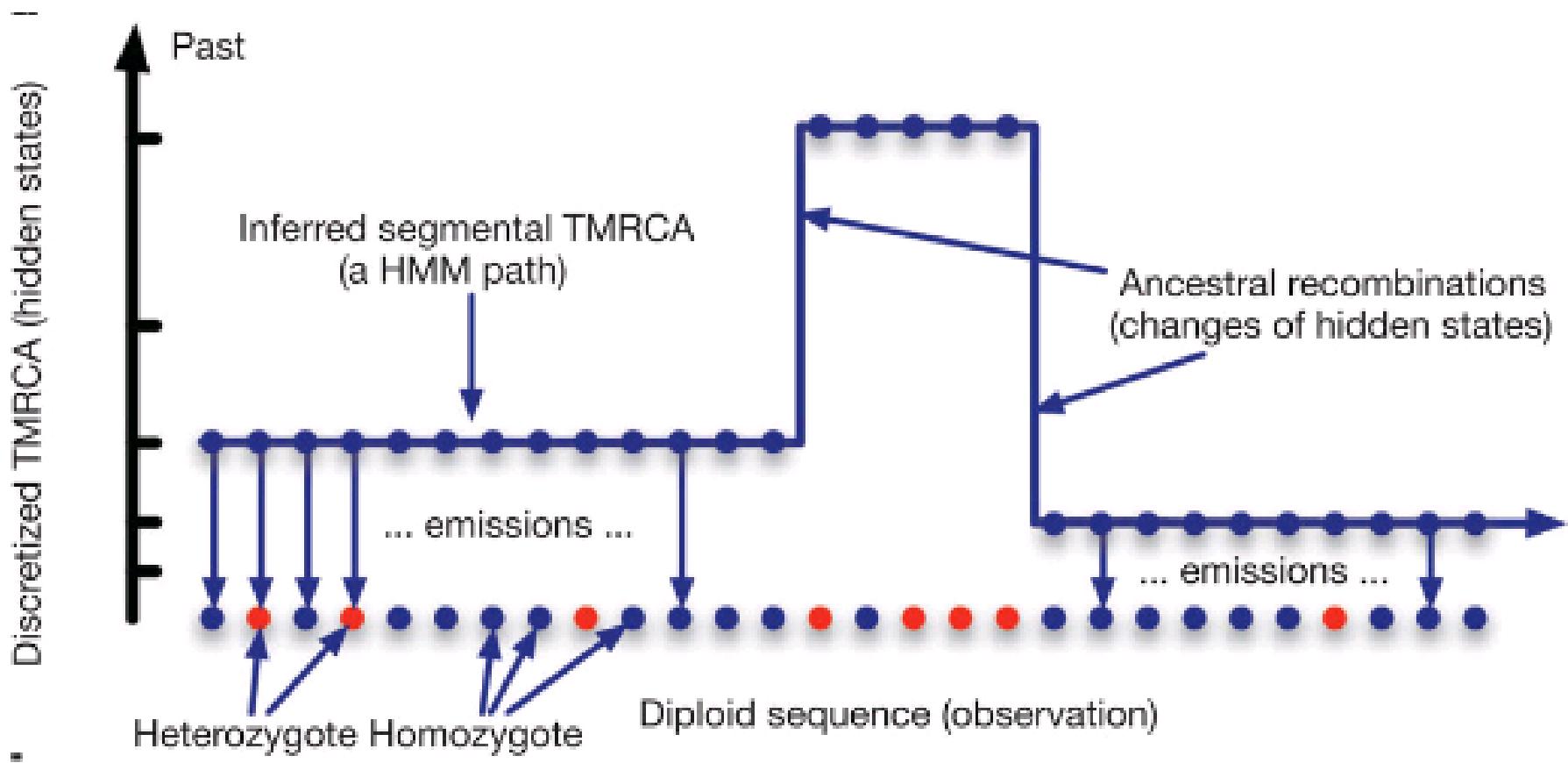
- Many more variants of disease classes.
- Frequency shifted to rarer variants.
- Rare variants are more likely to have deleterious effects.

**Genetic heterogeneity**

# Allelic heterogeneity of PAH

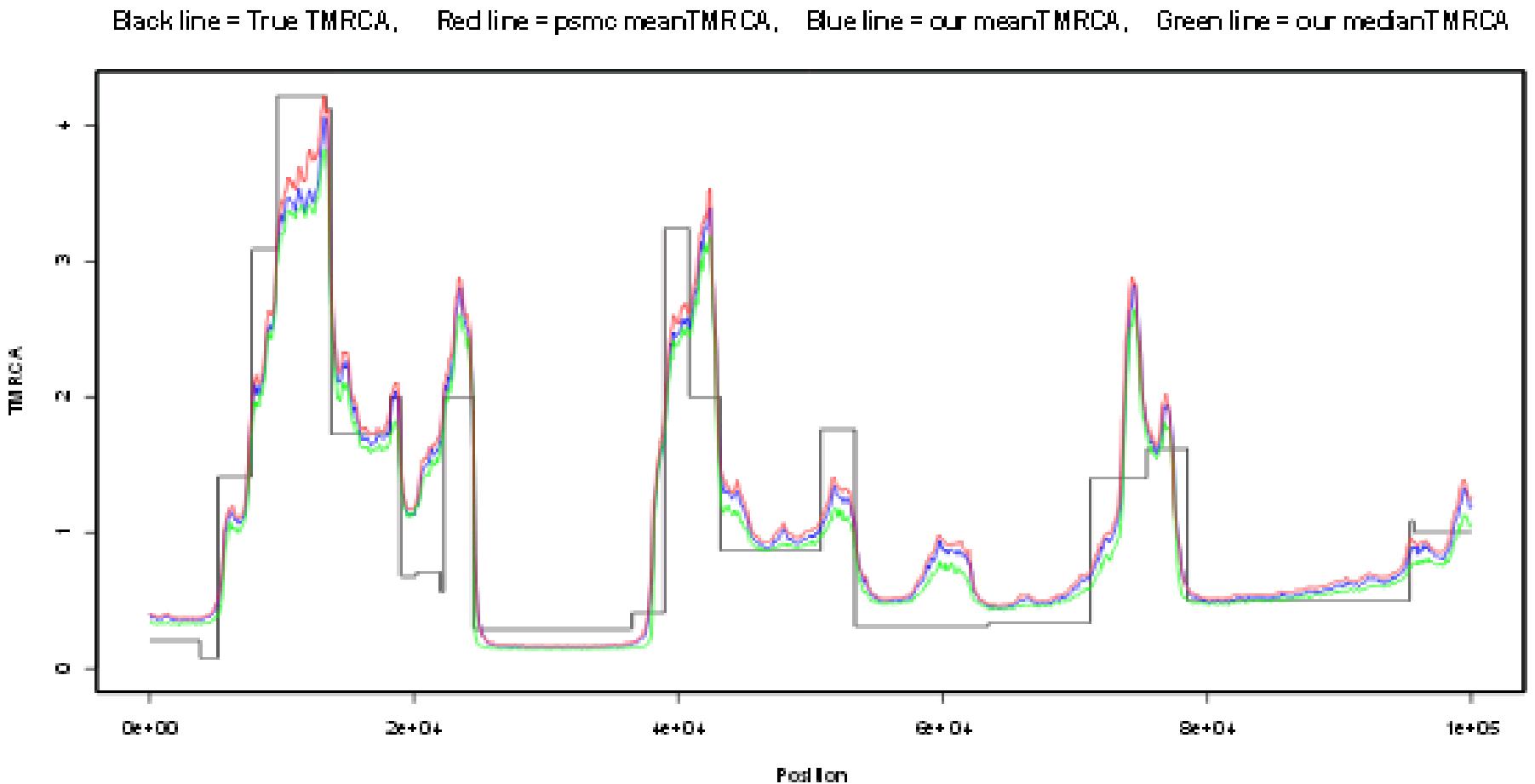


# Inferring demography from a single individual: Pairwise Sequentially Markovian Coalescent

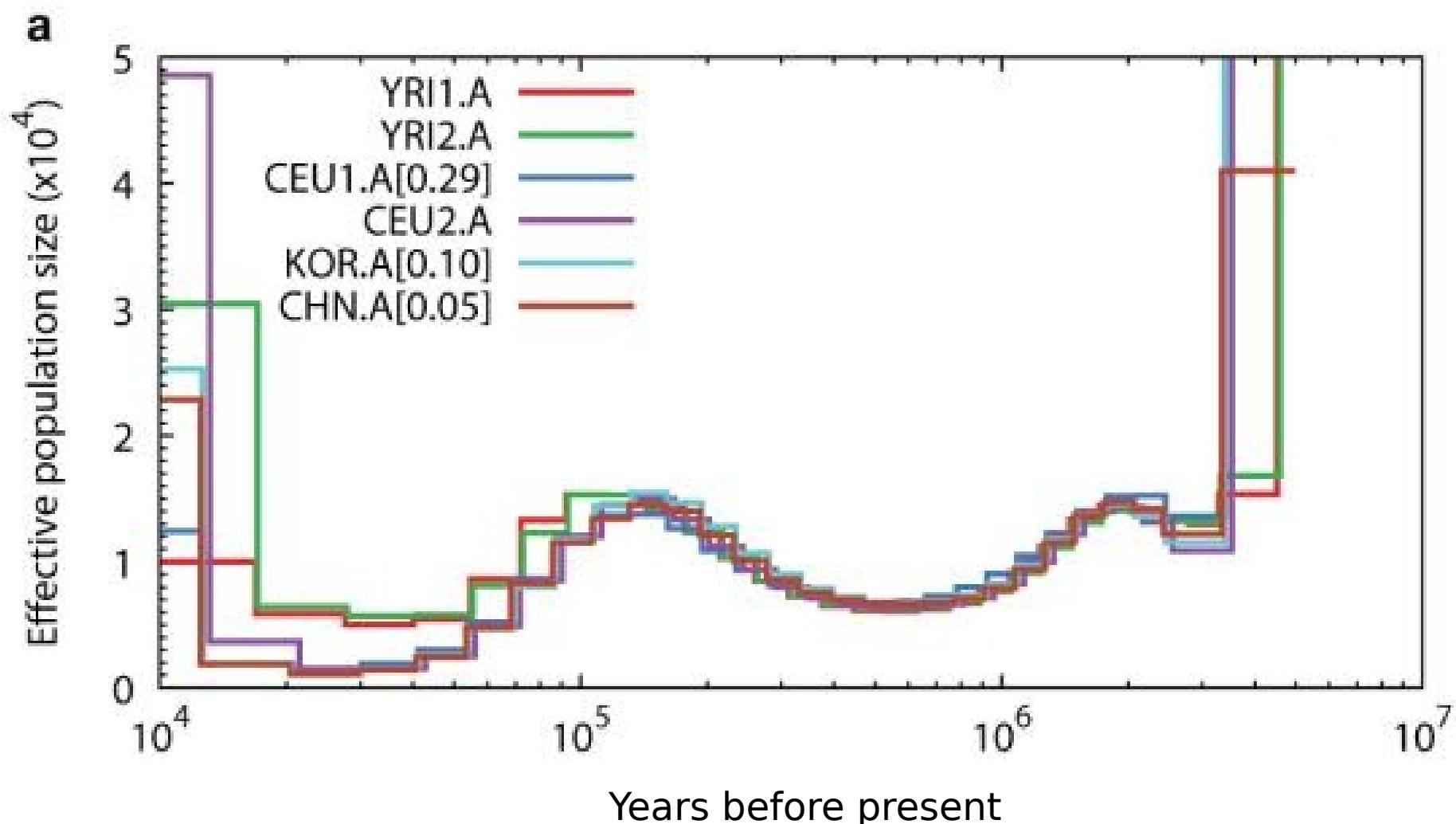


Li and Durbin 2011 Nature

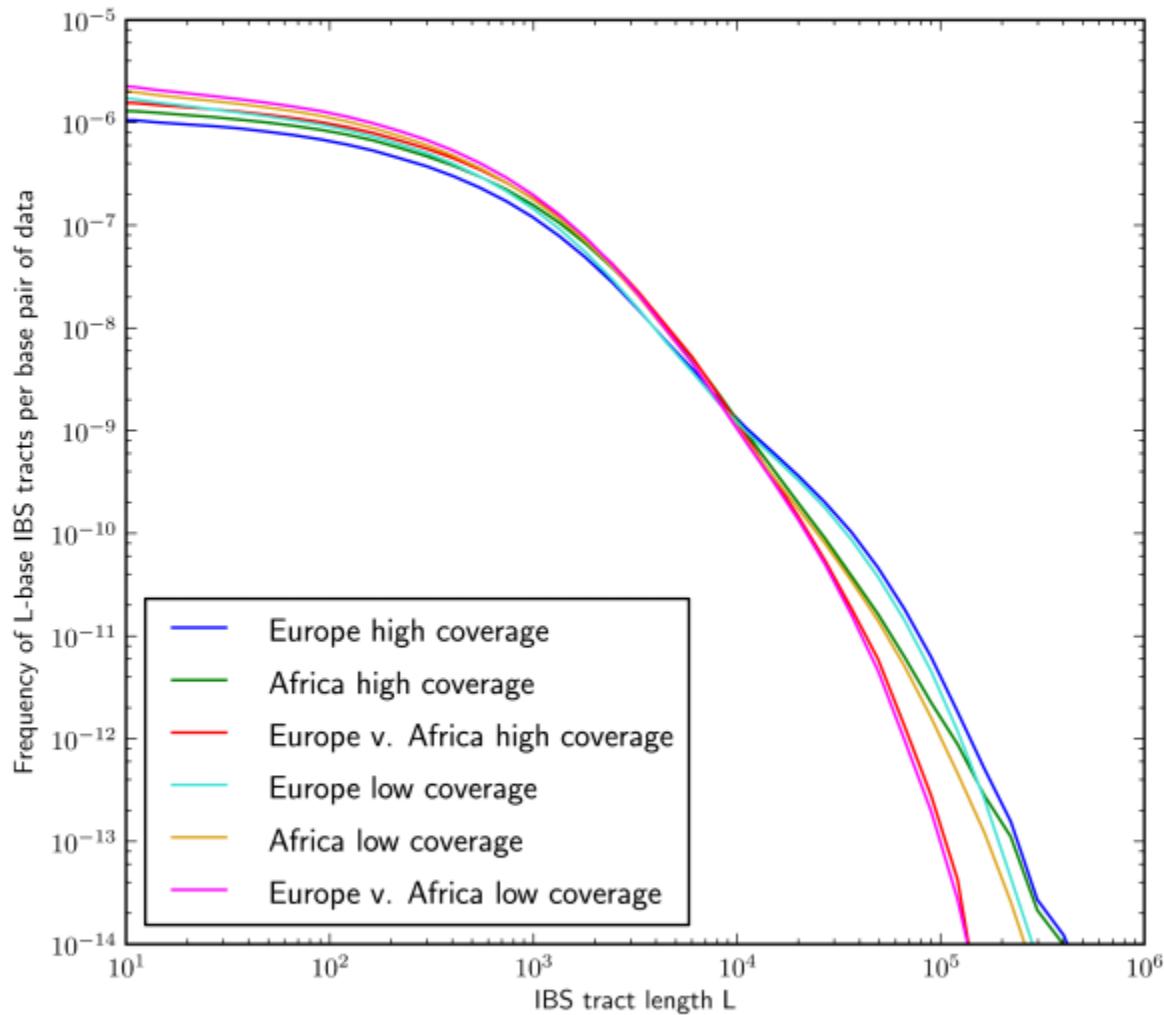
# Simulations of past population demography shows reasonably good accuracy of PSMC



# Application of Pairwise Sequentially Markovian Coalescent



# Identity-by-Descent tracts for inference of demography



**DEMOGRAPHY**

# (Population collapse)

What happens to genetic variation in genomes of populations that are crashing?

# Florida Scrub-Jay

(*Aphelocoma coerulescens*)

Cooperative breeder

Federally Threatened

Non-migratory

Highly territorial & philopatric

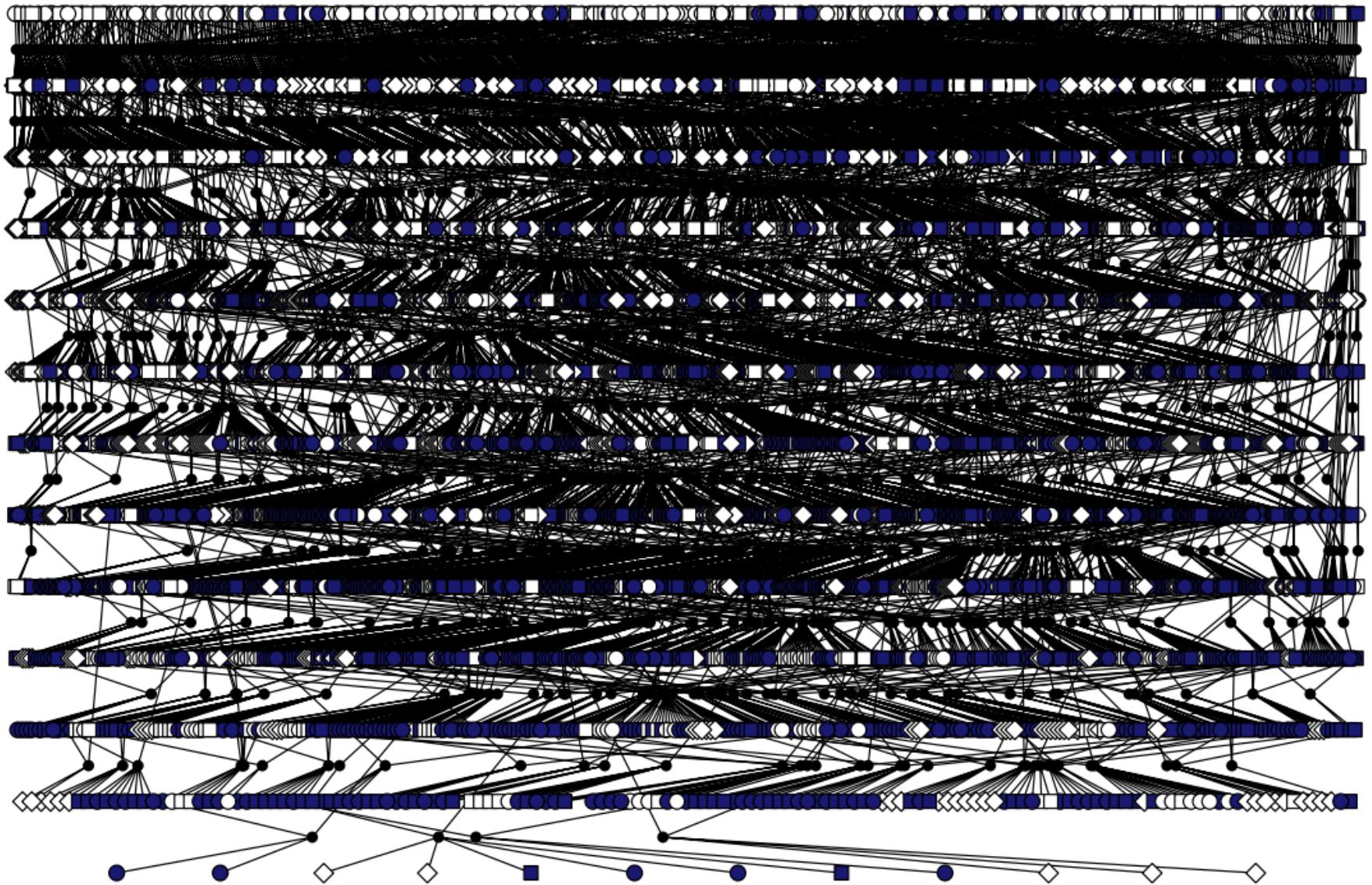
Socially & (mostly) genetically  
monogamous

**All individuals banded  
and tracked  
throughout life**

Quinn et al. 1999; Townsend et al. 2011; Woolfenden & Fitzpatrick 1984, 1996

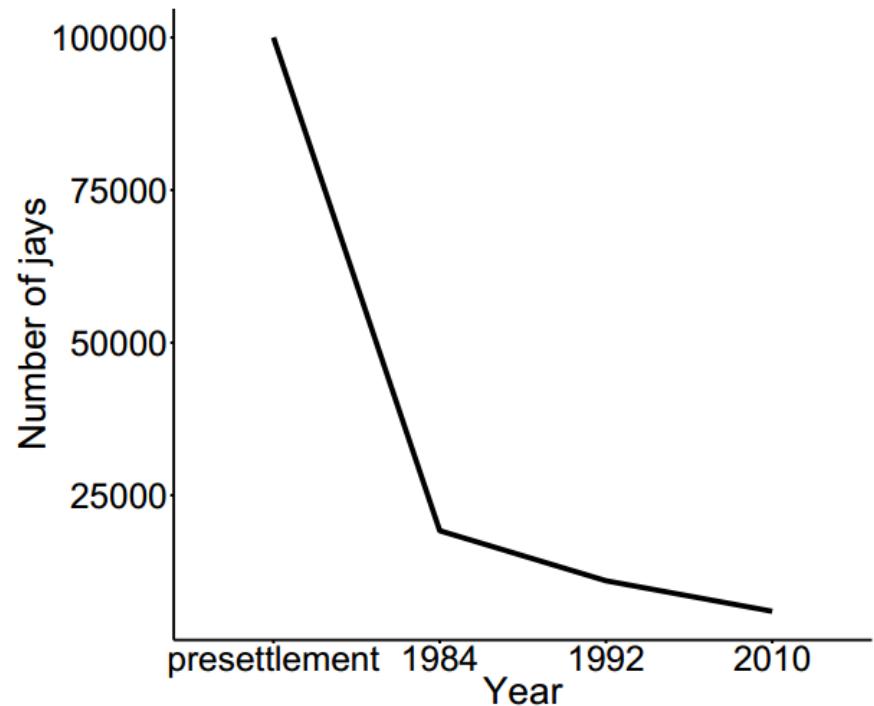
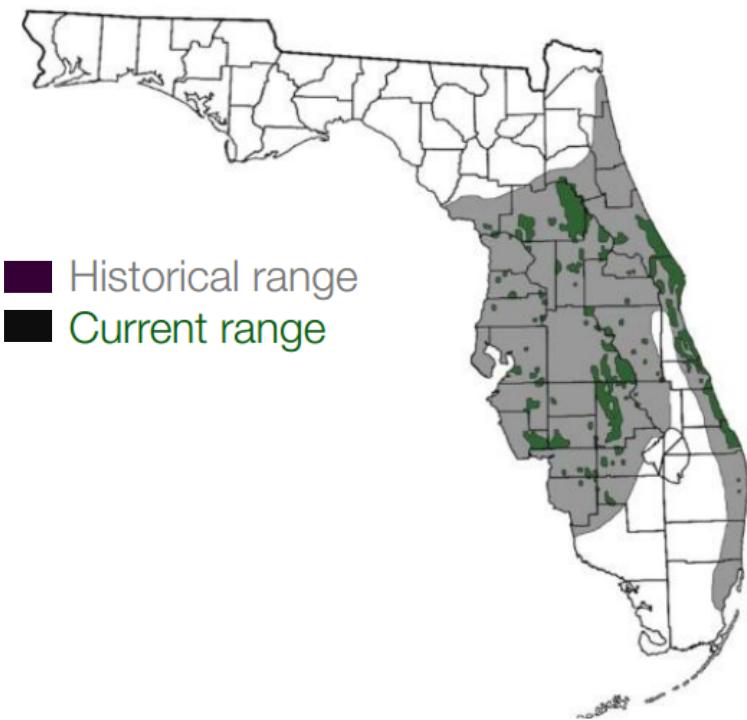


R. Bowman



We have blood samples as well as life history & morphological data for >4,000 pedigreed individuals

Florida Scrub-Jay populations have drastically declined due to habitat loss



97% decline in past century. 50% decline in past 20 years

# Florida Scrub-Jay genomic resources



Genome-wide SNPs

Genome assembly

Transcriptome assembly

Genome annotation

Linkage map construction

# Florida Scrub-Jay genomic resources



Genome-wide SNPs

Genome assembly

Transcriptome assembly

Genome annotation

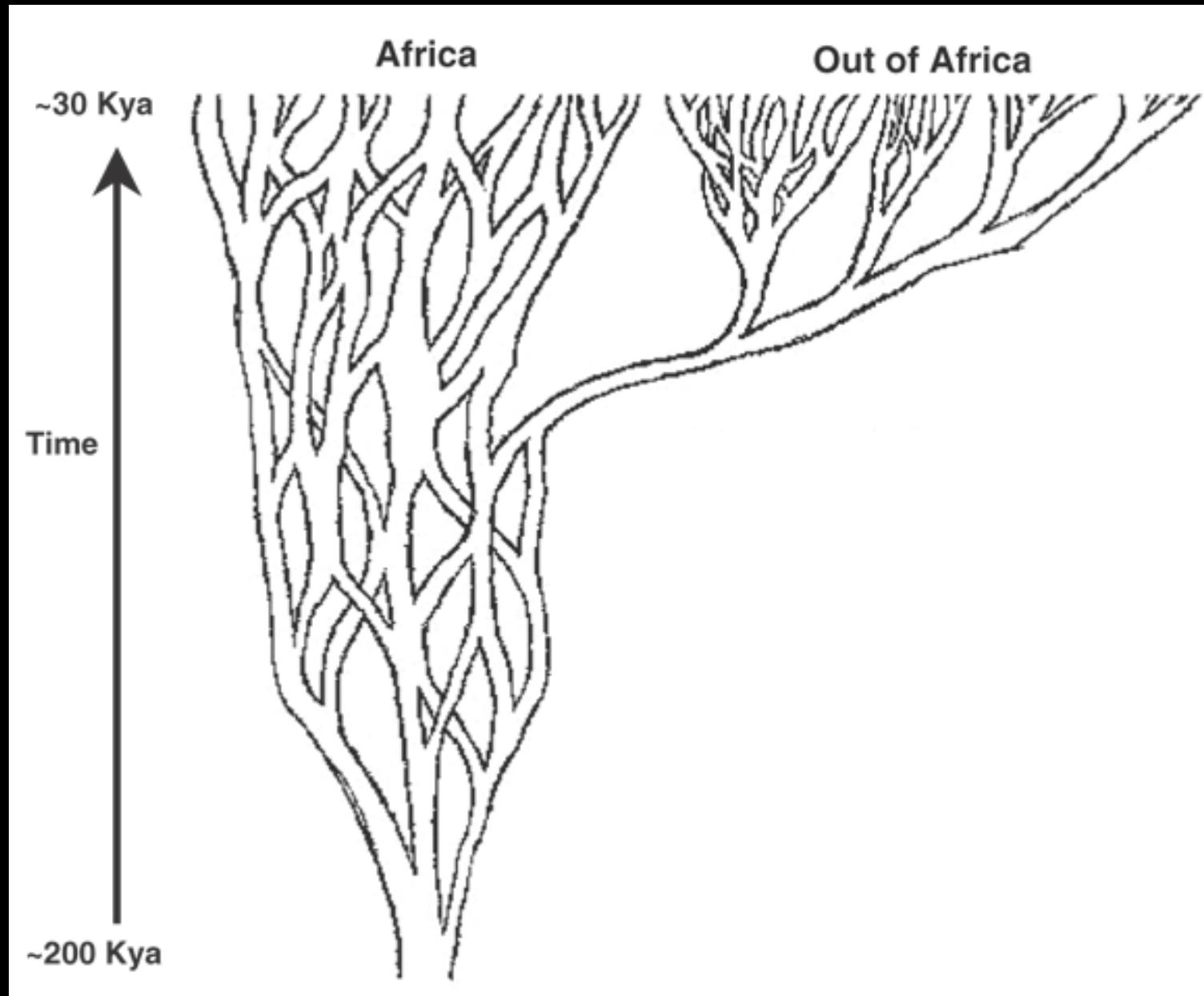
Linkage map construction

How does the recent population crash  
manifest itself in the structure of  
genetic variation?

# Population Structure

How can we infer past patterns of migration from genome sequence?

# Human population history



# Principal Components Analysis for population structure

OPEN  ACCESS Freely available online

PLOS GENETICS

## Population Structure and Eigenanalysis

Nick Patterson<sup>1\*</sup>, Alkes L. Price<sup>1,2</sup>, David Reich<sup>1,2</sup>

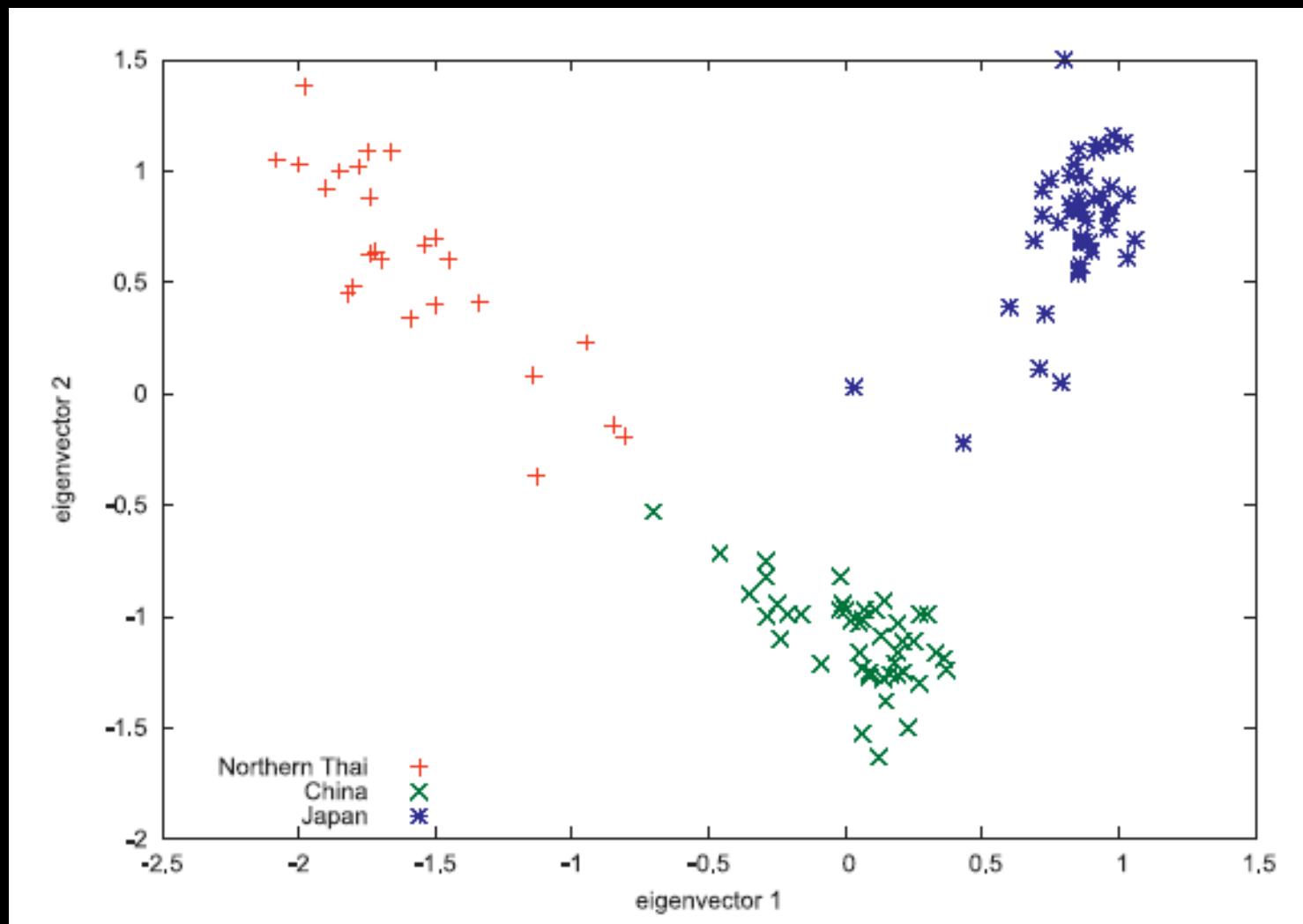
1 Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America, 2 Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America

Current methods for inferring population structure from genetic data do not provide formal significance tests for population differentiation. We discuss an approach to studying population structure (principal components analysis) that was first applied to genetic data by Cavalli-Sforza and colleagues. We place the method on a solid statistical footing, using results from modern statistics to develop formal significance tests. We also uncover a general “phase change” phenomenon about the ability to detect structure in genetic data, which emerges from the statistical theory we use, and has an important implication for the ability to discover structure in genetic data: for a fixed but large dataset size, divergence between two populations (as measured, for example, by a statistic like  $F_{ST}$ ) below a threshold is essentially undetectable, but a little above threshold, detection will be easy. This means that we can predict the dataset size needed to detect structure.

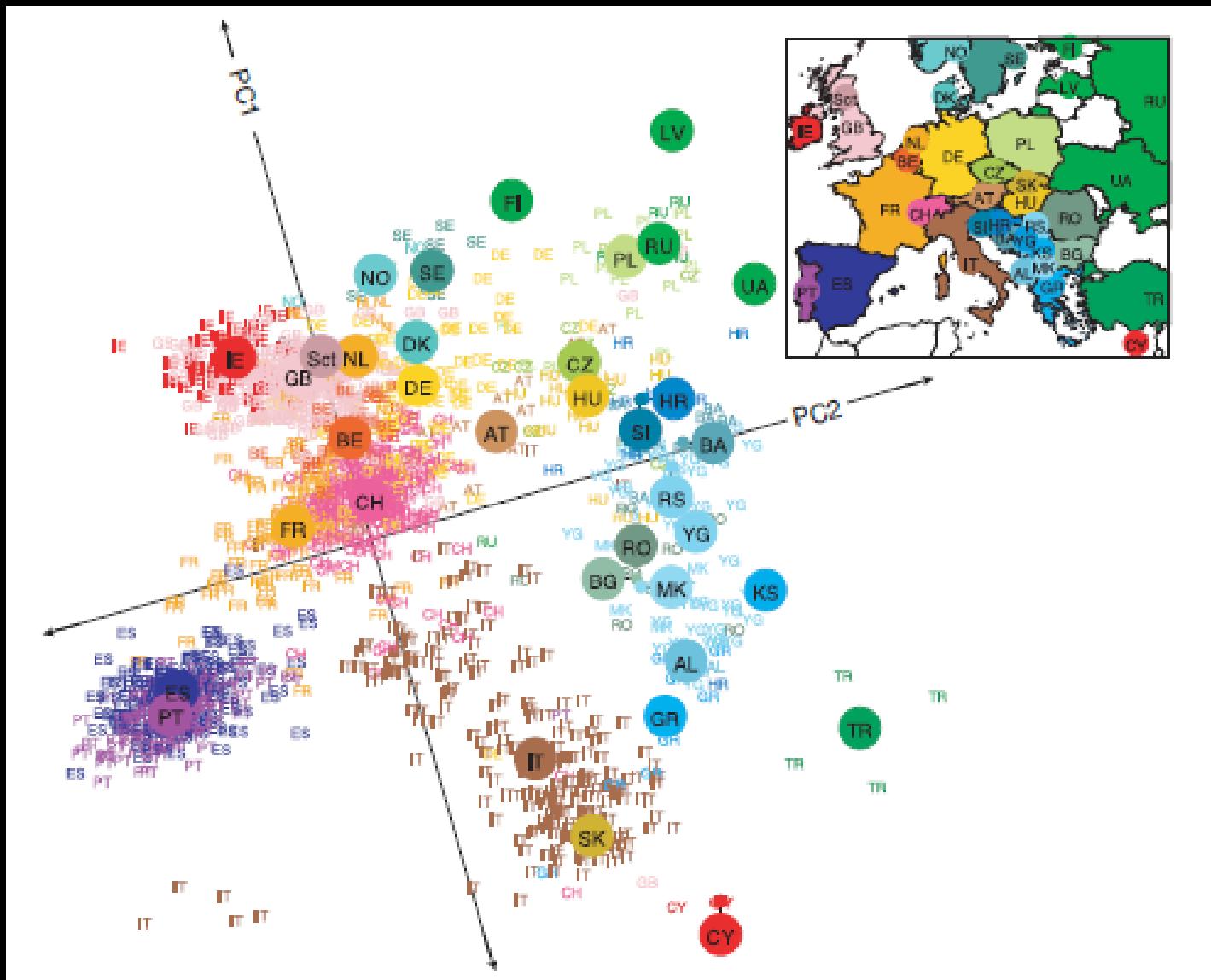
Citation: Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12): e190. doi:10.1371/journal.pgen.0020190

Patterson N, Price AL, Reich D (2006) PLoS Genet 2(12): e190.

# Three East Asian populations



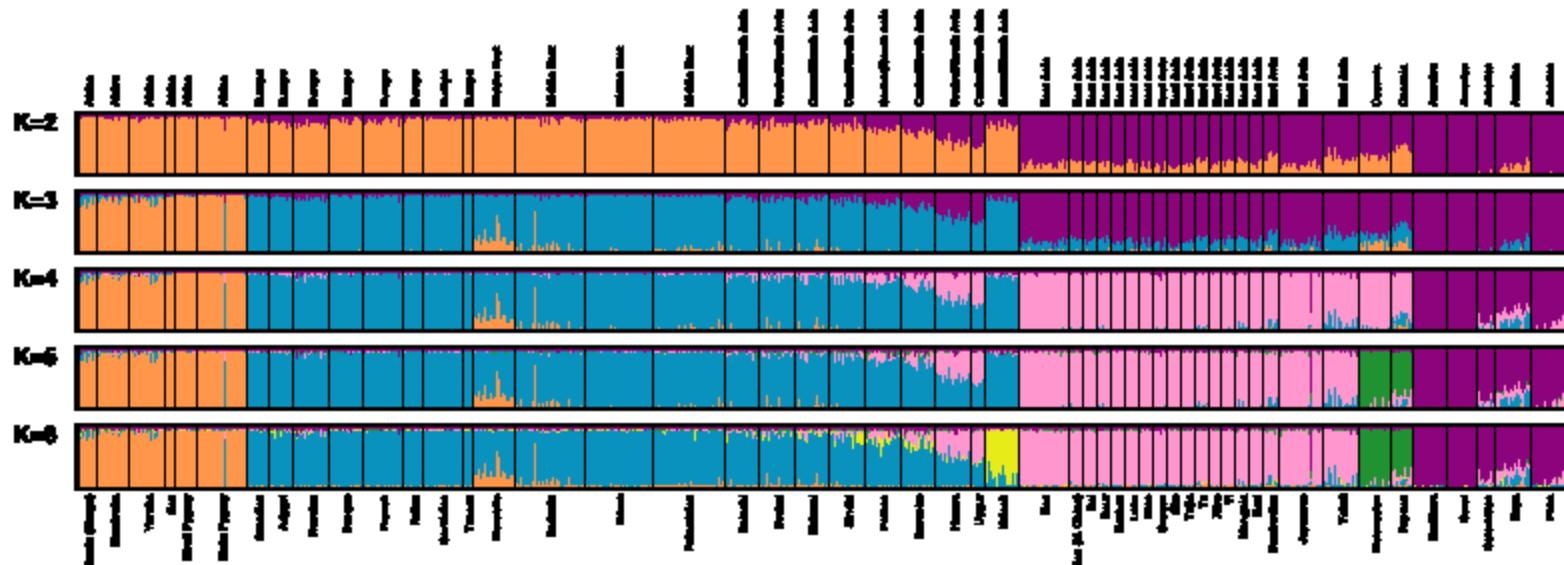
# European genetic differentiation using PCA



John Novembre

Novembre et al. 2008 Nature 456:274.

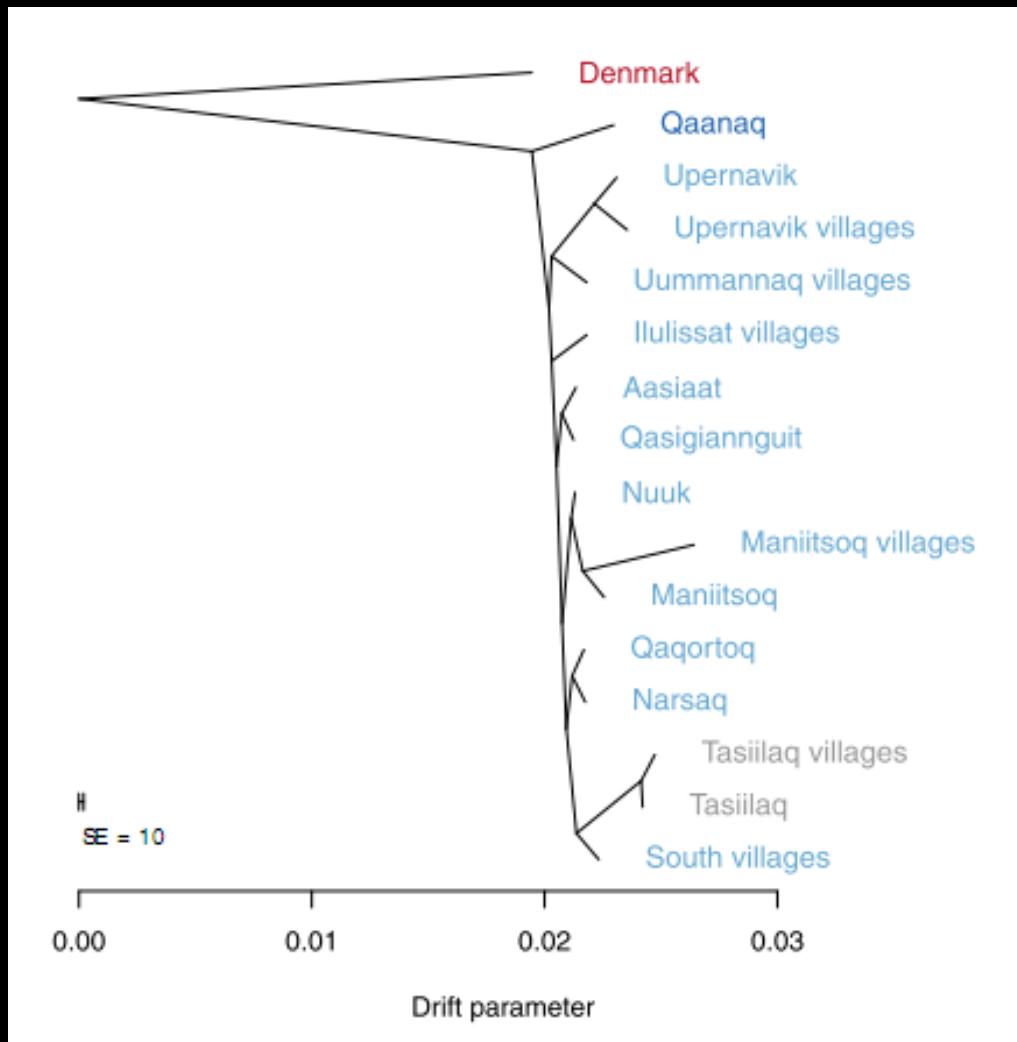
Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002  
Genetic structure of human populations.  
Science. 298:2381-2385.

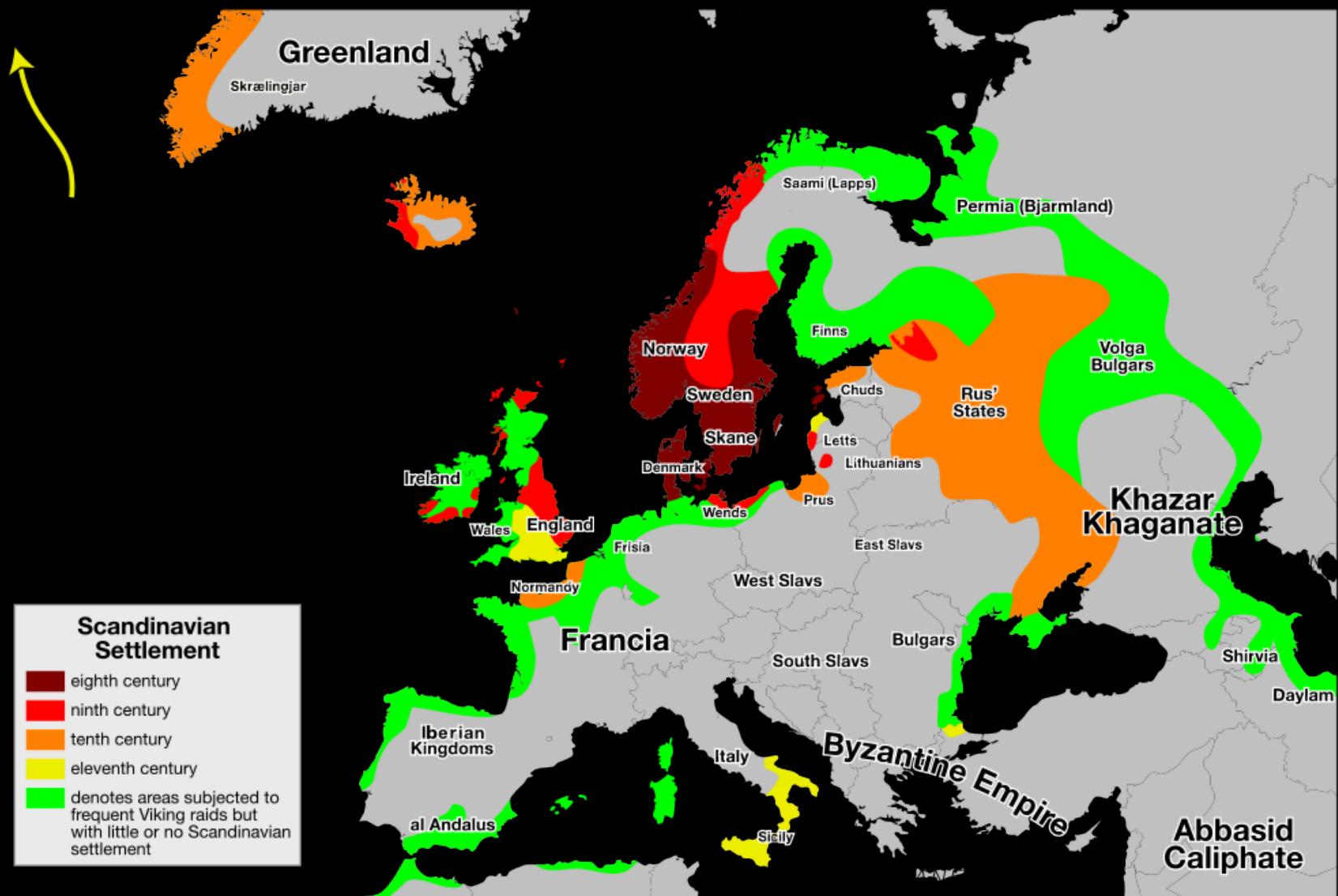


**Fig. 1.** Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into  $K$  colored segments that represent the individual's estimated membership fractions in  $K$  clusters. Black lines separate individuals of different populations. Populations are labeled below the figure, with their regional affiliations above it. Ten structure runs at each

$K$  produced nearly identical individual membership coefficients, having pairwise similarity coefficients above 0.97, with the exceptions of comparisons involving four runs at  $K = 3$  that separated East Asia instead of Eurasia, and one run at  $K = 6$  that separated Karitiana instead of Kalash. The figure shown for a given  $K$  is based on the highest probability run at that  $K$ .

# Pritchard 's TreeMix analysis

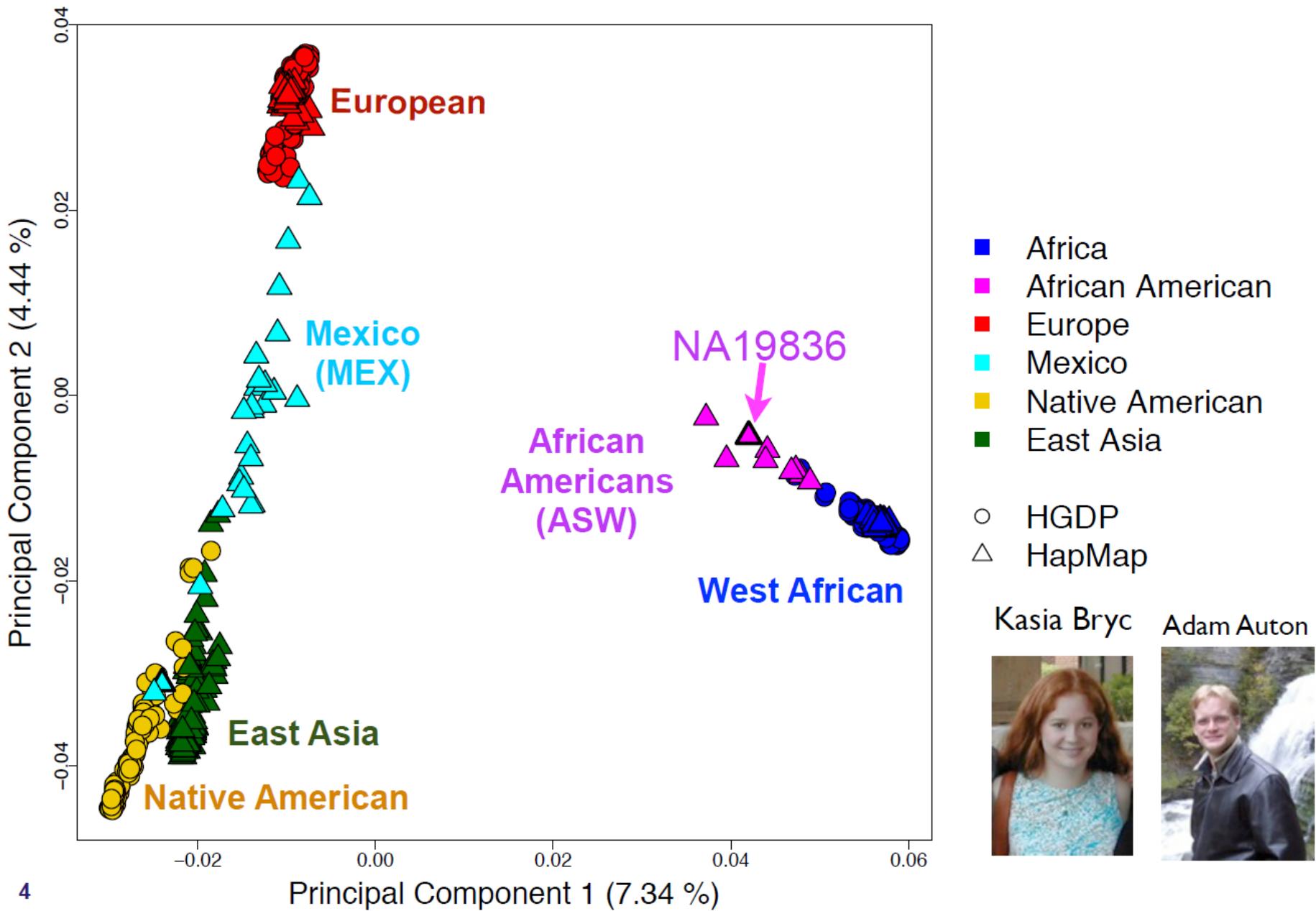




# **ADMIXTURE**

How can we infer consanguinity and  
admixture from genetic data?

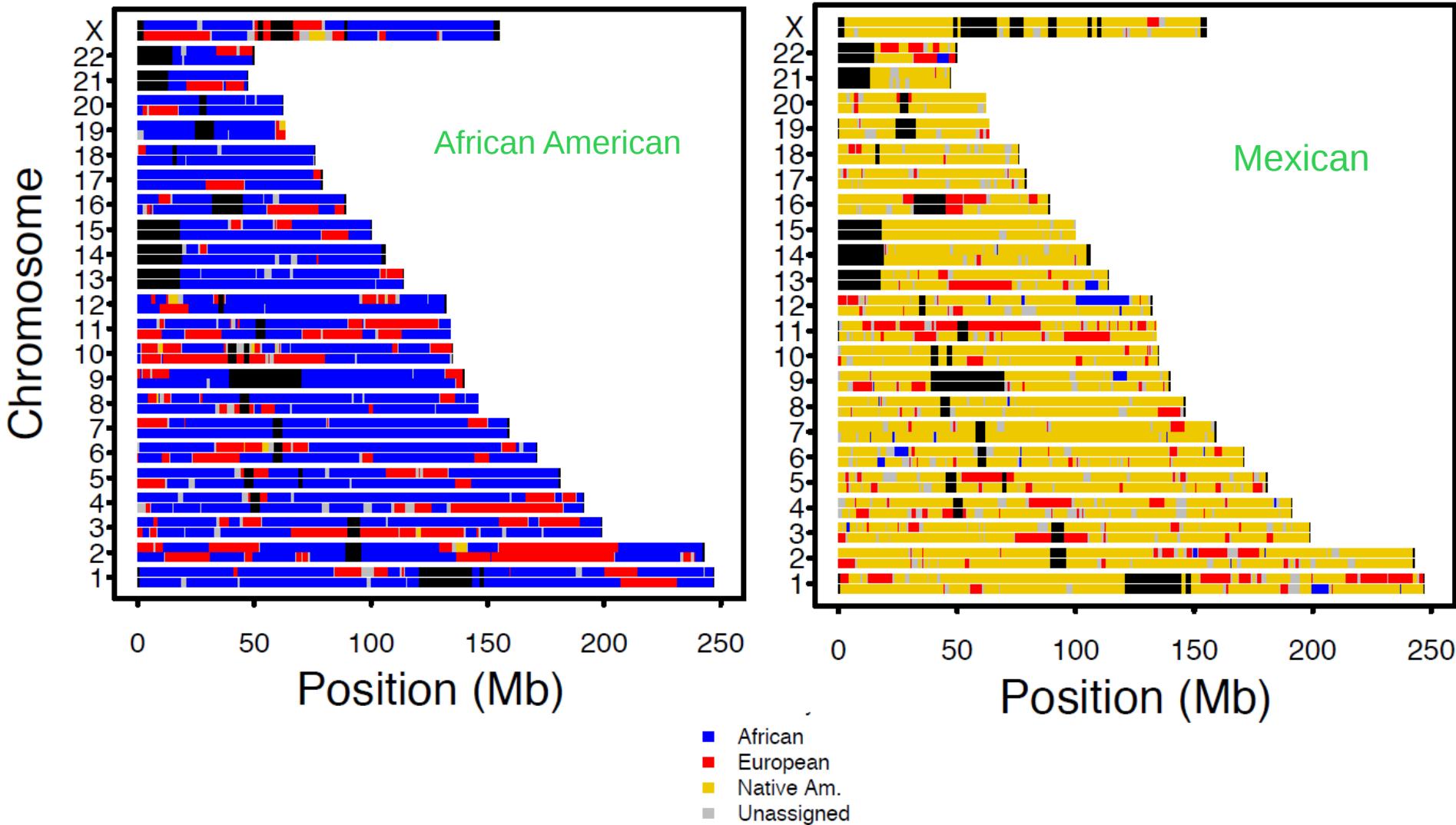
# Using PCA to infer admixed individuals



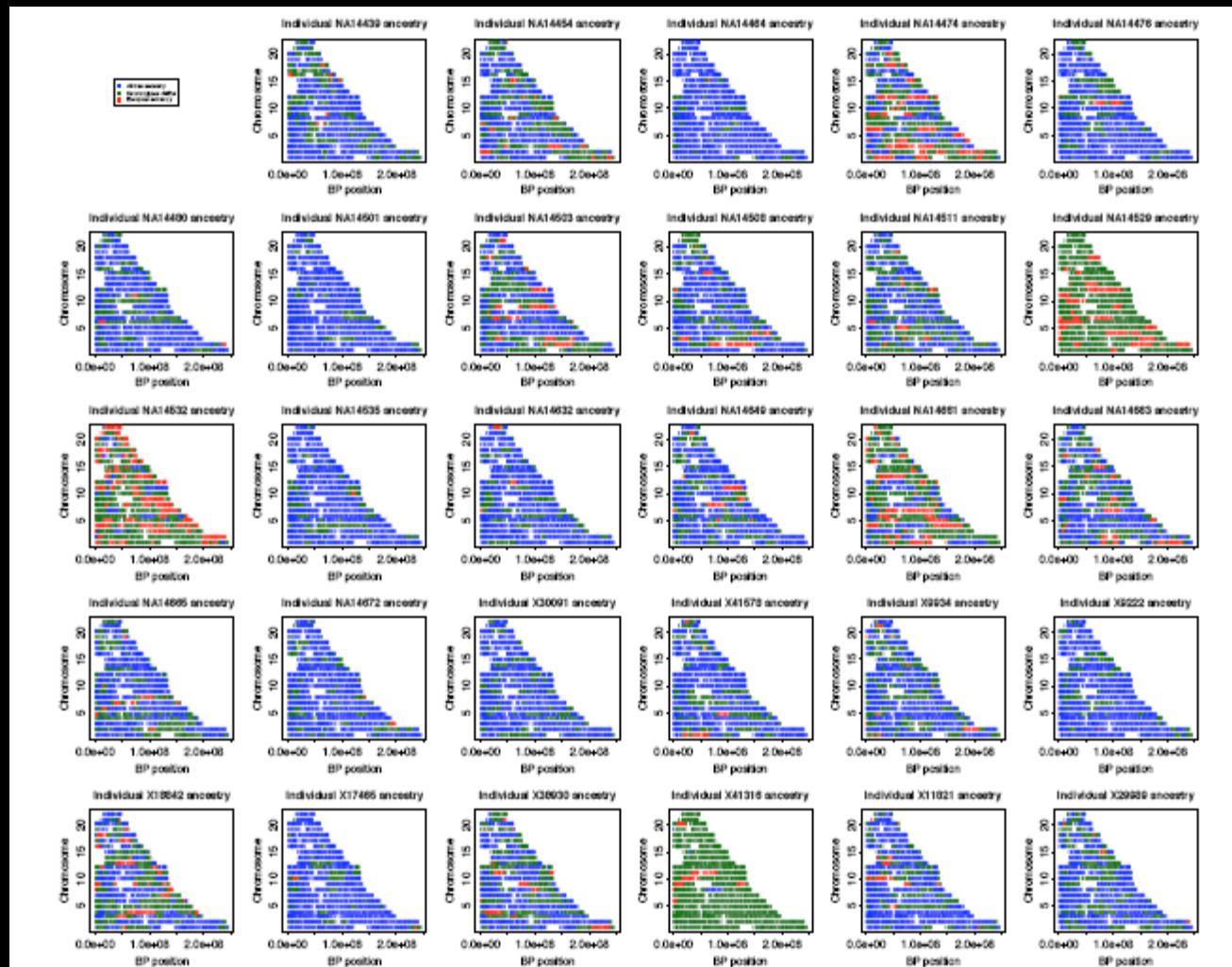
# Local ancestry inference

- Run PCA on African, European, and African American samples.
- From PC loadings of 15 SNP windows, infer 0, 1, or 2 copies of African ancestry.
- Slide along the genome.
- RESULT: Call of 0, 1, 2 copies of African ancestry for each chunk of the genome in each individual.

# PCAdmix can identify the population-of-origin of segments of the genome



# High variability among individuals in admixture patterns



An extreme case of admixture:

**modern humans x  
Neanderthals**

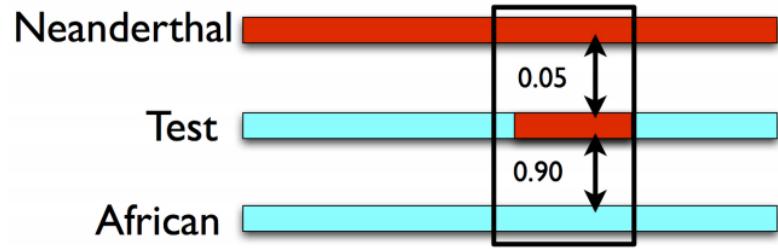
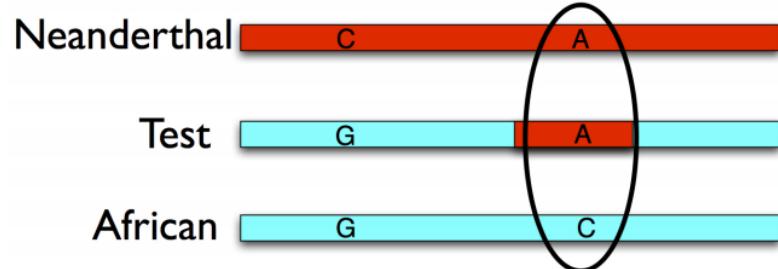
# The genomic landscape of Neanderthal ancestry in present-day humans

Sriram Sankararaman<sup>1,2</sup>, Swapna Mallick<sup>1,2</sup>, Michael Dannemann<sup>3</sup>, Kay Prüfer<sup>3</sup>, Janet Kelso<sup>3</sup>, Svante Pääbo<sup>3</sup>, Nick Patterson<sup>1,2</sup>  
& David Reich<sup>1,2,4</sup>

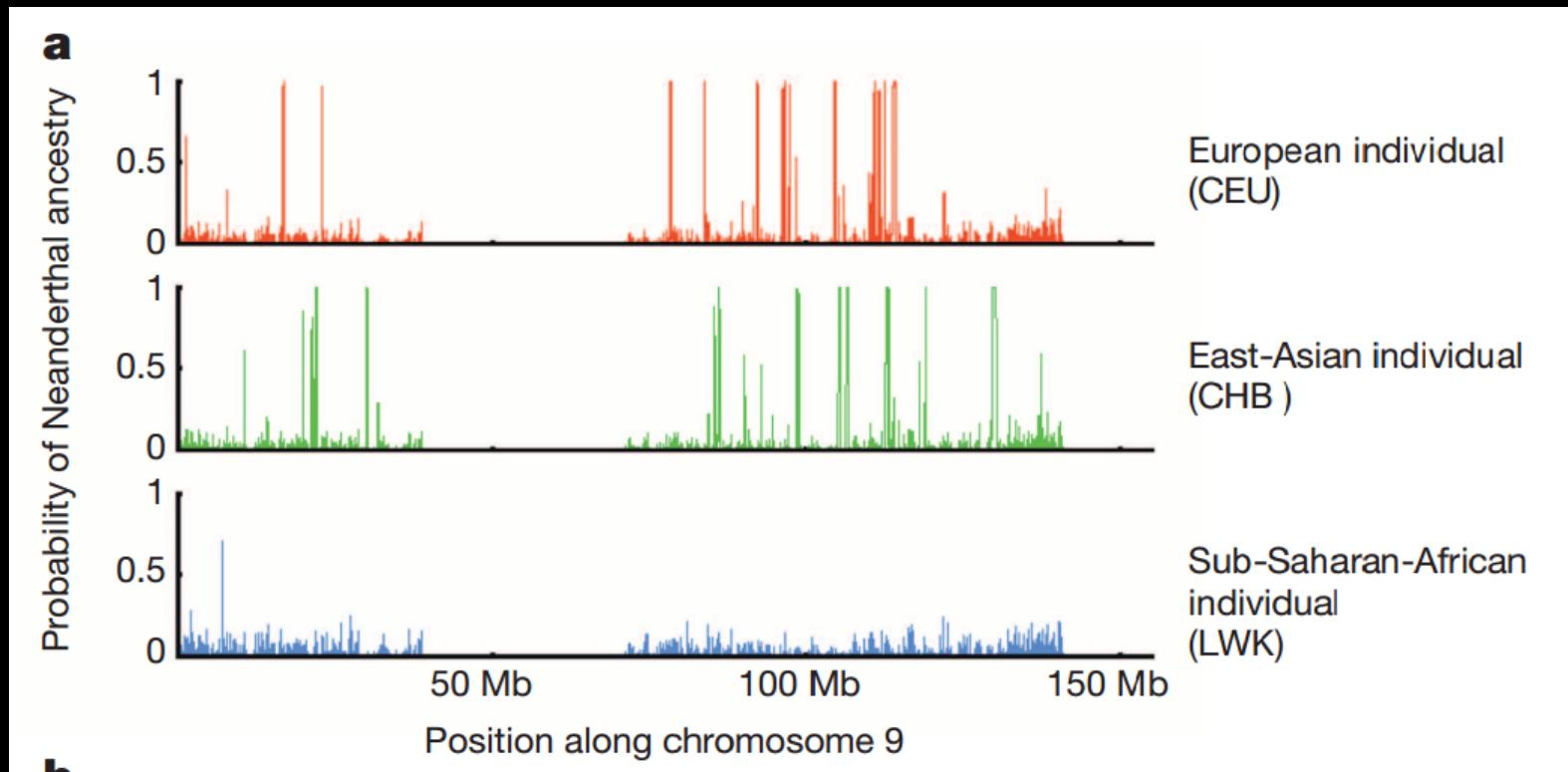
2014 Nature 507:354-357

## Three criteria: Introgressed region is:

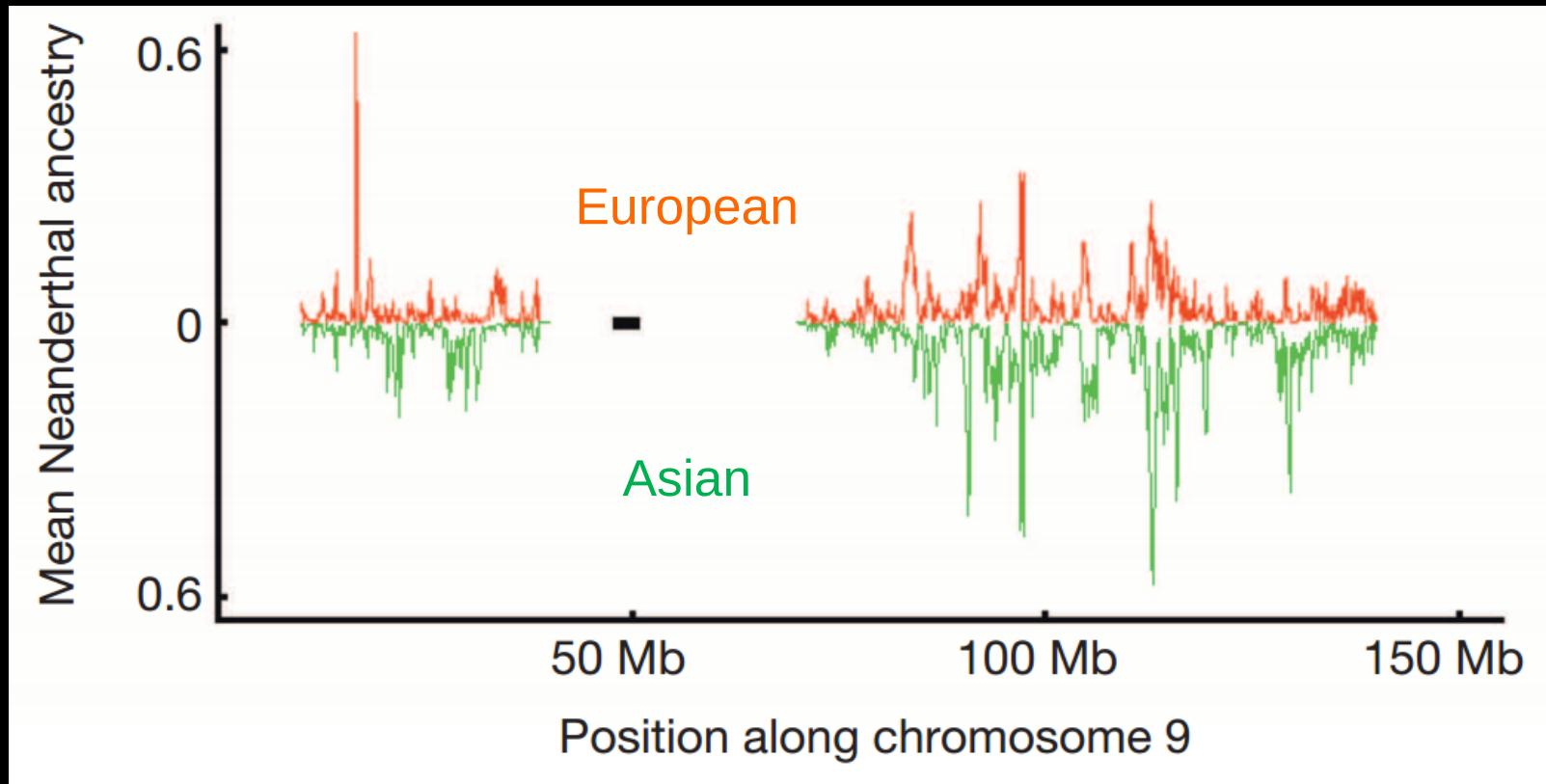
1. A highly diverged, derived haplotype in European individuals.
2. More similar to Neanderthal than African allele, and
3. Haplotype of expected span ( $0.05$  cM) given the time, 2000 generations.



# Only non-Africans have substantial Neanderthal ancestry



# Neanderthal ancestry proportion varies along the genome

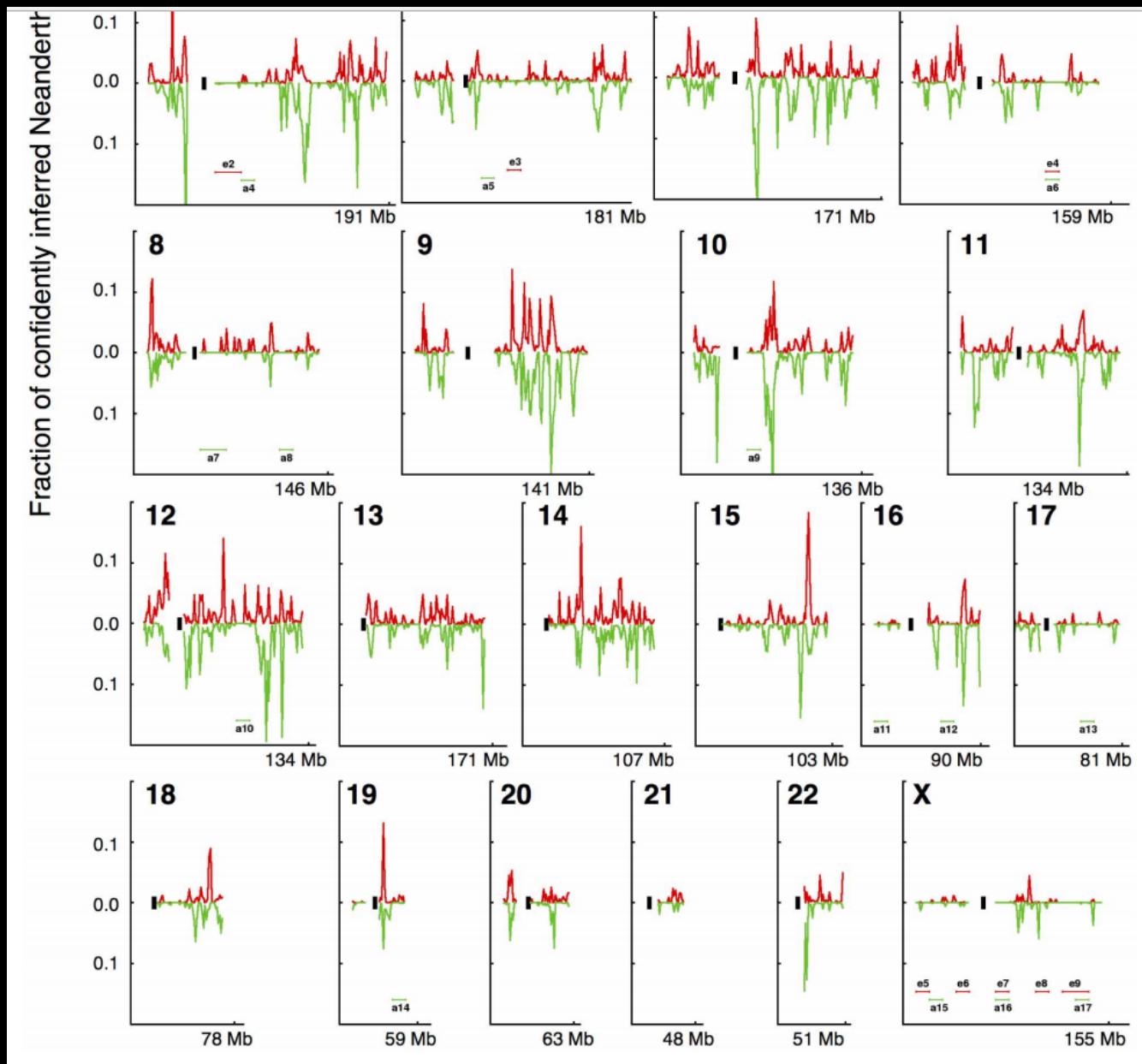


# Non-African genomes are up to 2% Neanderthal

**Table 1 | Genome-wide estimates of Neanderthal ancestry**

Region	Population	Number of individuals	Neanderthal ancestry on autosomes (%)
Europe	CEU	85	1.17 ± 0.08
	FIN	93	1.20 ± 0.07
	GBR	89	1.15 ± 0.08
	IBS	14	1.07 ± 0.06
	TSI	98	1.11 ± 0.07
East Asia	CHB	97	1.40 ± 0.08
	CHS	100	1.37 ± 0.08
	JPT	89	1.38 ± 0.10
America	CLM	60	1.14 ± 0.12
	MXL	66	1.22 ± 0.09
	PUR	55	1.05 ± 0.12
Africa	LWK	97	0.08 ± 0.02
	ASW	61	0.34 ± 0.22

# Note the low level of introgression of the X



# Several disease alleles appear to have come from Neanderthal

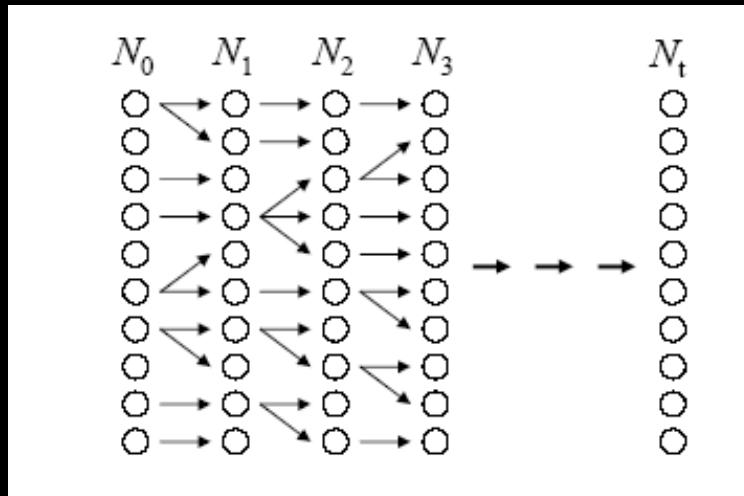
Extended Data Table 2 | Neanderthal-derived alleles that have been associated with phenotypes in genome-wide association studies

rs id	Coordinates	Derived allele	Derived allele frequency (%)		Phenotype
			Europeans	East Asians	
rs12531711	7:128,617,466	G	10.03	0.17	Systemic lupus erythematosus, Primary biliary cirrhosis
rs3025343	9:136,478,355	A	8.44	0.00	Smoking behavior
rs7076156	10:64,415,184	A	26.52	8.74	Crohn's disease
rs12571093	10:70,019,371	A	16.35	14.86	Optic disc size
rs1834481	11:112,023,827	G	21.50	0.35	Interleukin-18 levels
rs11175593	12:40,601,940	T	1.98	3.32	Crohn's disease
rs75493593	17:6,945,087	T	1.85	12.06	
rs75418188	17:6,945,483	T	1.85	11.54	Type-2 Diabetes
rs117767867	17:6,946,330	T	1.85	11.54	

# **Random Genetic Drift**

What can we infer from allele frequency dynamics of every nucleotide in the genome?

# The Wright-Fisher drift model: the Null model for Evolve-and-Resequence



- Selfing allowed
- Random mating
- Non-overlapping generations
- Constant population size
- No migration
- No selection

IGV

File View Tracks Help

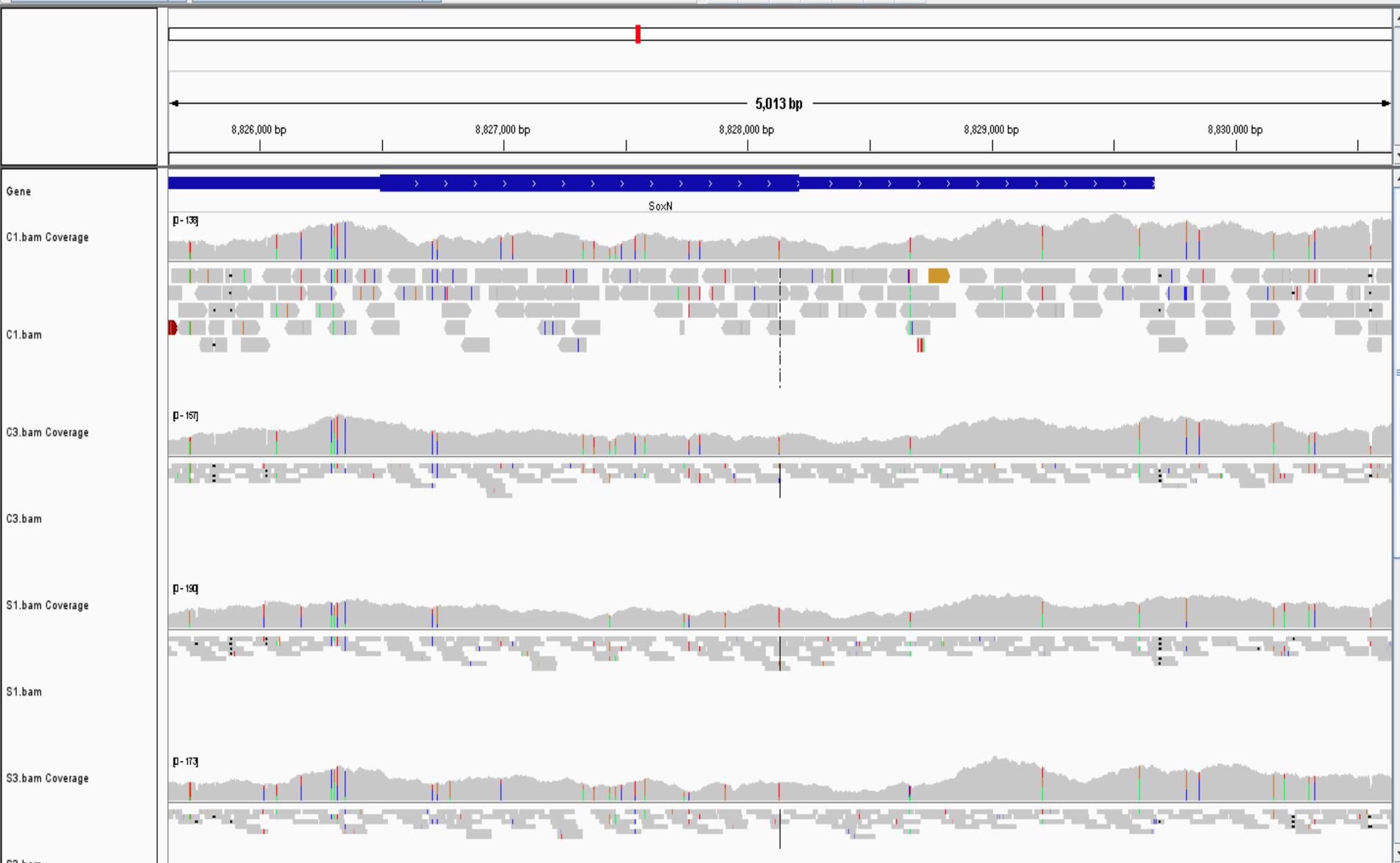
D. melanogaster (5.9) ▾ 2L

▼ 2L:8,825,625-8,829,670

Go



- +

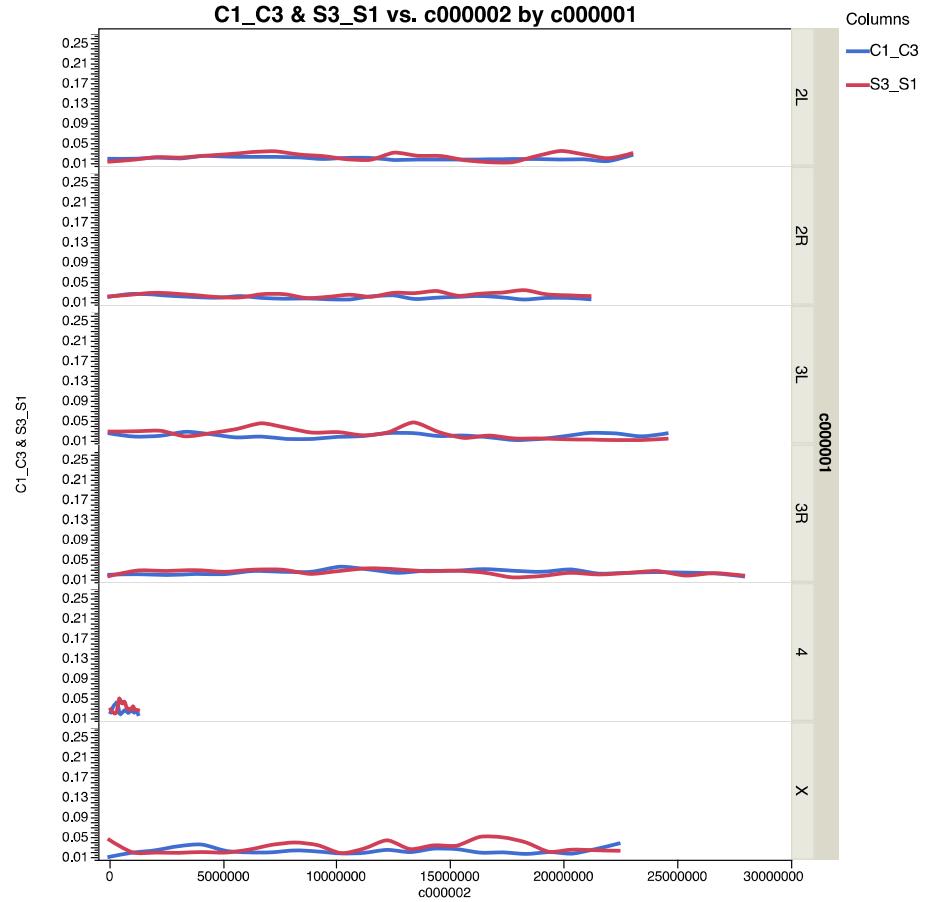
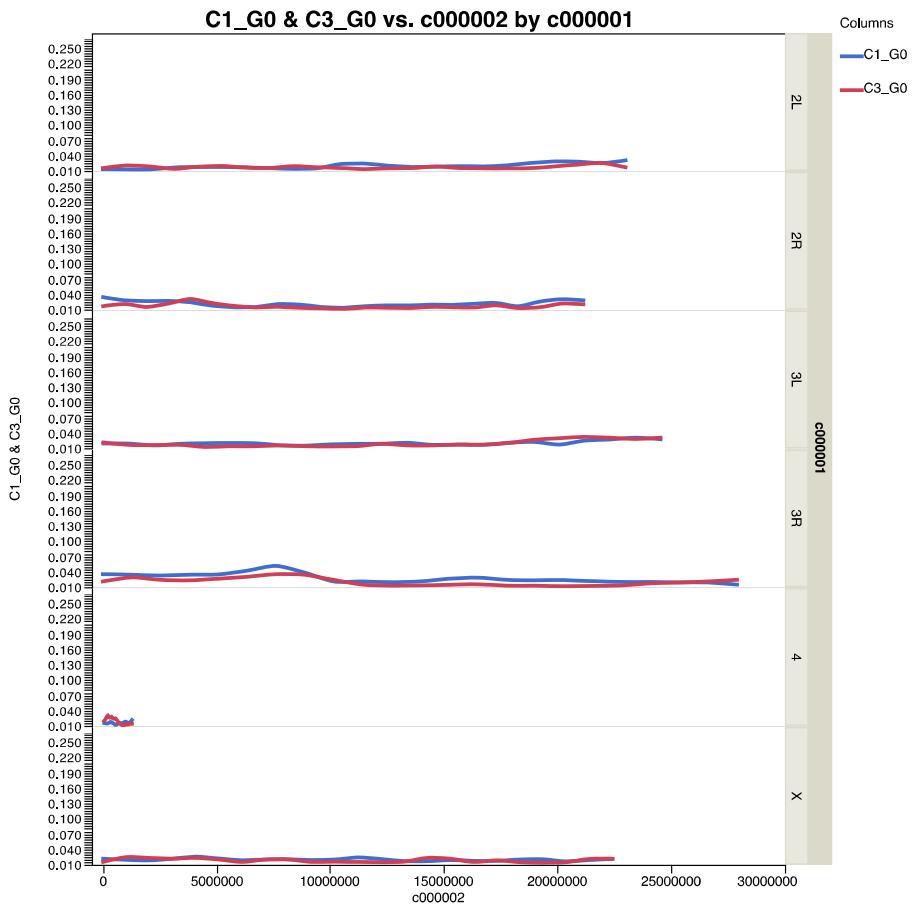


2L:8,827,848

410M of 700M

# mean allele frequencies under Wright-Fisher do not change

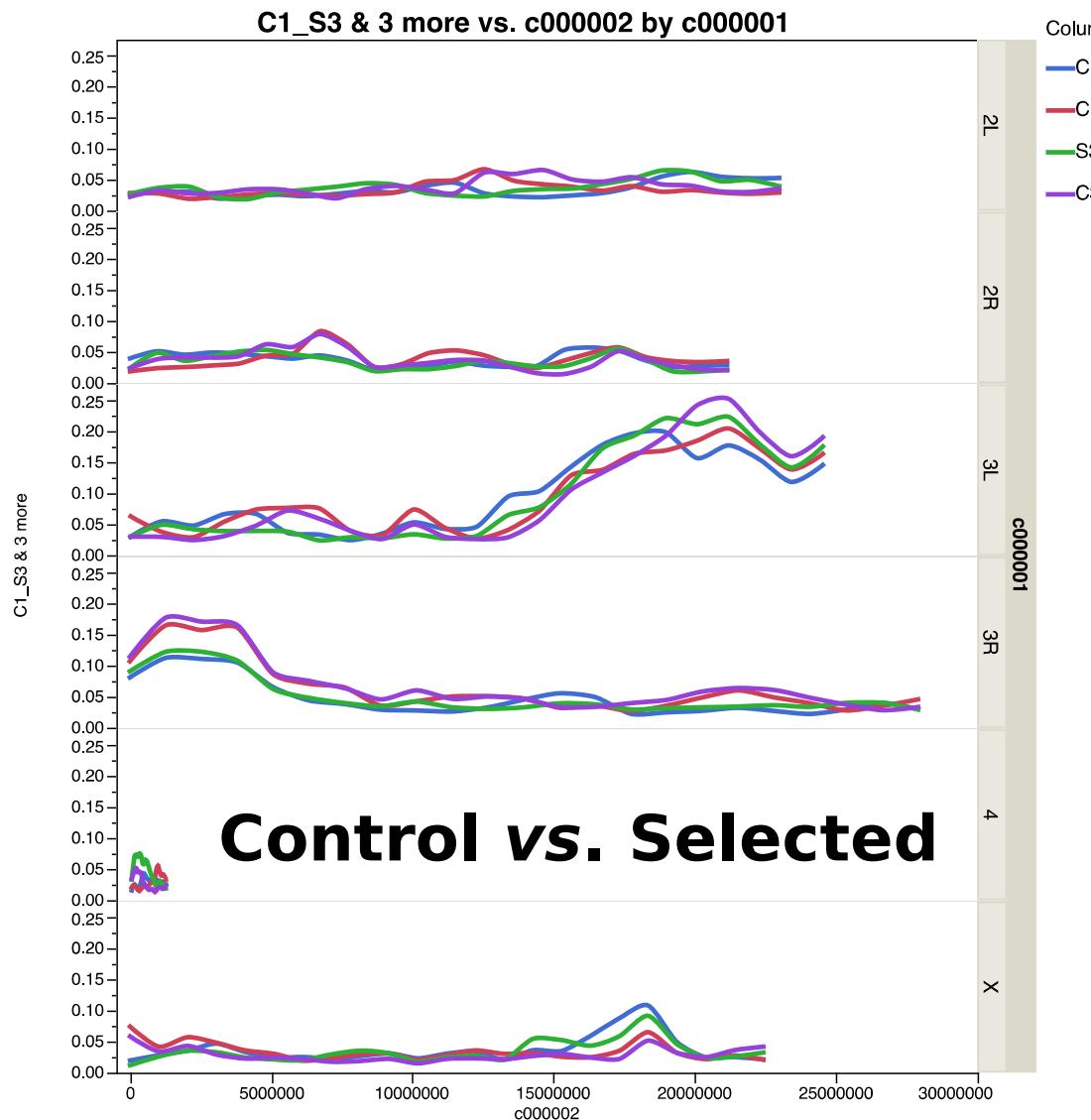
## Scored by $F_{ST}$ (in sliding window 500 bp)



**Control vs. G0**

**Control vs. Control  
Selected vs. Selected**

# Genomic regions show consistent changes in allele frequency



# Modeling the two sources of variation

Random genetic drift (Wright-Fisher)

Error in estimating allele frequencies

# Modeling the two sources of variation

Random genetic drift (Wright-Fisher)  
**Process error**

Error in estimating allele frequencies

# Modeling the two sources of variation

Random genetic drift (Wright-Fisher)  
**Process error**

Error in estimating allele frequencies  
**Measurement error**

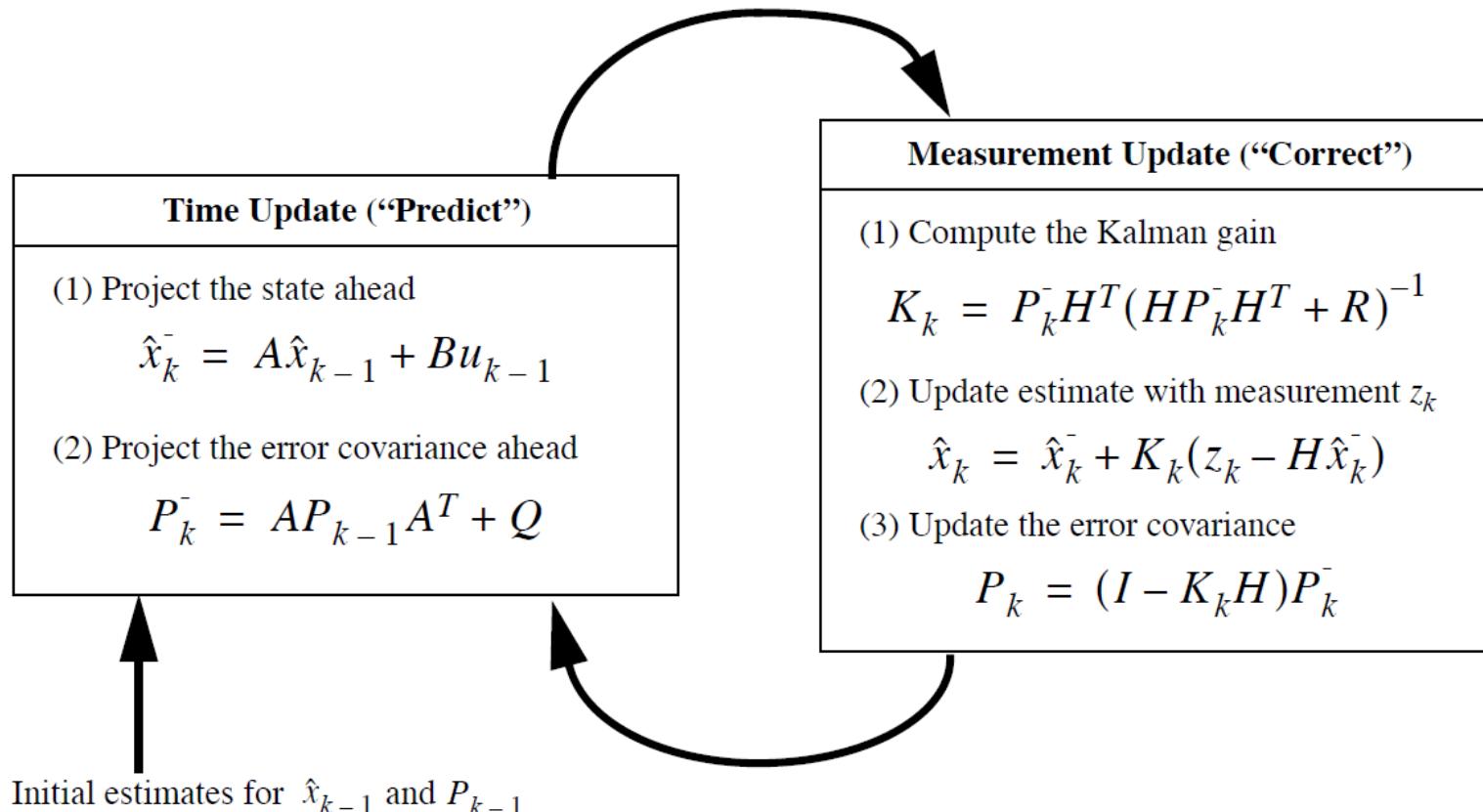
# Modeling the two sources of variation

Random genetic drift (Wright-Fisher)  
**Process error**

Error in estimating allele frequencies  
**Measurement error**

Fit this two-stage model, and identify a subset of sites that fit a model with selection better.

# Kalman filter finds the minimum squared error (MSE) solution

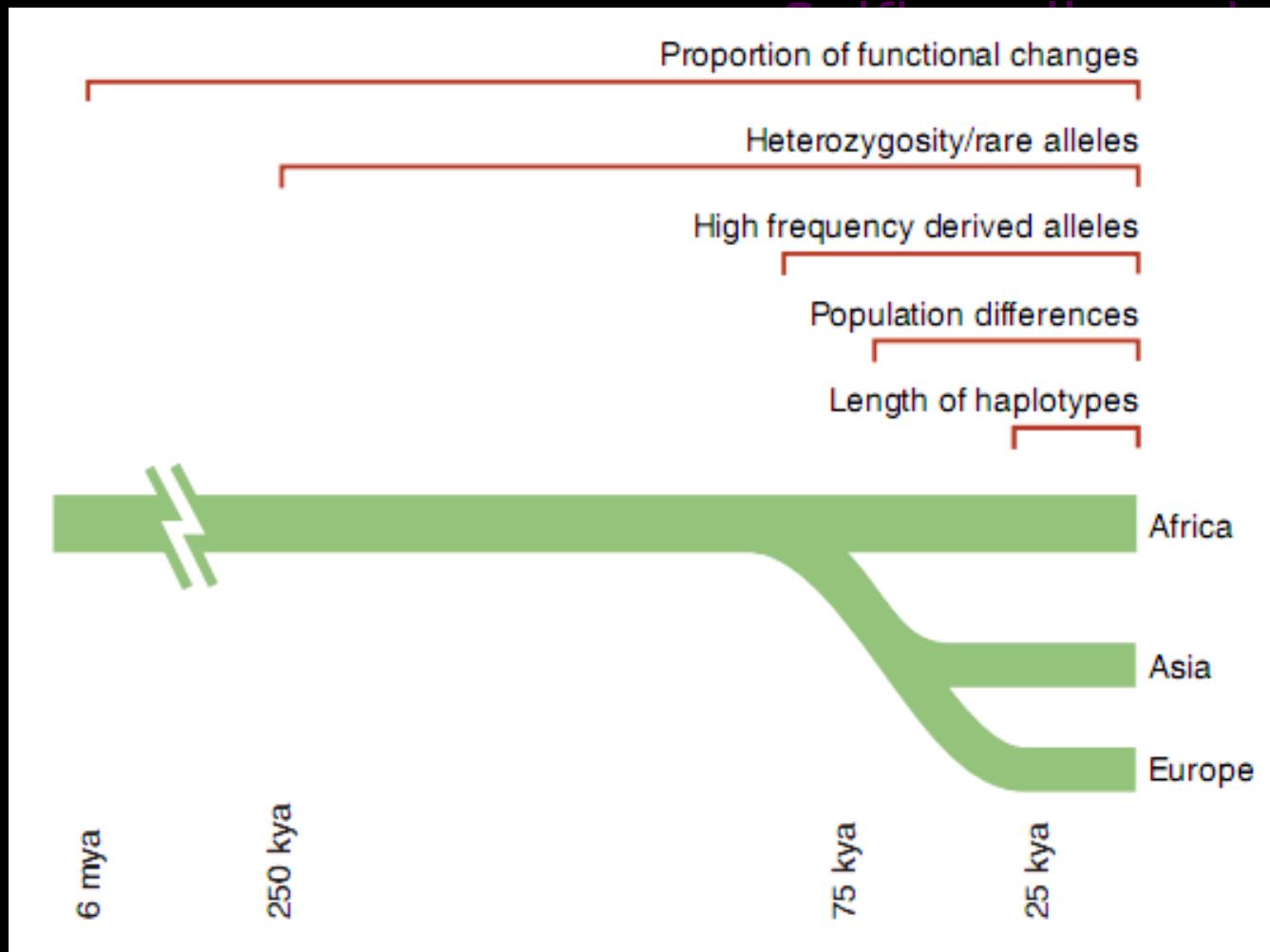


# Natural Selection

How best to do inference of selection  
genome-wide?

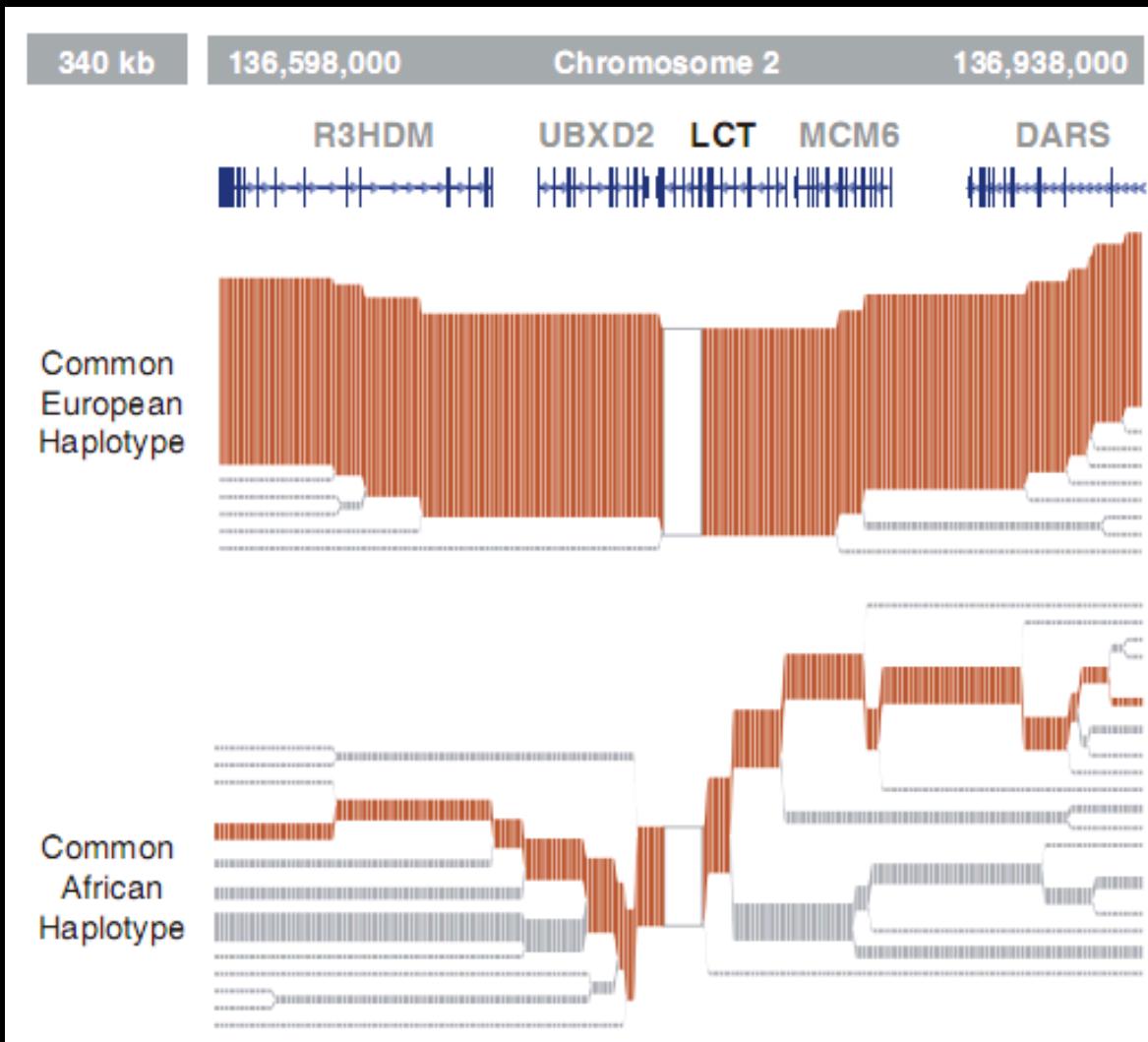
# Scans for selection

## Different signals at different time depths



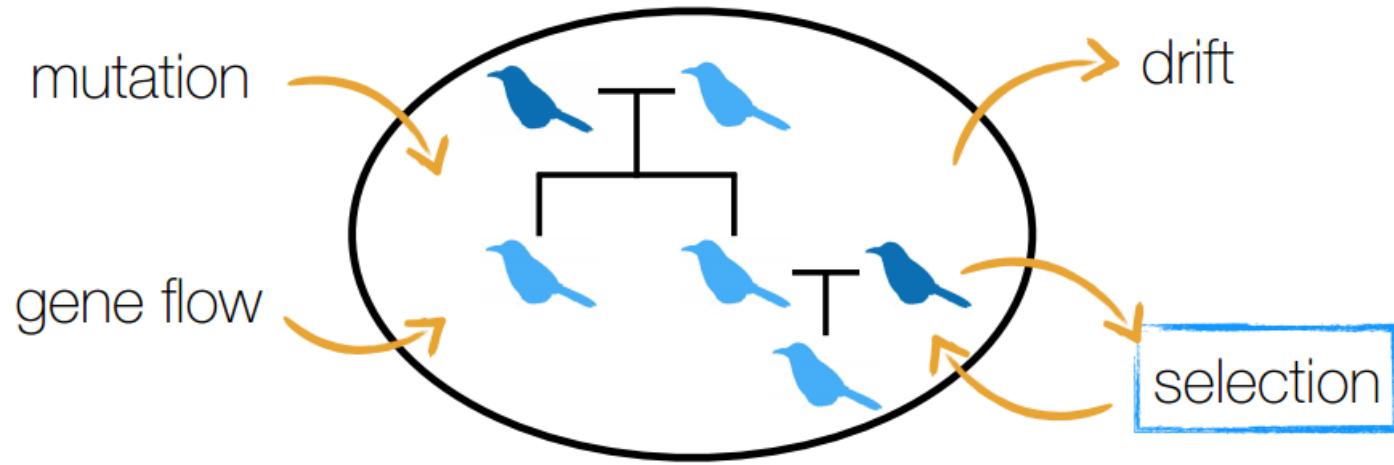
- Sabeti et al. 2006 Science

# Impressive selective sweeps (are rare in humans)

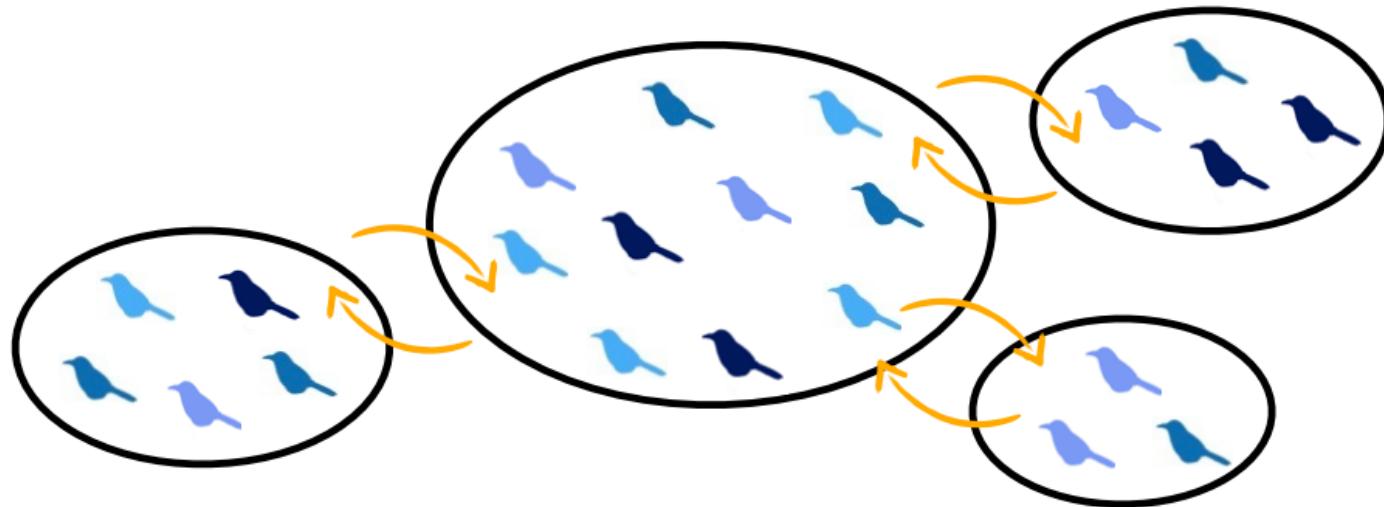


- Sabeti et al. 2006 Science

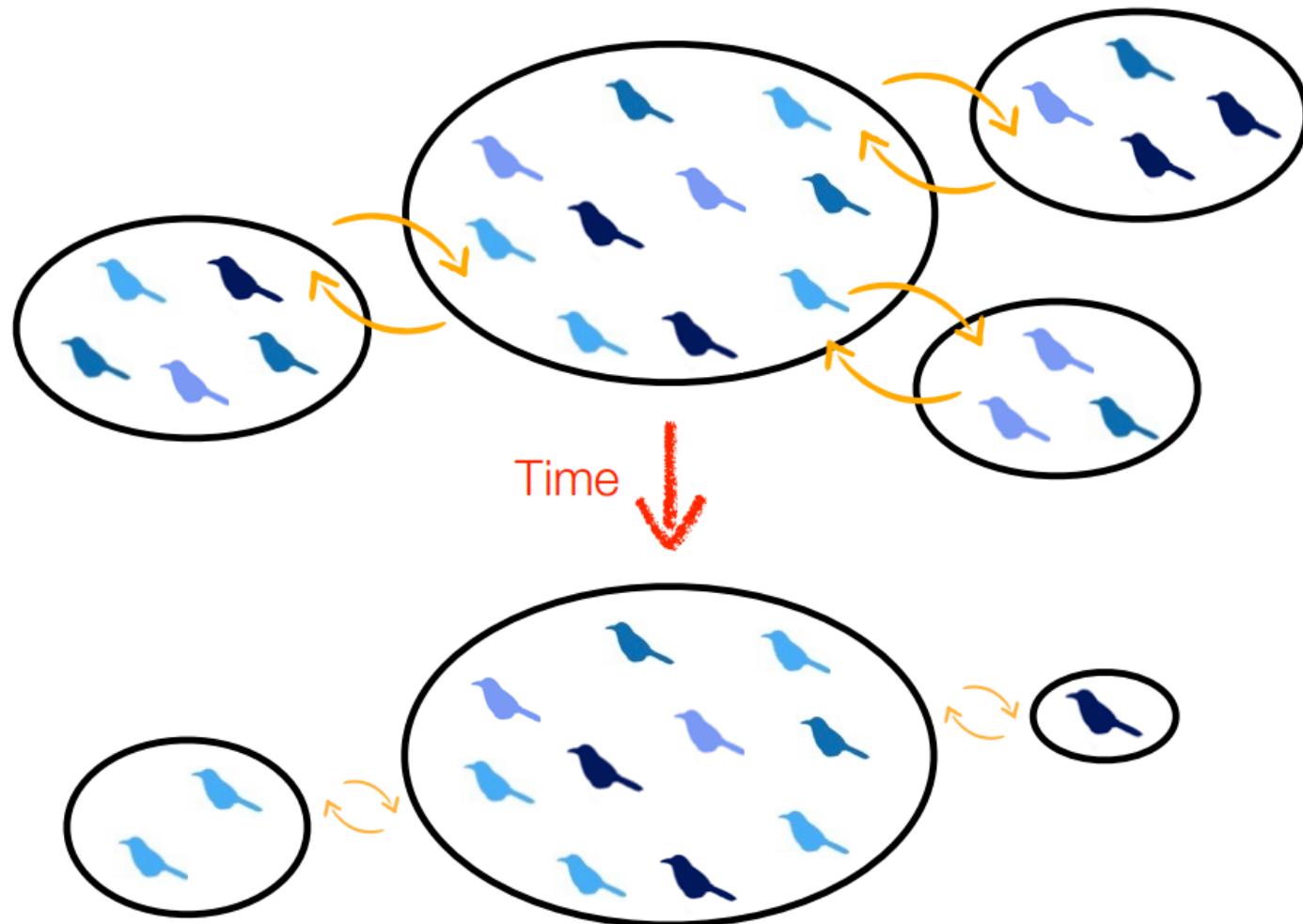
# The genetic basis of contemporary evolution in nature



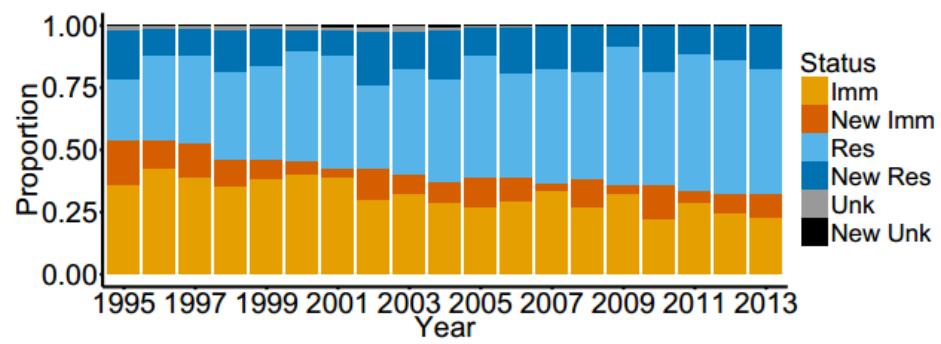
# Consequences of habitat fragmentation



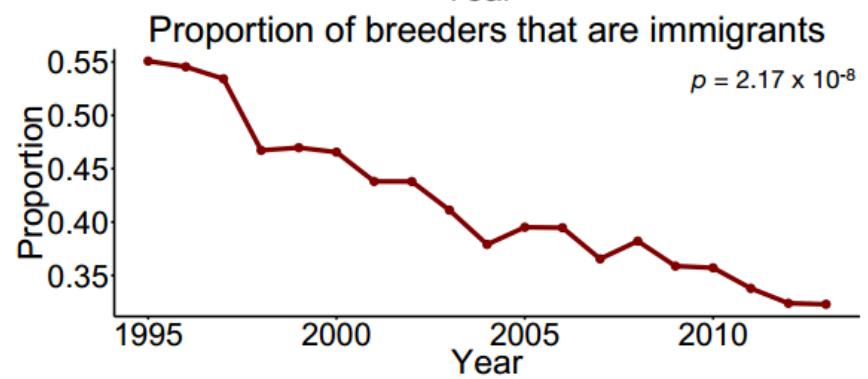
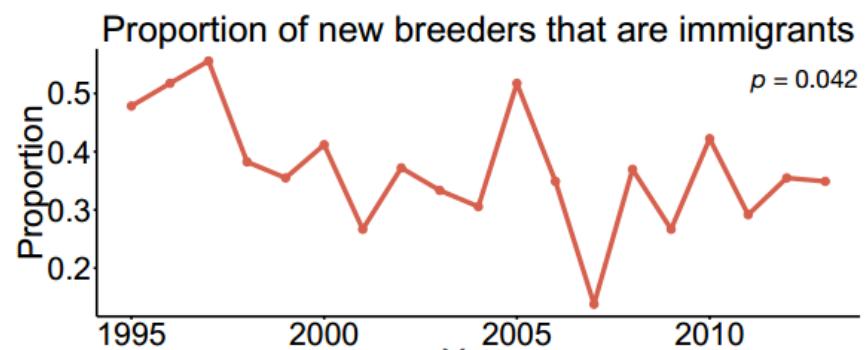
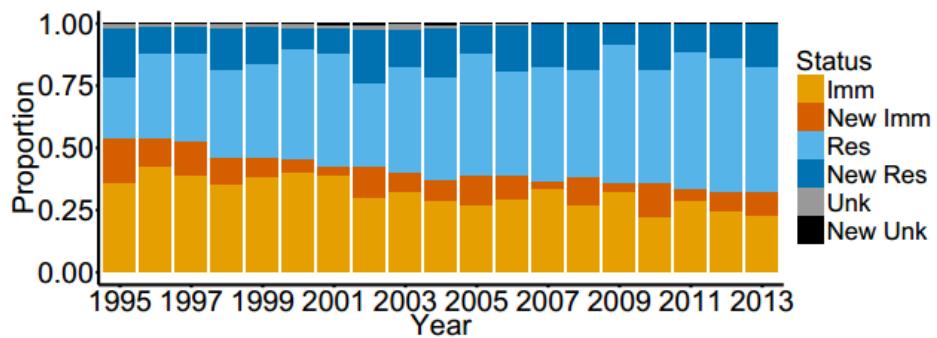
# Consequences of habitat fragmentation



# Immigration rate decreased over time



# Immigration rate decreased over time



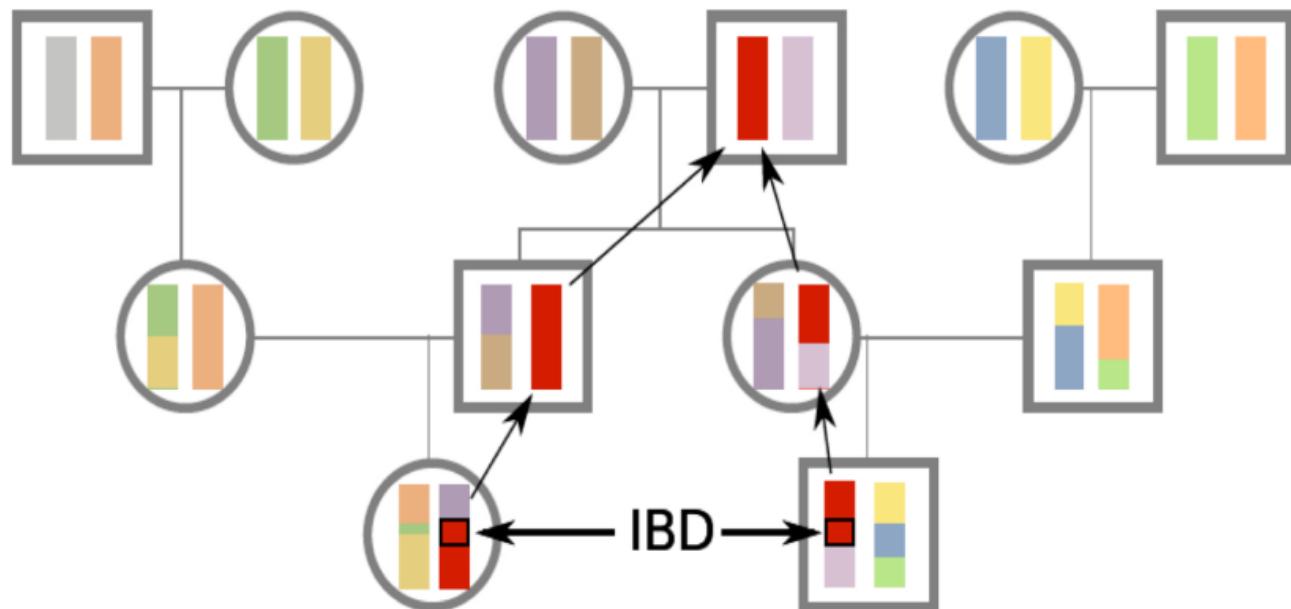
# Measures of genetic variation

7,834 autosomal SNPs in linkage equilibrium

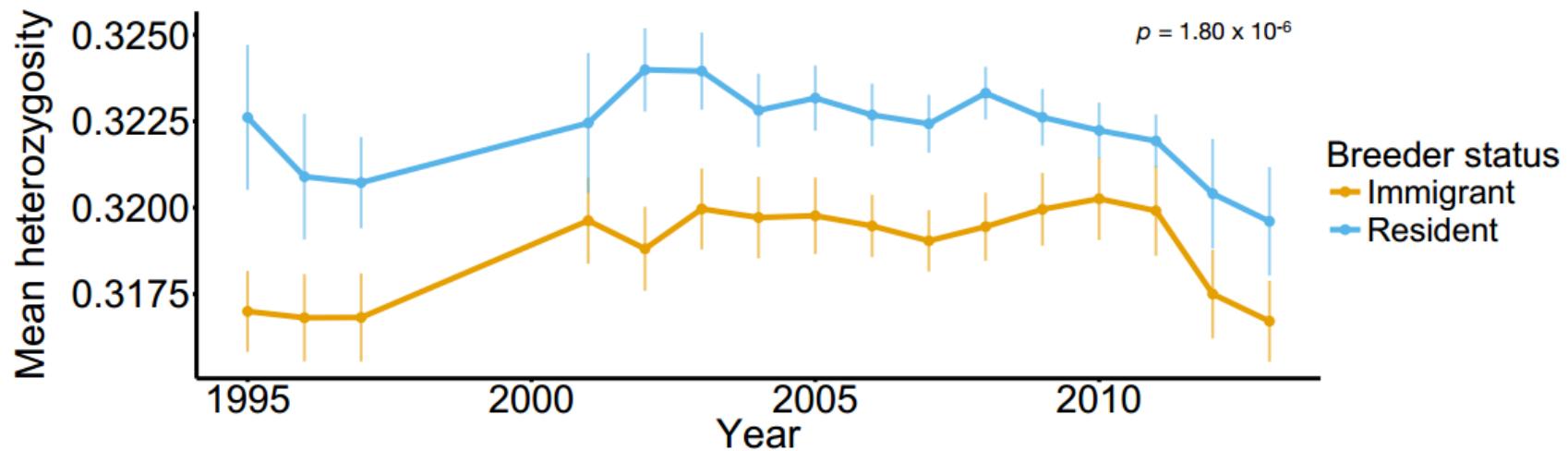
Mean observed heterozygosity

Inbreeding coefficient

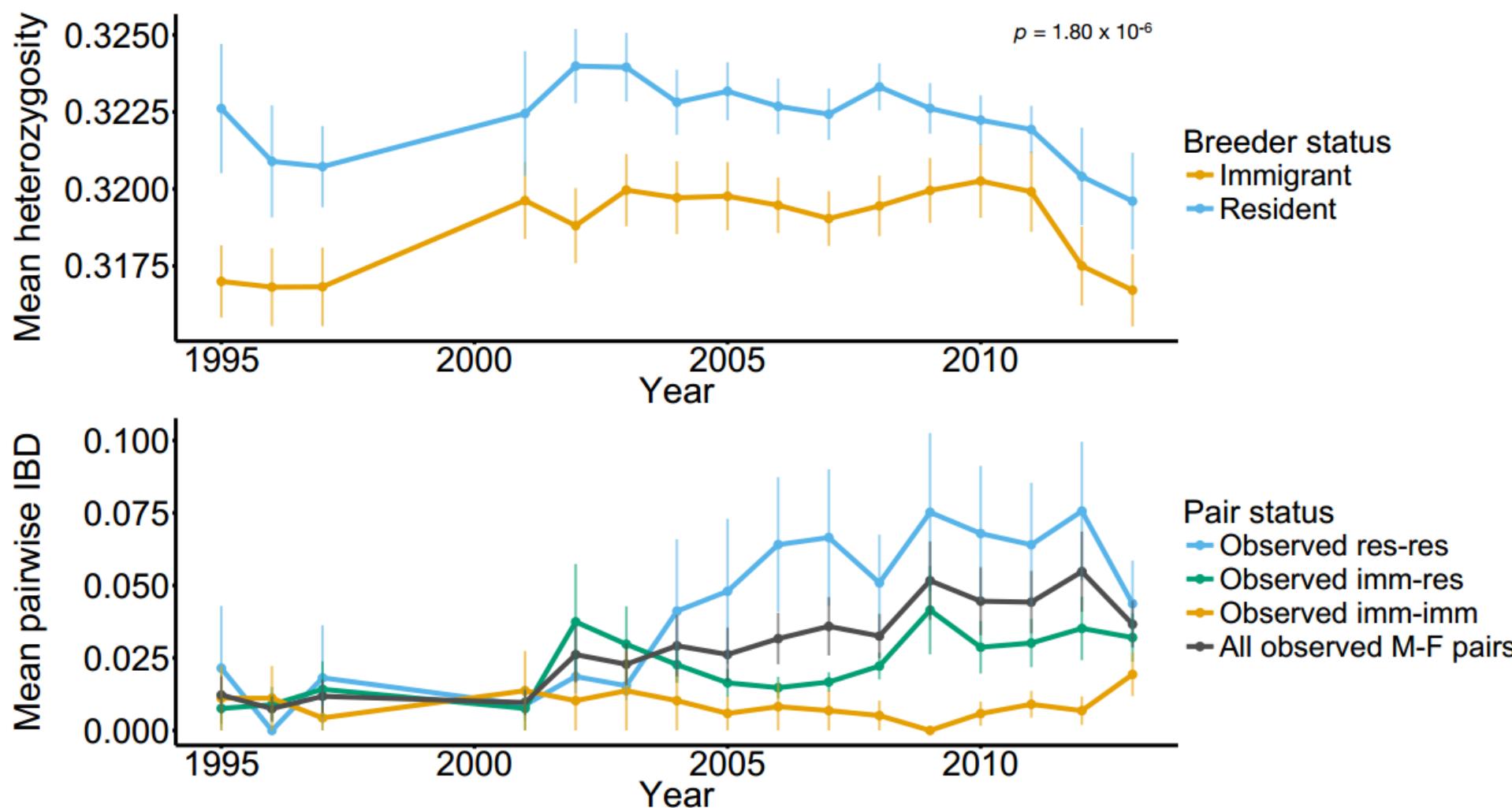
Pairwise Identity By Descent (IBD) values



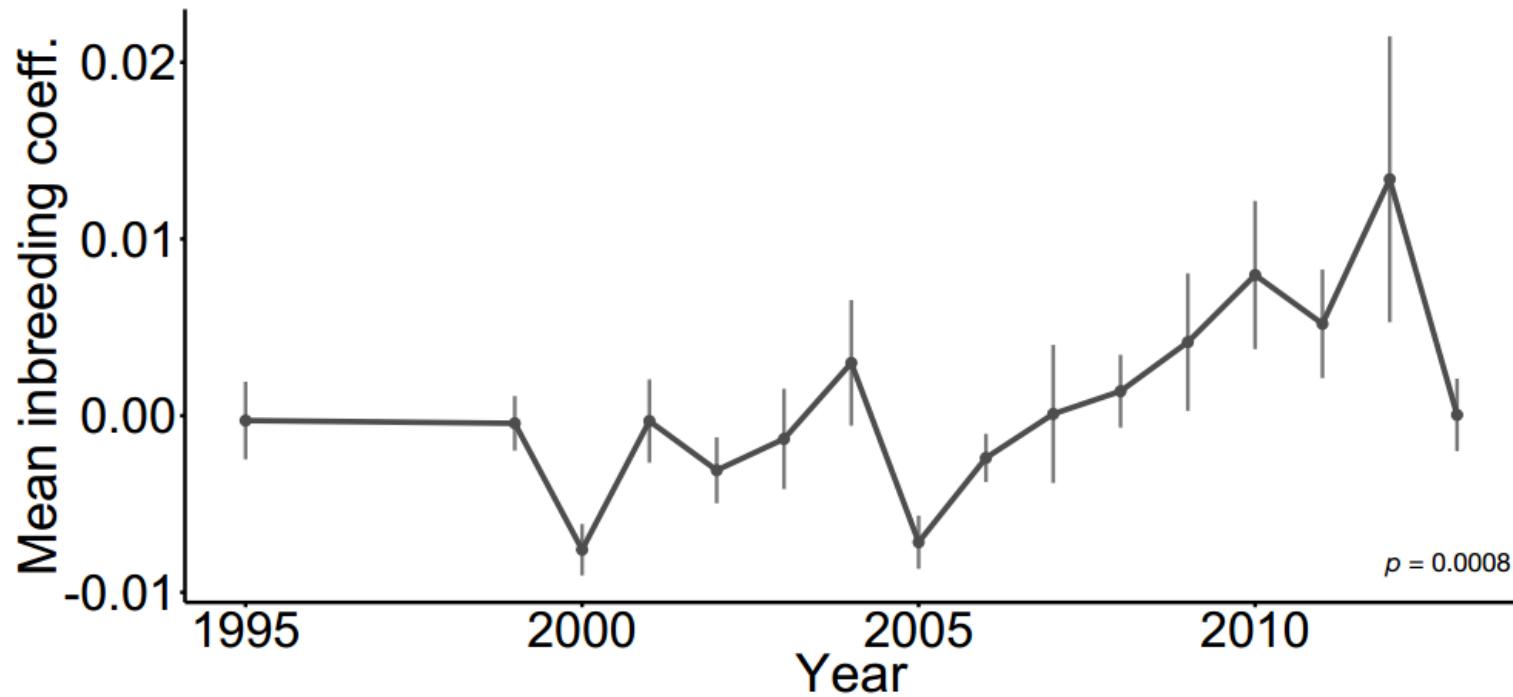
Immigrants were less heterozygous than residents



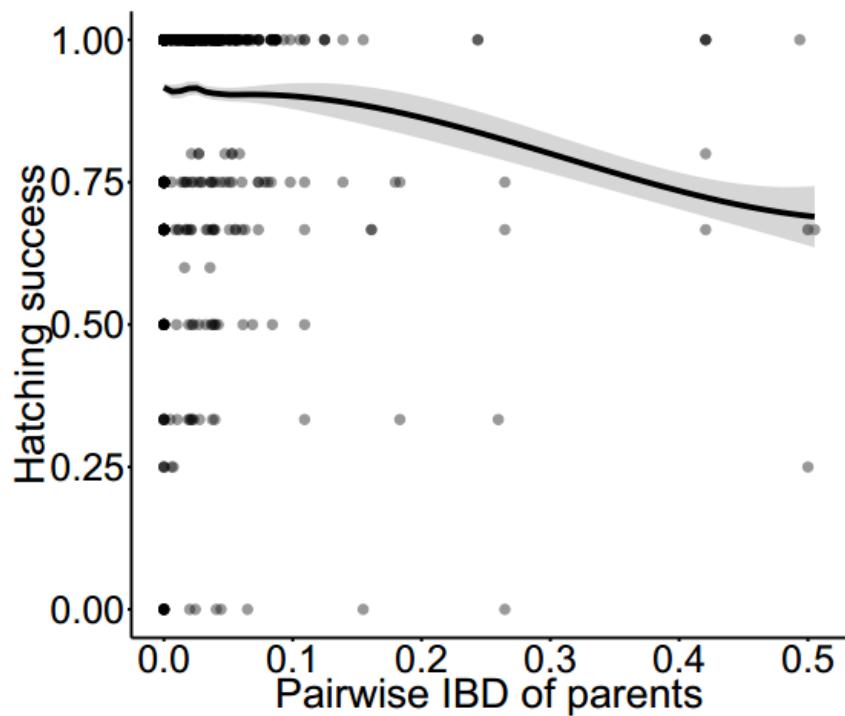
Immigrants were less heterozygous than residents  
but still contributed genetic variation



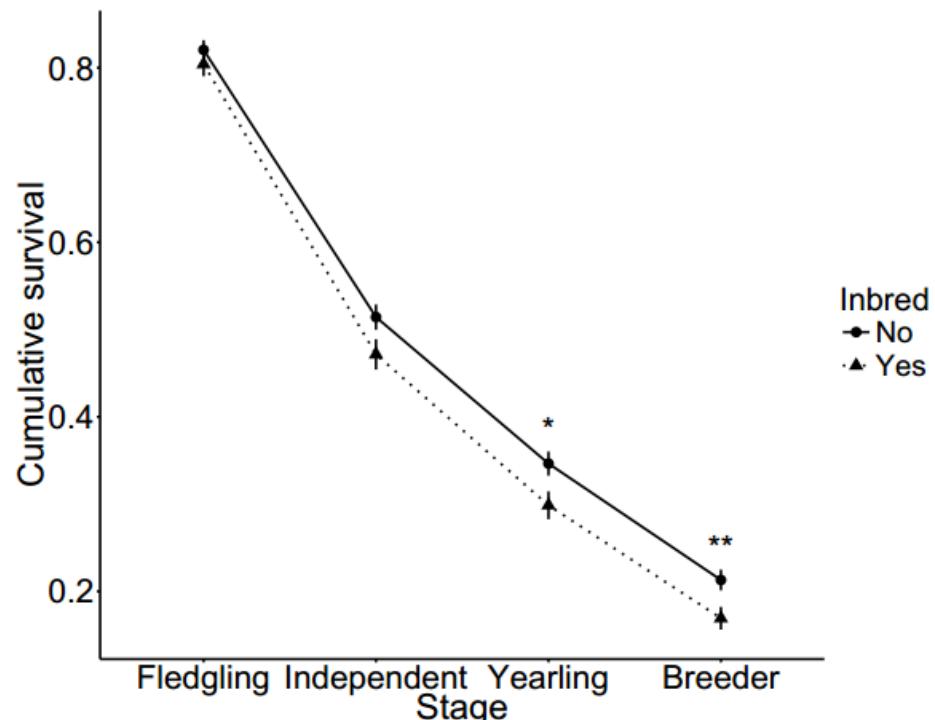
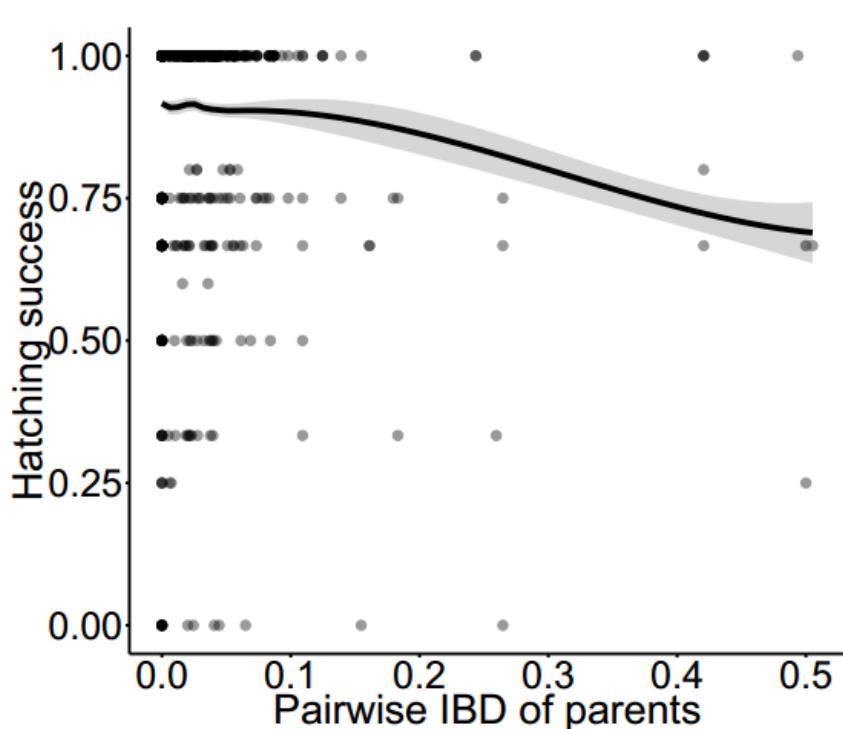
# Mean inbreeding coefficient of the birth cohort increased over time



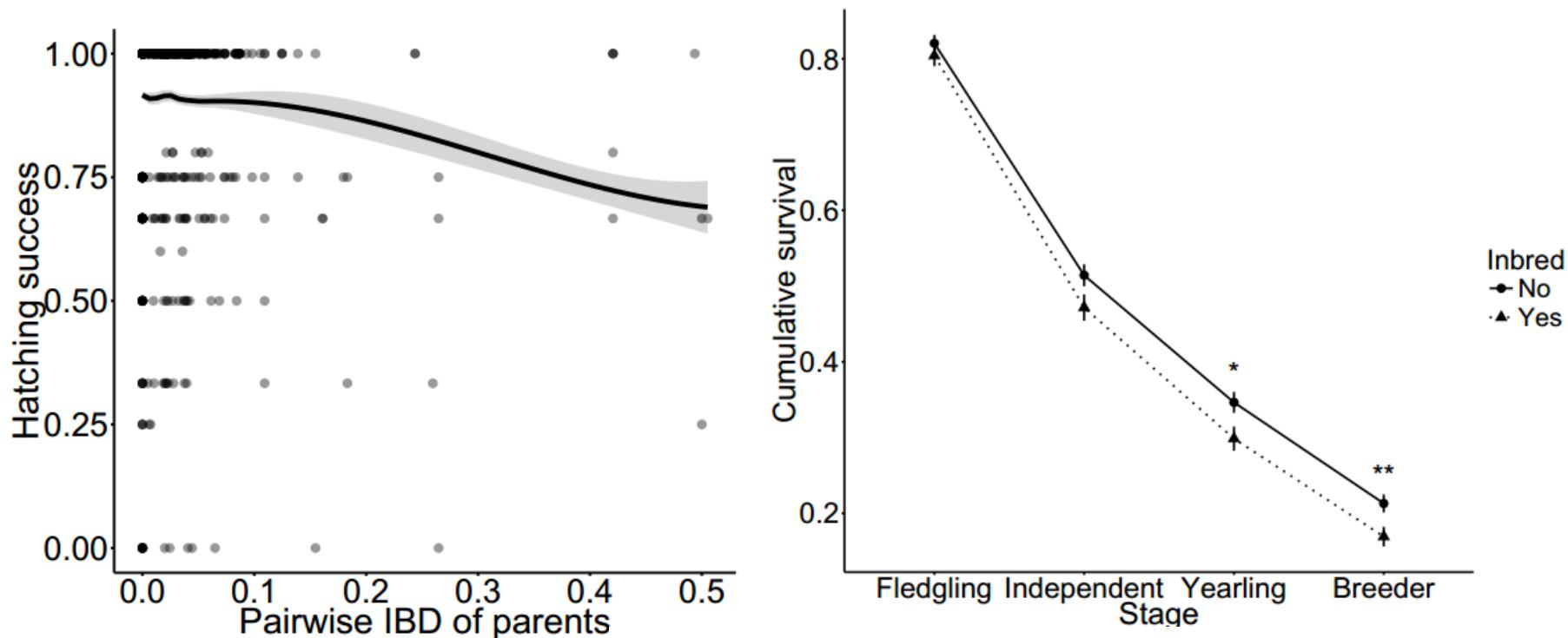
# Inbreeding depression in multiple life-history stages



# Inbreeding depression in multiple life-history stages



# Inbreeding depression in multiple life-history stages



- ✓ Hatching success
- ✓ Nestling weight
- ✓ Juvenile survival
- ✓ Breeder lifespan
- ✓ Lifetime reproductive success

# Isolation-by-distance is a consequence of dispersal

ISOLATION BY DISTANCE\*

SEWALL WRIGHT

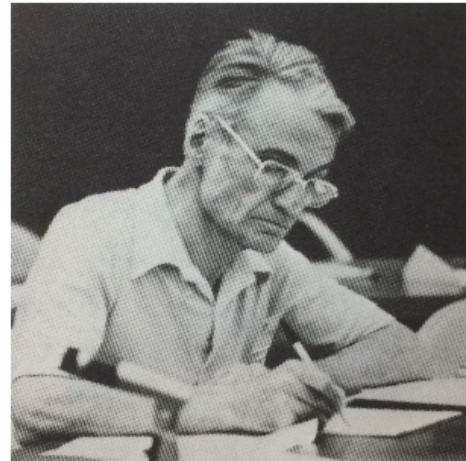
*The University of Chicago*<sup>1</sup>

Received November 9, 1942

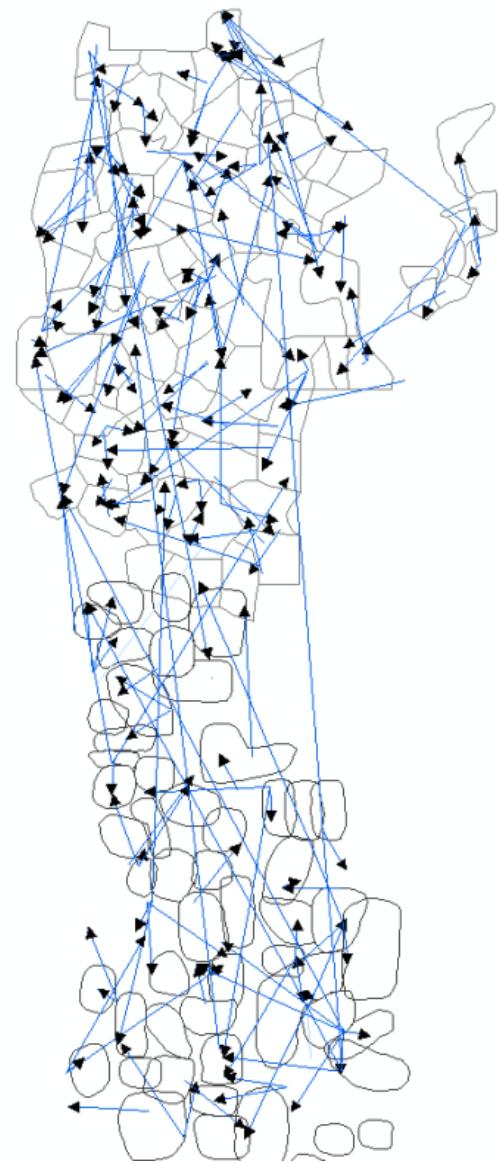
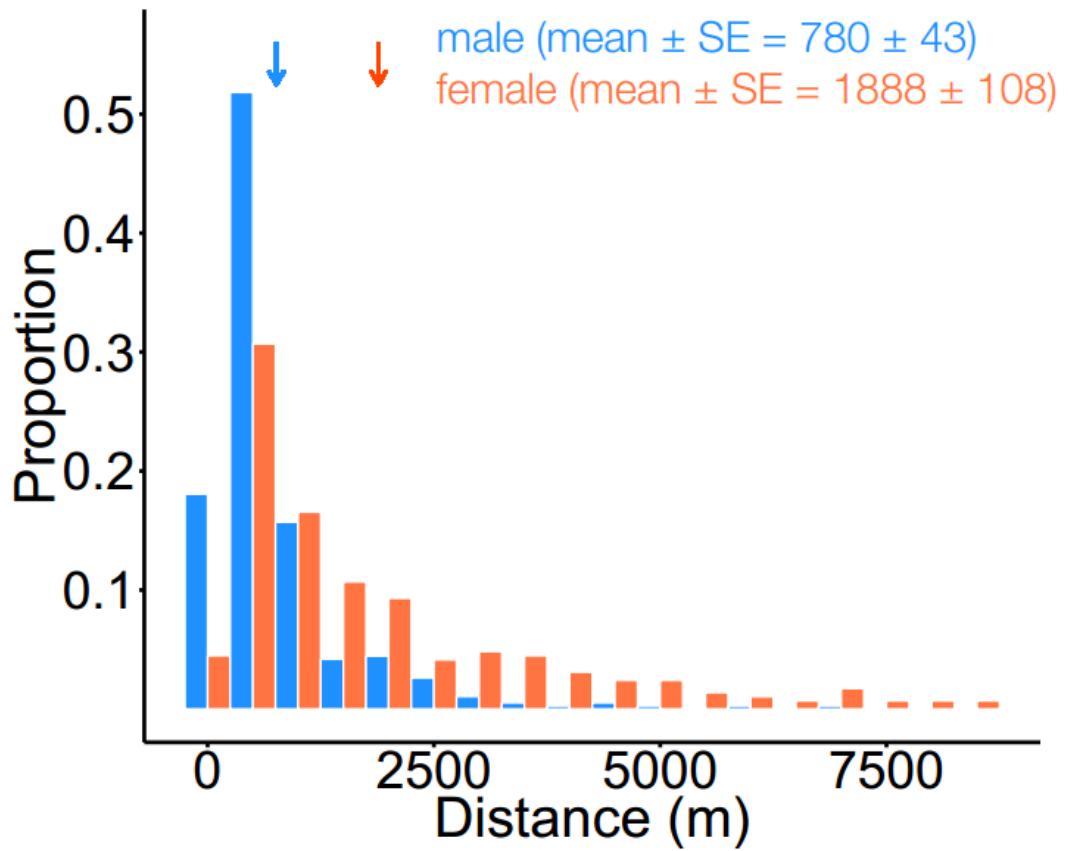
GÉNÉTIQUE DES POPULATIONS DIPLOÏDES  
NATURELLES DANS LE CAS D'UN SEUL LOCUS

III. — PARENTÉ, MUTATIONS ET MIGRATION

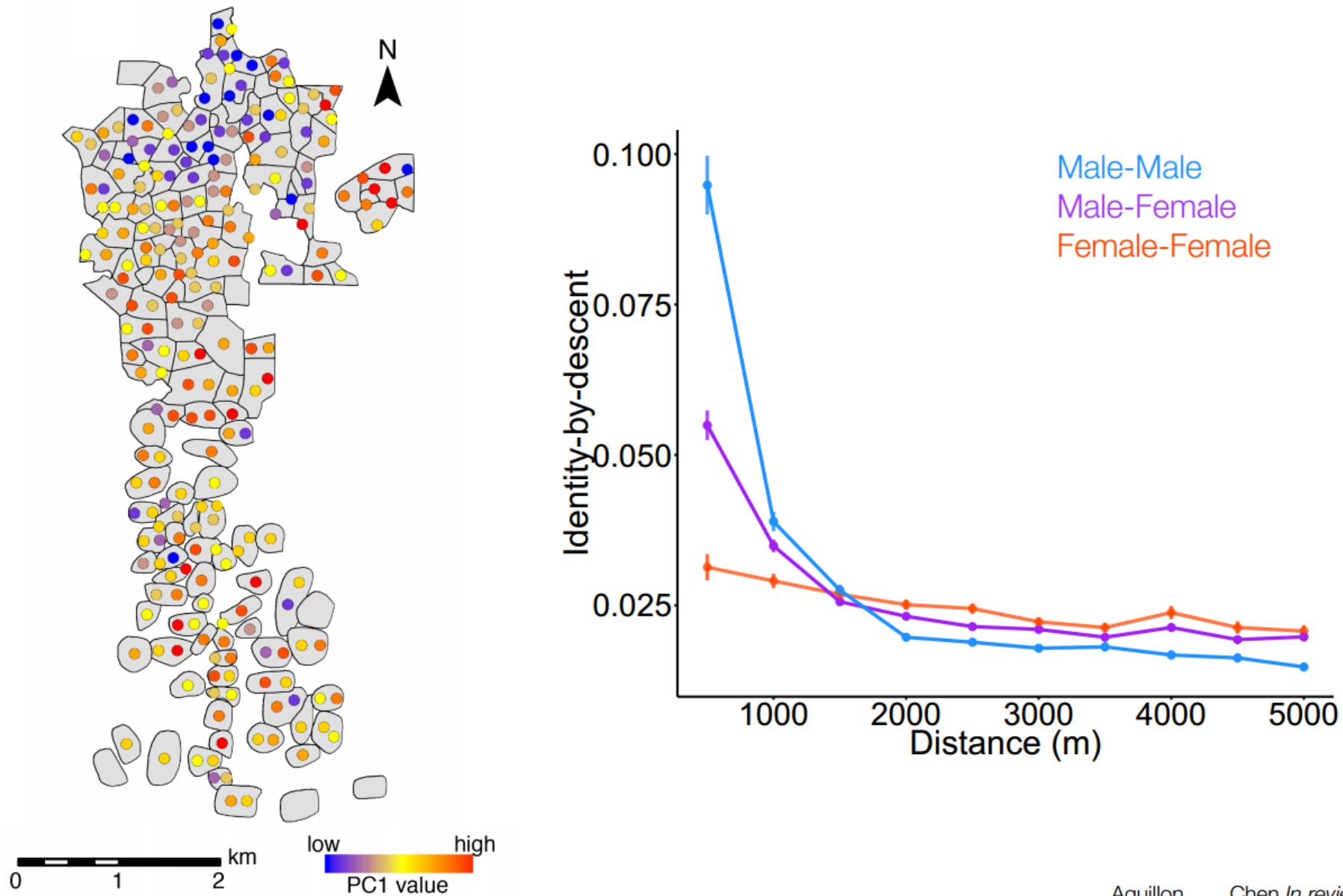
G. MALÉCOT



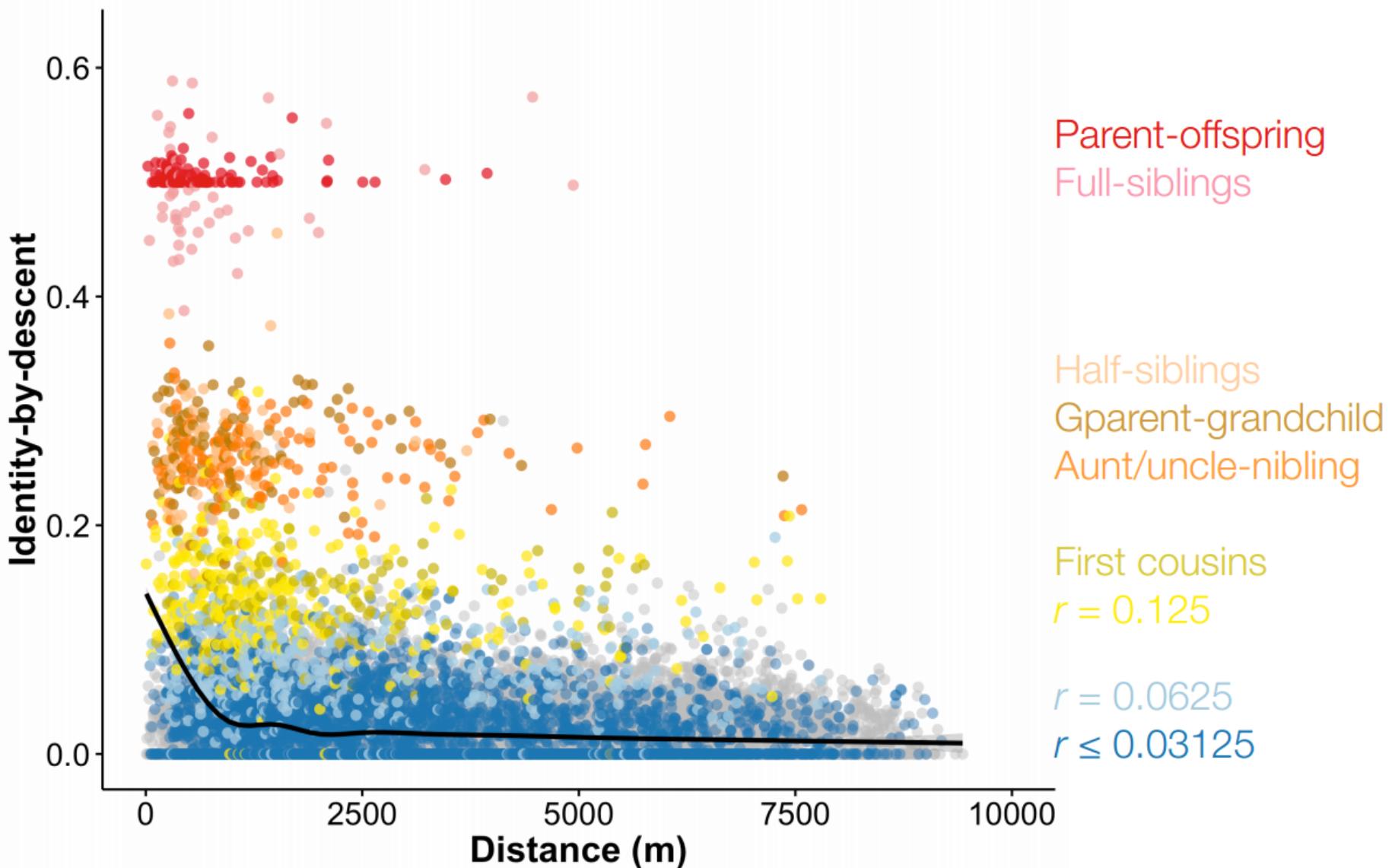
Females disperse farther than males



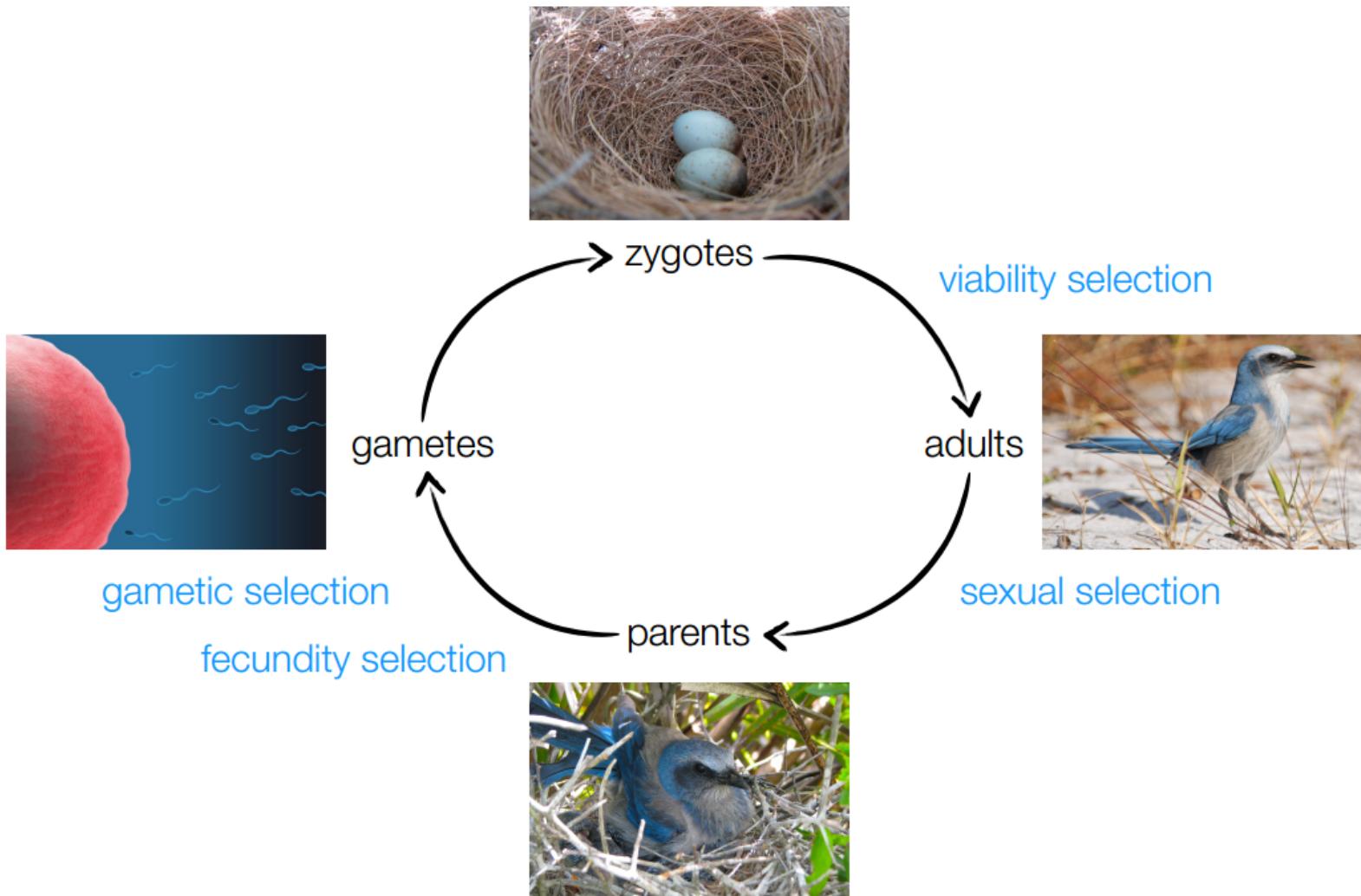
# Limited dispersal leads to isolation-by-distance



# A closer look at the isolation-by-distance pattern



# Selection can act at different stages of the life cycle



# Mutation Spectrum

What can we learn about mutation from sequence data?

# How rapidly do rates and patterns of mutations evolve?

Kelley Harris: tally rare variants with flanking base context (16 flanking pairs of bases for each of 6 transitions and transversions)

Frequencies of the 96 mutational classes differ significantly across populations.

TCC → TAC was largest difference

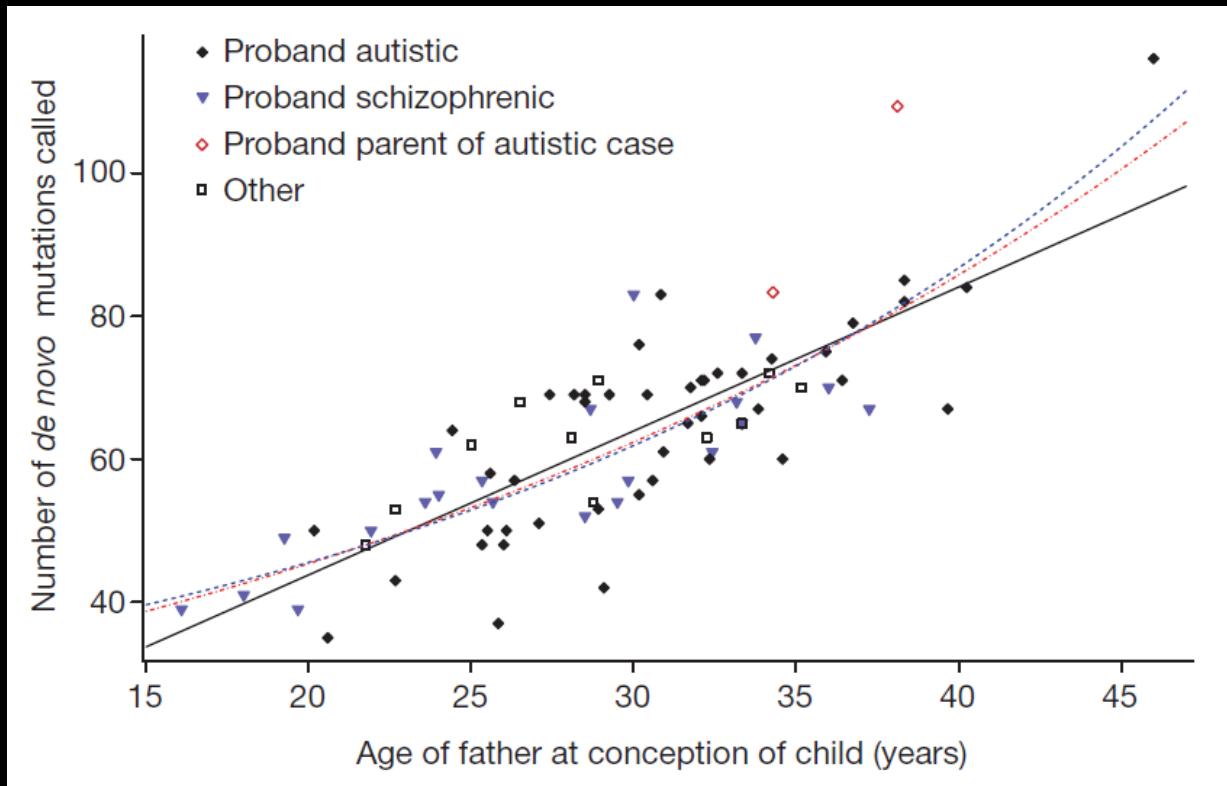
## Rate of *de novo* mutations and the importance of father's age to disease risk

Augustine Kong<sup>1</sup>, Michael L. Frigge<sup>1</sup>, Gisli Masson<sup>1</sup>, Soren Besenbacher<sup>1,2</sup>, Patrick Sulem<sup>1</sup>, Gisli Magnusson<sup>1</sup>, Sigurjon A. Gudjonsson<sup>1</sup>, Asgeir Sigurdsson<sup>1</sup>, Aslaug Jonasdottir<sup>1</sup>, Adalbjorg Jonasdottir<sup>1</sup>, Wendy S. W. Wong<sup>3</sup>, Gunnar Sigurdsson<sup>1</sup>, G. Bragi Walters<sup>1</sup>, Stacy Steinberg<sup>1</sup>, Hannes Helgason<sup>1</sup>, Gudmar Thorleifsson<sup>1</sup>, Daniel F. Gudbjartsson<sup>1</sup>, Agnar Helgason<sup>1,4</sup>, Olafur Th. Magnusson<sup>1</sup>, Unnur Thorsteinsdottir<sup>1,5</sup> & Kari Stefansson<sup>1,5</sup>

Mutations generate sequence diversity and provide a substrate for selection. The rate of *de novo* mutations is therefore of major importance to evolution. Here we conduct a study of genome-wide mutation rates by sequencing the entire genomes of 78 Icelandic parent-offspring trios at high coverage. We show that in our samples, with an average father's age of 29.7, the average *de novo* mutation rate is  $1.20 \times 10^{-8}$  per nucleotide per generation. Most notably, the diversity in mutation rate of single nucleotide polymorphisms is dominated by the age of the father at conception of the child. The effect is an increase of about two mutations per year. An exponential model estimates paternal mutations

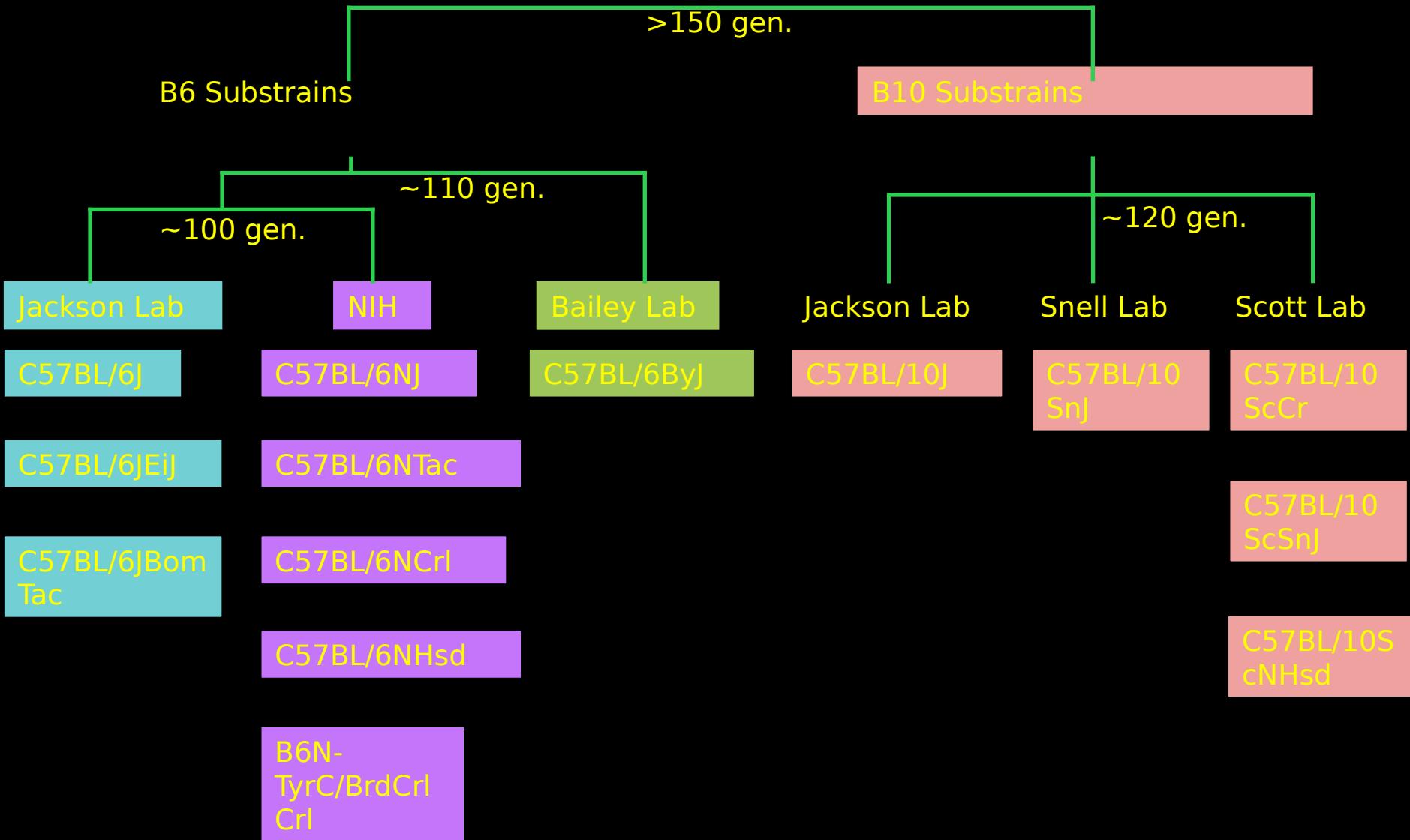
**Table 1 | De novo mutations observed with parental origin assigned**

	Father's age (yr)	Mother's age (yr)	Paternal chromosome	Maternal chromosome	Combined
Trio 1	21.8	19.3	39	9	48
Trio 2	22.7	19.8	43	10	53
Trio 3	25.0	22.1	51	11	62
Trio 4	36.2	32.2	53	26	79
Trio 5	40.0	39.1	91	15	106
Mean	29.1	26.5	55.4	14.2	69.6
s.d.	8.4	8.8	20.7	7.0	23.5
Variance	70.2	77.0	428.8	48.7	555.3

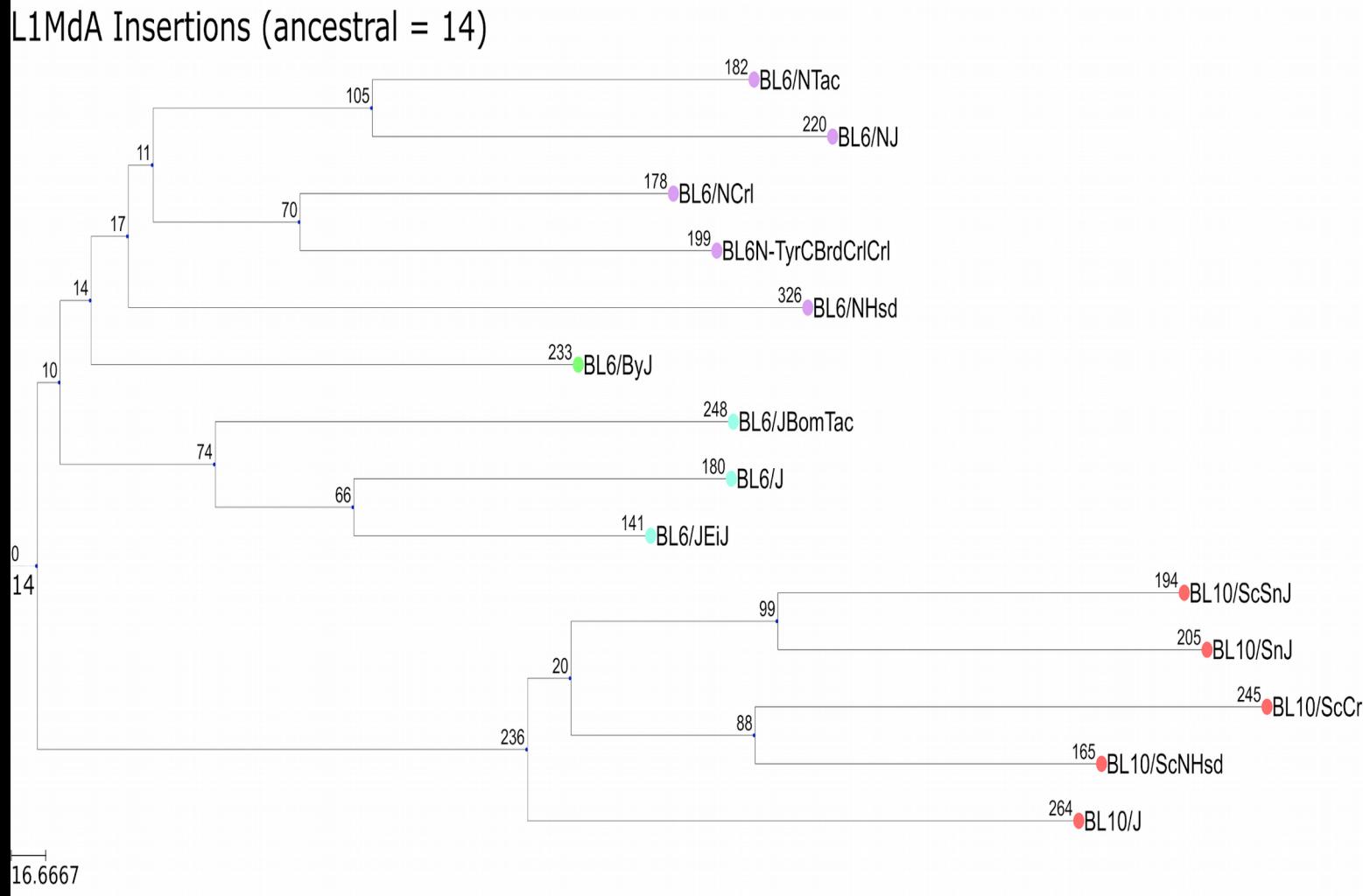


# I. Genealogy of Black6 and Black10 mice

Laboratory mice are a “natural” mutation accumulation experiment



# Gain of L1MdA insertions looks more uniform than SINEs, but still biased toward BL10



# Disease Association

How can variation in genome sequence be used to infer disease risk?

# **Population genetics and disease-causing variants**

Early onset disease-risk enhancing alleles ought to have lower allele frequency.

Loss-of-function alleles in particular ought to be rare.

Include these simple population genetic principles in approaches that infer likelihood of disease-causation.

# Variance explained



Peter Visscher

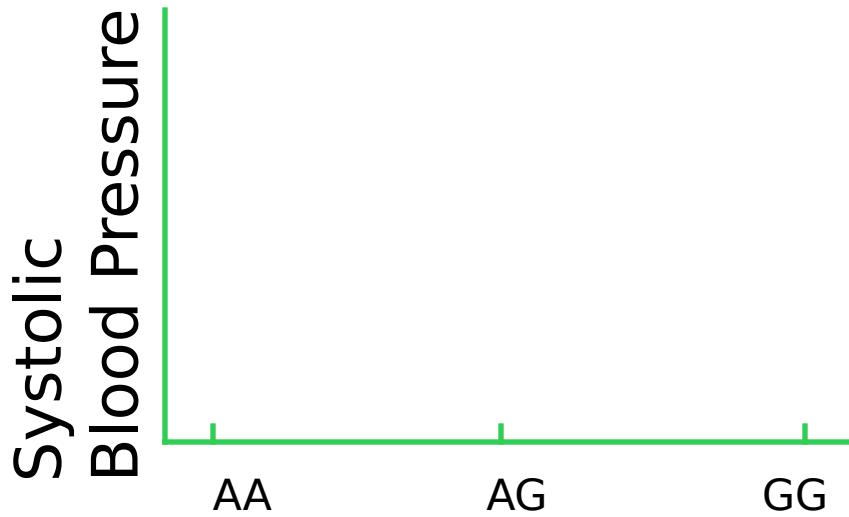
Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index

Jian Yang<sup>1,2,24</sup>, Andrew Bakshi<sup>1</sup>, Zhihong Zhu<sup>1</sup>, Gibran Hemani<sup>1,3</sup>, Anna A E Vinkhuyzen<sup>1</sup>, Sang Hong Lee<sup>1,4</sup>, Matthew R Robinson<sup>1</sup>, John R B Perry<sup>5</sup>, Ilja M Nolte<sup>6</sup>, Jana V van Vliet-Ostaptchouk<sup>6,7</sup>, Harold Snieder<sup>6</sup>, The LifeLines Cohort Study<sup>8</sup>, Tonu Esko<sup>9-12</sup>, Lili Milani<sup>9</sup>, Reedik Mägi<sup>9</sup>, Andres Metspalu<sup>9,13</sup>, Anders Hamsten<sup>14</sup>, Patrik K E Magnusson<sup>15</sup>, Nancy L Pedersen<sup>15</sup>, Erik Ingelsson<sup>16,17</sup>, Nicole Soranzo<sup>18,19</sup>, Matthew C Keller<sup>20,21</sup>, Naomi R Wray<sup>1</sup>, Michael E Goddard<sup>22,23</sup> & Peter M Visscher<sup>1,2,24</sup>

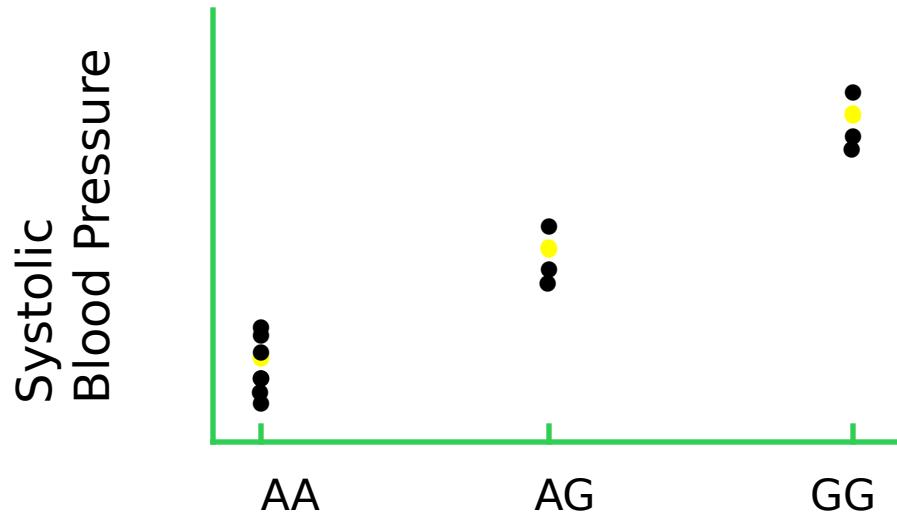
GCTA

The sum of infinitesimal effects of 17 M imputed SNPs yields 56% of variance.

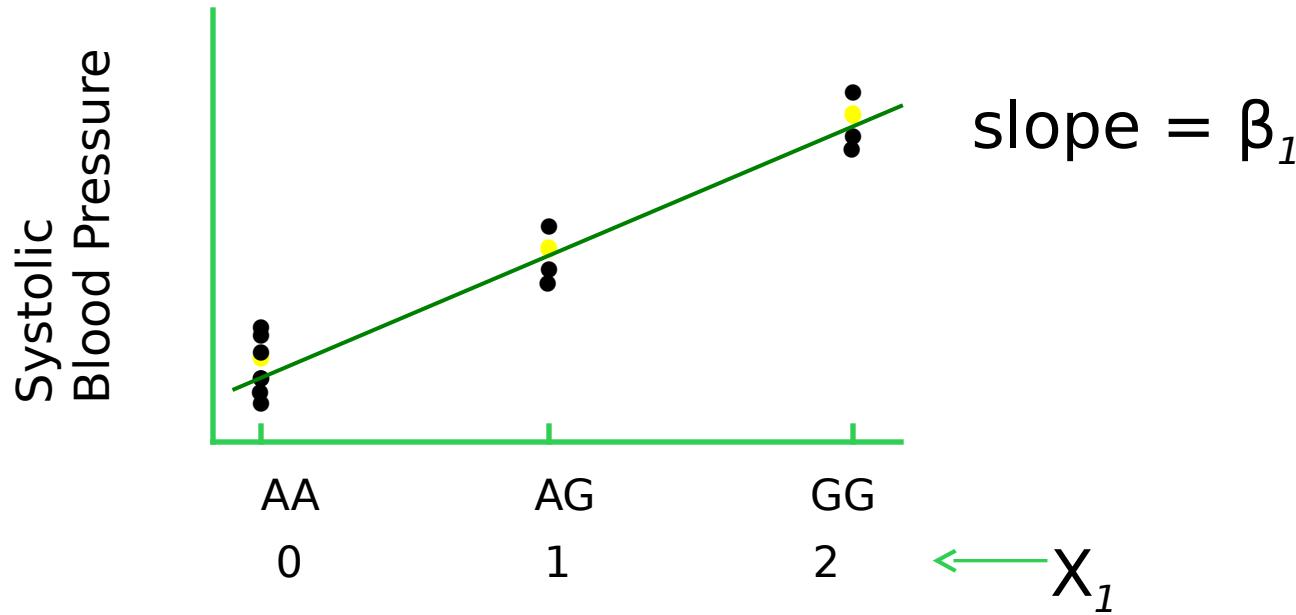
# PREDICTING disease risk from genomes



# PREDICTING disease risk from genomes

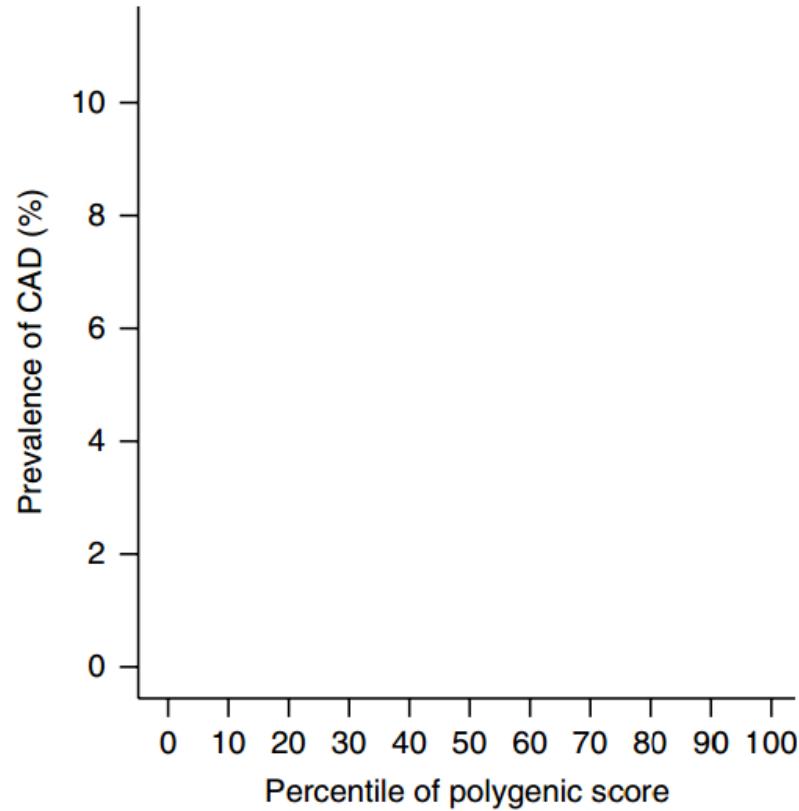


# PREDICTING disease risk from genomes

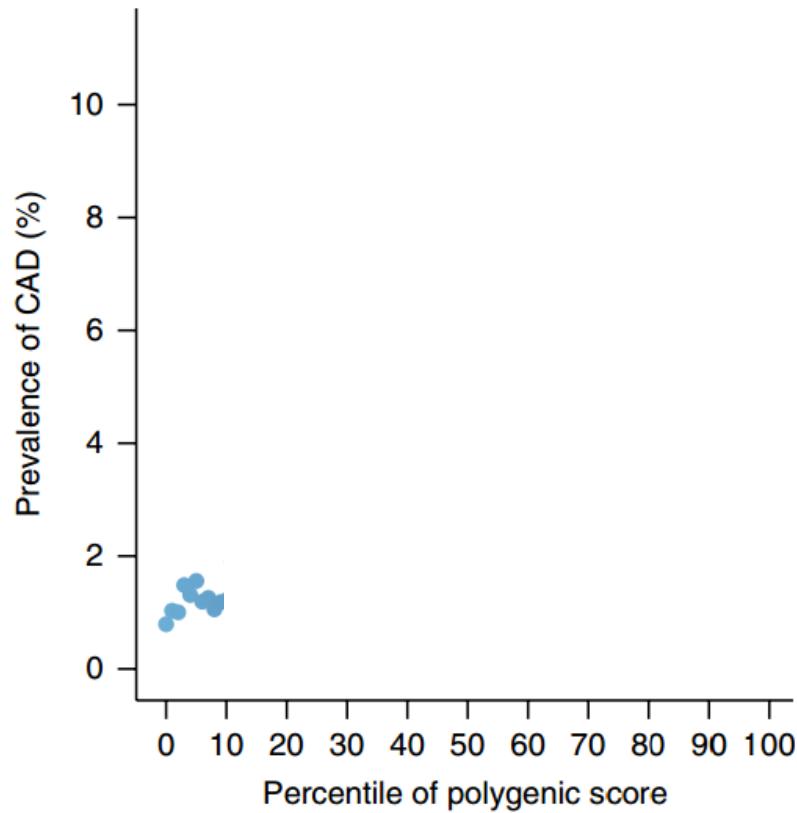


$$\text{Blood pressure} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \dots$$

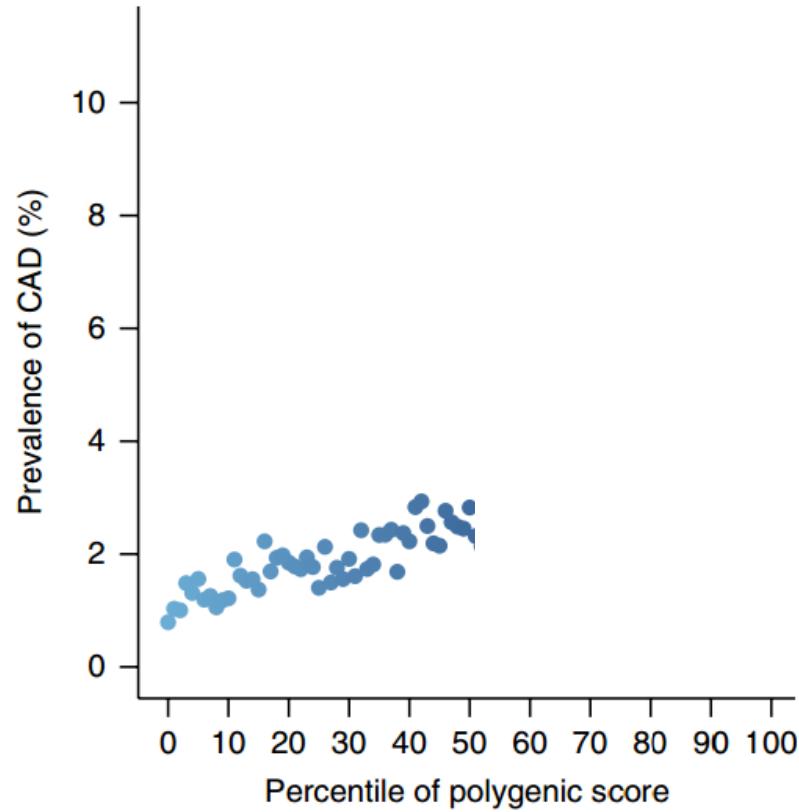
# Polygenic Risk Score



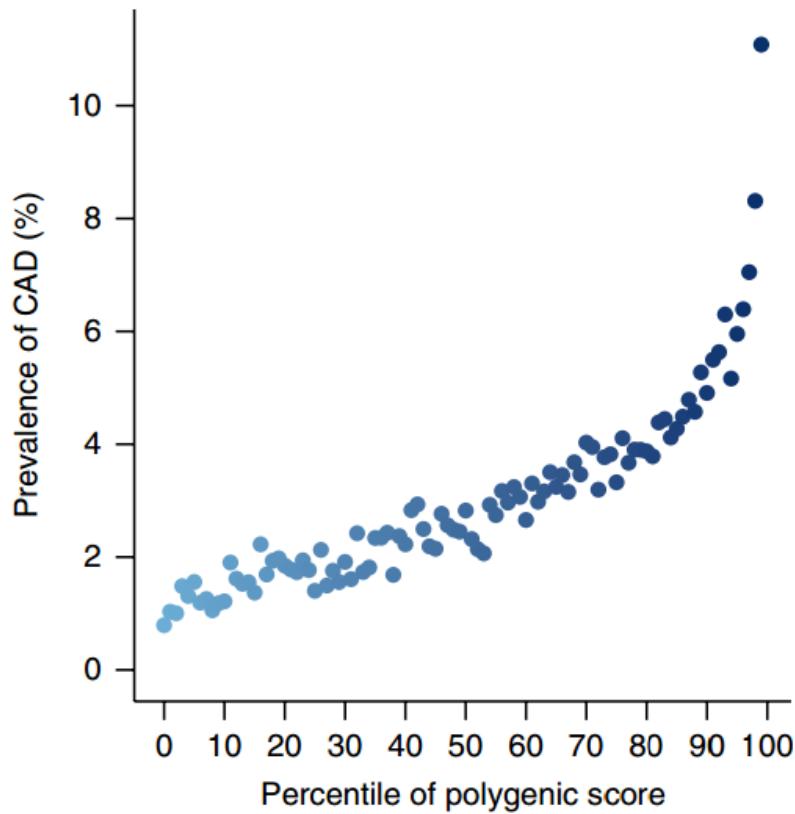
# Polygenic Risk Score



# Polygenic Risk Score



# Polygenic Risk Score

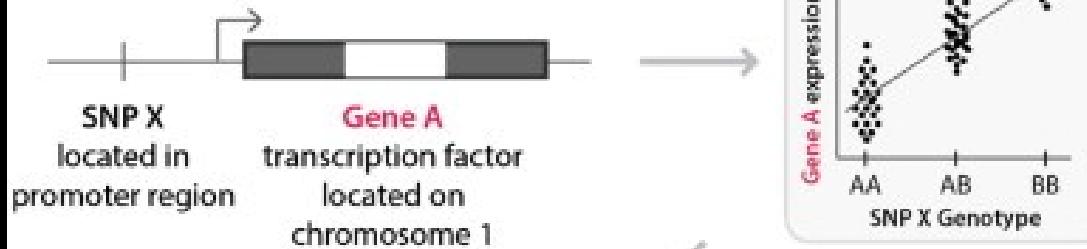


# Genome functional variation

What do we learn from analysis of:  
eQTL (expression)  
mQTL (methylation)  
sQTL (splicing) and  
rtQTL (replication timing)?

## Cis-eQTL

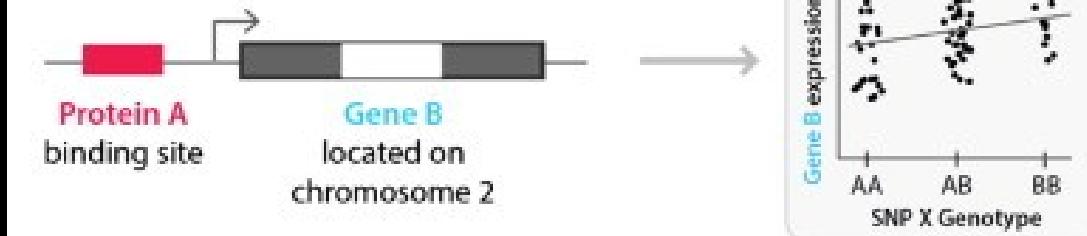
SNP X has an effect on local Gene A



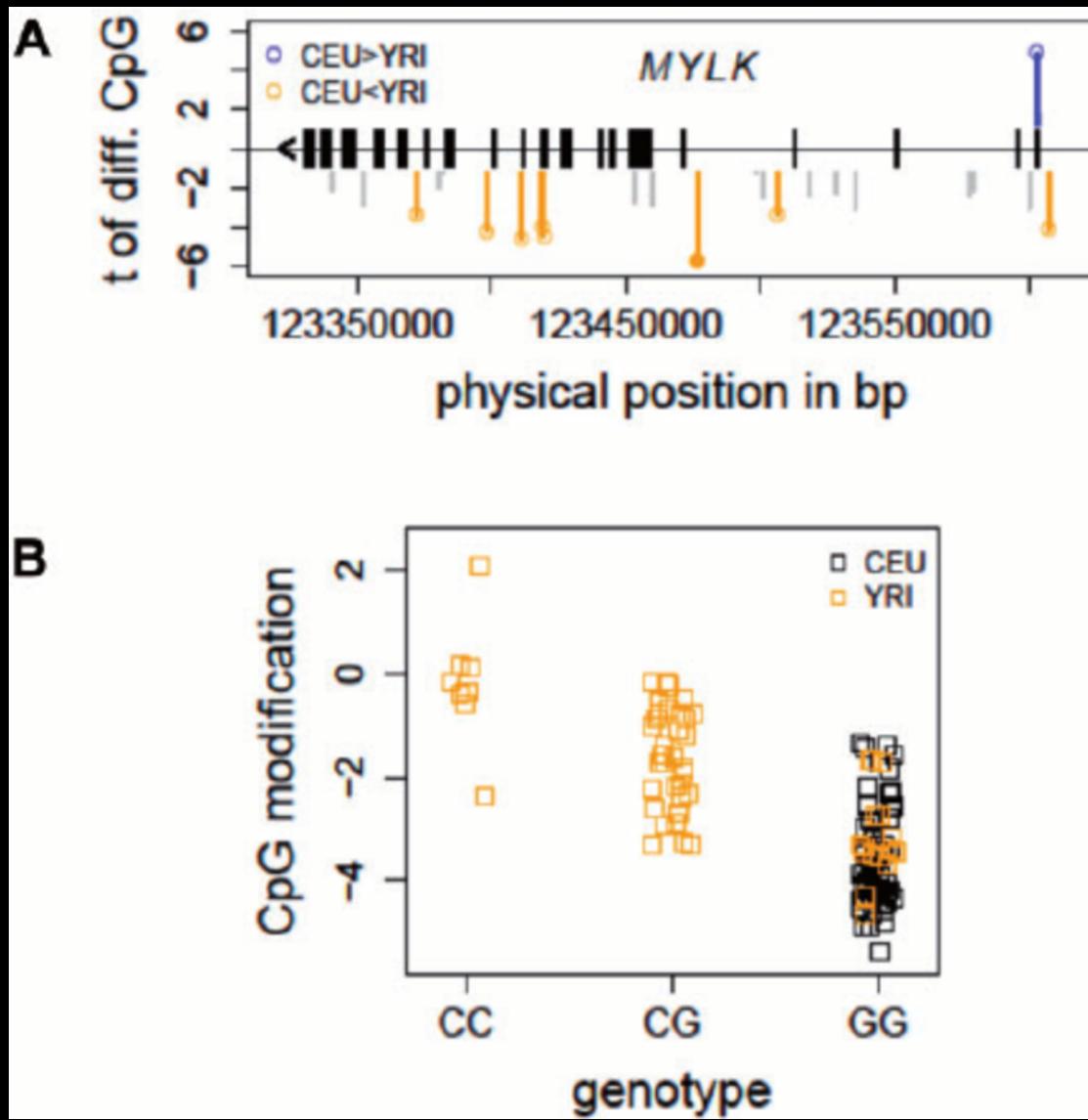
Altered **Protein A** levels,  
effect on the binding to  
the transcription factor  
binding sites of  
downstream genes

## Trans-eQTL

SNP X has an effect on distant Gene B through an intermediary factor (such as a transcription factor)

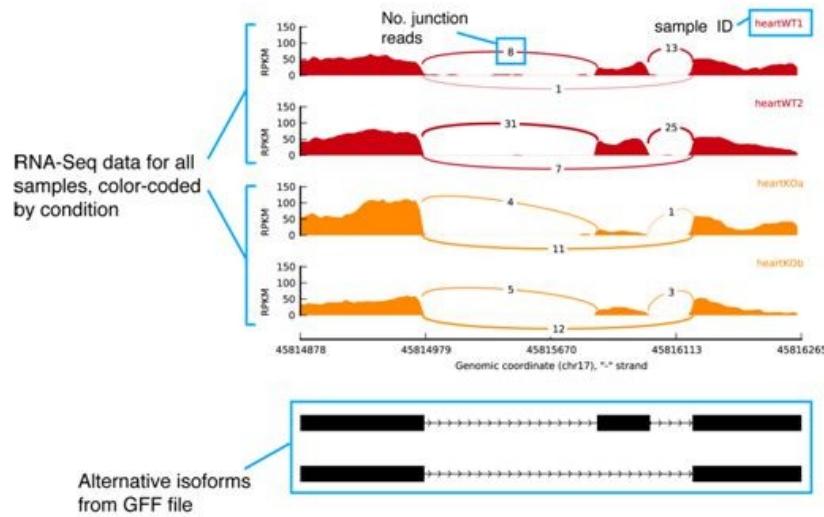


## Methylation QTL (mQTL)



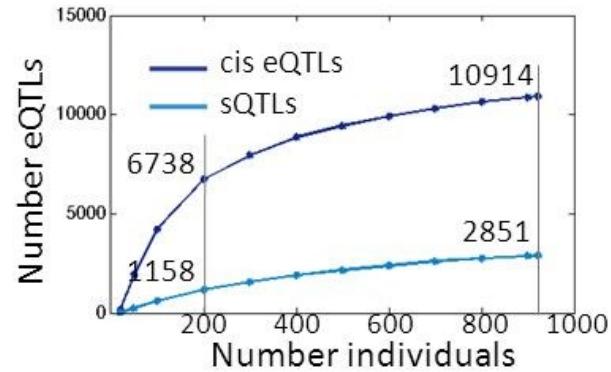
# Splicing eQTL

Can investigate relative transcript ratios or reads across junctions.

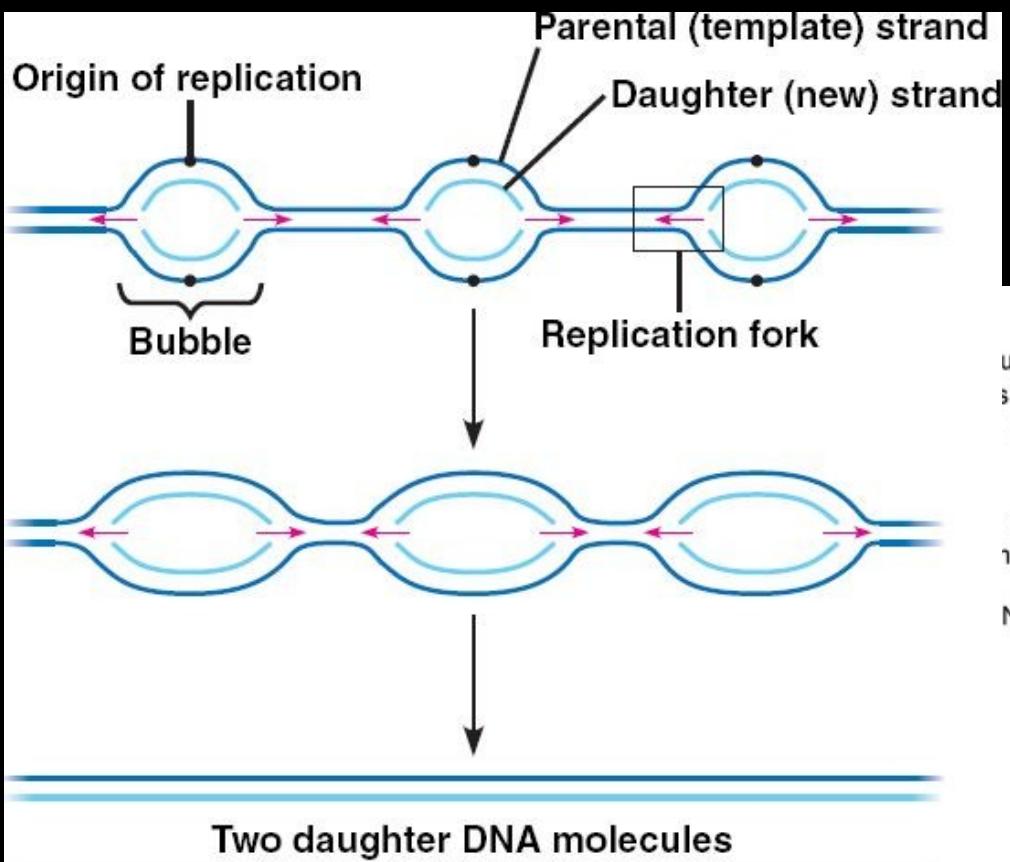


Katz et al, Nature Methods, 2010

- Splicing also affected for many genes



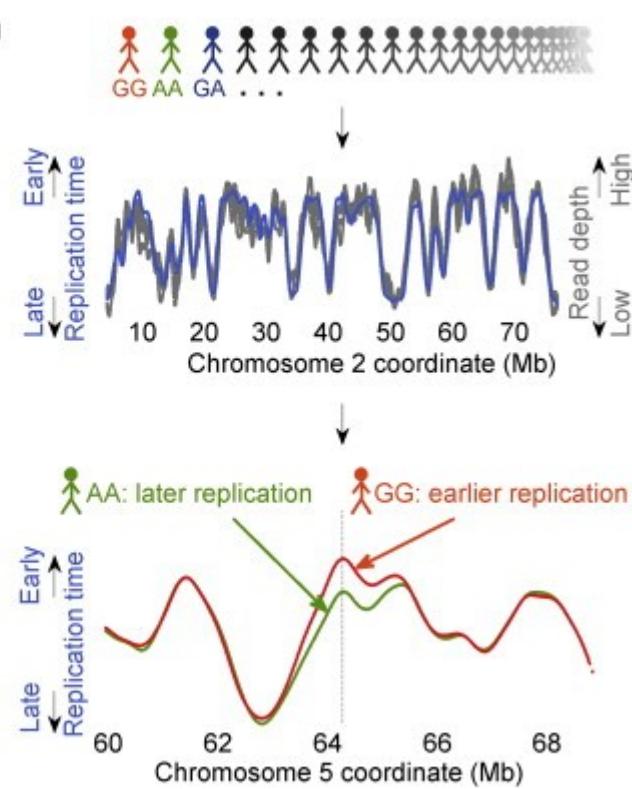
Battle et al, Genome Research, 2014



Population sequencing  
using proliferating  
cell cultures

Read depth  
along chromosomes  
~  
DNA replication  
timing

Replication timing  
quantitative trait loci  
(rtQTLs)



# Outline

- Demographic inference
- Population structure and history
- Admixture/ Introgression
- Random genetic drift
- Natural selection
- Mutation spectrum
- Disease association
- Genome functional variation