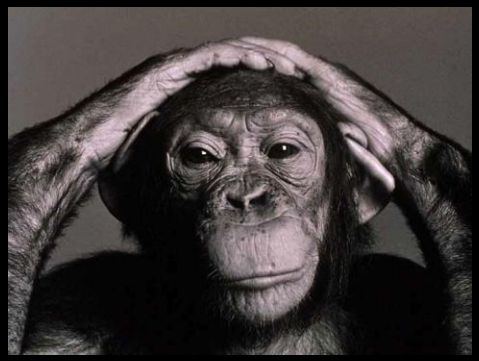
Population Genomics – Where are we going (in 60 minutes….)



Andrew Clark
Cornell University

EMBO short course
Napoli 17-22 May 2017

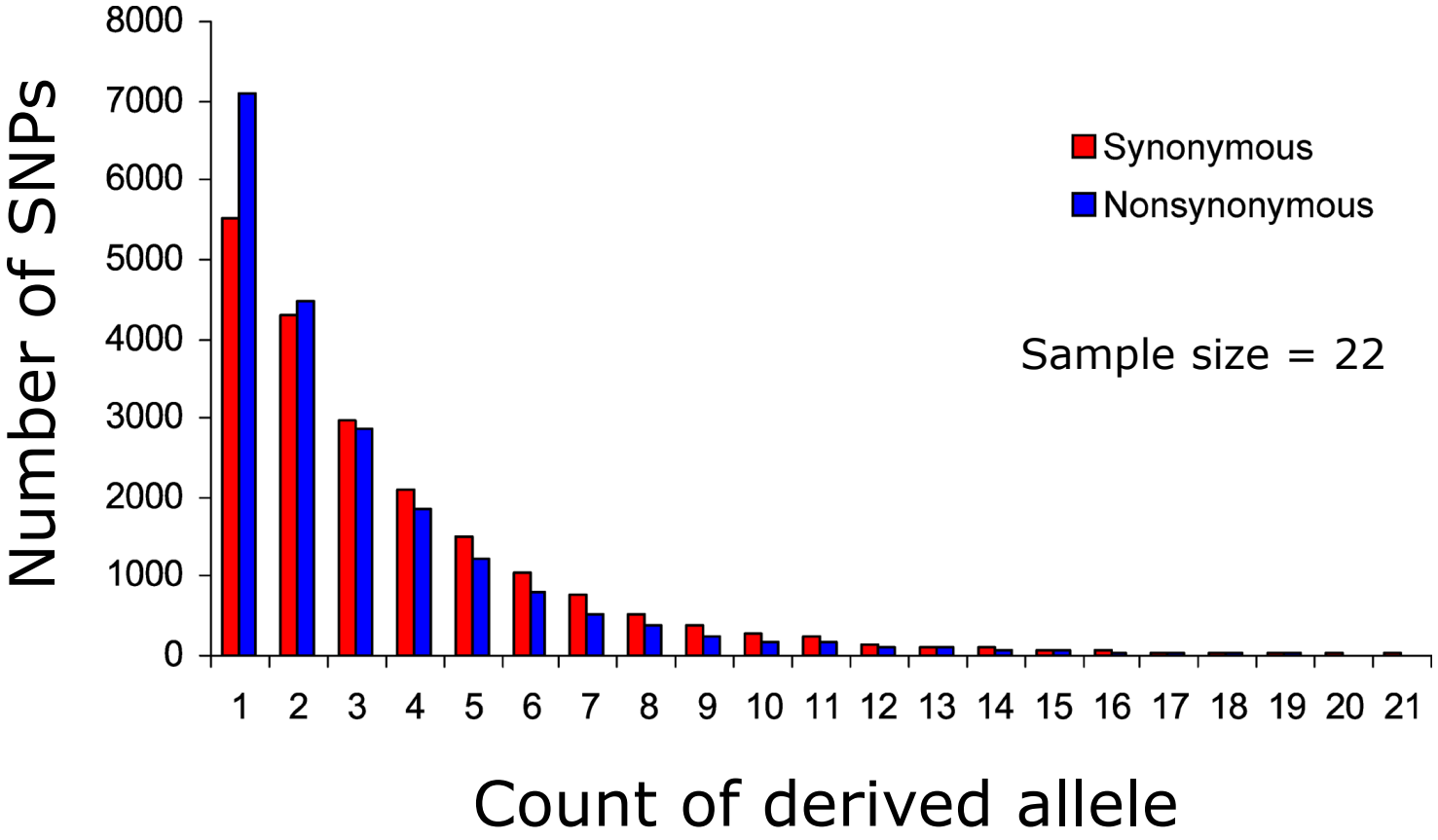
Outline

- Demographic inference
- Population structure and history
- Admixture/ Introgression
- Random genetic drift
- Natural selection
- Recombination
- Mutation spectrum
- Cryptic relatives
- Disease association
- Genome function
- Gene drive

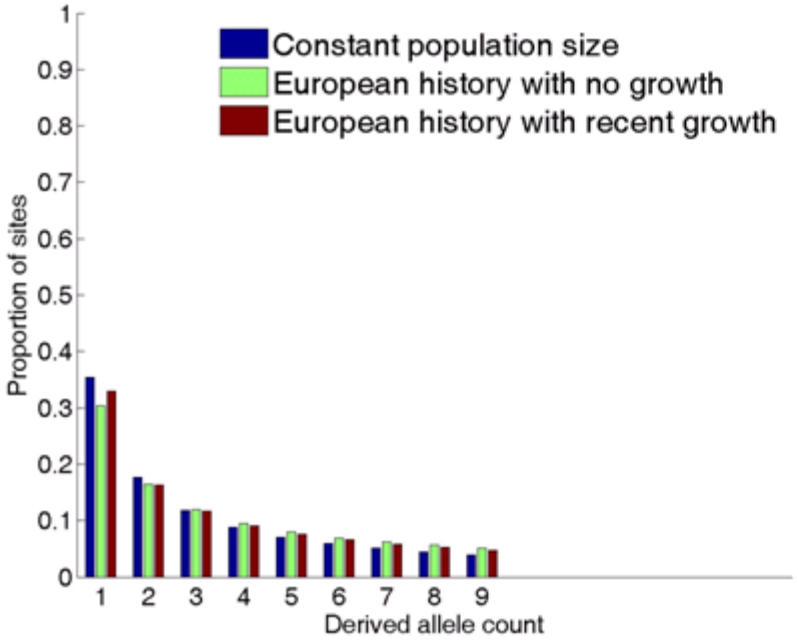
DEMOGRAPHY

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

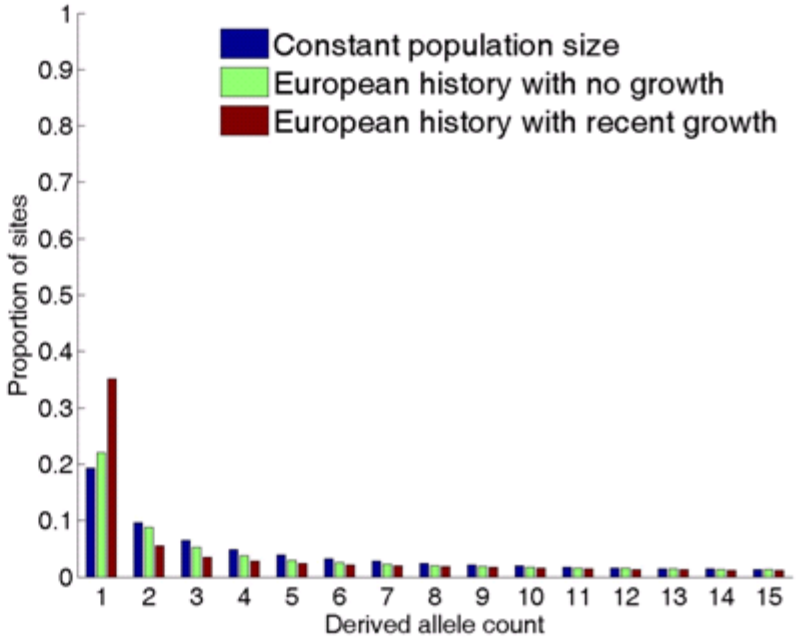
The Site Frequency Spectrum



Large samples are needed to see extent of skew to SFS



n = 10

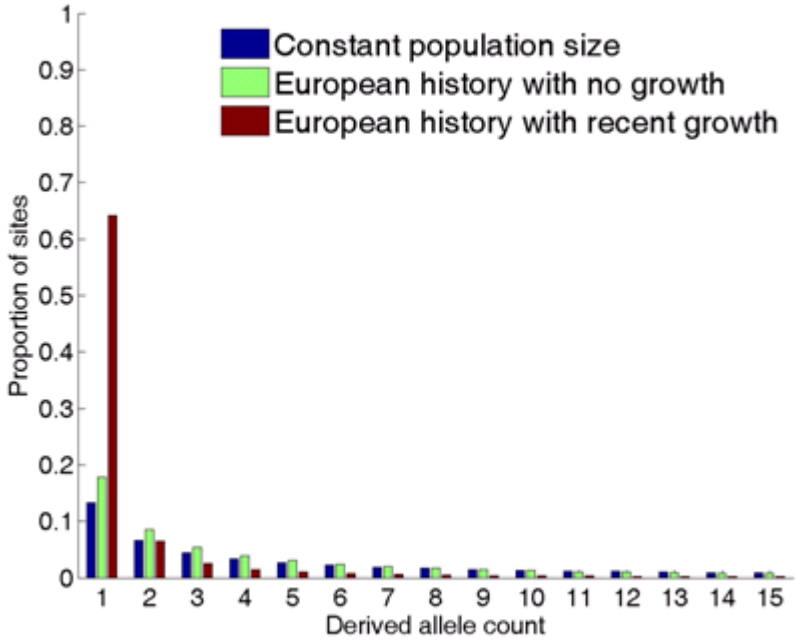


n = 100

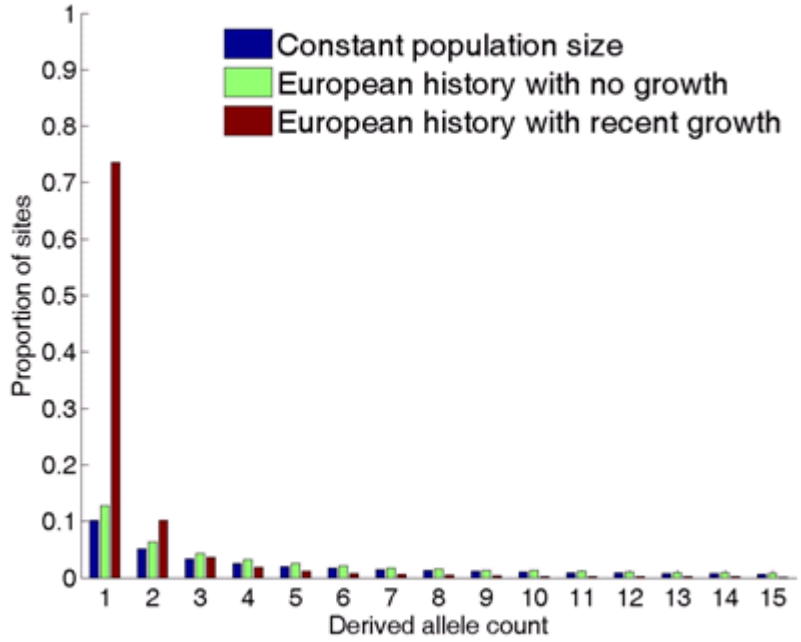
Simulation results

Keinan and Clark 2012 *Science*

Large samples are needed to see extent of skew to SFS



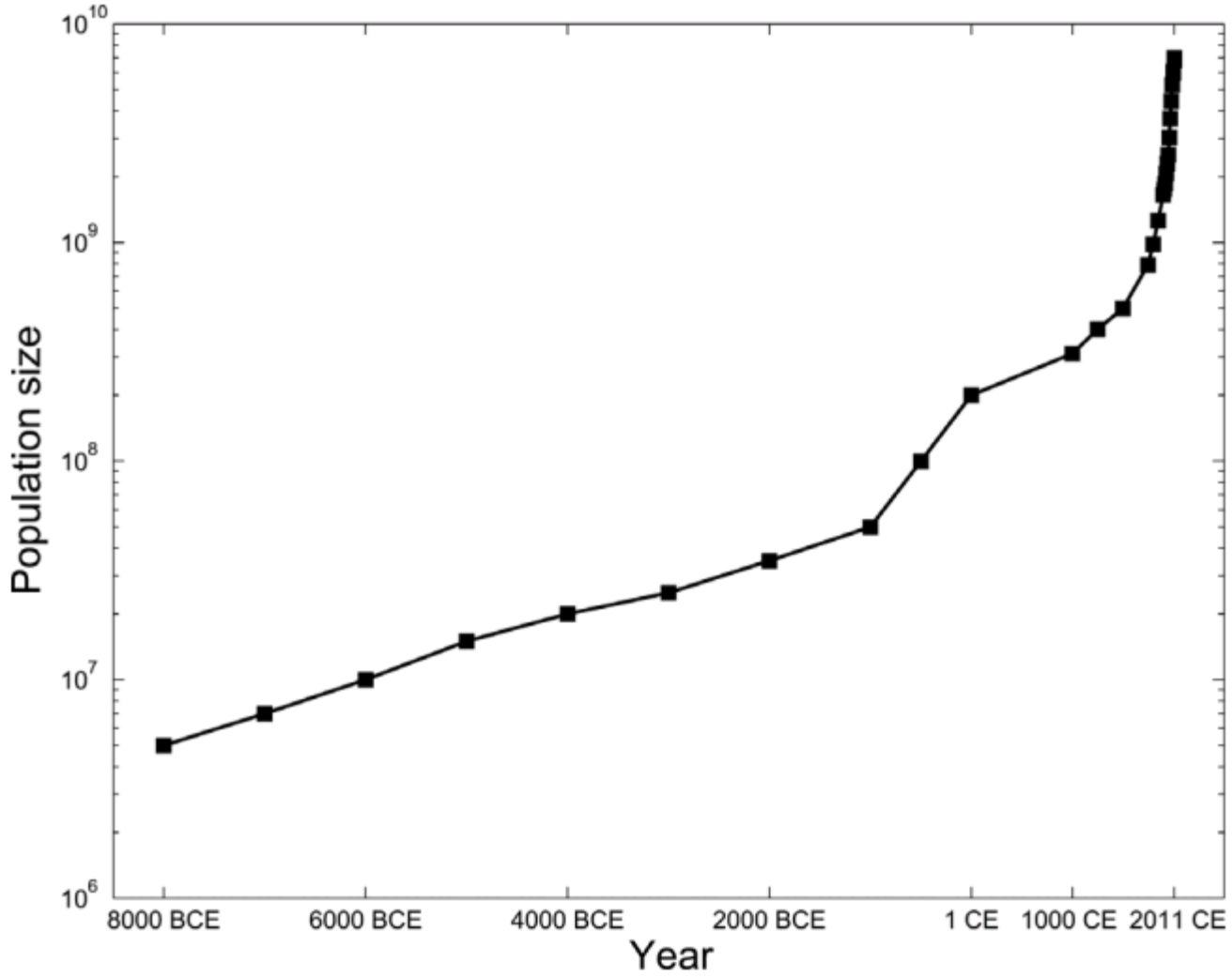
n = 1000



n = 10,000

Keinan and Clark 2012 *Science*

The human population has grown super-exponentially



Cohen J 1995 *How Many People can the Earth Support*?

Science
AAAS

An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People
Matthew R. Nelson *et al.*
Science **337**, 100 (2012);
DOI: 10.1126/science.1217876

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

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

Science
AAAS

An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People
Matthew R. Nelson *et al.*
Science **337**, 100 (2012);
DOI: 10.1126/science.1217876

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

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

Science
AAAS

Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes
Jacob A. Tennessen *et al.*
Science **337**, 64 (2012);
DOI: 10.1126/science.1219240

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

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

Coventry *et al.* 2010 *Nature Comm.* 1:131

These studies all show massive excesses of rare variation.

Implications of recent growth for complex disease studies

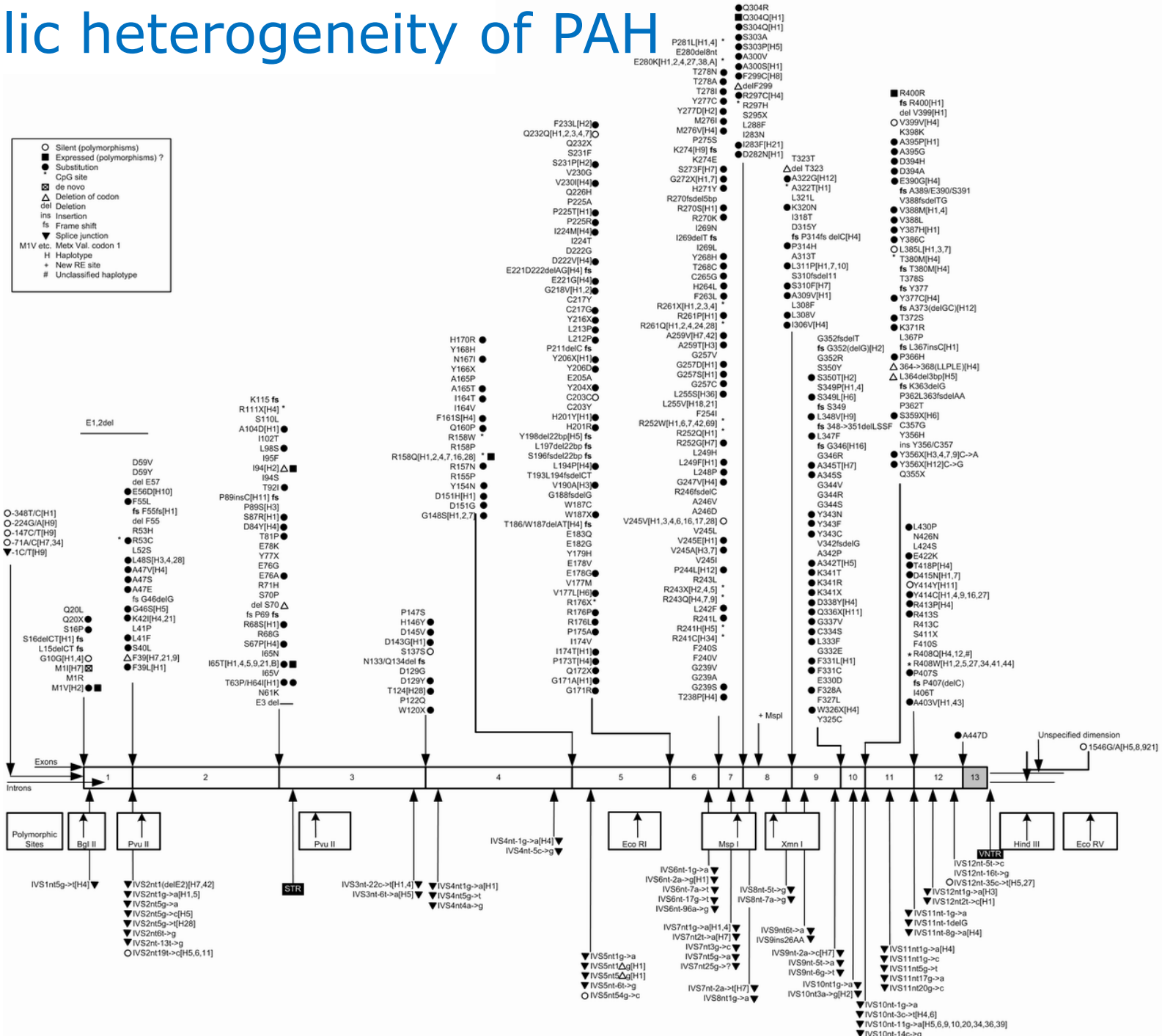
- Many more variants of all classes.
- Massive excess of rare variants.
- Rare variants are more likely to have deleterious effects.

Implications of recent growth for complex disease studies

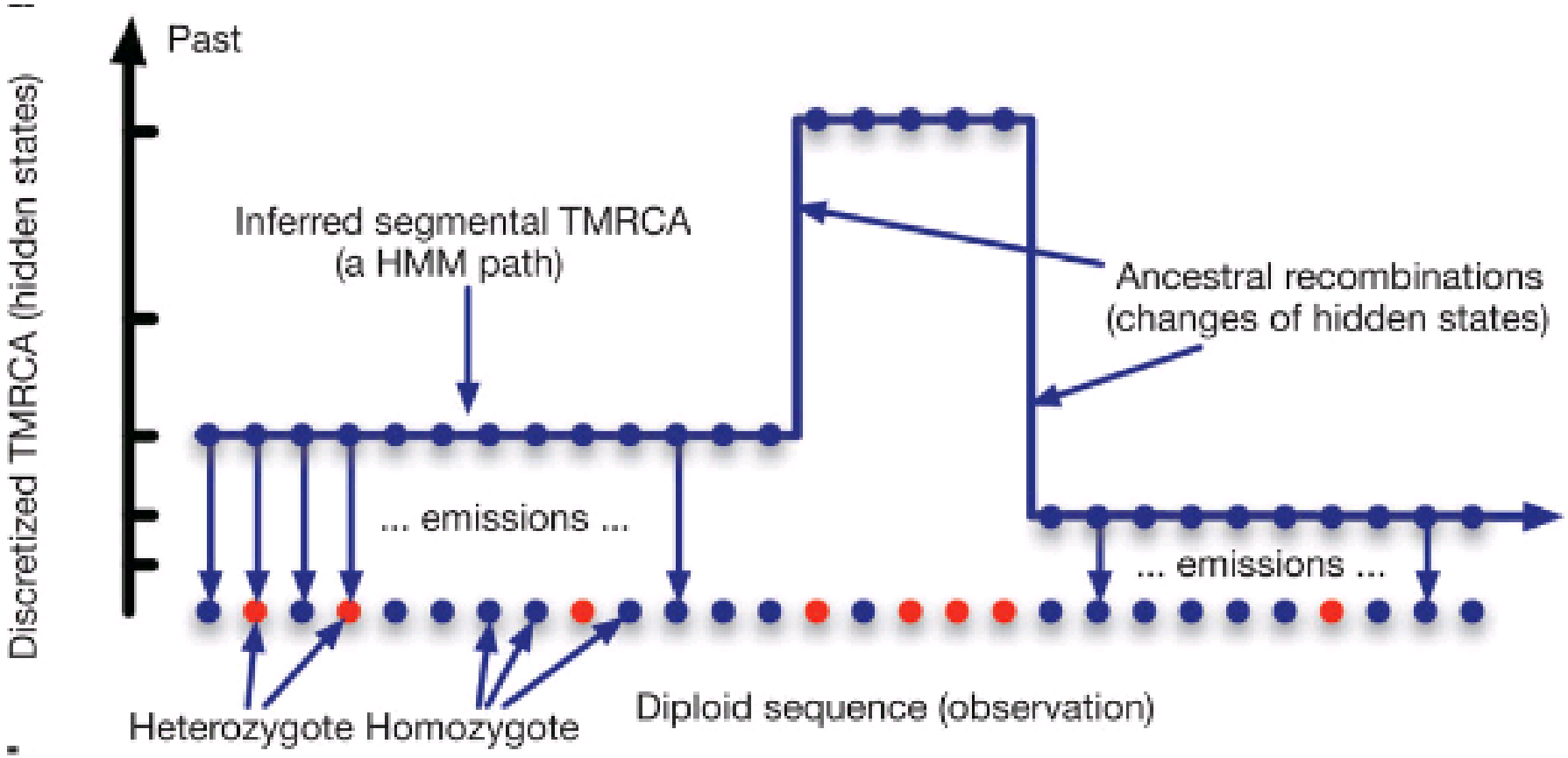
- Many more variants of [illegible]asses.
- Frequency shifted [illegible]er variants.
- Rare variant[illegible] more likely to have deleteri[illegible]ects.

Genetic heterogeneity

Allelic heterogeneity of PAH

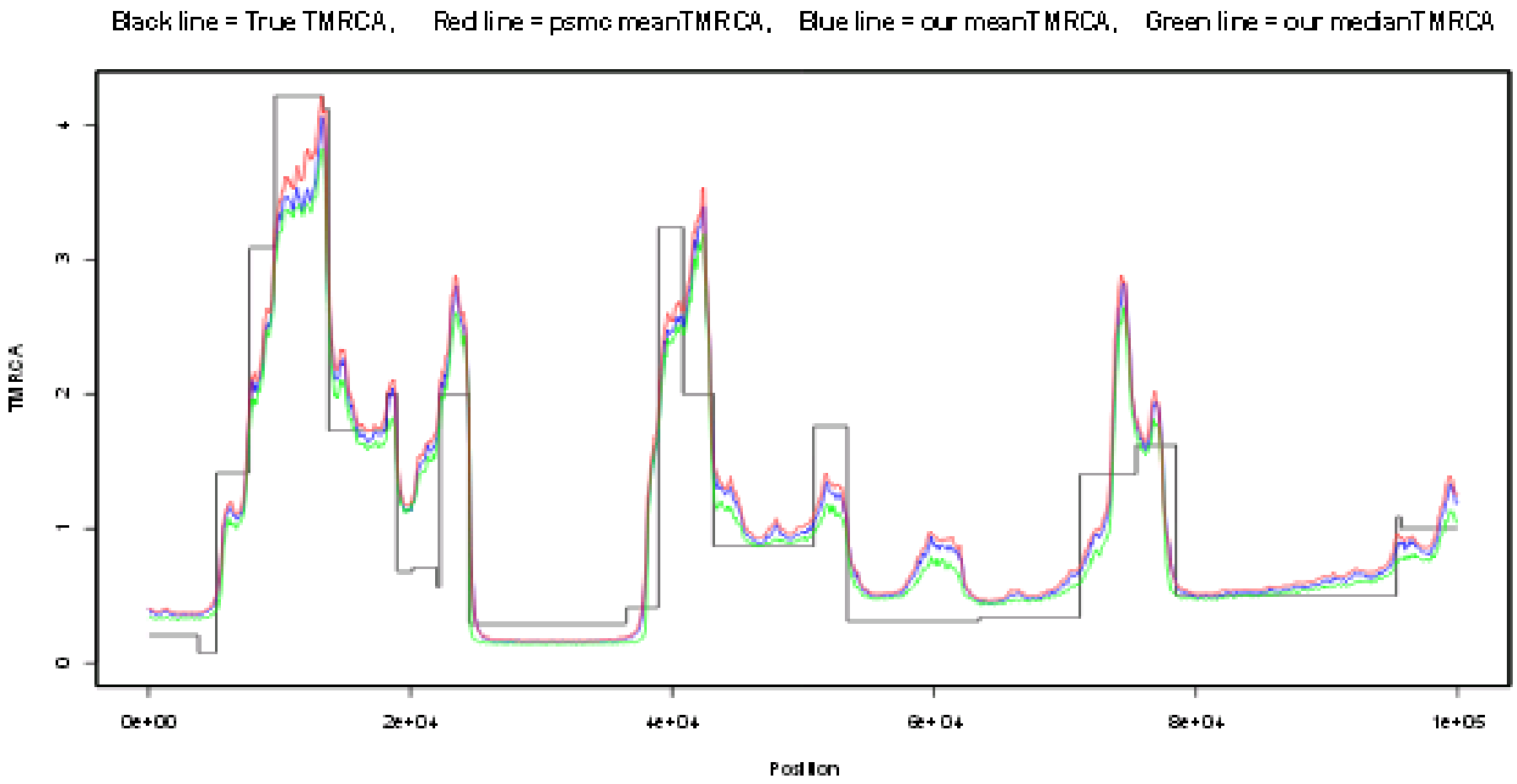


Inferring demography from a single individual: Pairwise Sequentially Markovian Coalescent



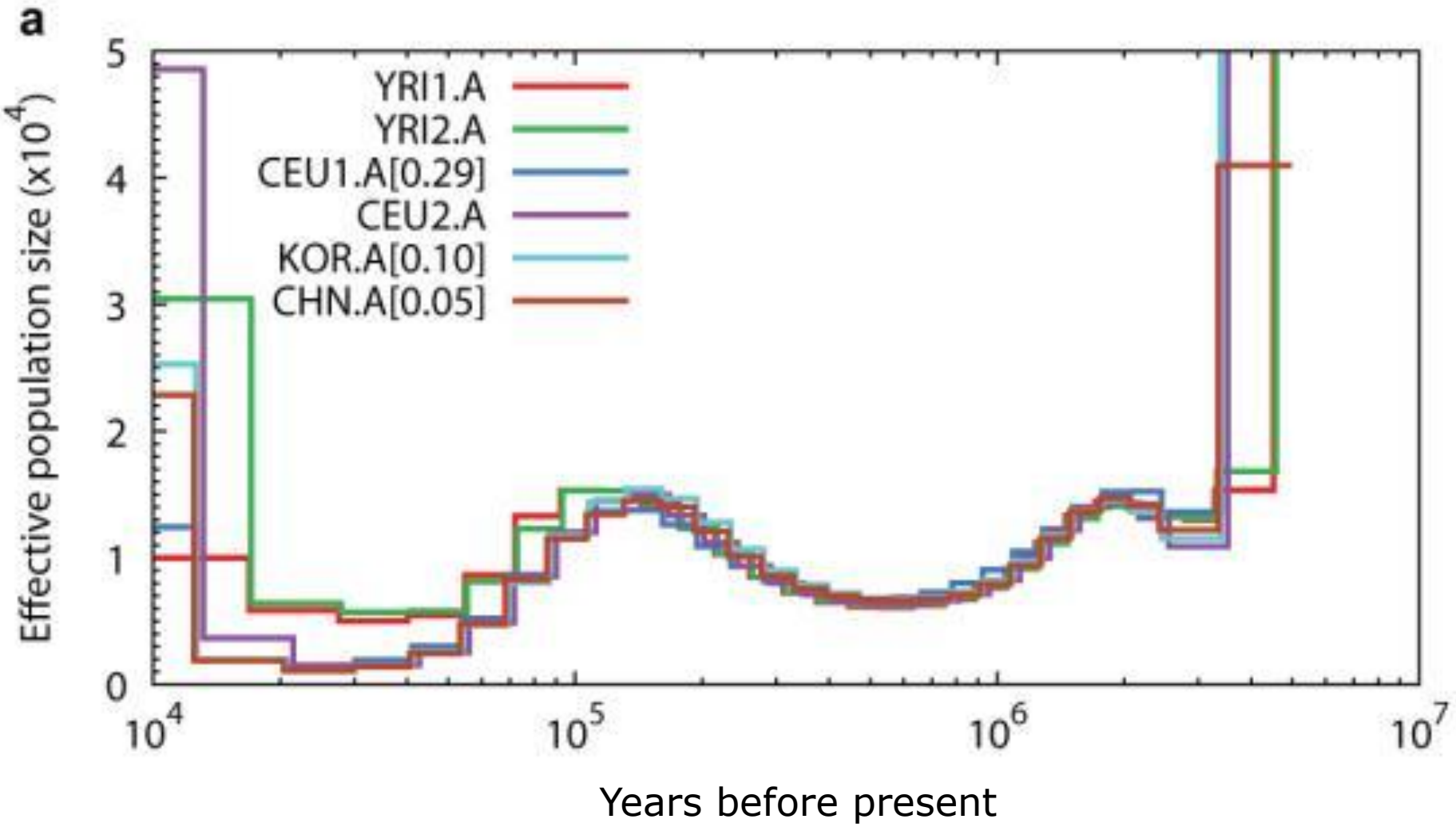
Li and Durbin 2011 *Nature*

Simulations of past population demography shows reasonably good accuracy of PSMC



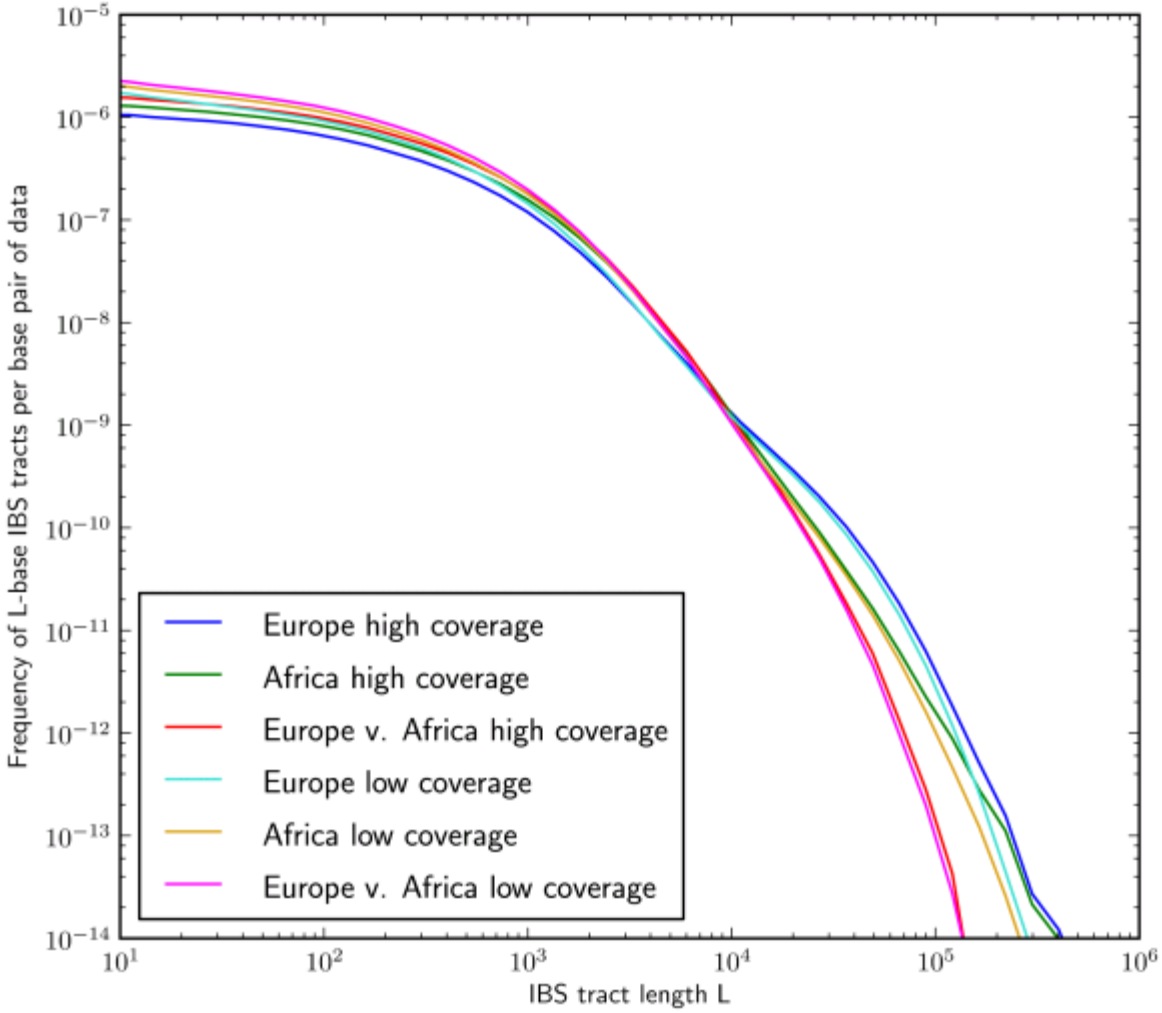
Li and Durbin 2011 *Nature*

Application of Pairwise Sequentially Markovian Coalescent



Li and Durbin 2011 *Nature*

Identity-by-Descent tracts for inference of demography



Harris and Nielsen 2013 *PLoS Genetics*

DEMOGRAPHY (Population collapse)

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

Florida Scrub-Jay

(*Aphelocoma coerulescens*)

Cooperative breeder

Federally Threatened

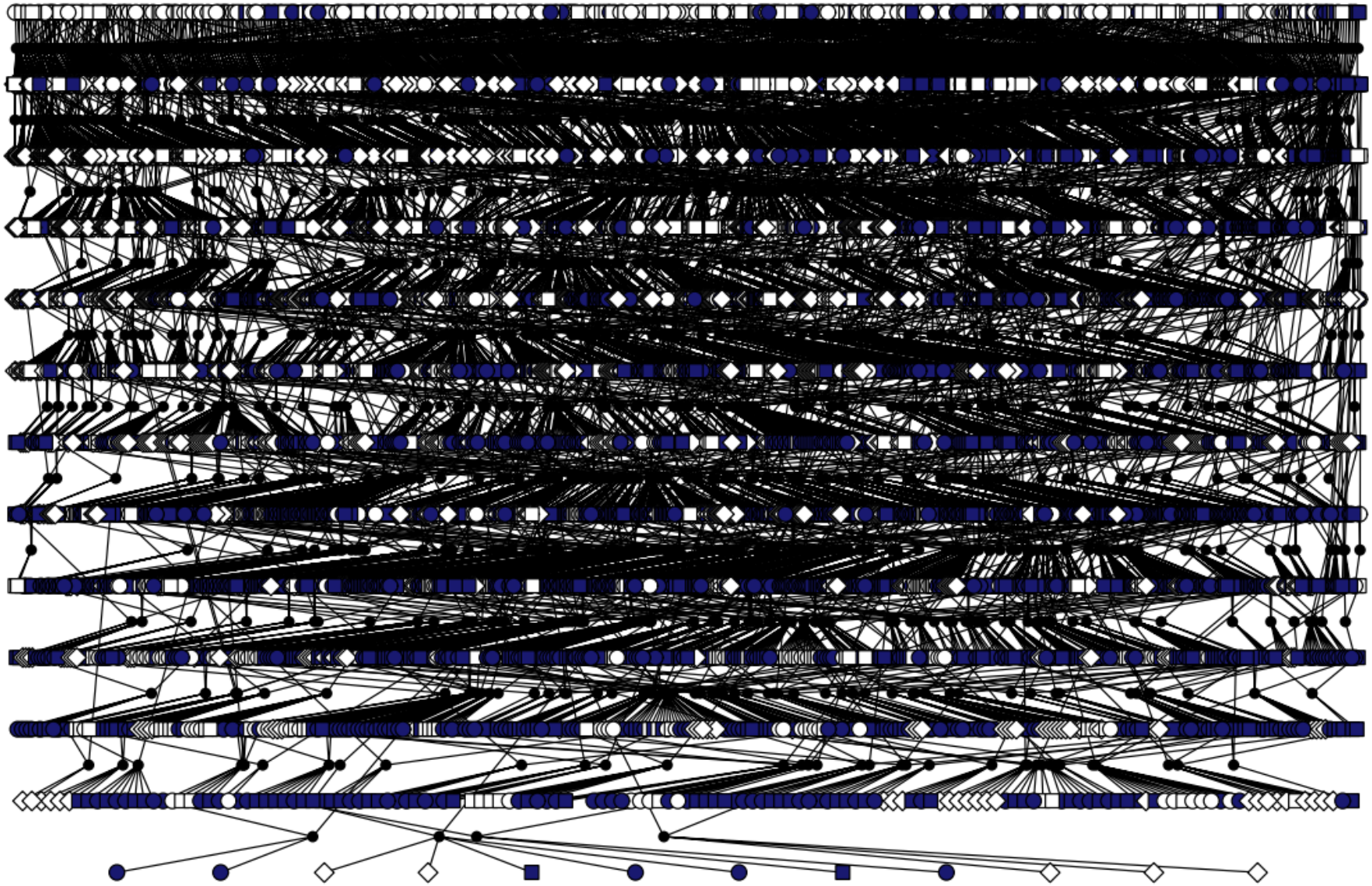
Non-migratory

Highly territorial & philopatric

Socially & (mostly) genetically monogamous

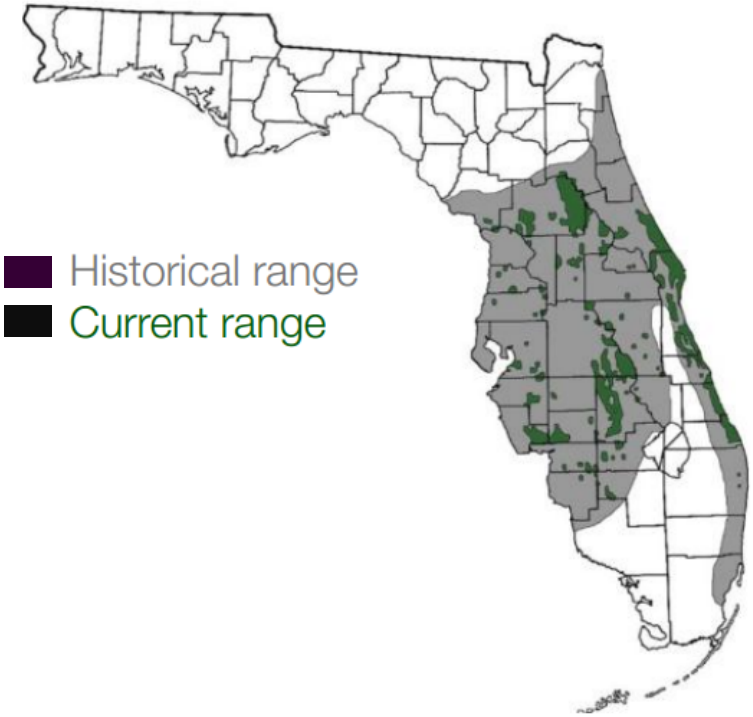
All individuals banded and tracked throughout life

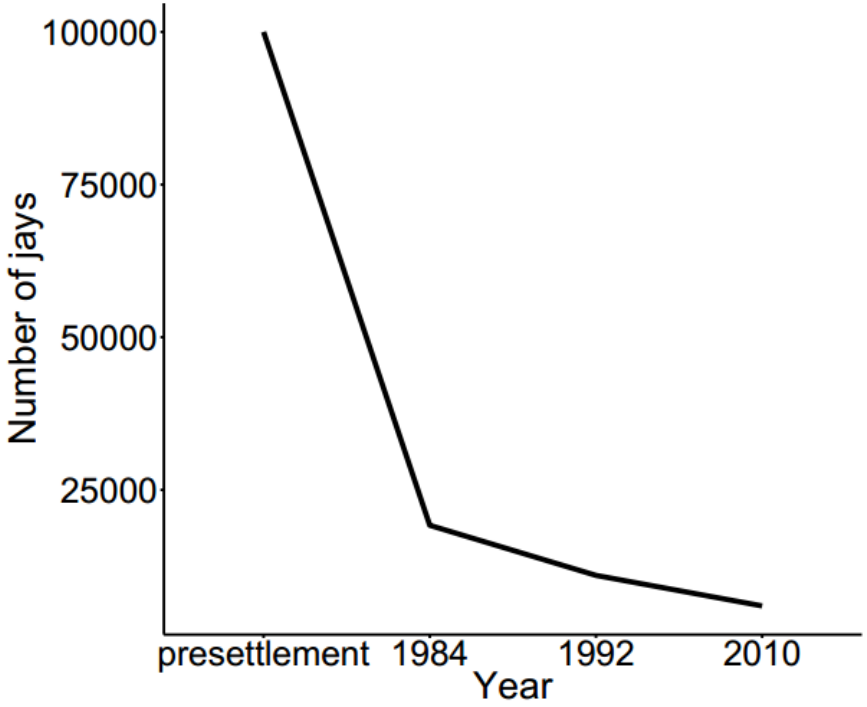
Quinn *et al.* 1999; Townsend *et al.* 2011; Woolfenden & Fitzpatrick 1984, 1996



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

map by Monica E. McGarrit

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

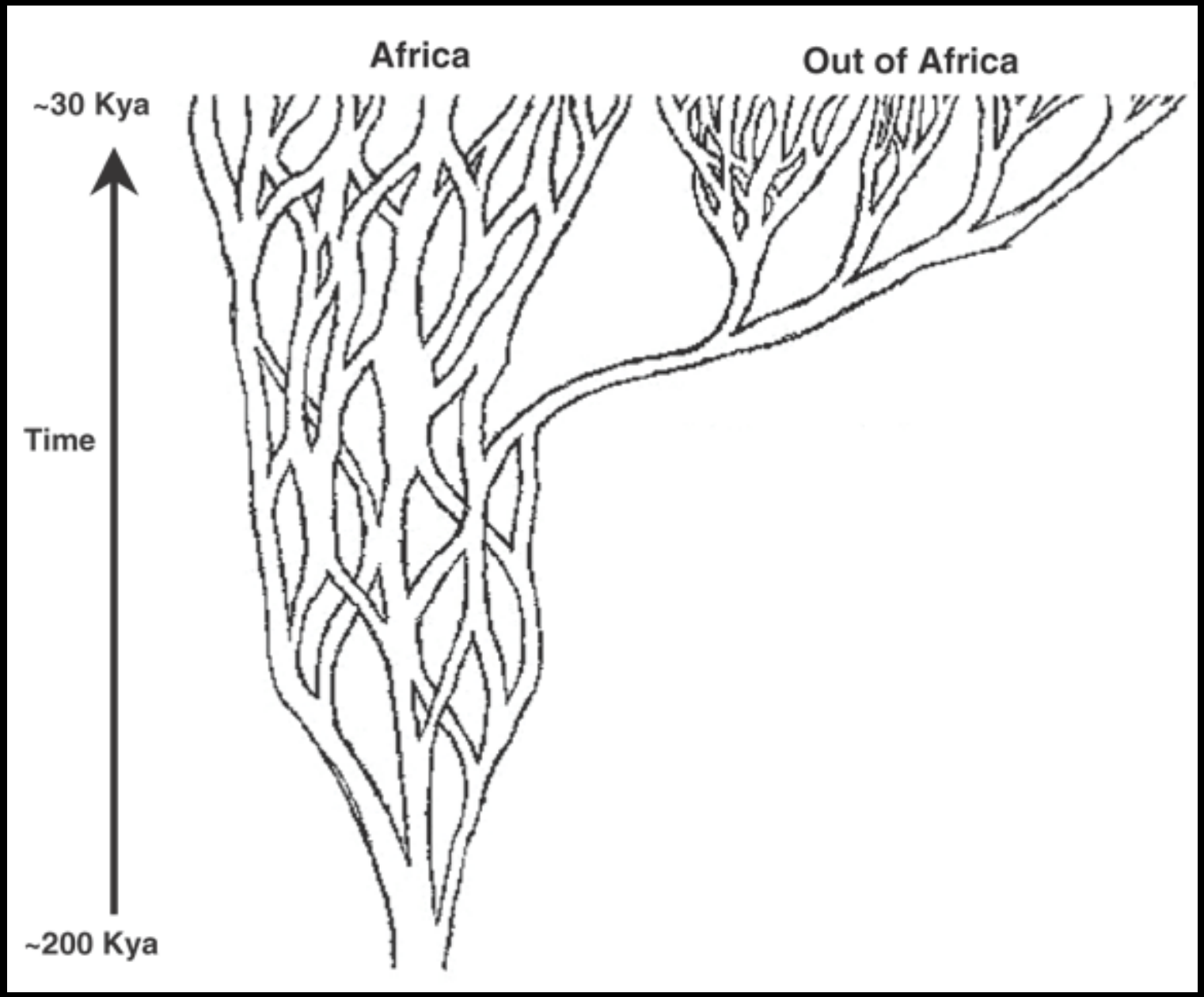
Linkage map construction

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[1*], Alkes L. Price[1,2], David Reich[1,2]

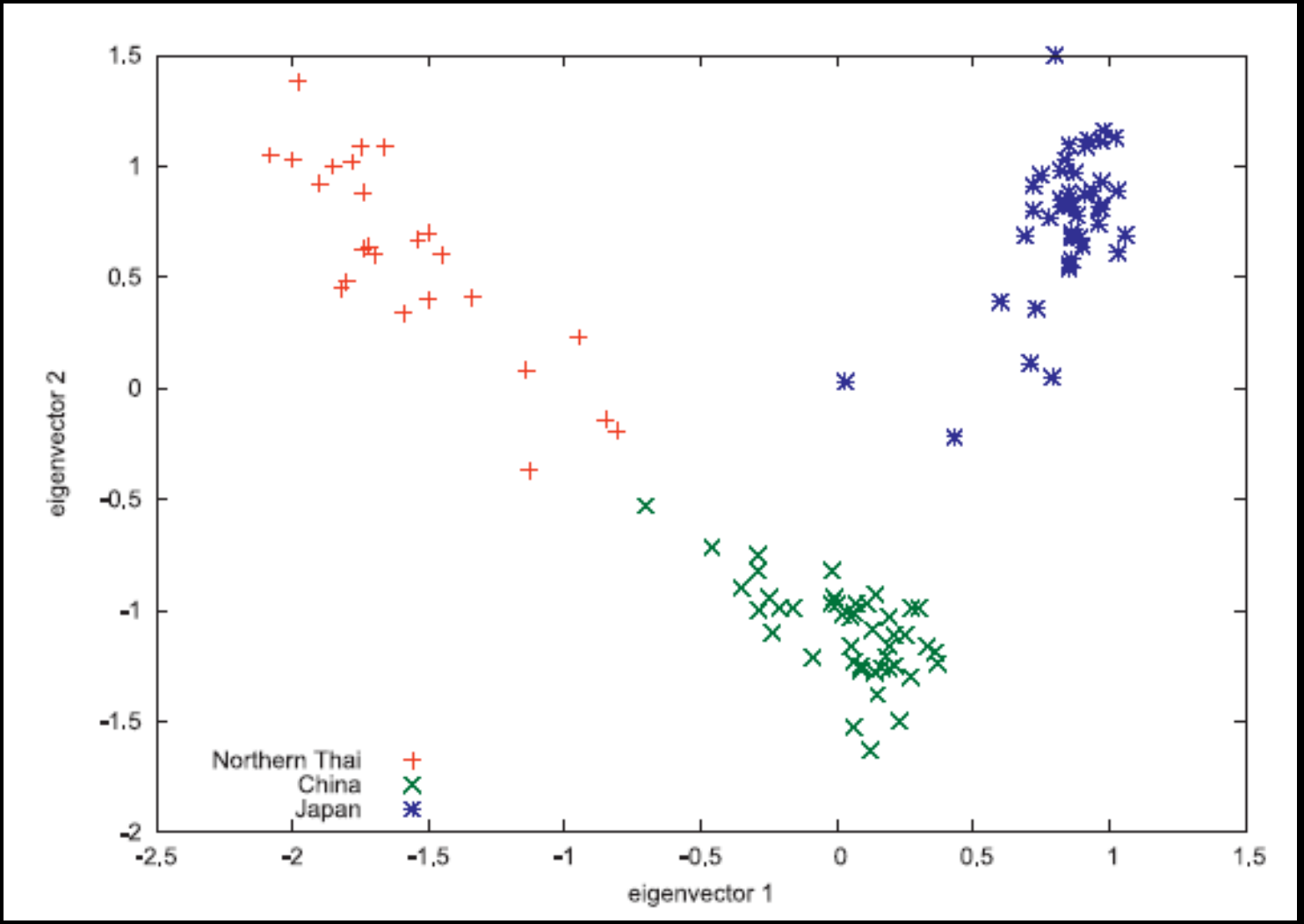
1 Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America, 2 Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America

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.

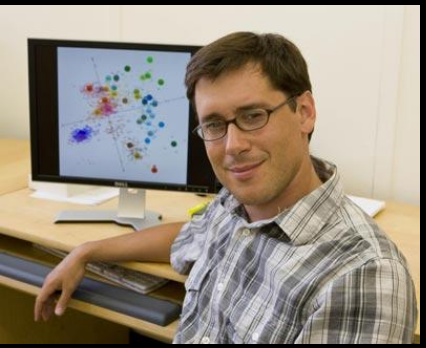
Citation: Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12): e190. doi:10.1371/journal.pgen.0020190

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.

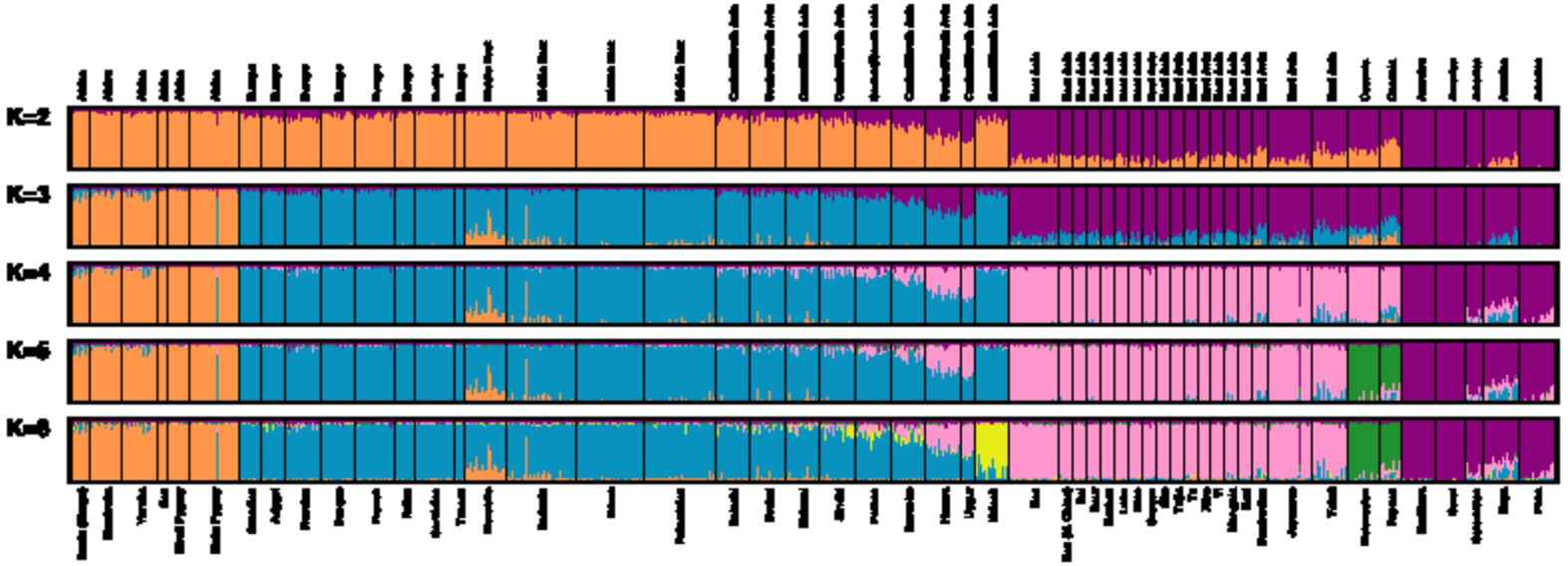
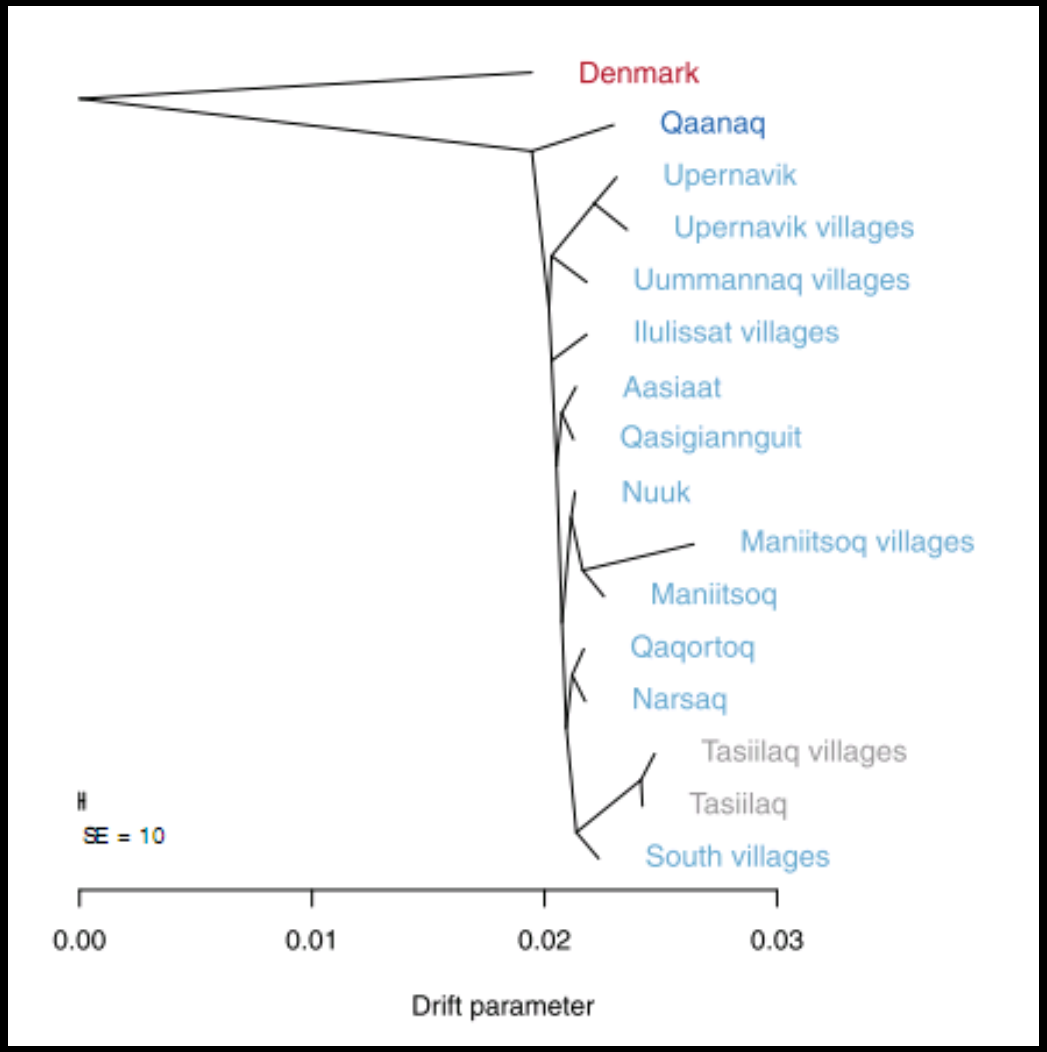


Fig. 1. Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into *K* colored segments that represent the individual's estimated membership fractions in *K* clusters. Black lines separate individuals of different populations. Populations are labeled 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

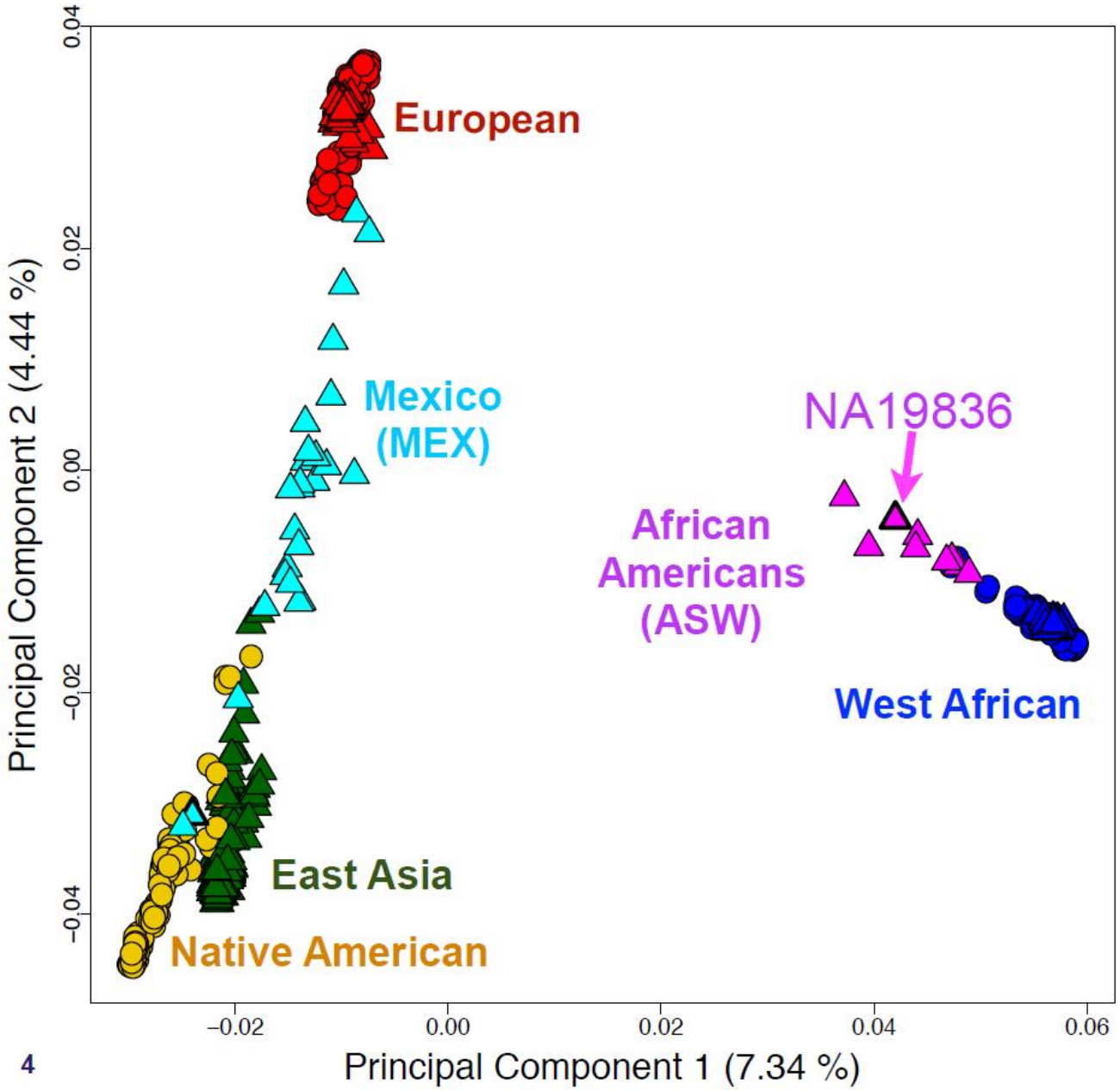


Moltke *et al.* 2015 *AJHG*

ADMIXTURE

How can we infer consanguinity and admixture from genetic data?

Using PCA to infer admixed individuals



- Africa
- African American
- Europe
- Mexico
- Native American
- East Asia

- ○ HGDP
- △ HapMap

Kasia Bryc

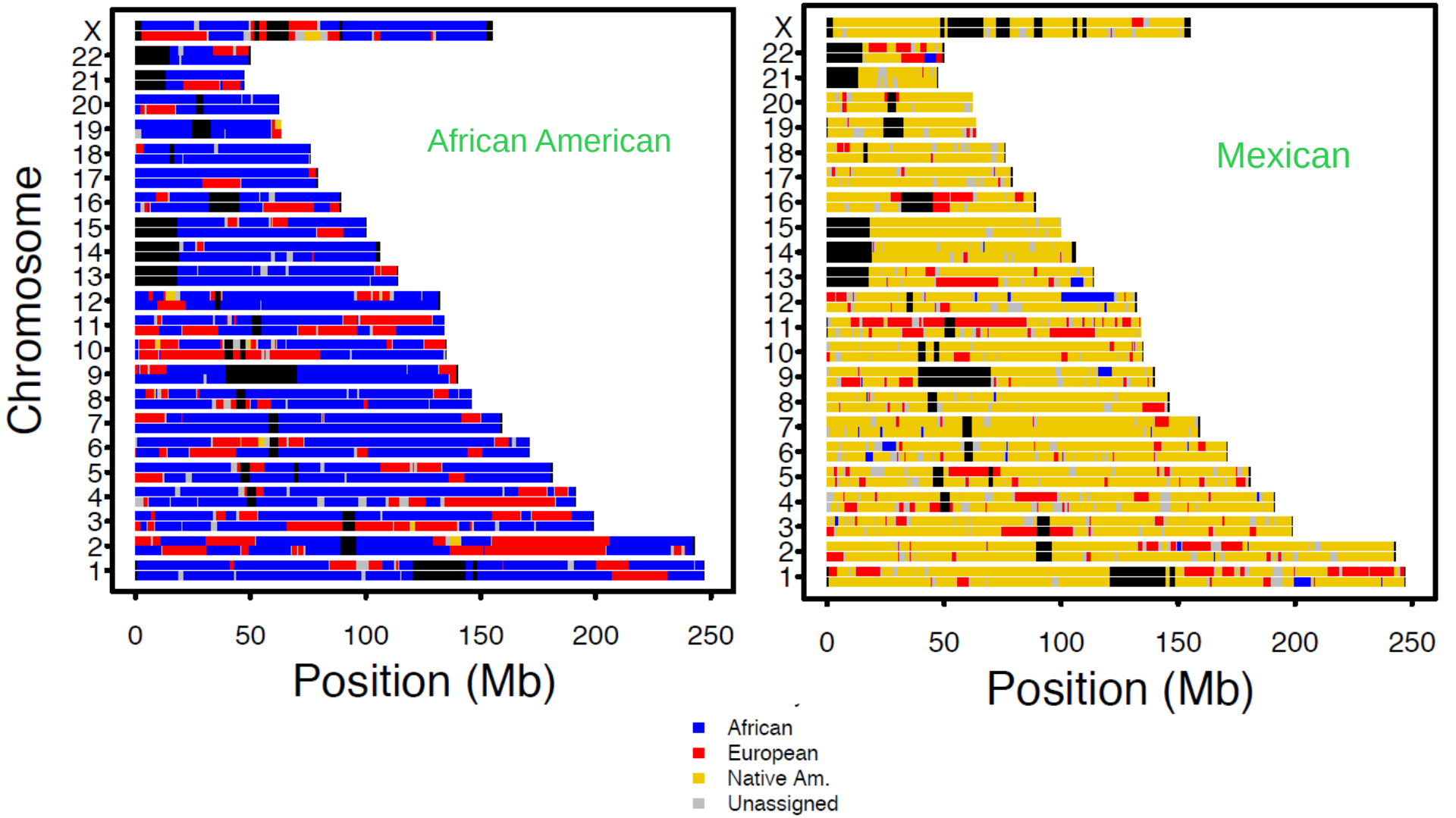
Adam Auton

Local ancestry inference

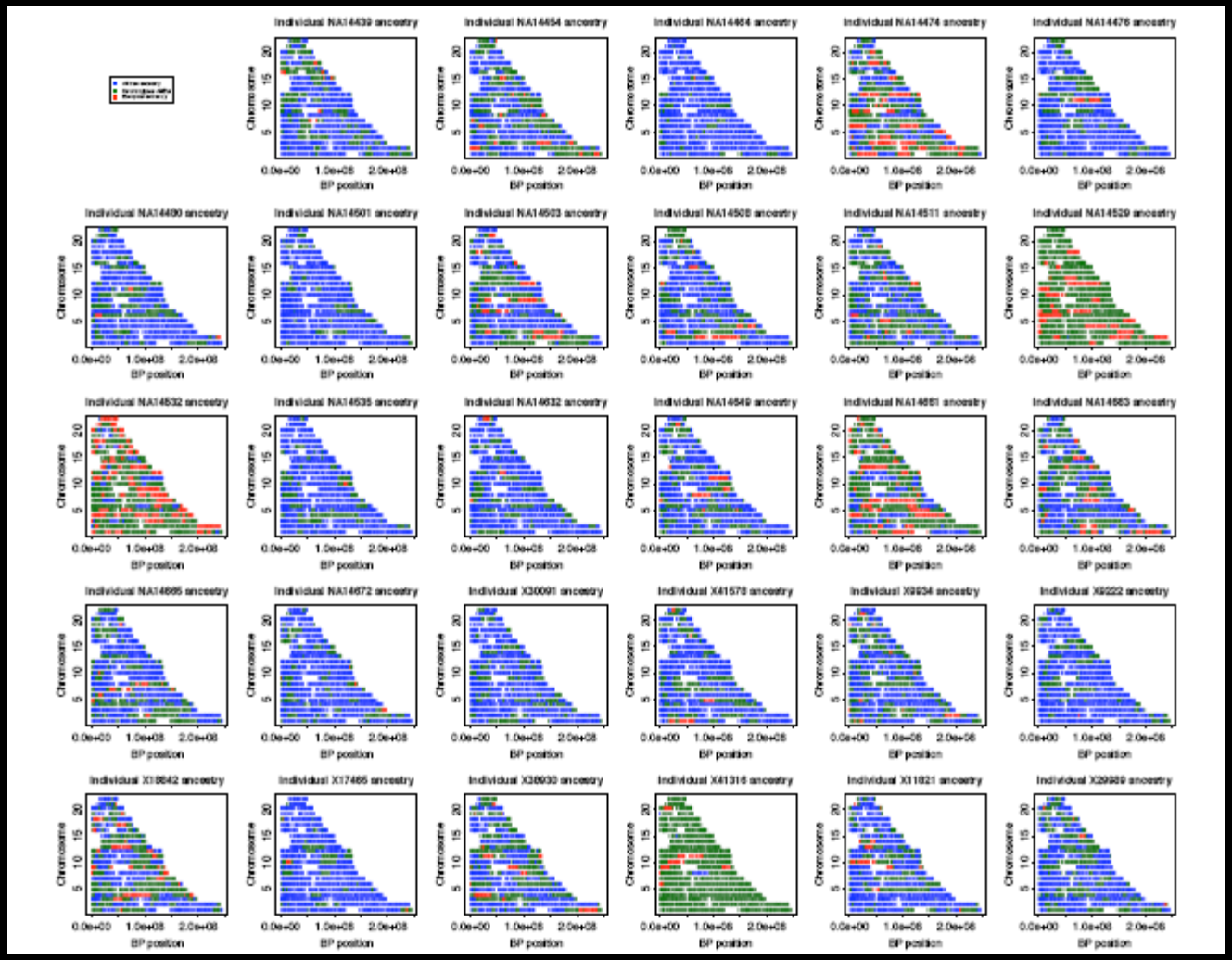
- Run PCA on African, European, and African American samples.
- From PC loadings of 15 SNP windows, infer 0, 1, or 2 copies of African ancestry.
- Slide along the genome.
- RESULT: Call of 0, 1, 2 copies of African ancestry for each chunk of the genome in each individual.

Kasia Bryc

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



High variability among individuals in admixture patterns



Kasia Bryc

An extreme case of admixture:

modern humans x Neanderthals

LETTER

doi:10.1038/nature12961

The genomic landscape of Neanderthal ancestry in present-day humans

Sriram Sankararaman[1,2], Swapan Mallick[1,2], Michael Dannemann[3], Kay Prüfer[3], Janet Kelso[3], Svante Pääbo[3], Nick Patterson[1,2] & David Reich[1,2,4]

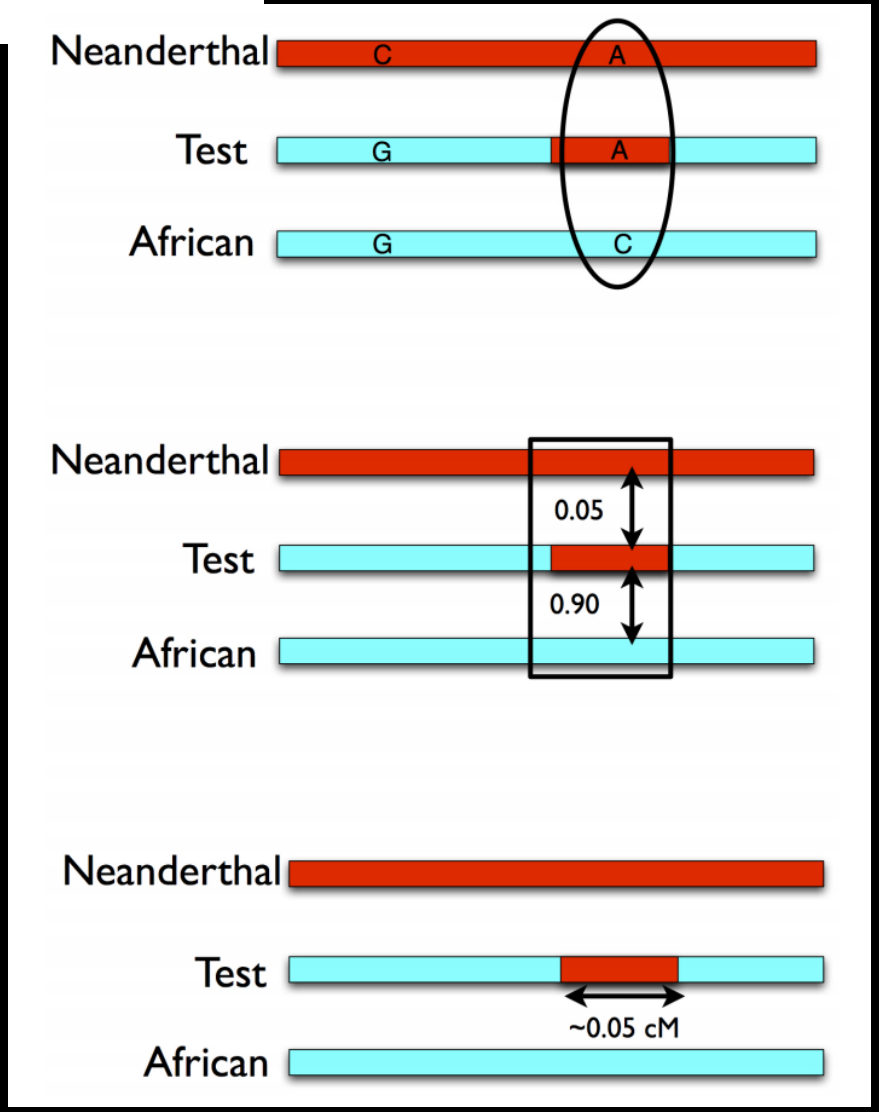
2014 Nature 507:354-357

Three criteria:
Introgressed region is:

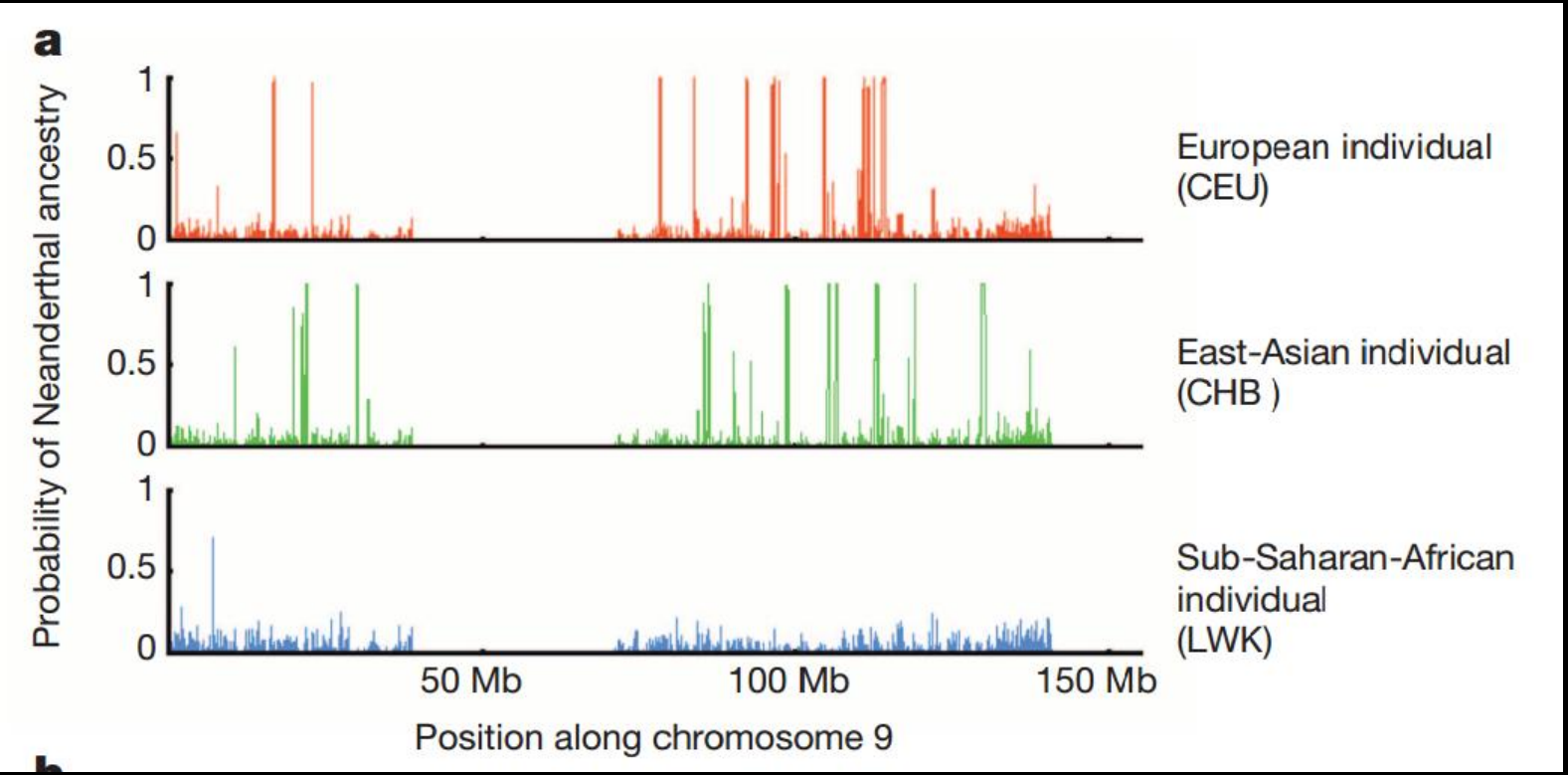
1. A highly diverged, derived haplotype in European individuals.

2. More similar to Neanderthal than African allele, and

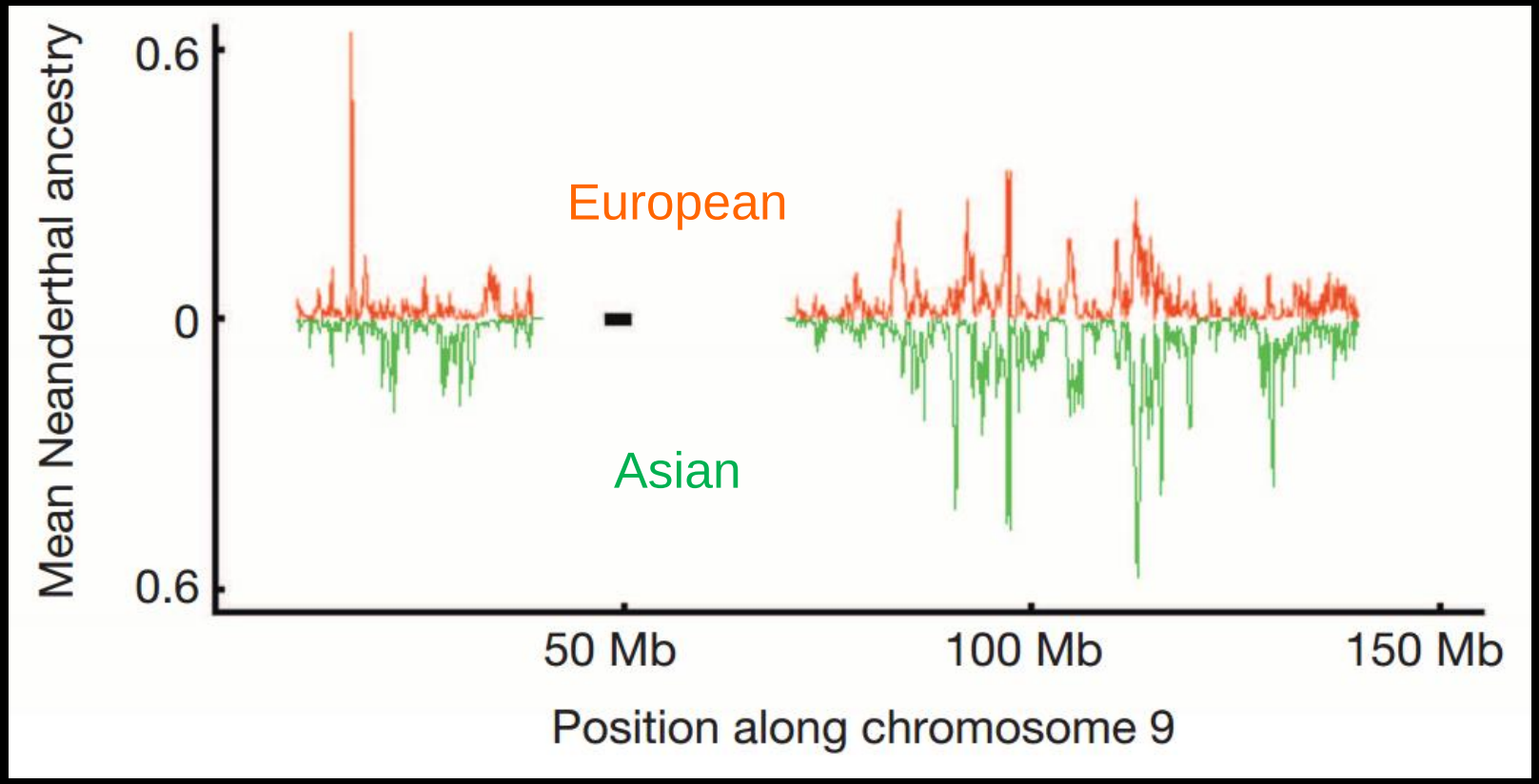
3. Haplotype of expected span (0.05 cM) given the time, 2000 generations.



Only non-Africans have substantial Neanderthal ancestry



Neanderthal ancestry proportion varies along the genome

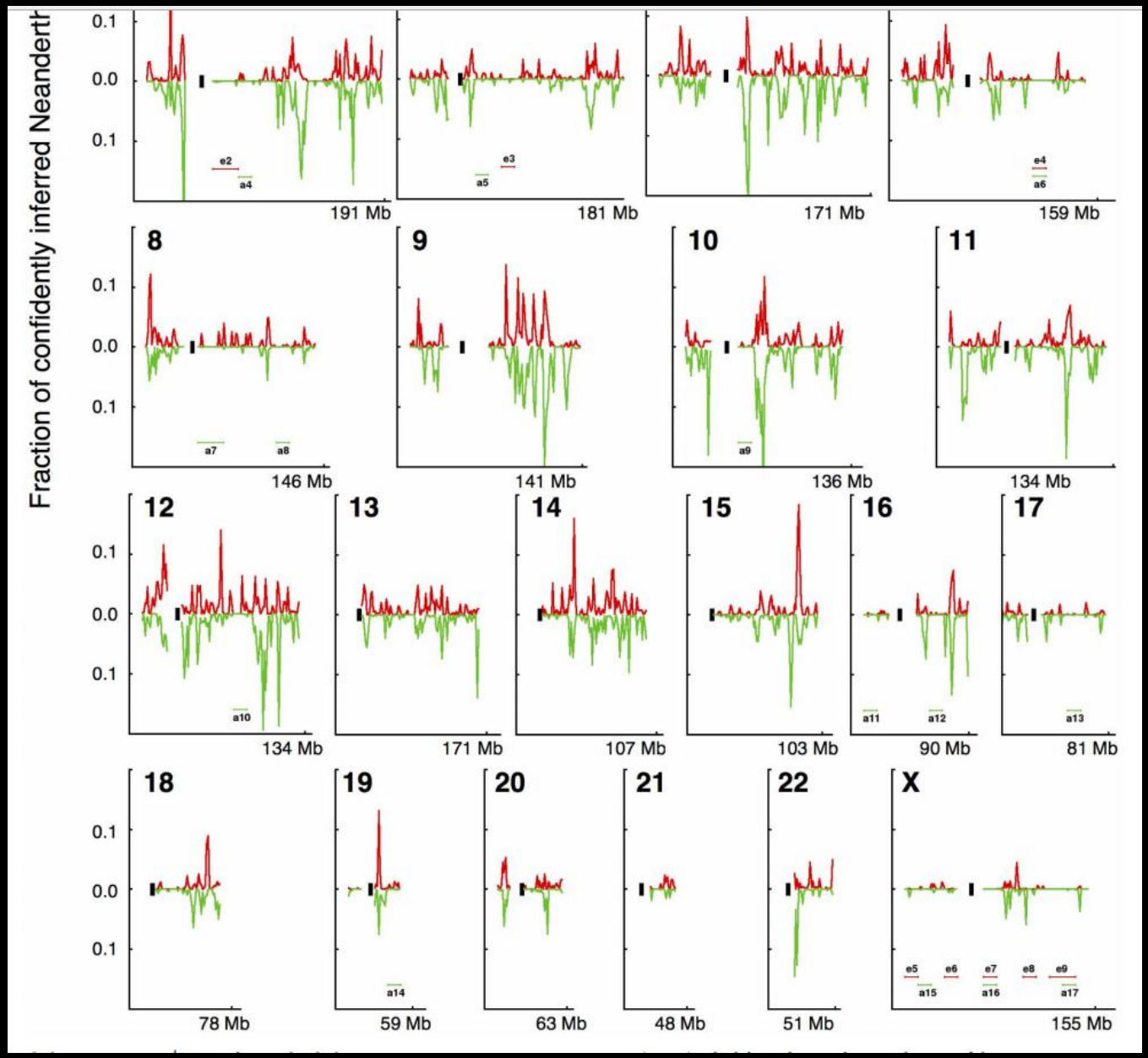


Non-African genomes are up to 2% Neanderthal

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

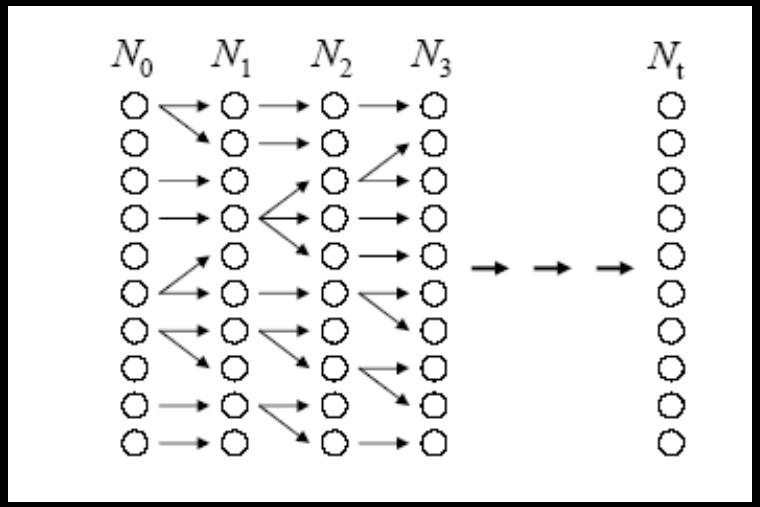
rs id	Coordinates	Derived allele	Derived allele frequency (%)		Phenotype
			Europeans	East Asians	
rs12531711	7:128,617,466	G	10.03	0.17	Systemic lupus erythematosus, Primary biliary cirrhosis
rs3025343	9:136,478,355	A	8.44	0.00	Smoking behavior
rs7076156	10:64,415,184	A	26.52	8.74	Crohn's disease
rs12571093	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

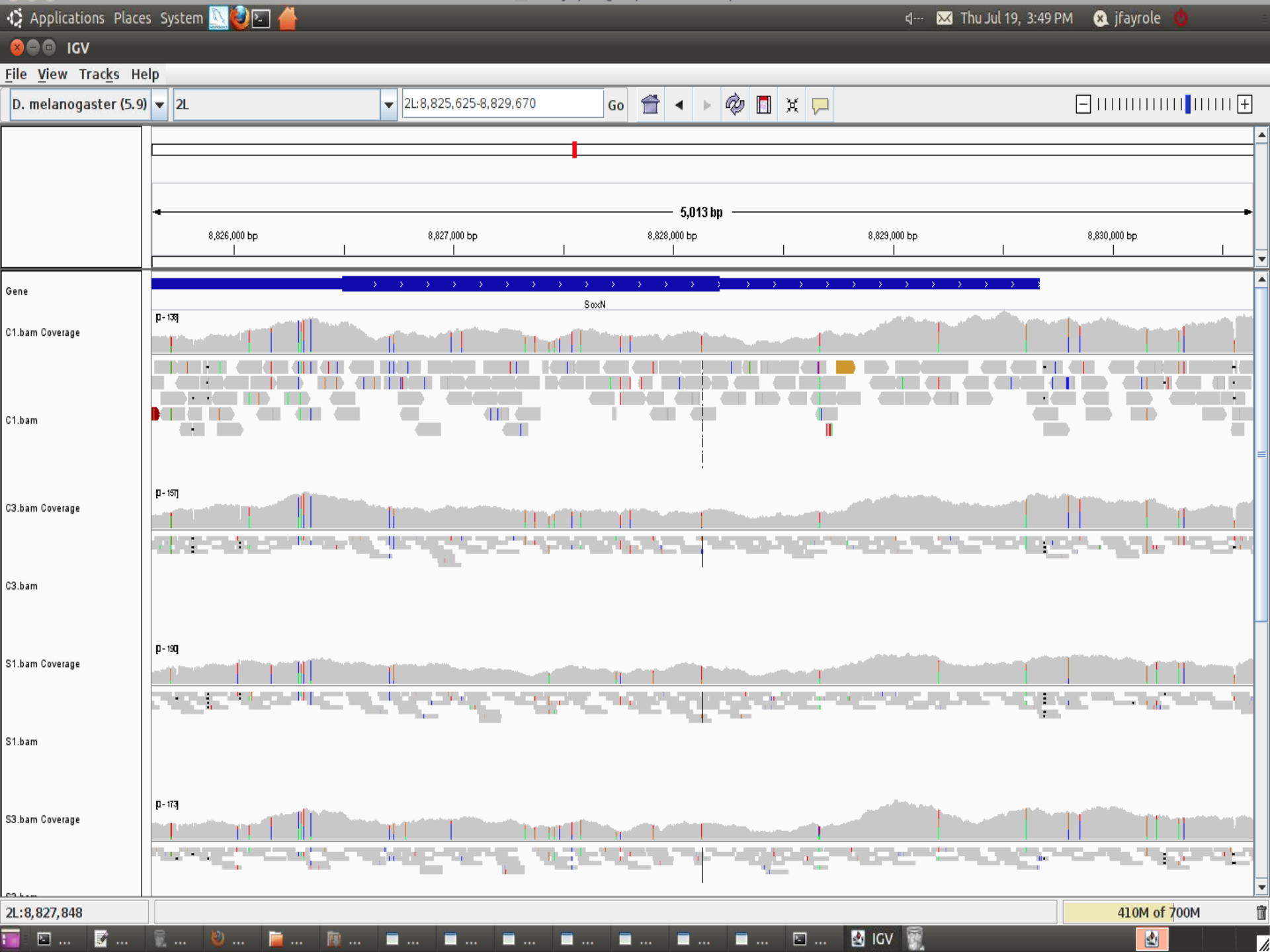
What can we infer from allele frequency dynamics of every nucleotide in the genome?

The Wright-Fisher drift model:

the Null model for Evolve-and-Resequence

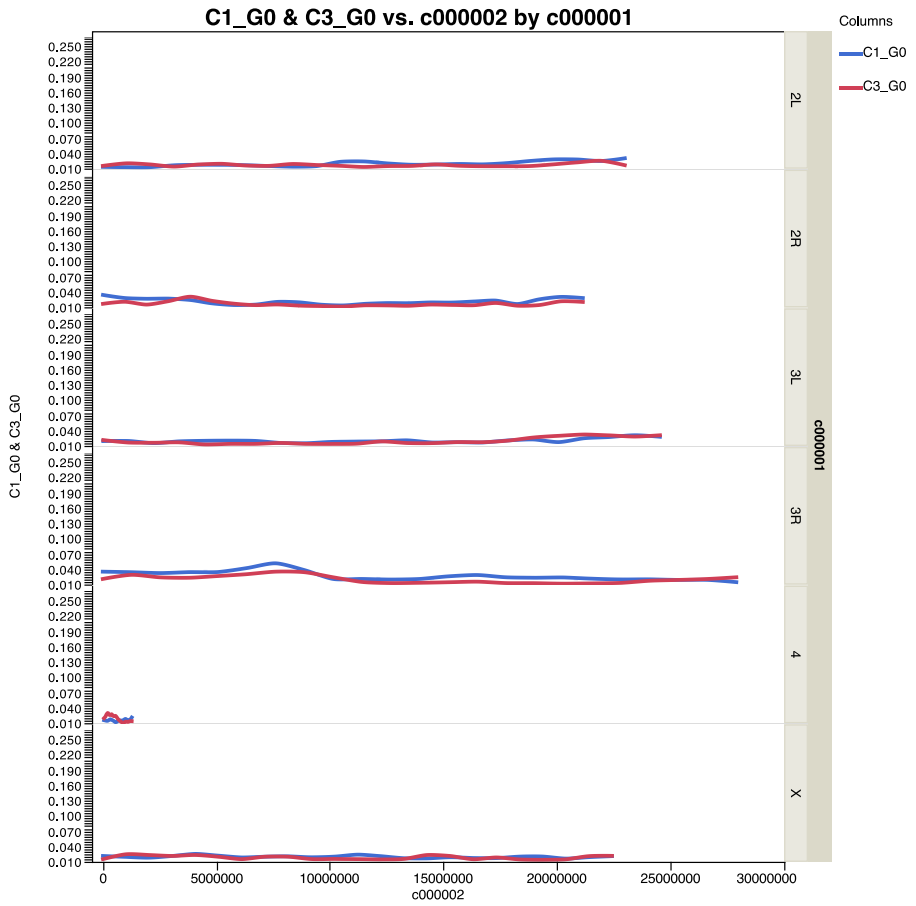


- Selfing allowed
- Random mating
- Non-overlapping generations
- Constant population size
- No migration
- No selection

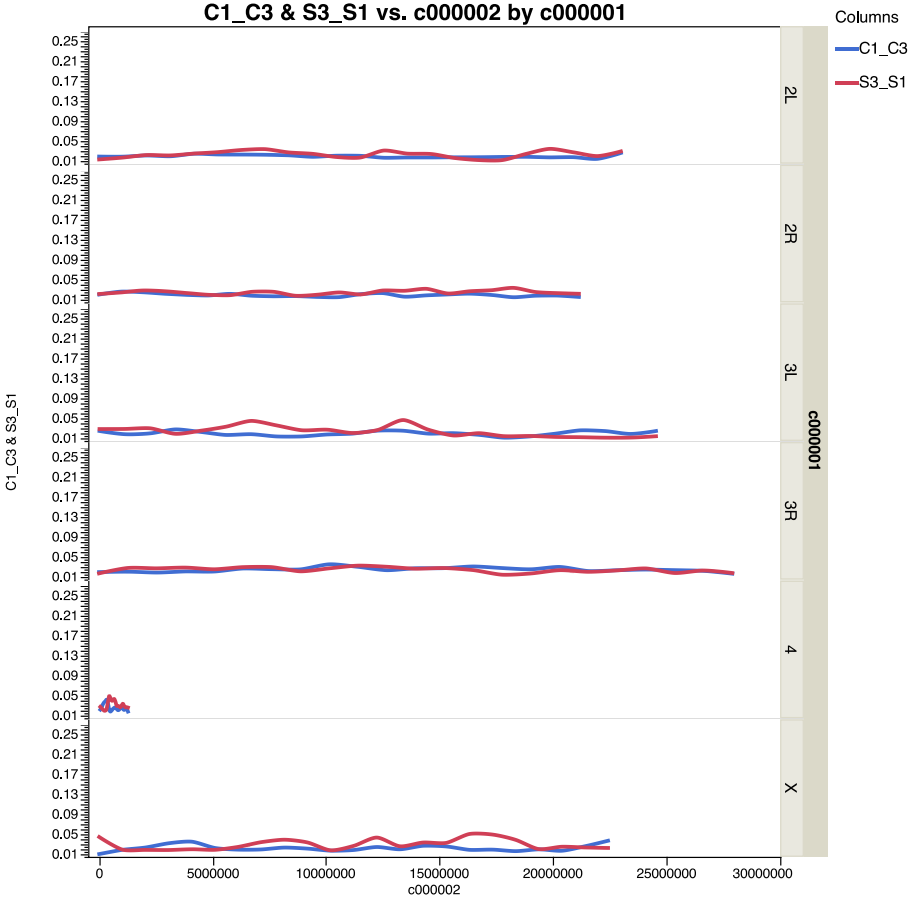

Applications Places System
Thu Jul 19, 3:49 PM
jfayrole
IGV
File View Tracks Help
D. melanogaster (5.9)
2L
2L:8,825,625-8,829,670
Go
5,013 bp
8,826,000 bp
8,827,000 bp
8,828,000 bp
8,829,000 bp
8,830,000 bp
Gene
SoxN
C1.bam Coverage
C1.bam
C3.bam Coverage
C3.bam
S1.bam Coverage
S1.bam
S3.bam Coverage
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)

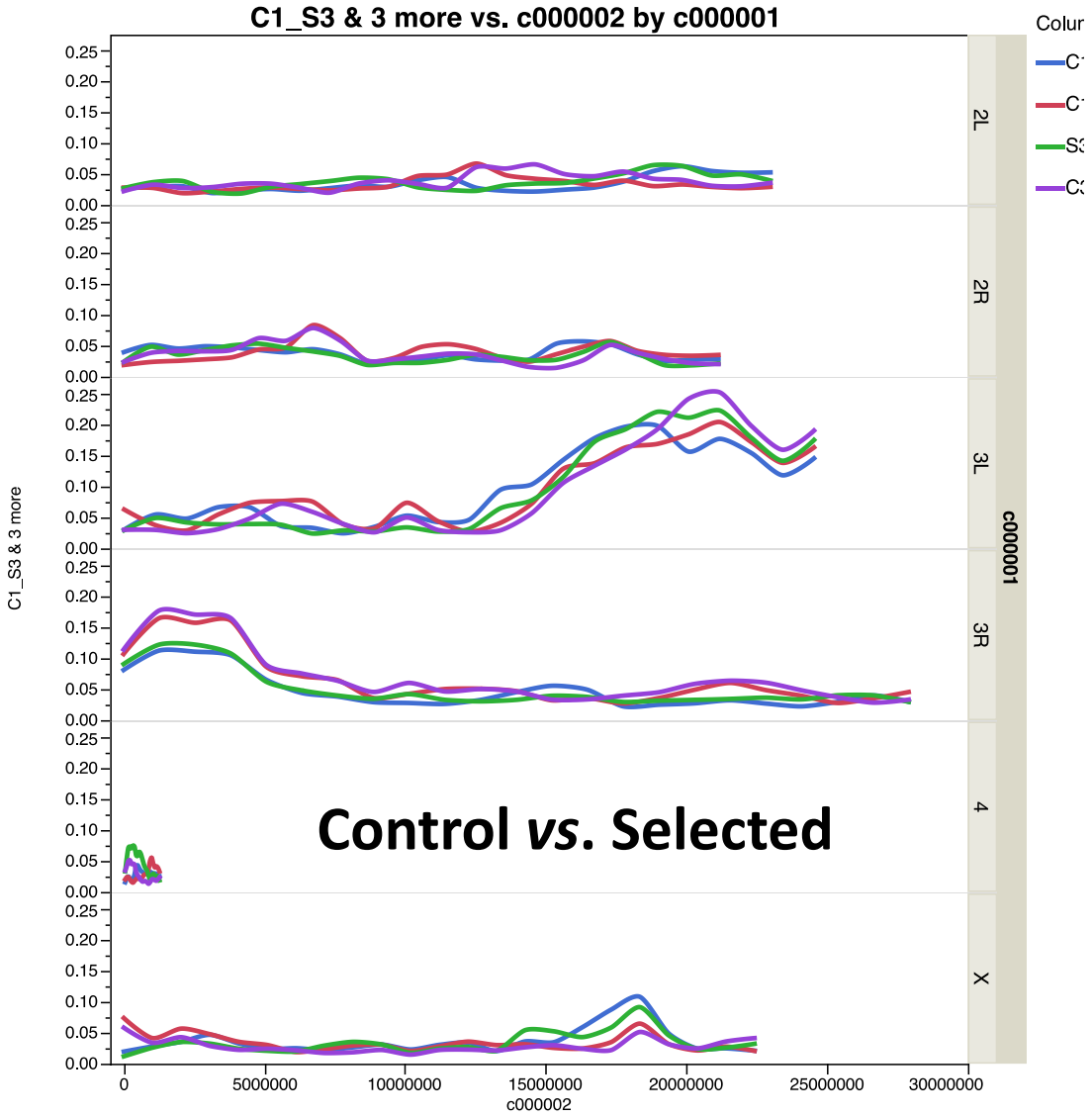


Control *vs*. G0



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)

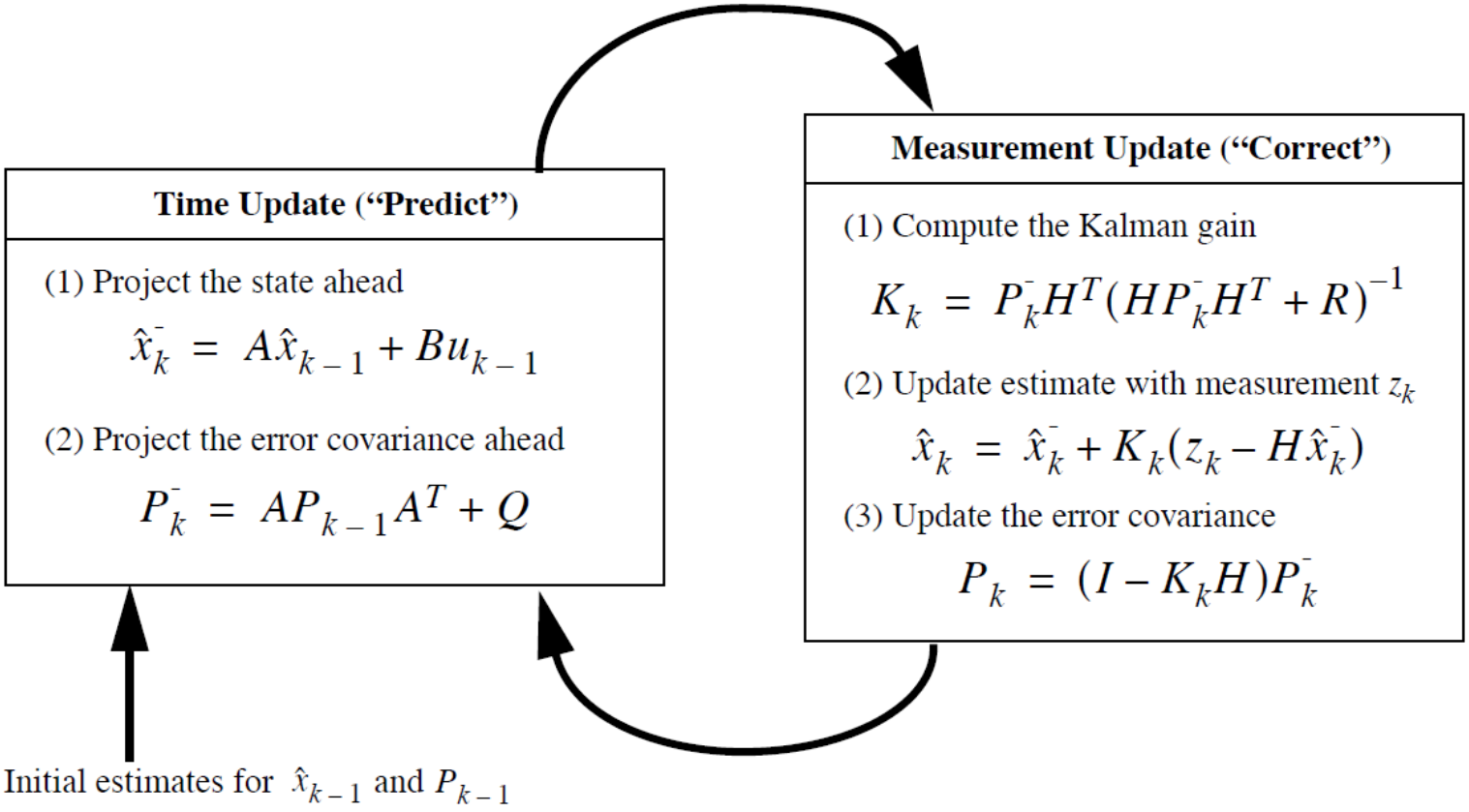
Process error

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



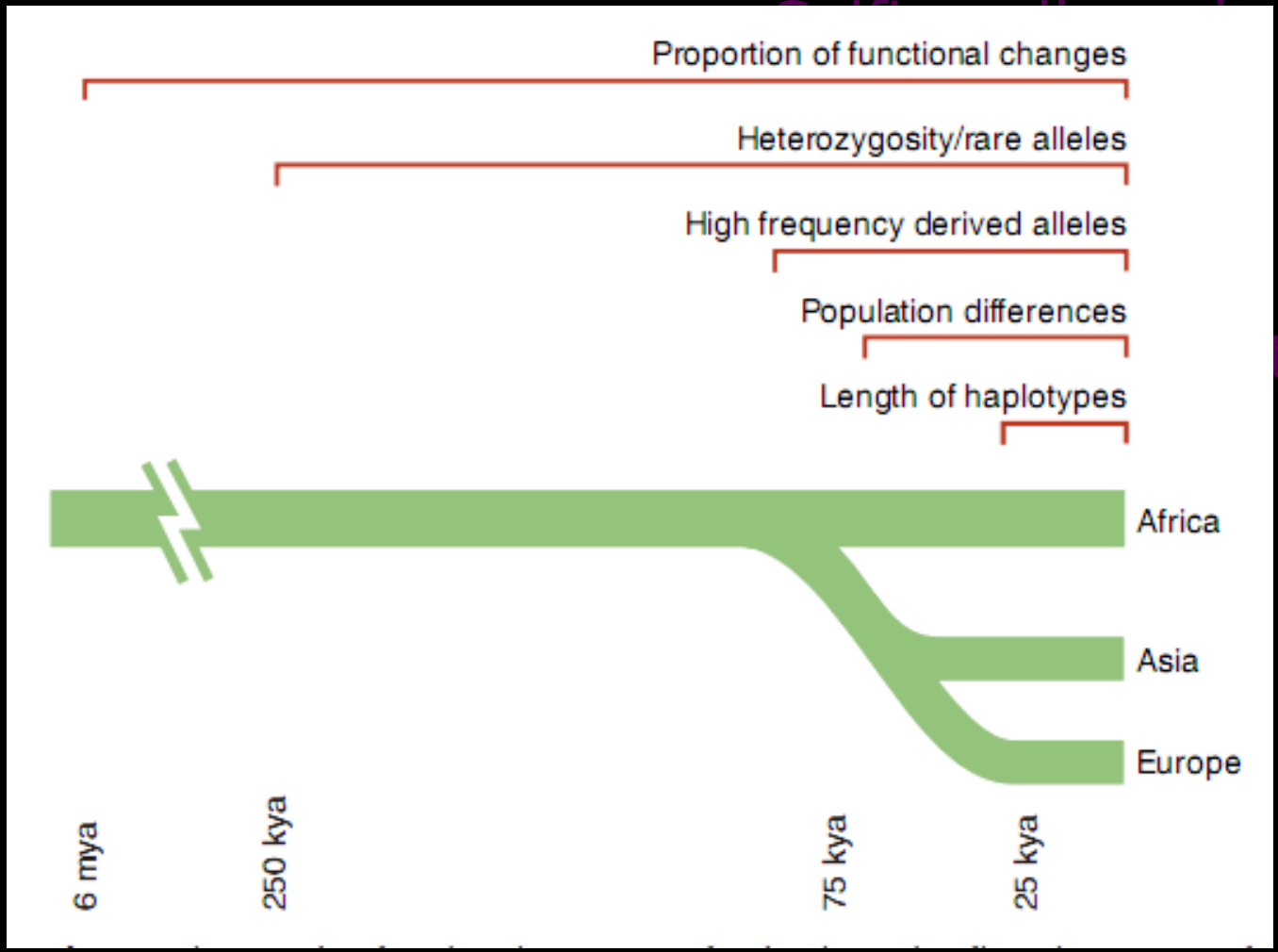
Welsh and Bishop 2006. An Introduction to the Kalman Filter
UNC-Chapel Hill, TR 95-041

Natural Selection

How best to do inference of selection genome-wide?

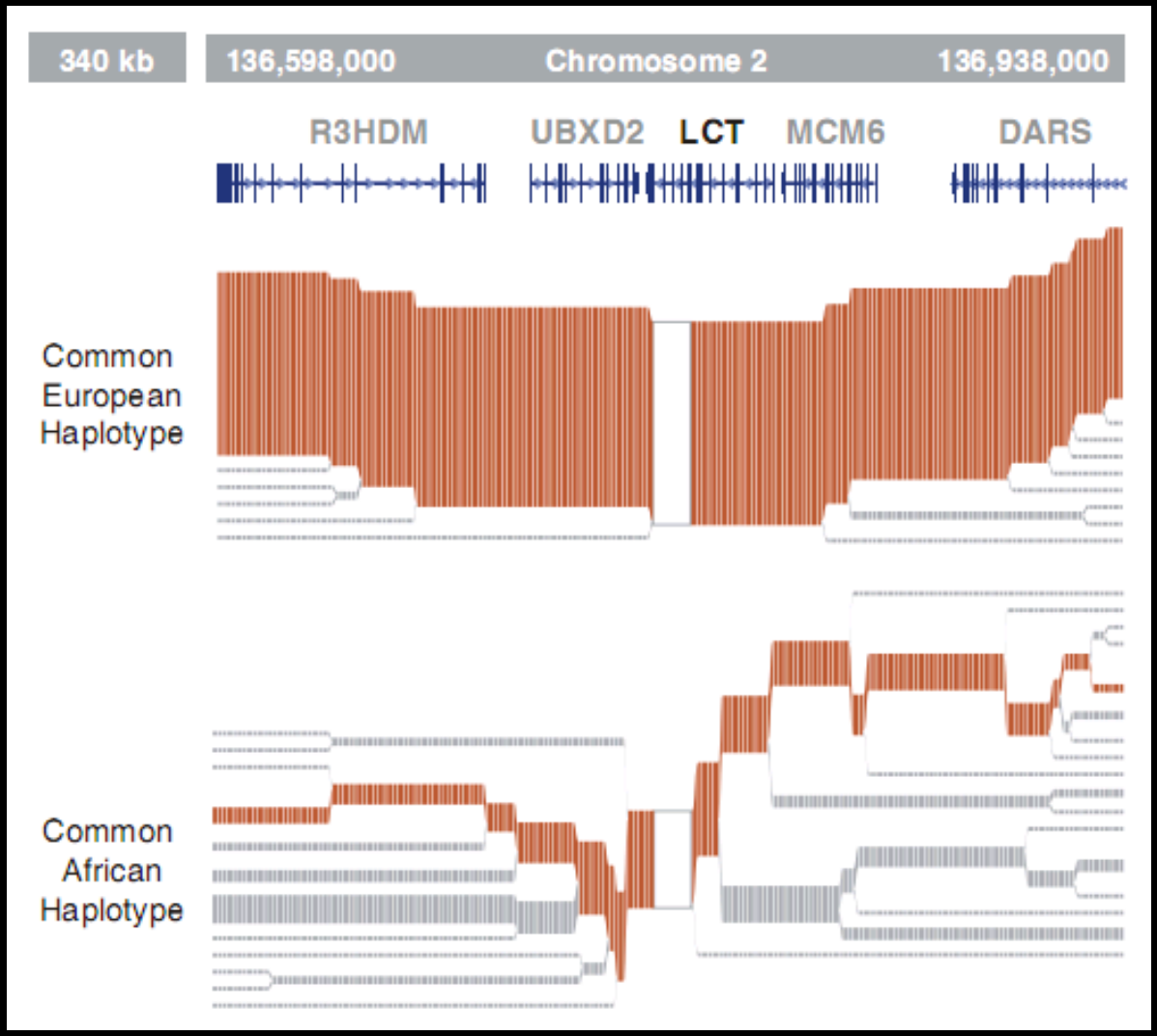
Scans for selection

Different signals at different time depths



- Sabeti *et al.* 2006 *Science*

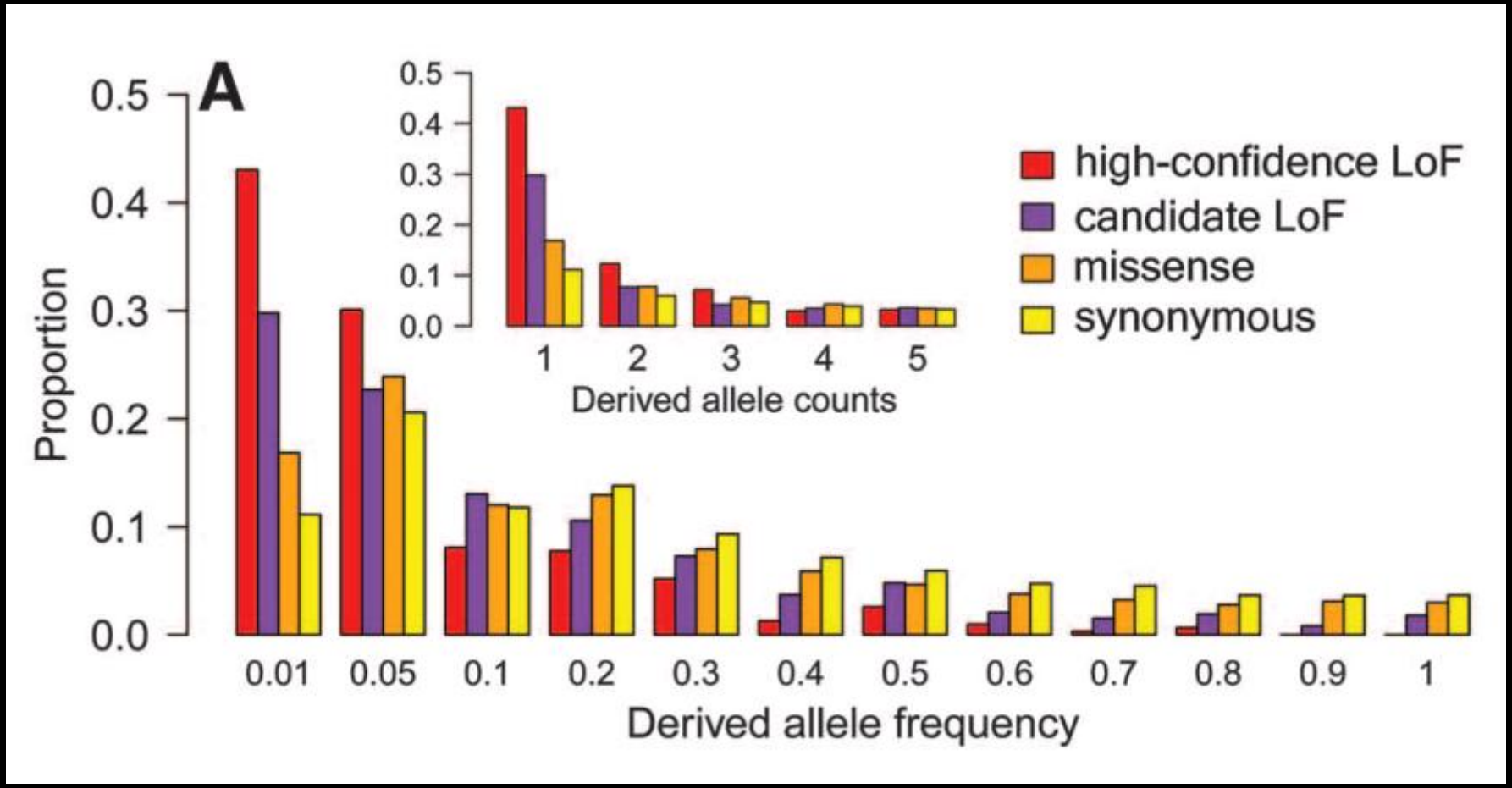
Impressive selective sweeps (are rare in humans)



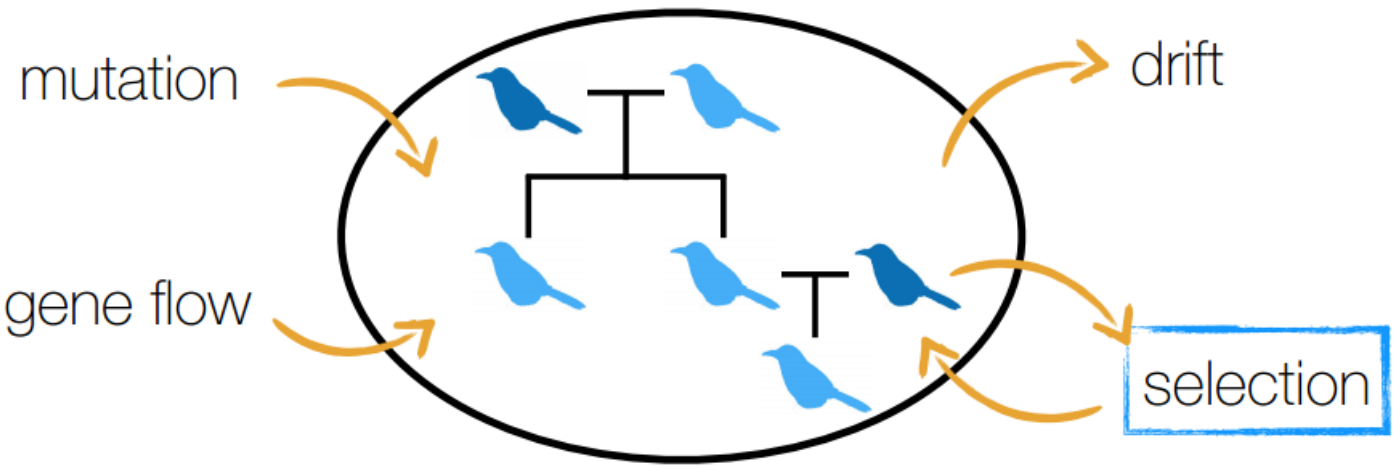
- Sabeti *et al.* 2006 *Science*

ExAc and enormous sample sizes

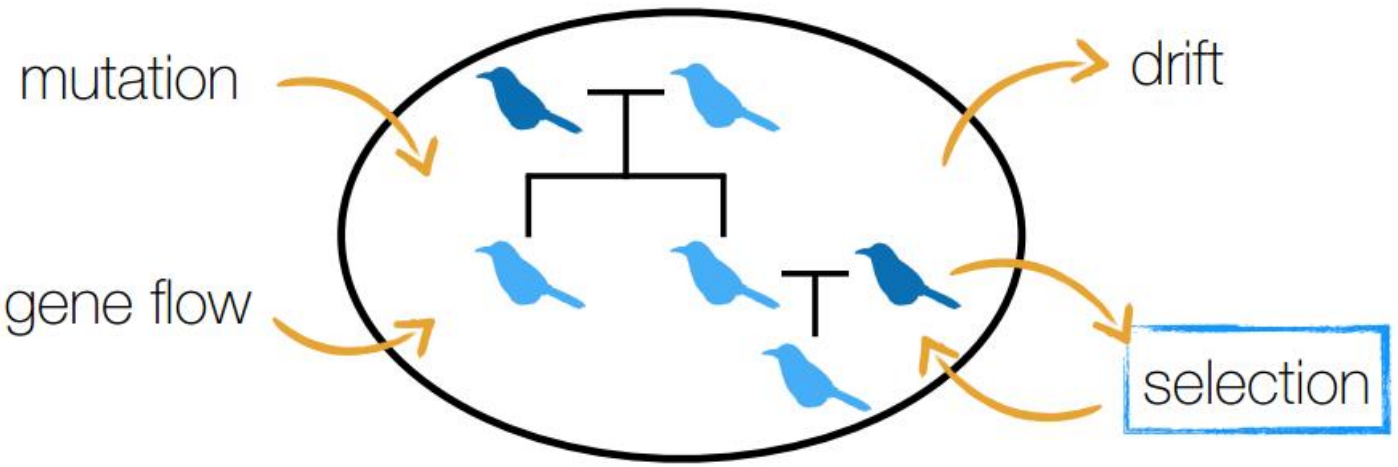
(60,000 whole exome sequences)



The genetic basis of contemporary evolution in nature



The genetic basis of contemporary evolution in nature

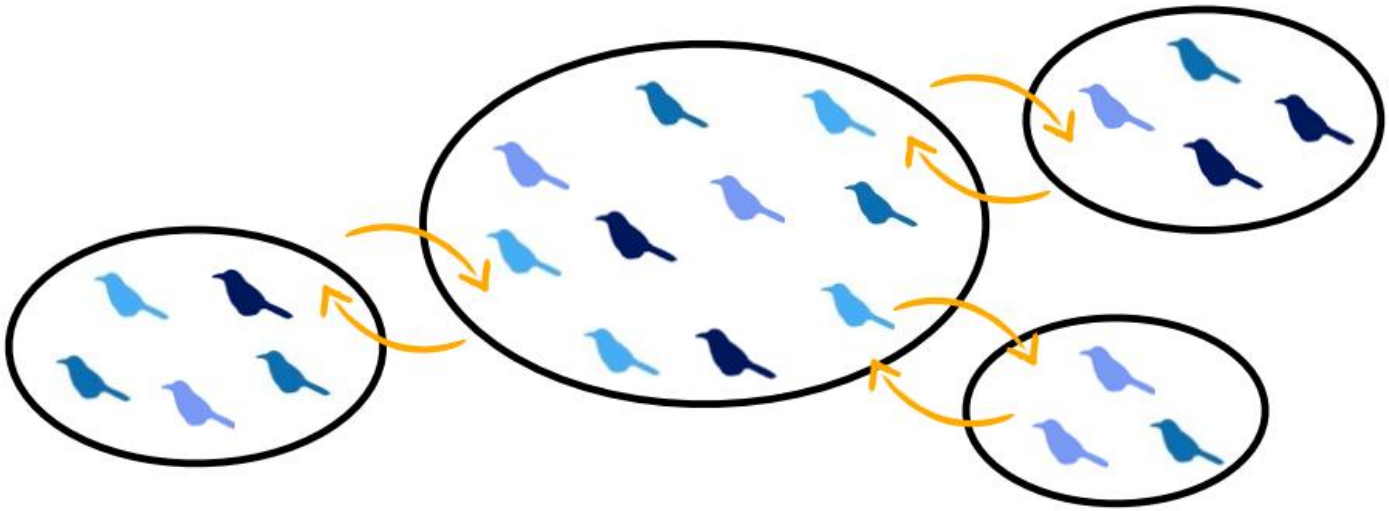


1. What are the genetic & fitness consequences of fragmentation?

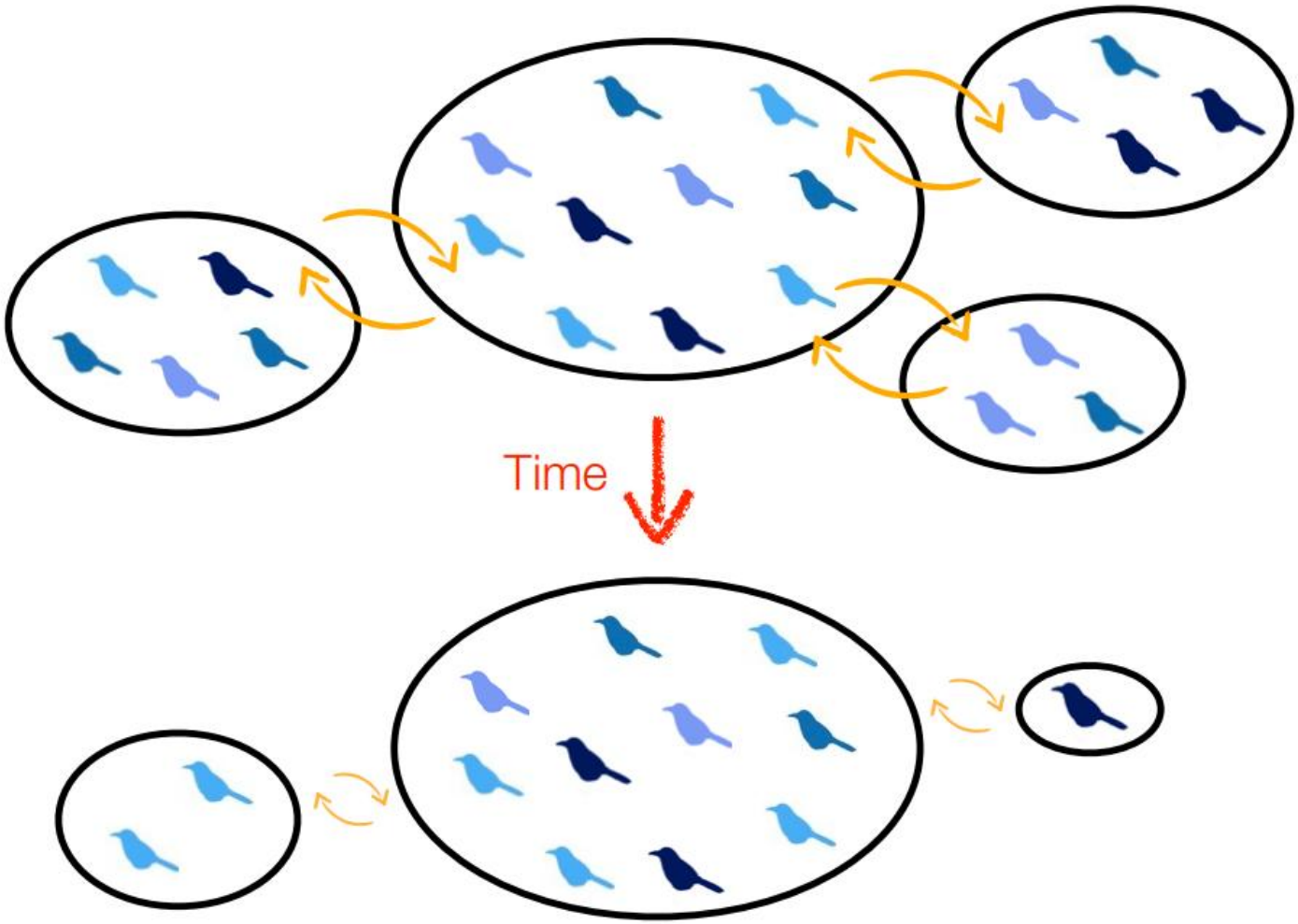
2. What are the genomic consequences of limited dispersal?

3. Can we detect short-term selection in the genome?

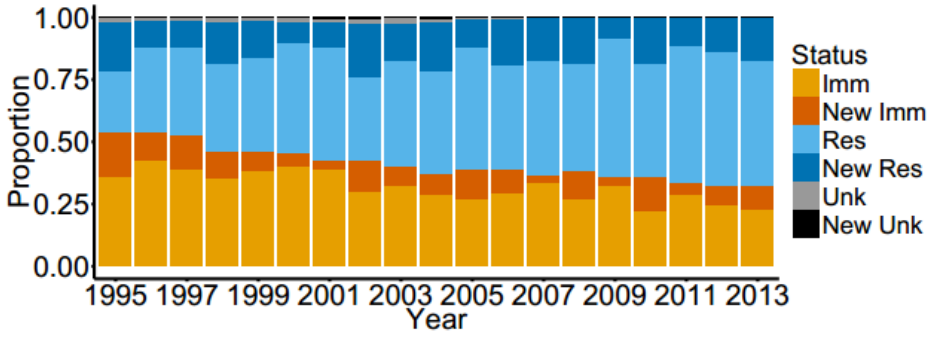
Consequences of habitat fragmentation



Consequences of habitat fragmentation

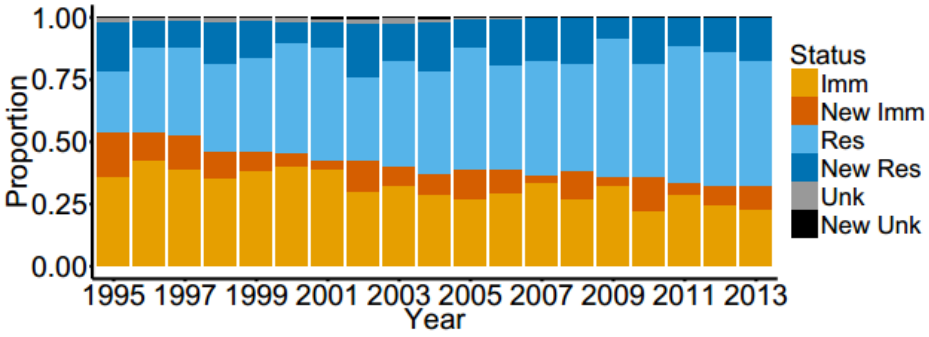


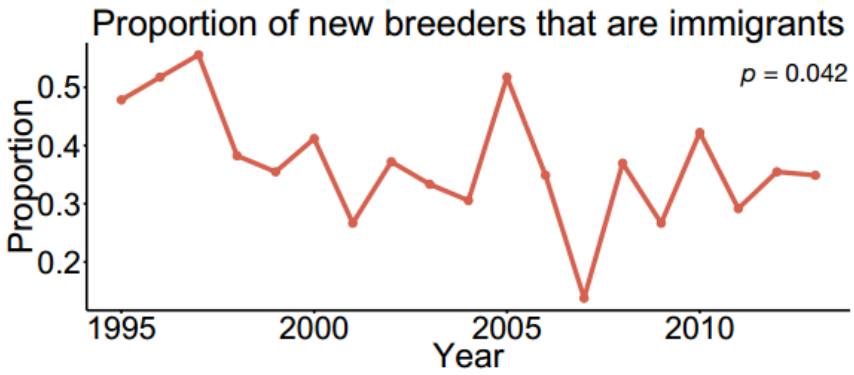
Immigration rate decreased over time



Chen *et al.* 2016 *Current Biology*

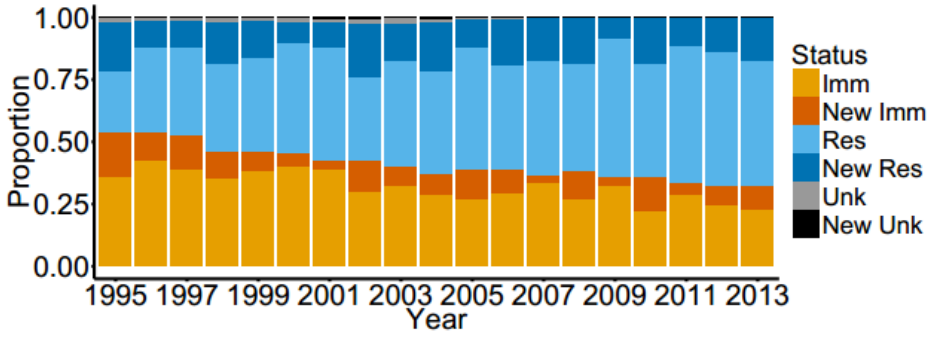
Immigration rate decreased over time

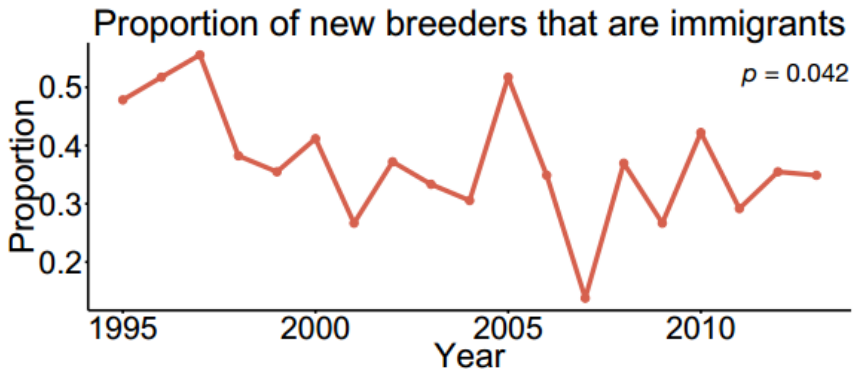


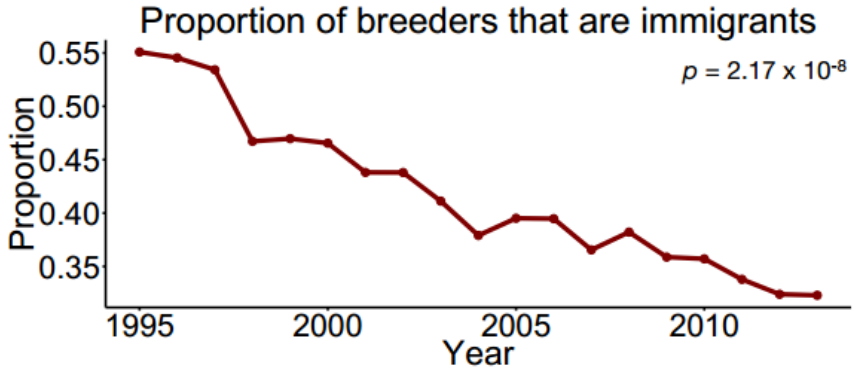


Chen *et al.* 2016 *Current Biology*

Immigration rate decreased over time







Chen *et al.* 2016 *Current Biology*

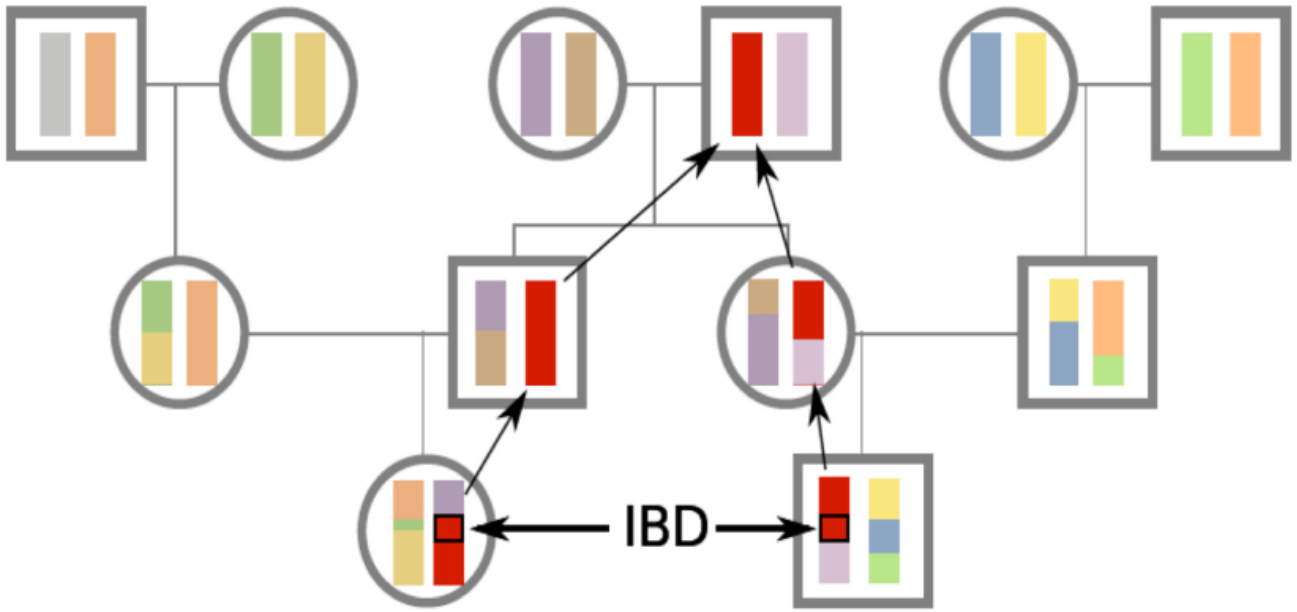
Measures of genetic variation

7,834 autosomal SNPs in linkage equilibrium

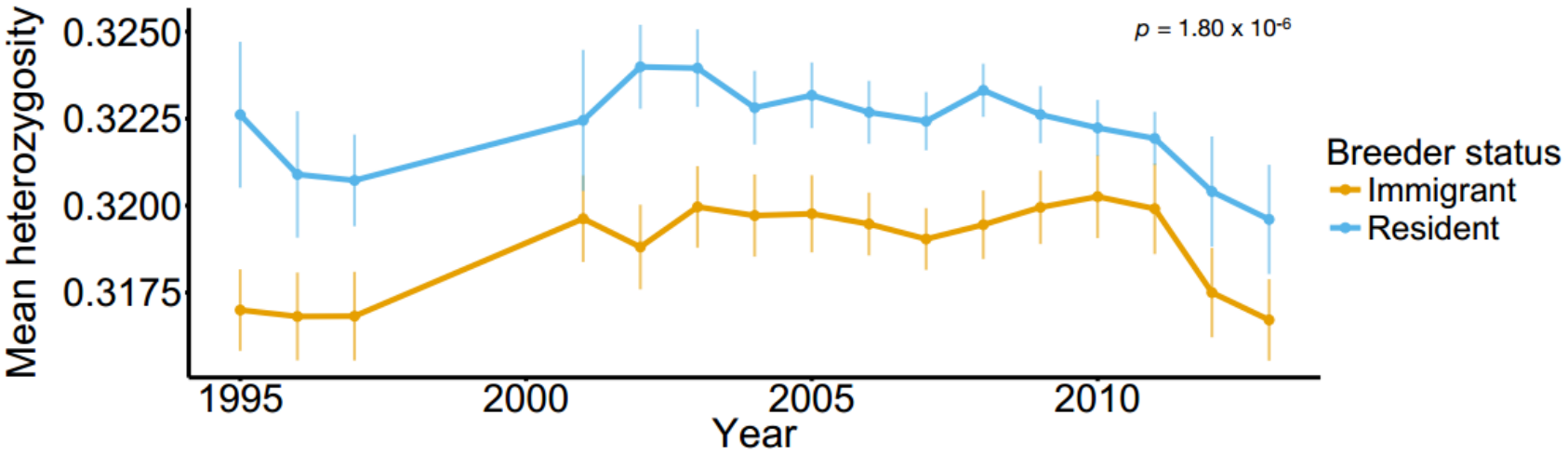
Mean observed heterozygosity

Inbreeding coefficient

Pairwise Identity By Descent (IBD) values

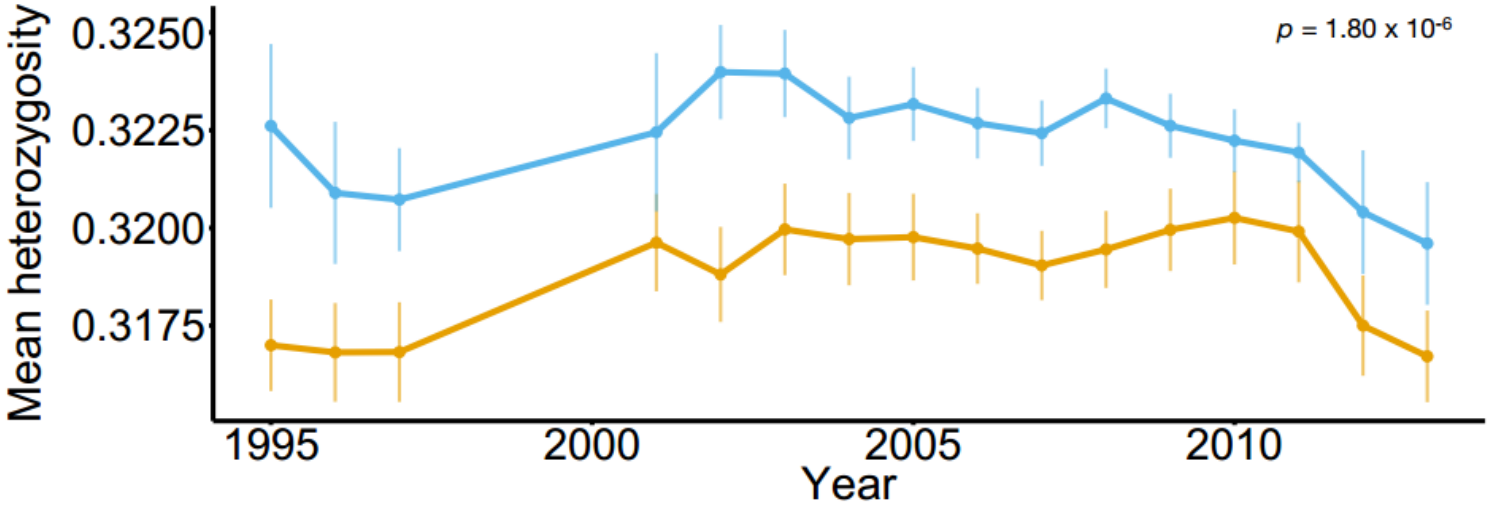


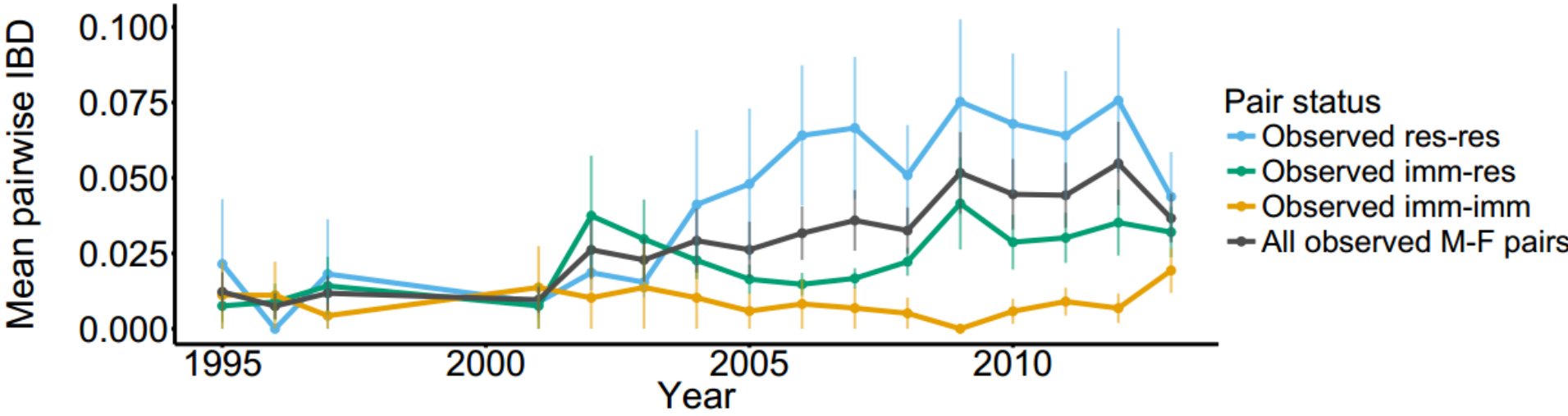
Immigrants were less heterozygous than residents



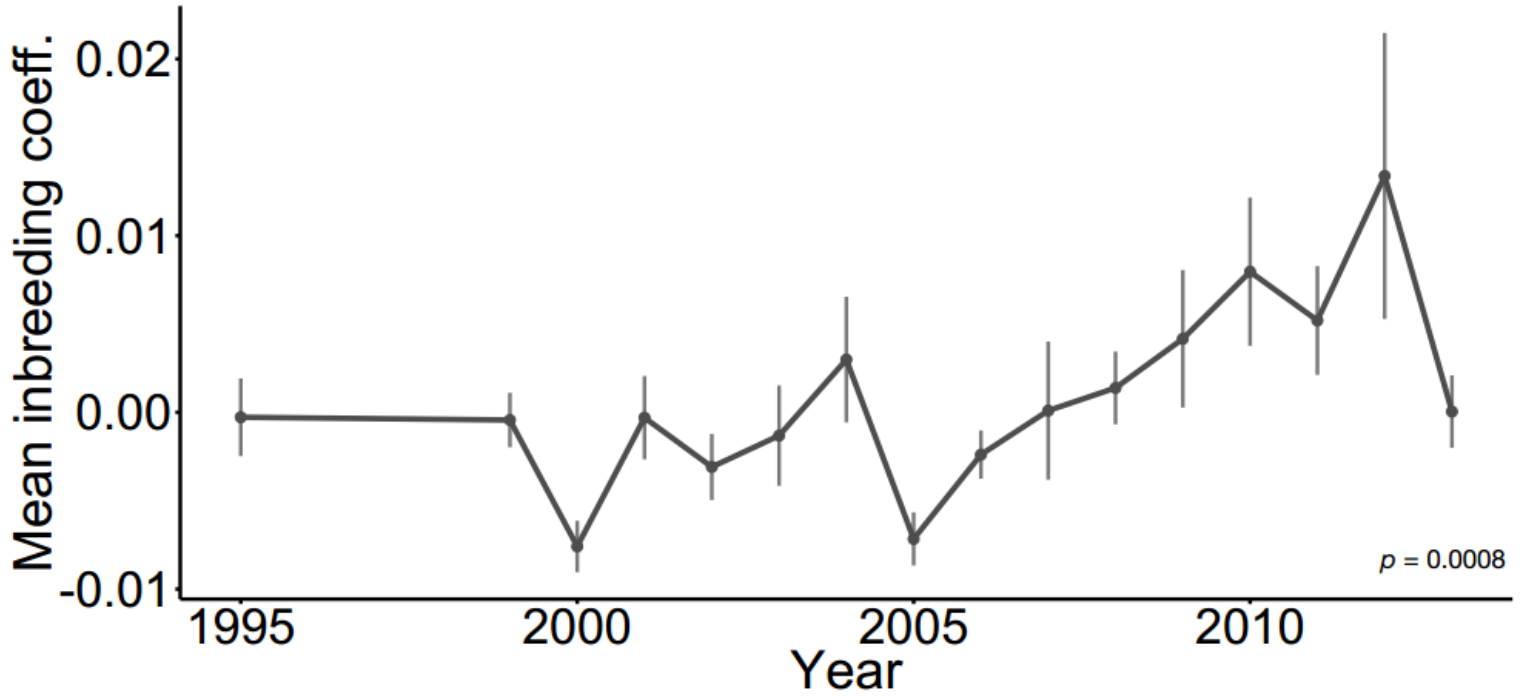
Immigrants were less heterozygous than residents

but still contributed genetic variation



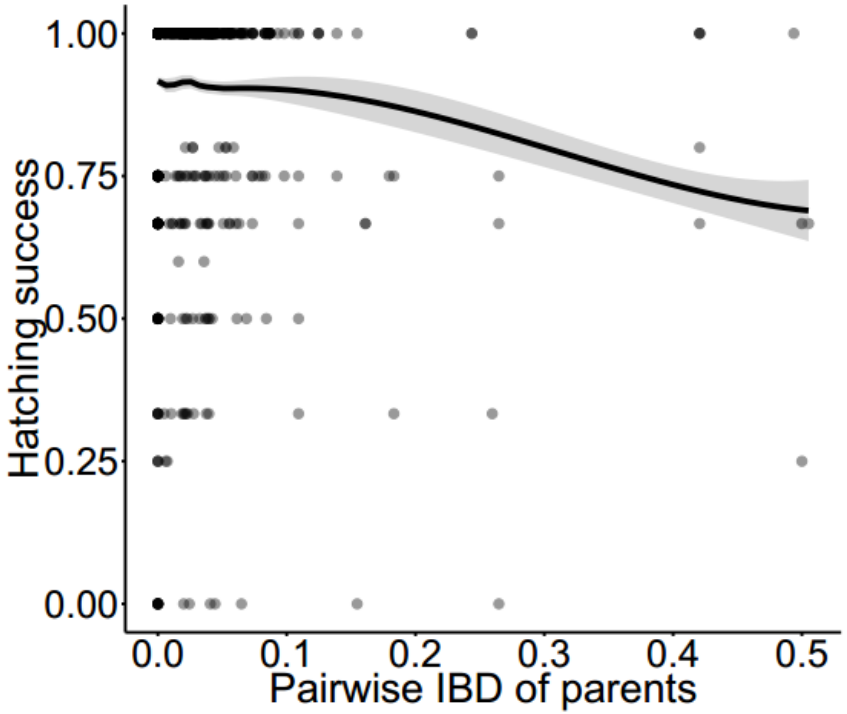


Mean inbreeding coefficient of the birth cohort increased over time

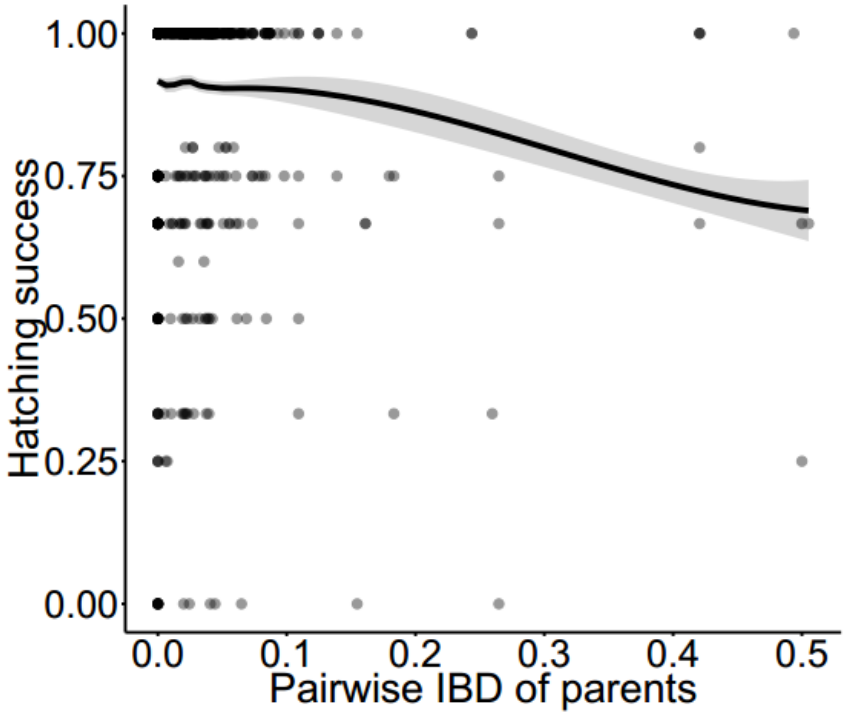


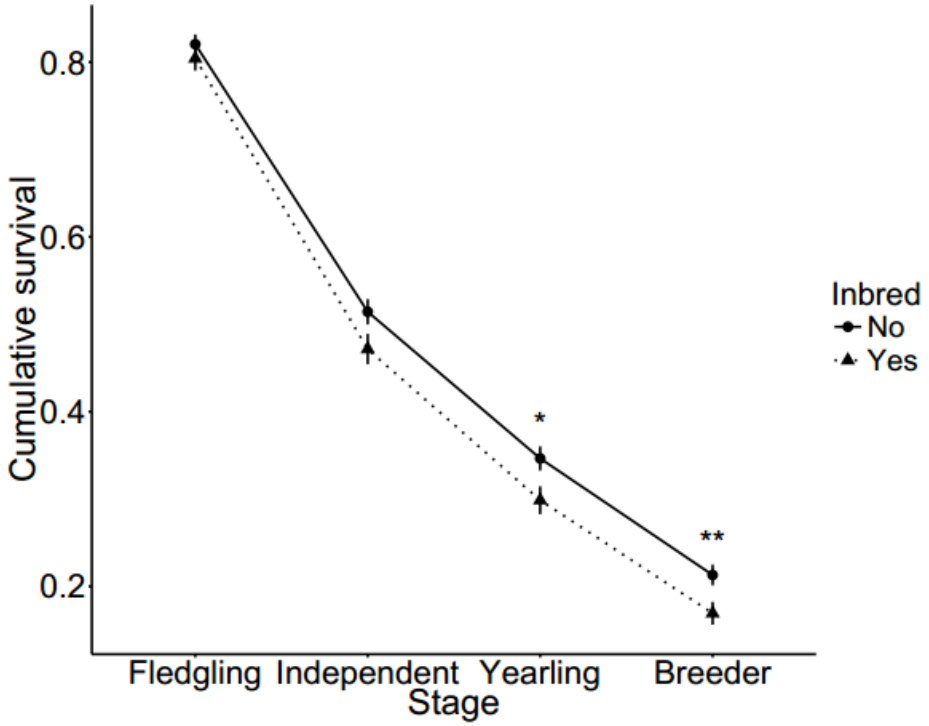
Chen *et al.* 2016 *Current Biology*

Inbreeding depression in multiple life-history stages

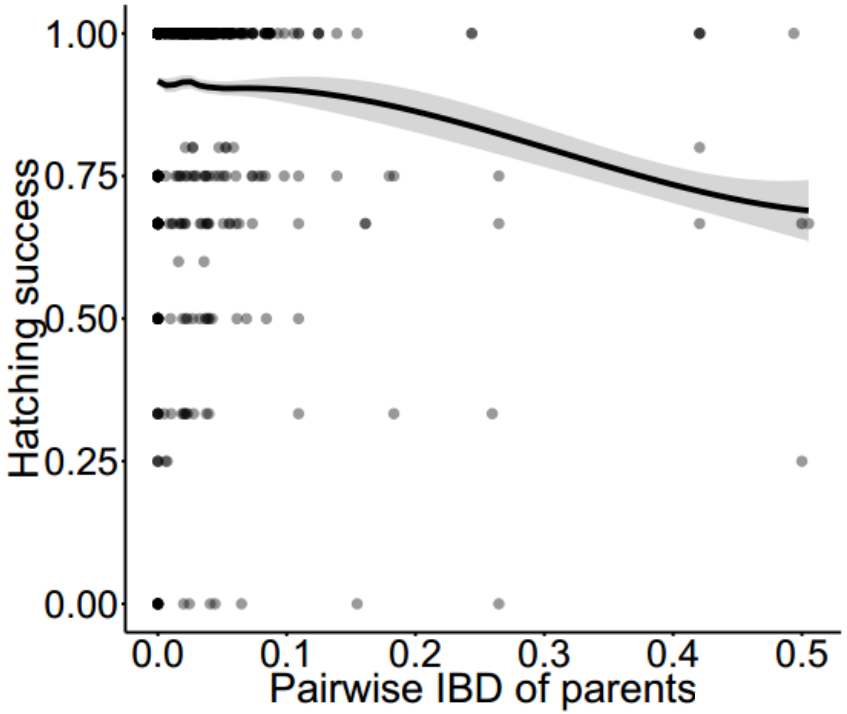


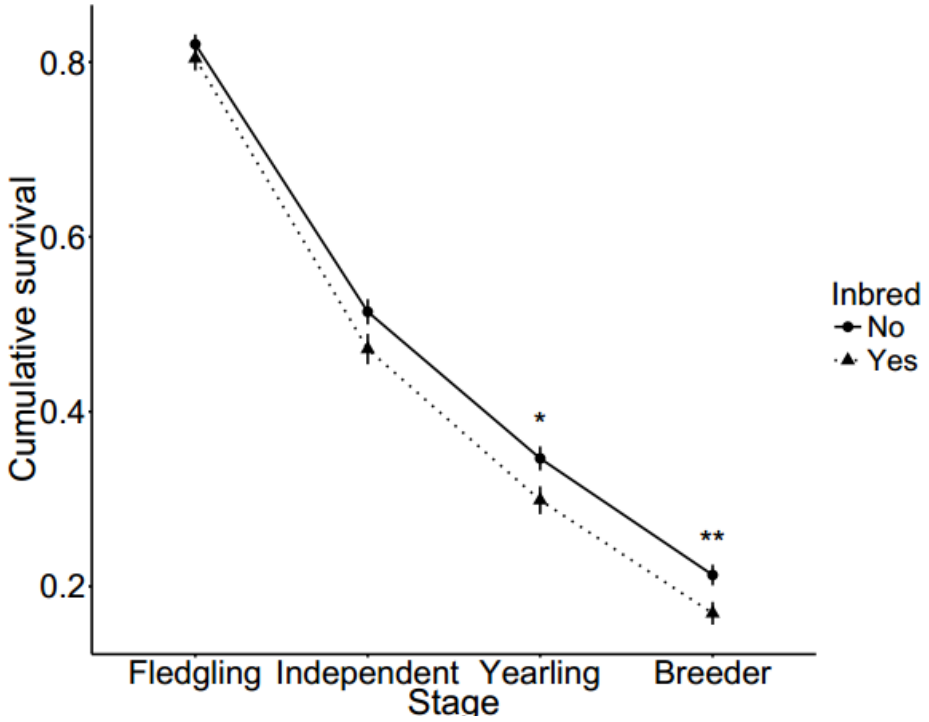
Inbreeding depression in multiple life-history stages





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*

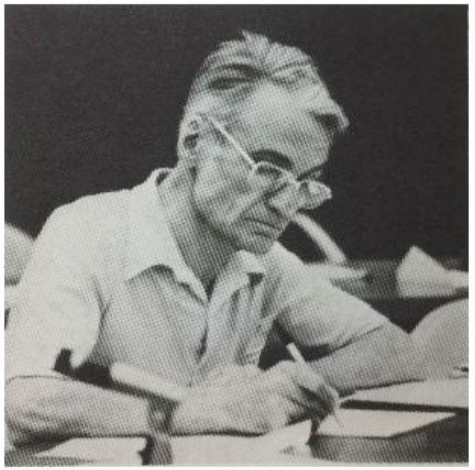
SEWALL WRIGHT
The University of Chicago[1]

Received November 9, 1942

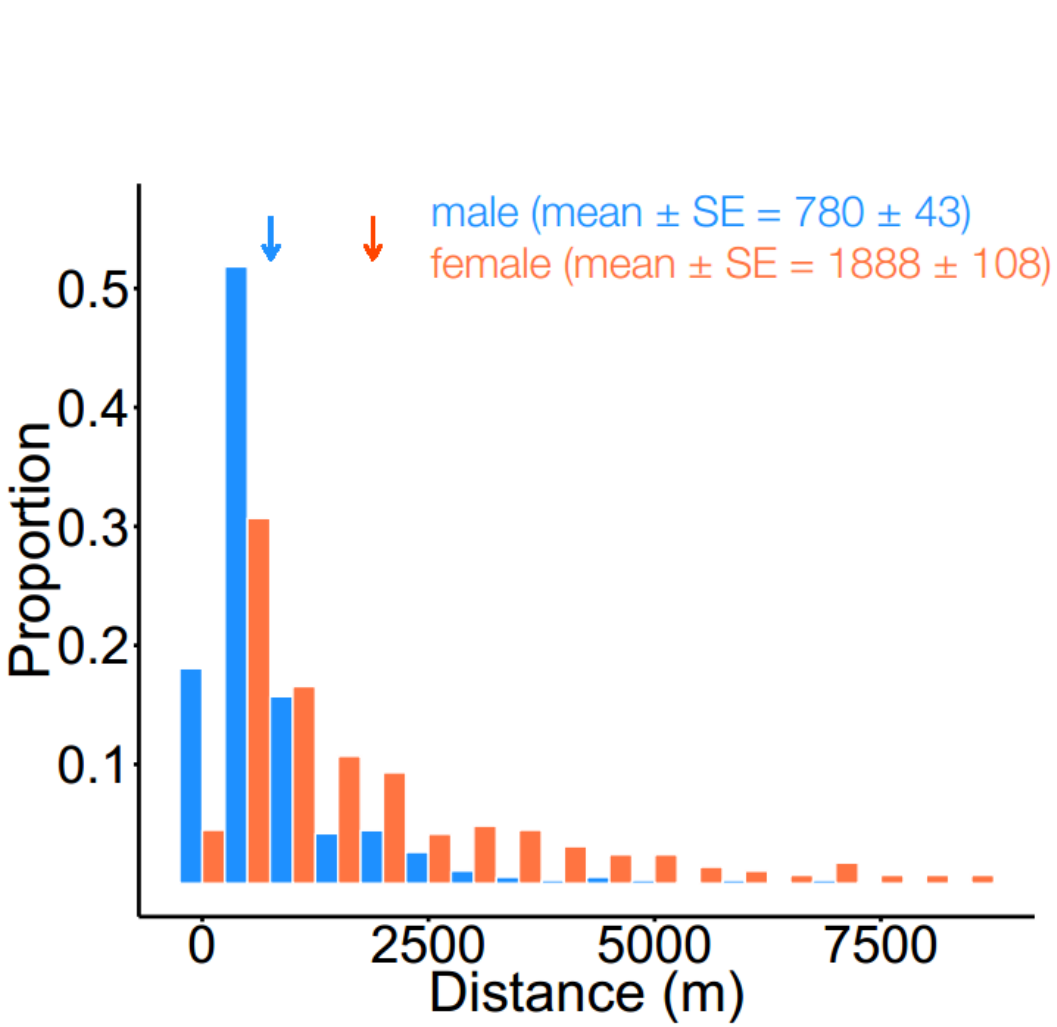
GÉNÉTIQUE DES POPULATIONS DIPLOÏDES NATURELLES DANS LE CAS D'UN SEUL LOCUS

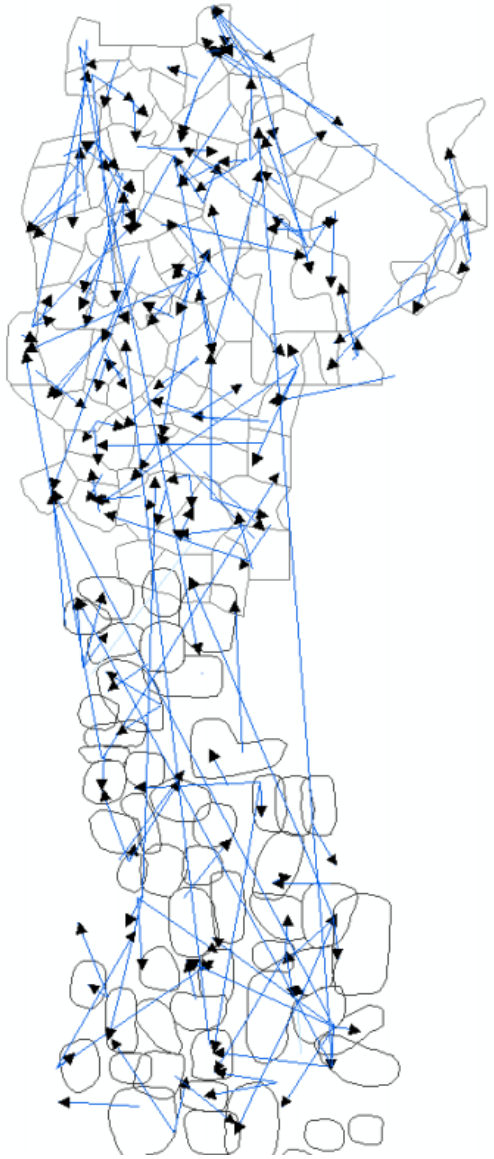
III. — PARENTÉ, MUTATIONS ET MIGRATION

G. MALÉCOT

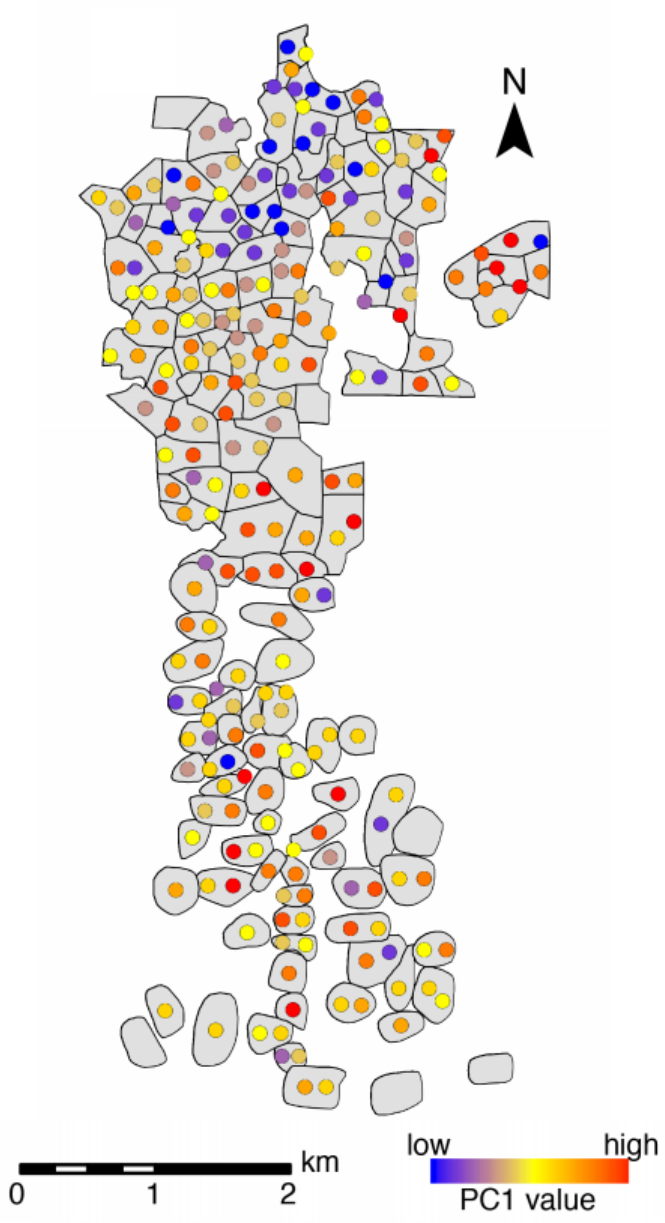


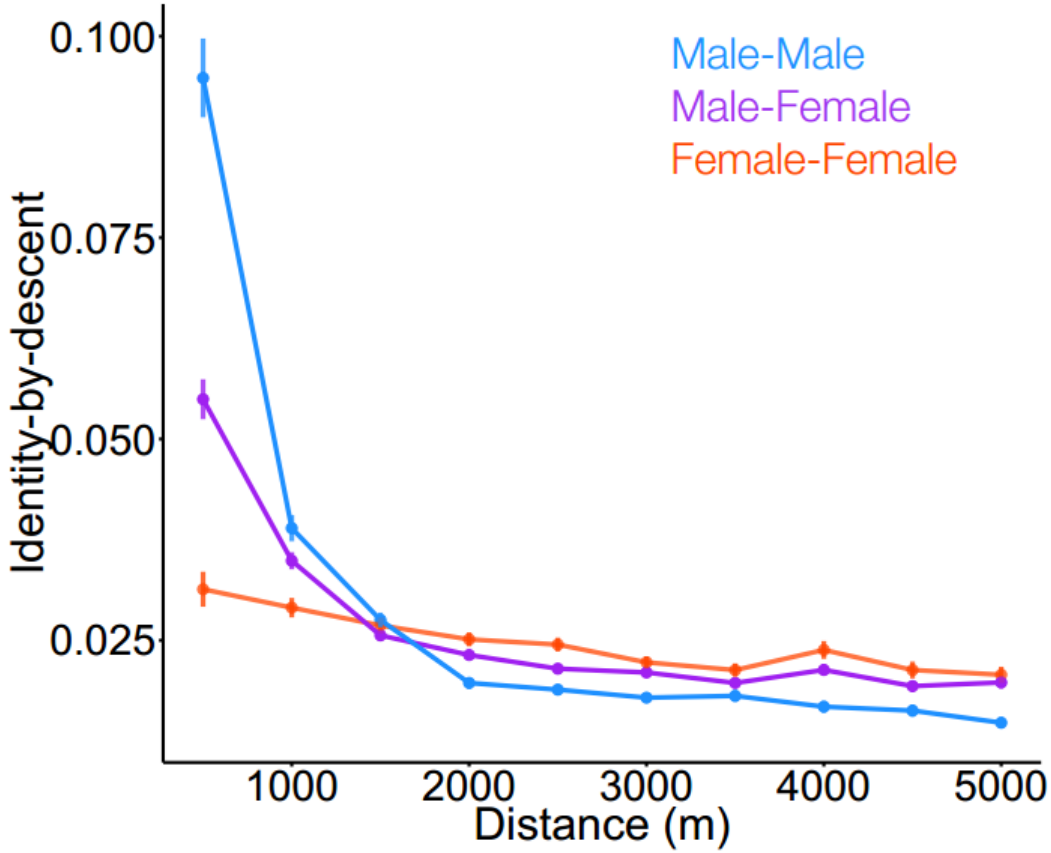
Females disperse farther than males





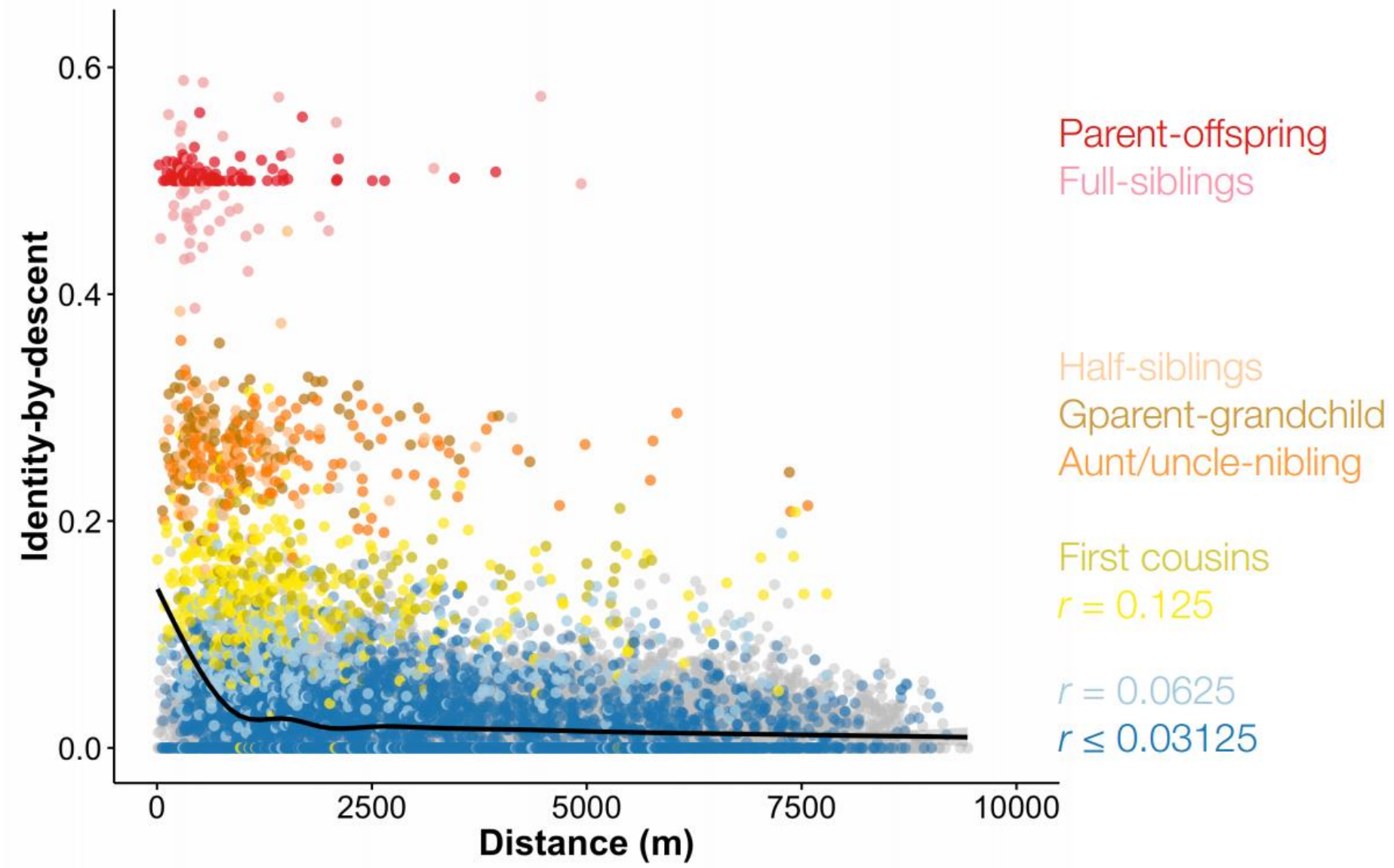
Limited dispersal leads to isolation-by-distance



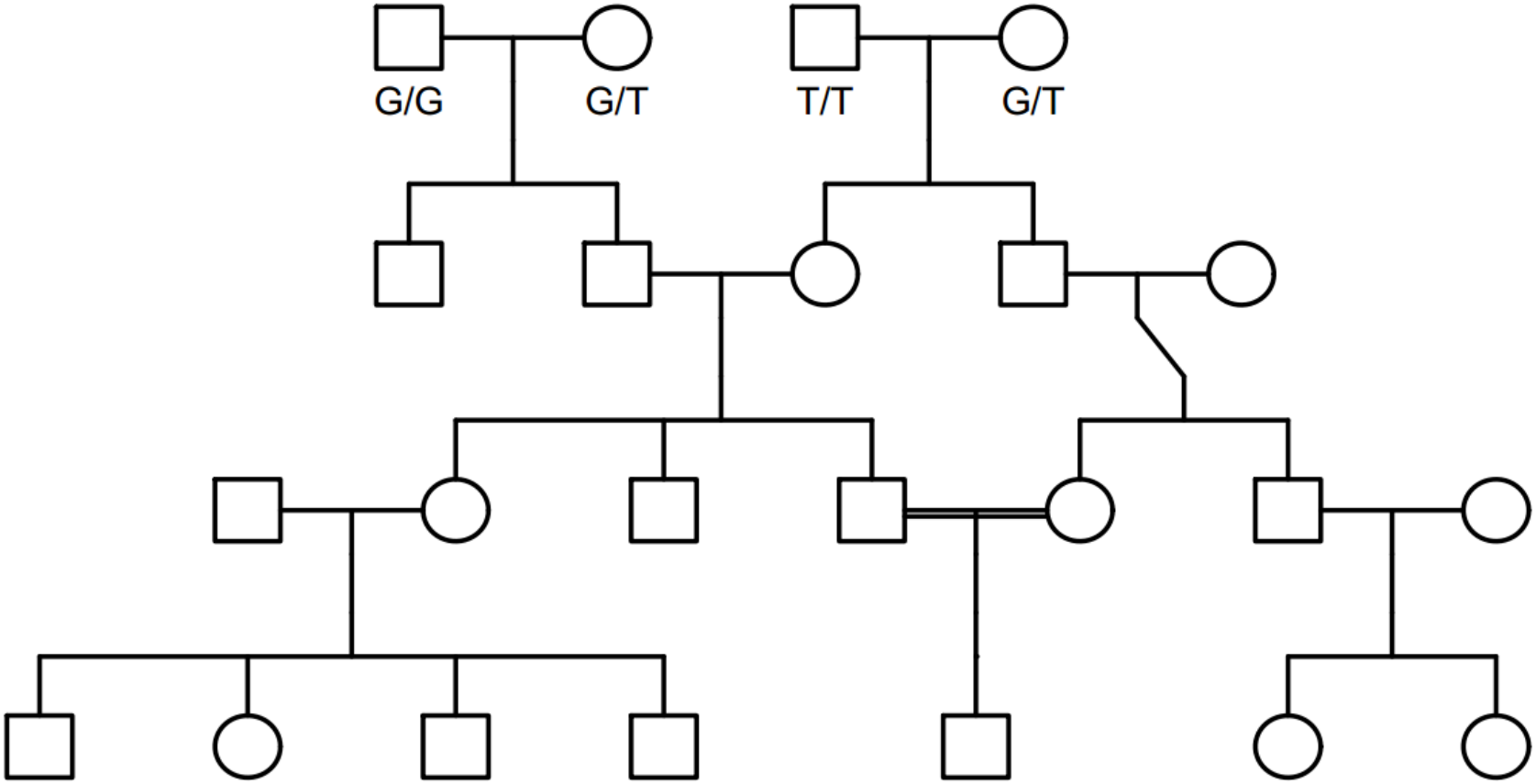


Aguillon, …, Chen *In review*

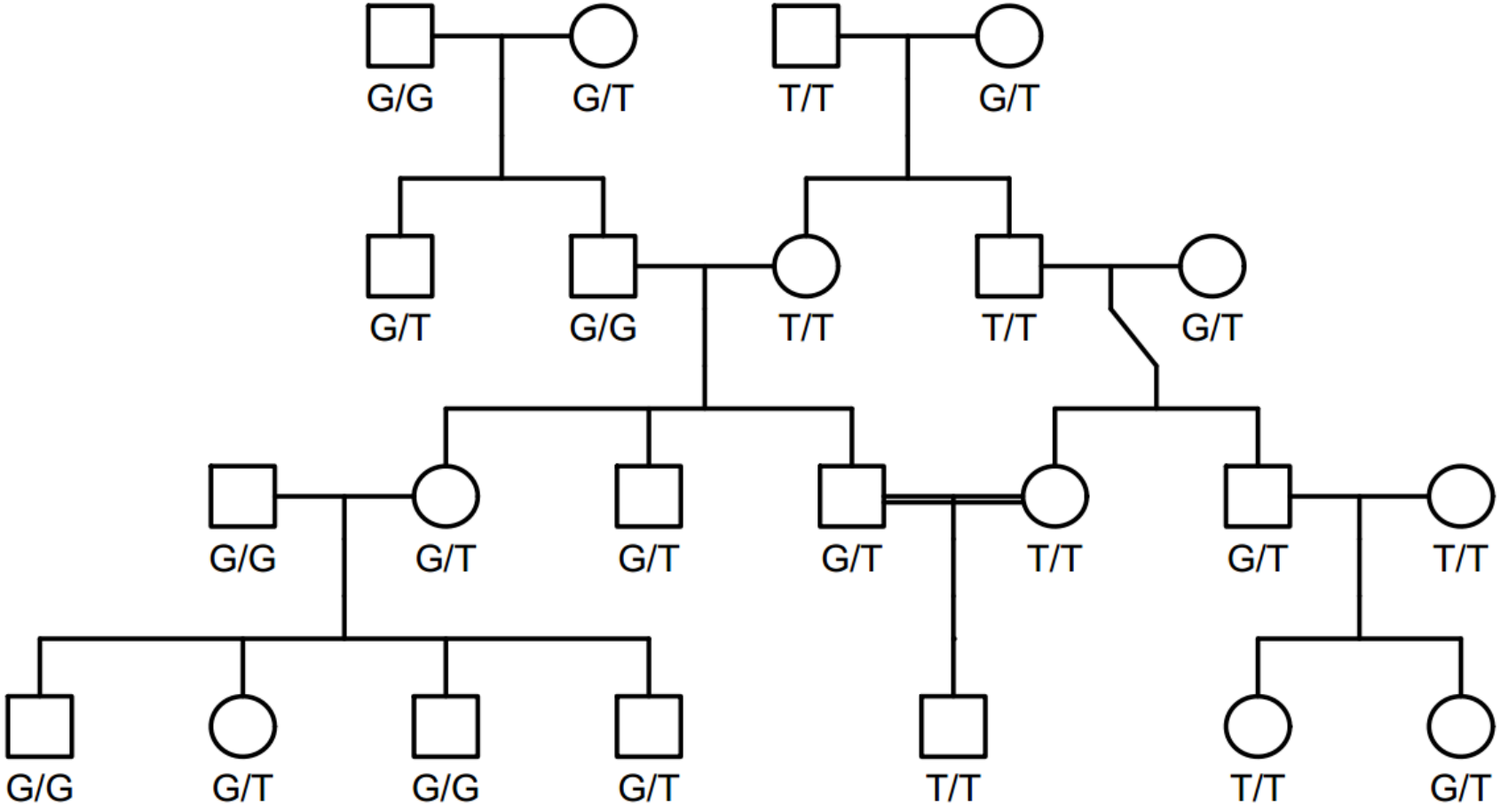
A closer look at the isolation-by-distance pattern



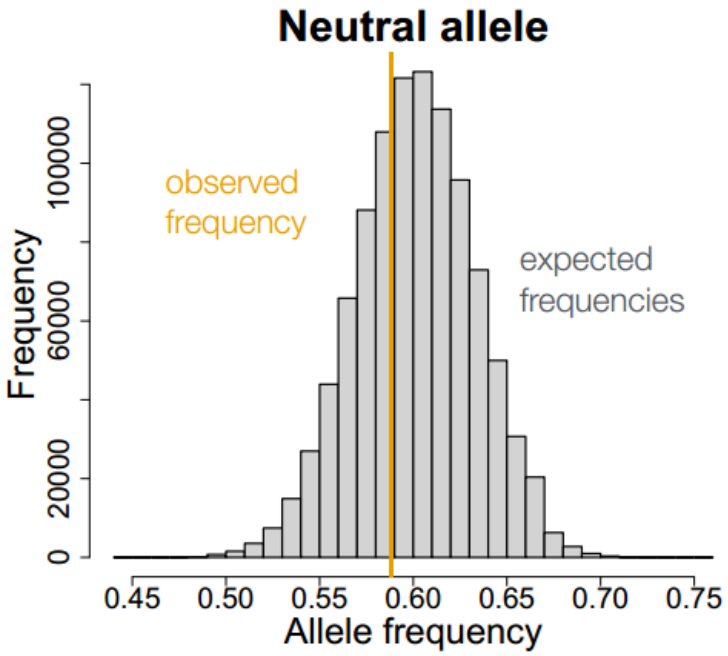
Testing for selection using pedigree information



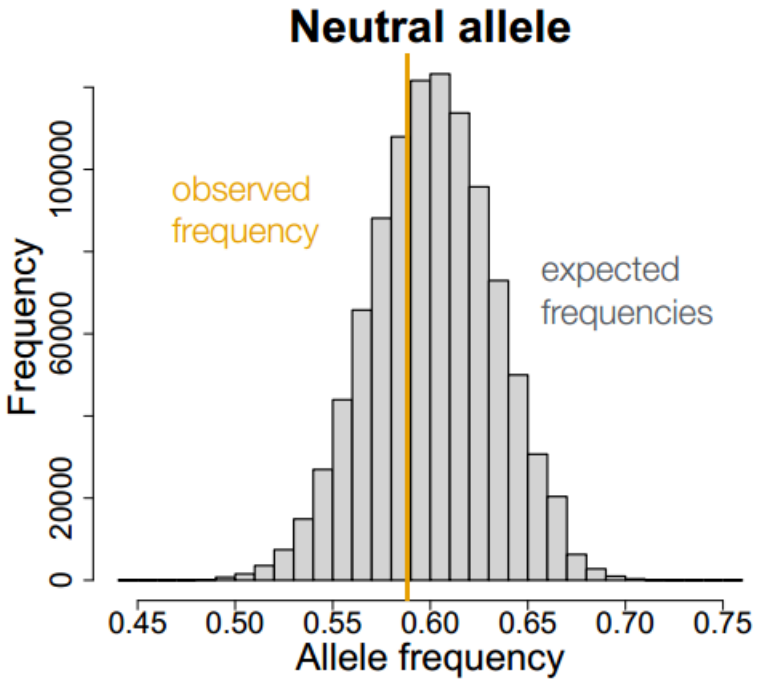
Gene-dropping to generate neutral expectations

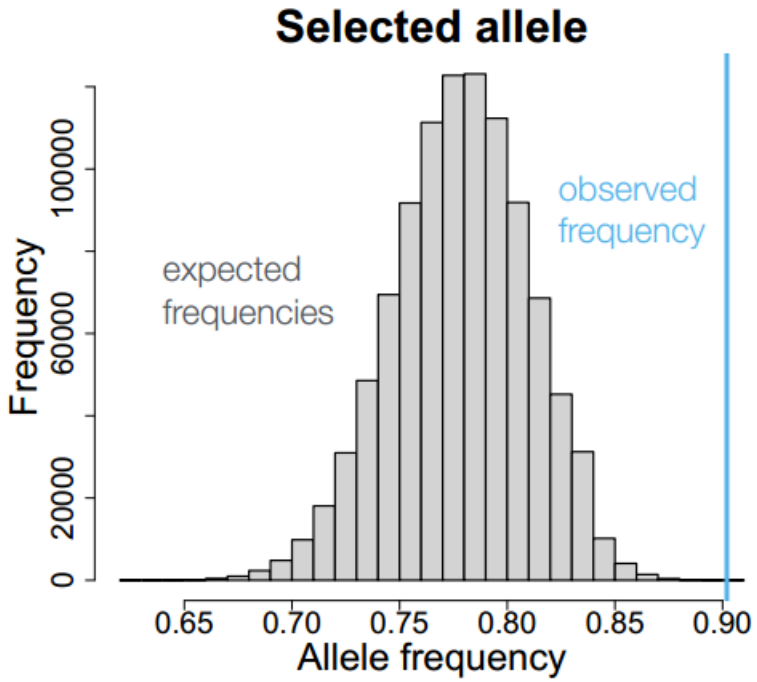


Most SNPs conform to the null hypothesis

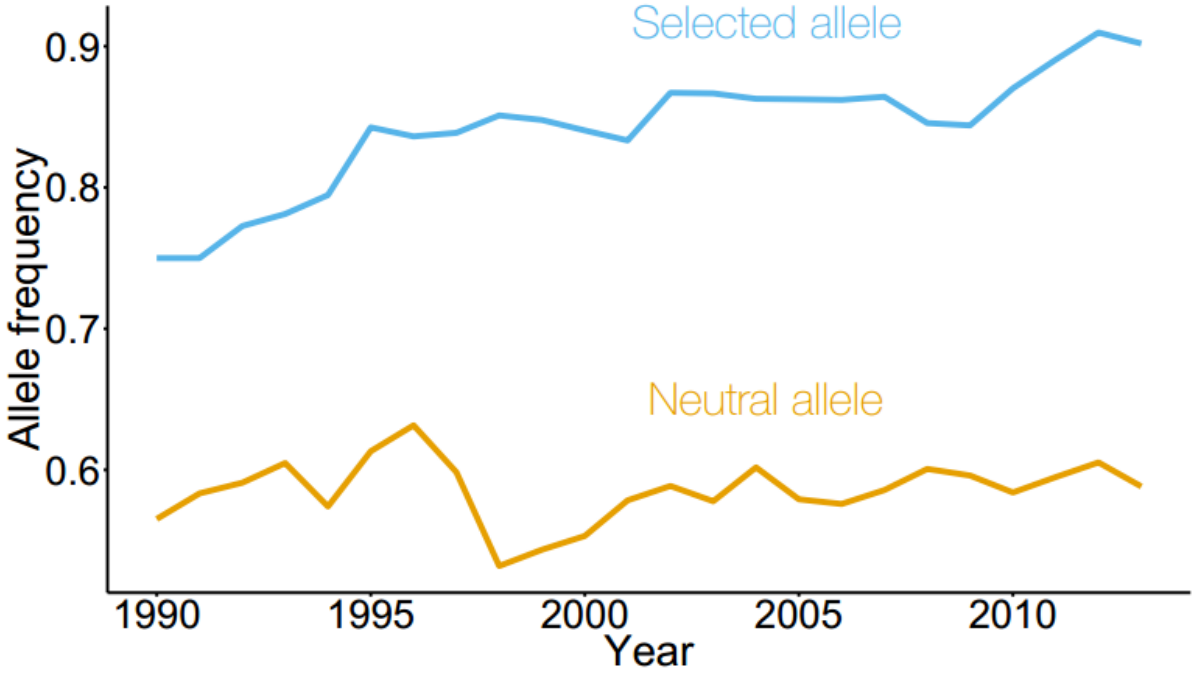


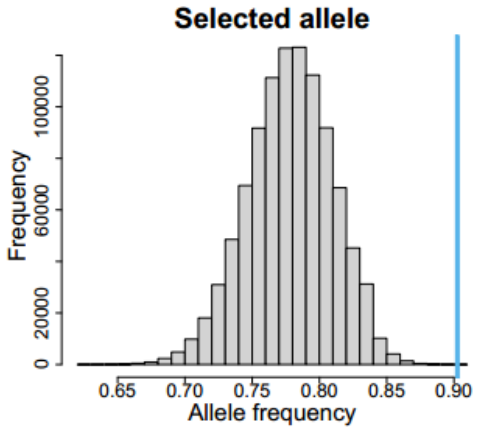
… but 67 SNPs display a significant departure.

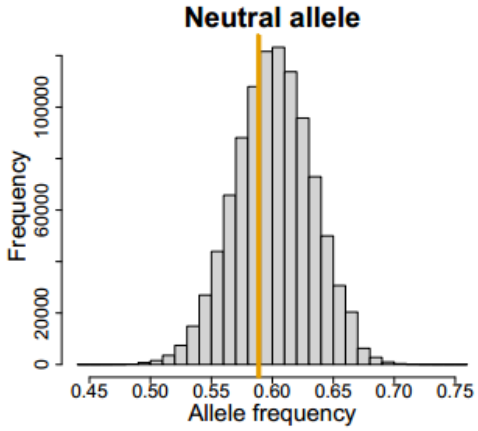




Evidence of selection at 67 SNPs







Selection can act at different stages of the life cycle

zygotes

viability selection

adults

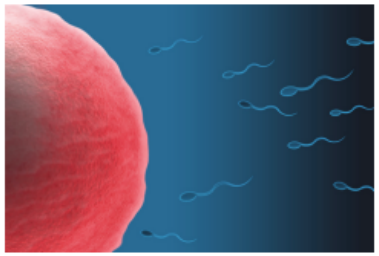
sexual selection

parents

fecundity selection

gametes

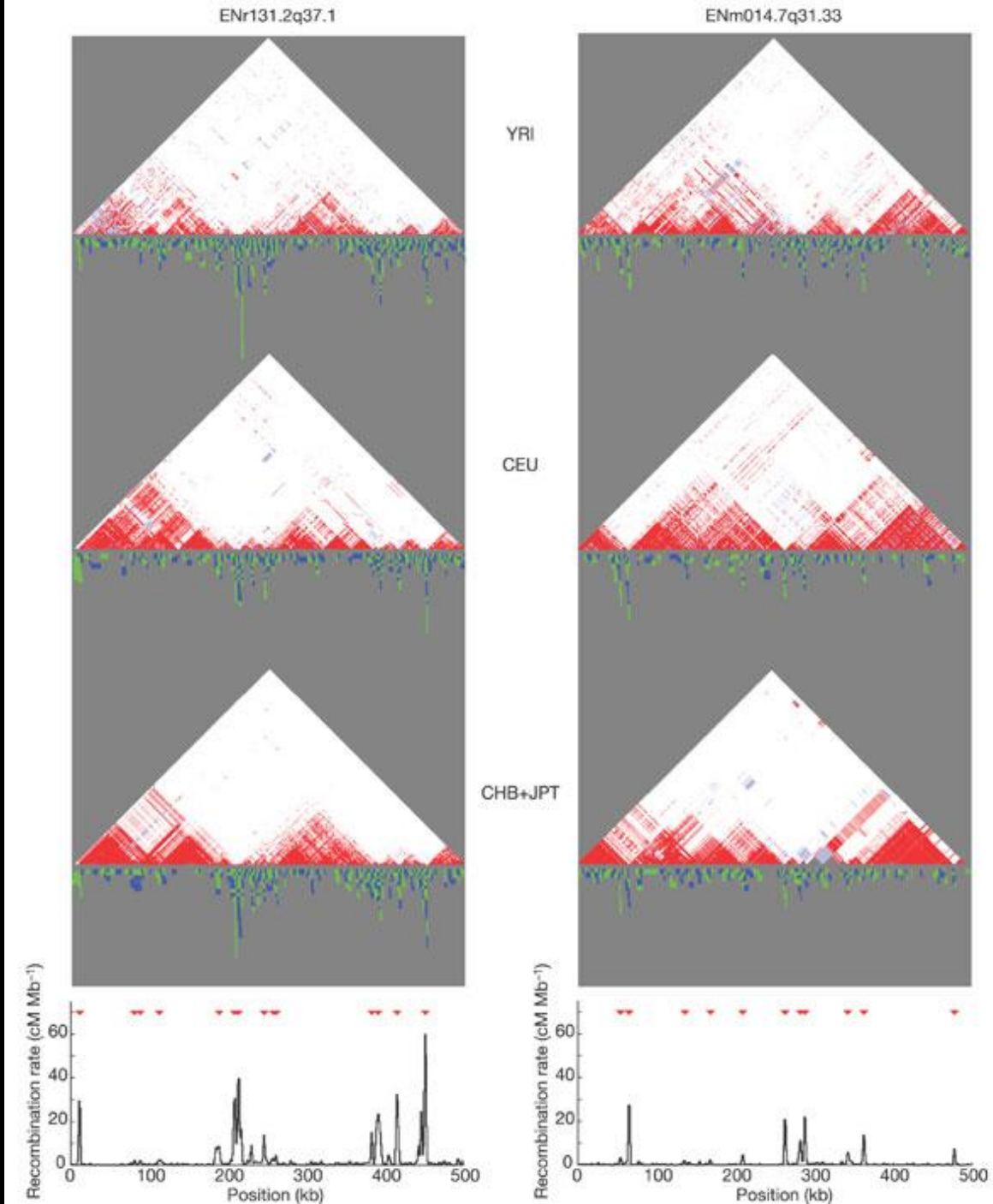
gametic selection



Recombination

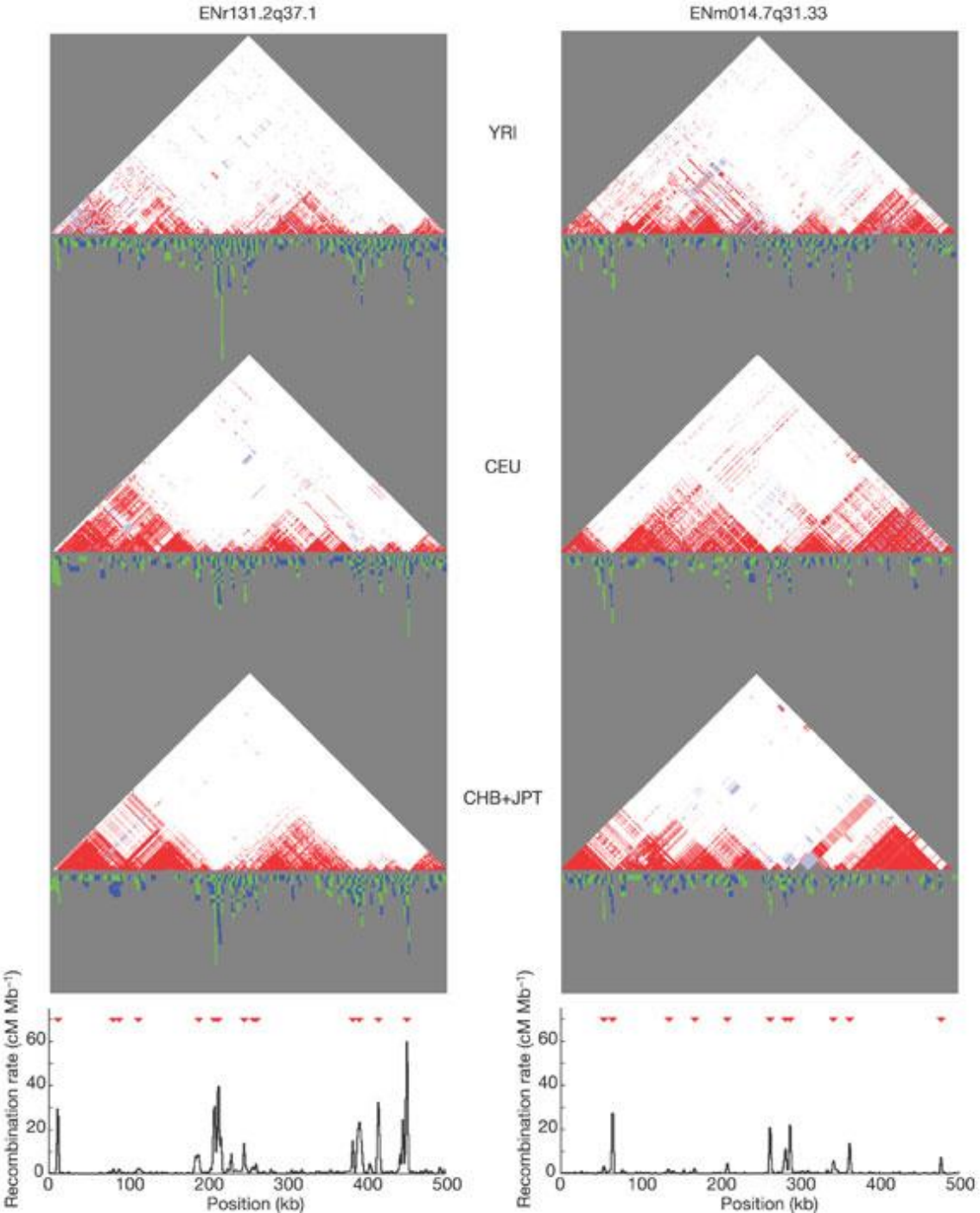
What can we infer about patterns of recombination from sequence data?

Inference of recombination hotspots from local LD

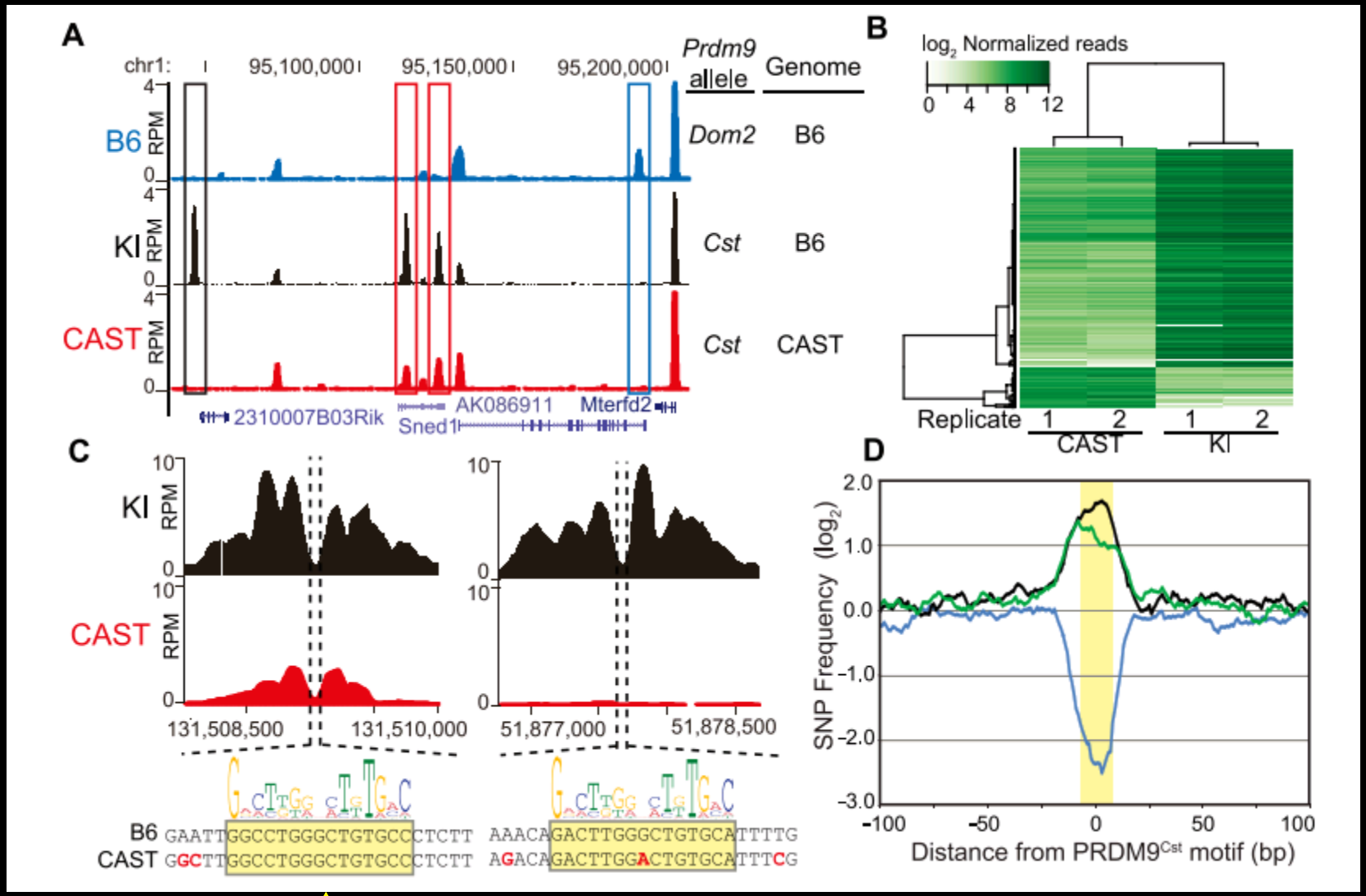


Inference of recombination hotspots from local LD

Is there among-individual variation in recombination rate?



Polymorphism in hotspot motifs → variation in hotspot usage



PRDM9 binding motif

Baker *et al*. 2015 *PLoS Genetics*

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

ARTICLE

doi:10.1038/nature11396

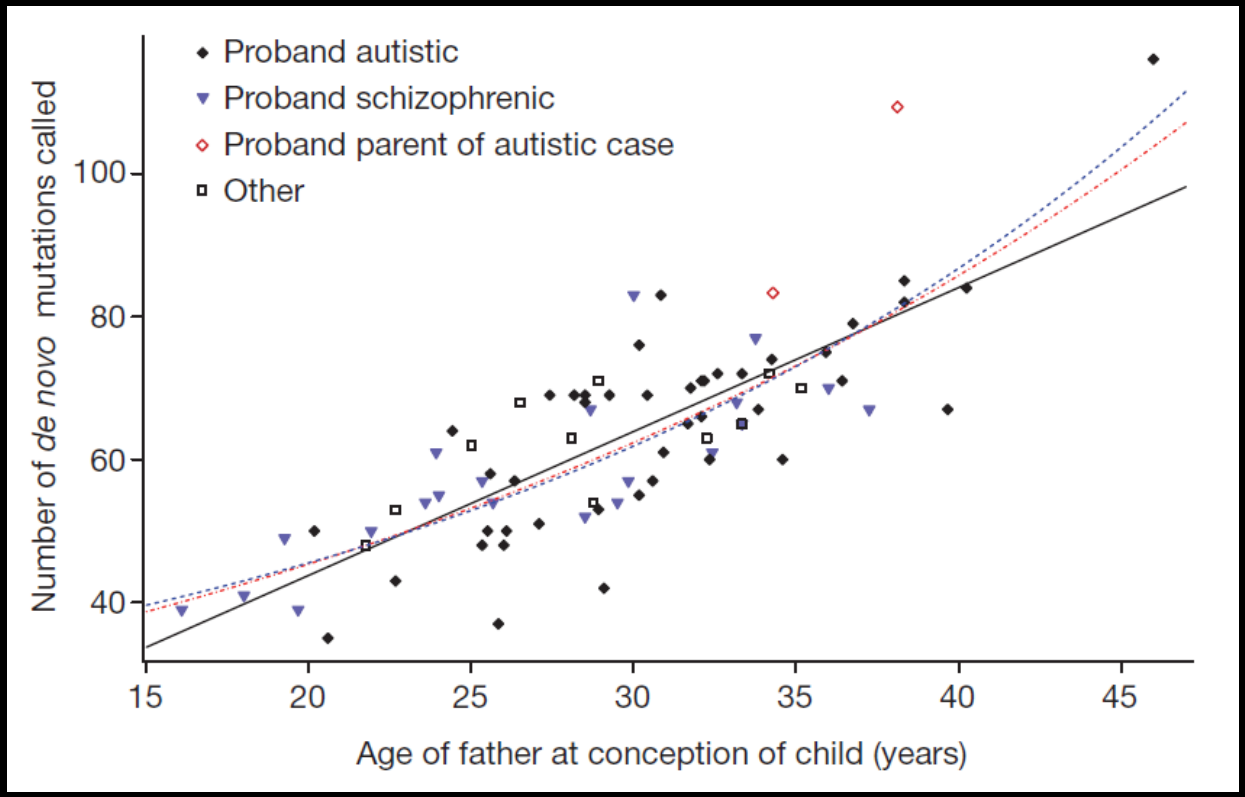
Rate of *de novo* mutations and the importance of father's age to disease risk

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

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 1.20×10^{-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)	Number of *de novo* mutations in proband: 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



Kong *et al.* 2012 *Nature* 488:471

Cryptic Relatedness

From random population samples, how well can we infer first, second, third degree relatives? How can we use this information?

Disease Association

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

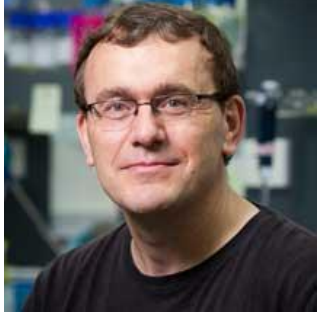
Population genetics and disease-causing variants

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

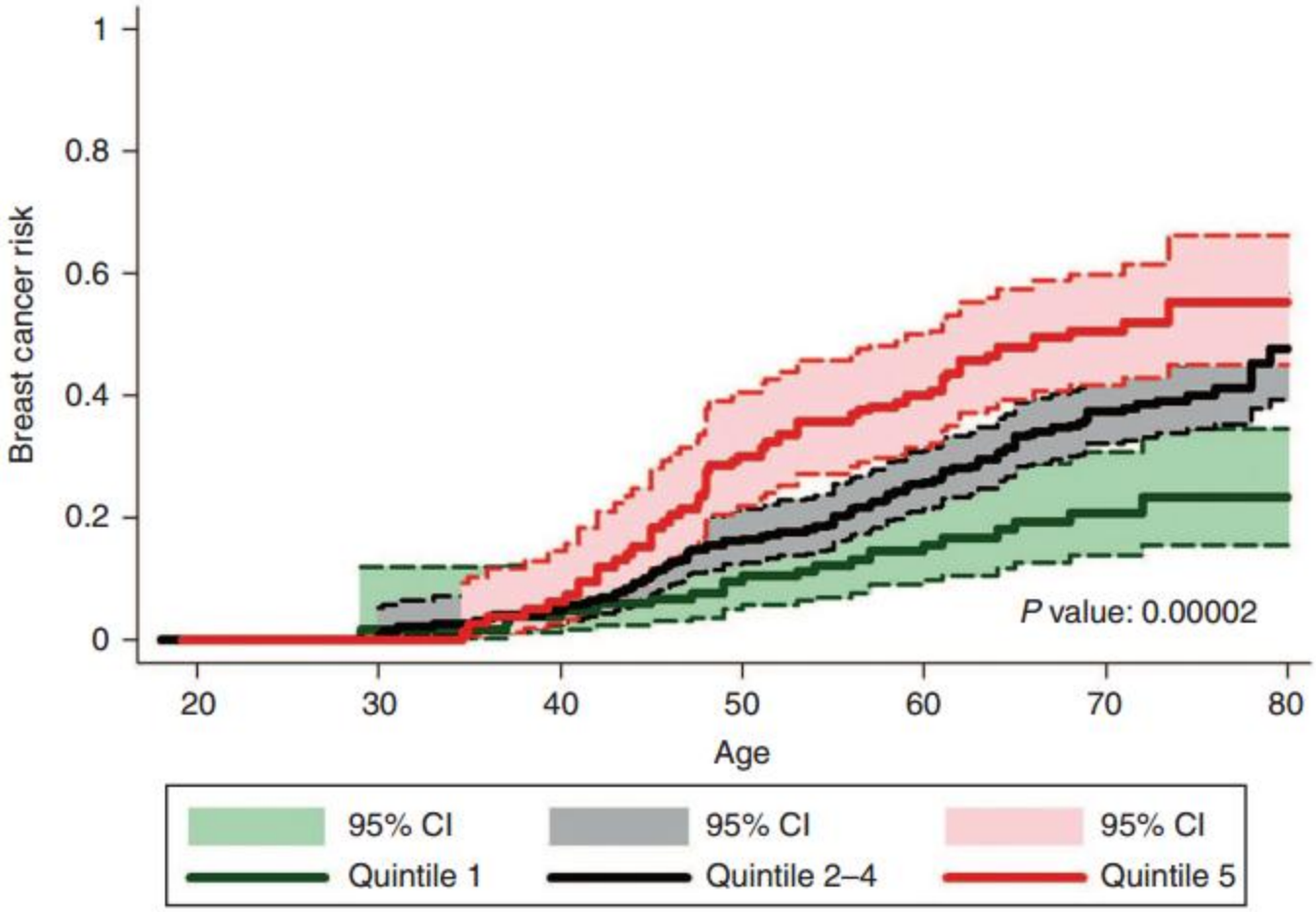
GCTA

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

GLM and prediction – polygenic risk score

- Purcell *et al*. (*Nature* 2009)
  - Polygenic Risk Score (schizophrenia R^2 = 3%)
- Chen et al. (*Genet. Epi*. 2015)
  - Big improvement with ancestry correction
  - hair color R^2 = 4-7%, tanning ability R^2 = 1-3%, basal cell carcinoma R^2 = 1-2%
- Vilhalmsson *et al.* (2015 *AJHG*)
  - Direct modeling LD increases accuracy of PRS

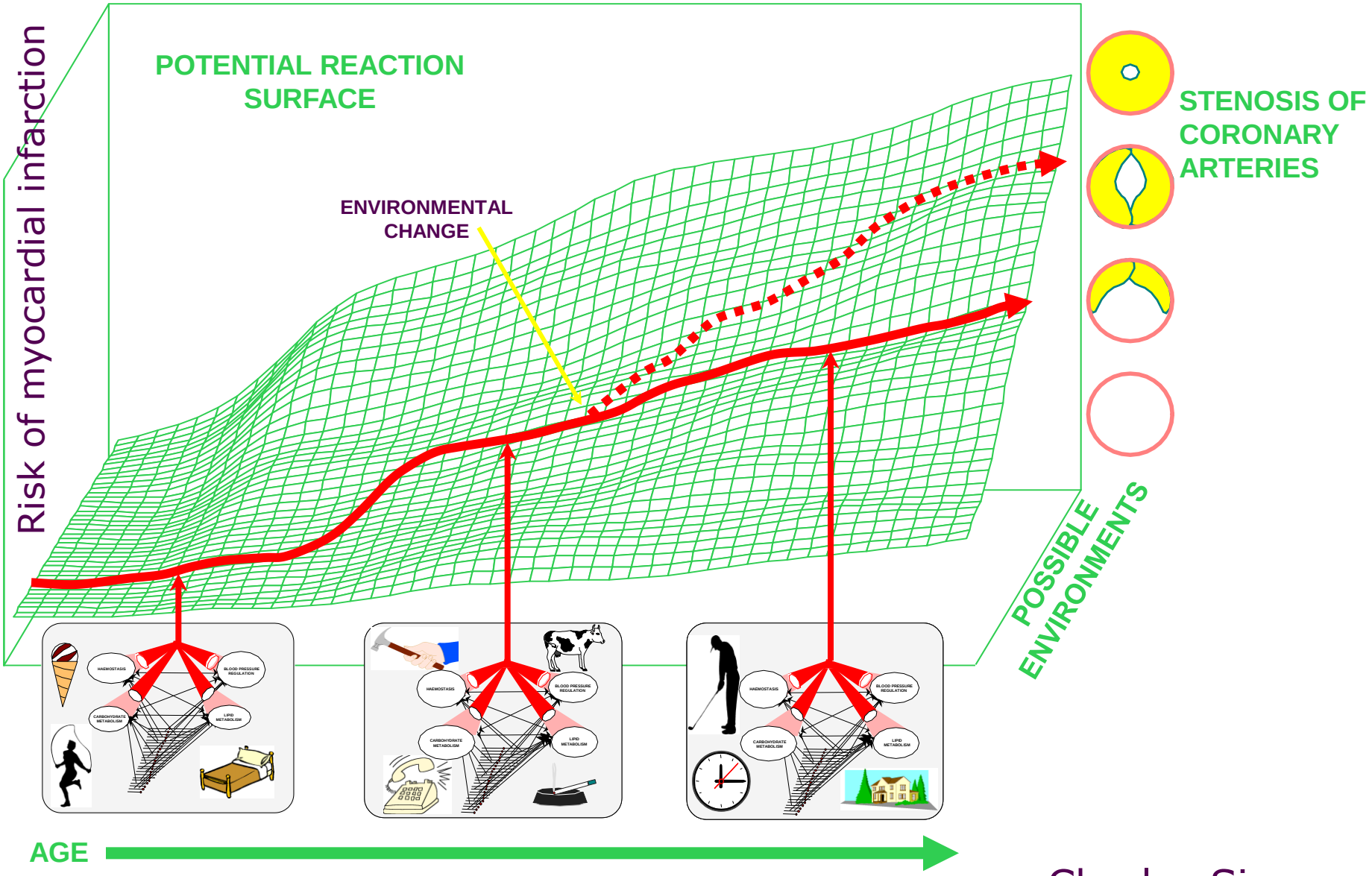
Polygenic Risk Score and Breast Cancer Risk



Li et al. 2016 *Genetics in Medicine*

What is still missing?

Adequate consideration of G x E.



Charles Sing

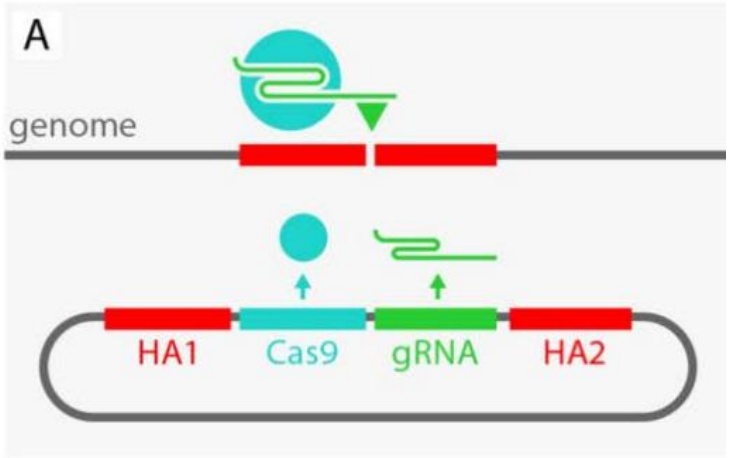
Genome functional variation

What do we learn from analysis of:

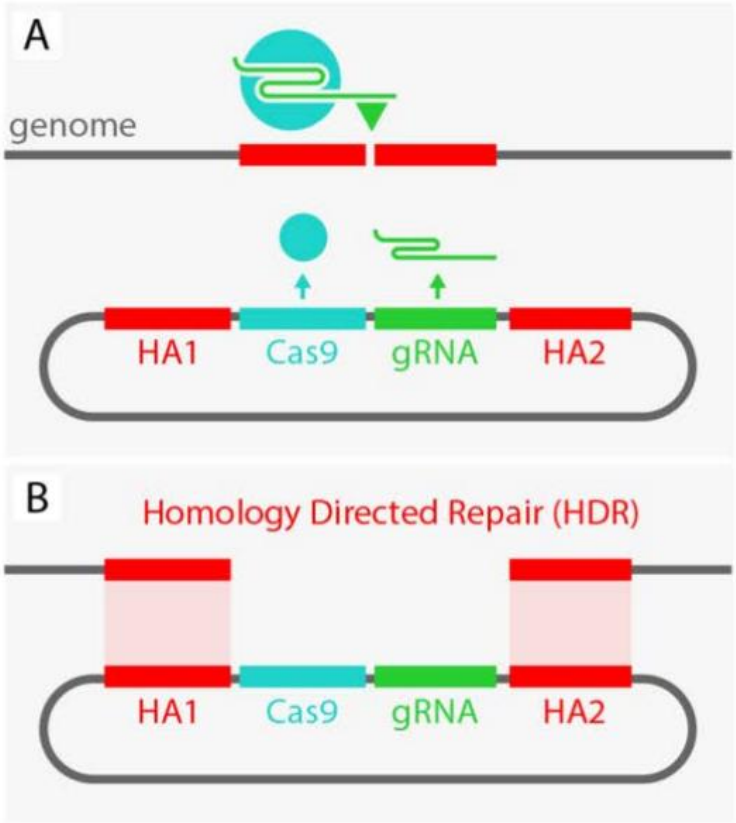
- eQTL (expression)
- mQTL (methylation)
- sQTL (splicing)
- rtQTL (replication timing)?

Gene drives

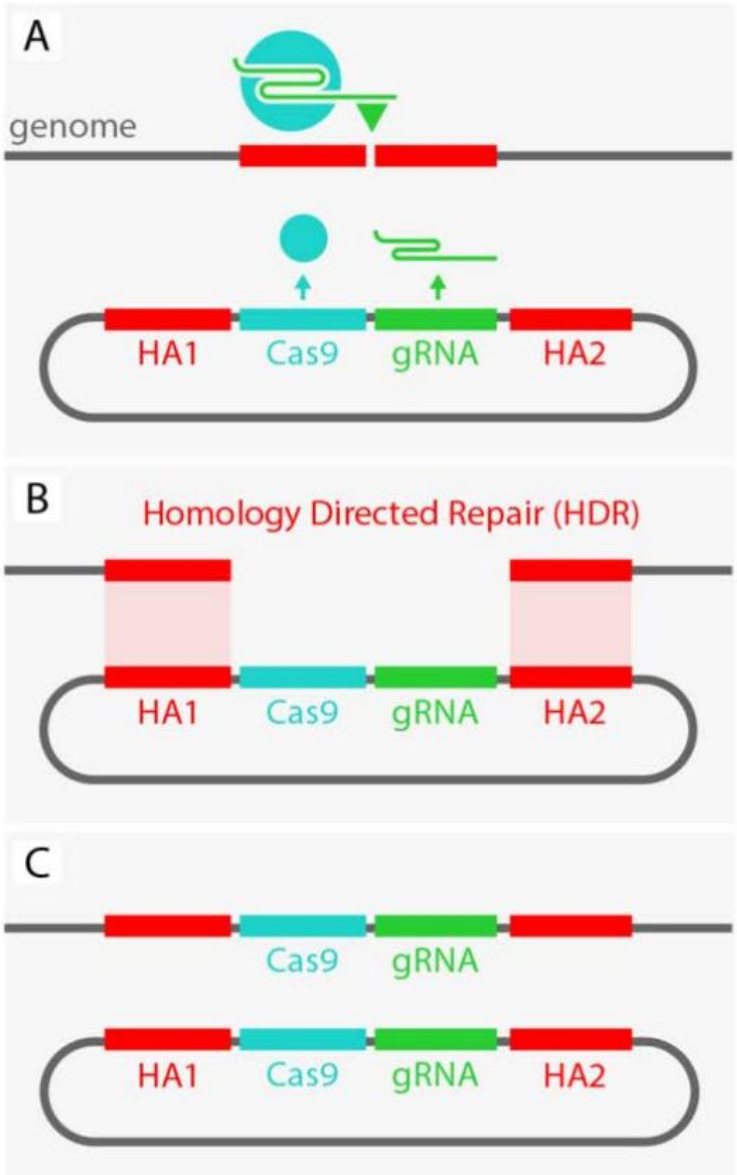
What are the population genetic consequences of introduction of a CRISPR gene drive allele?



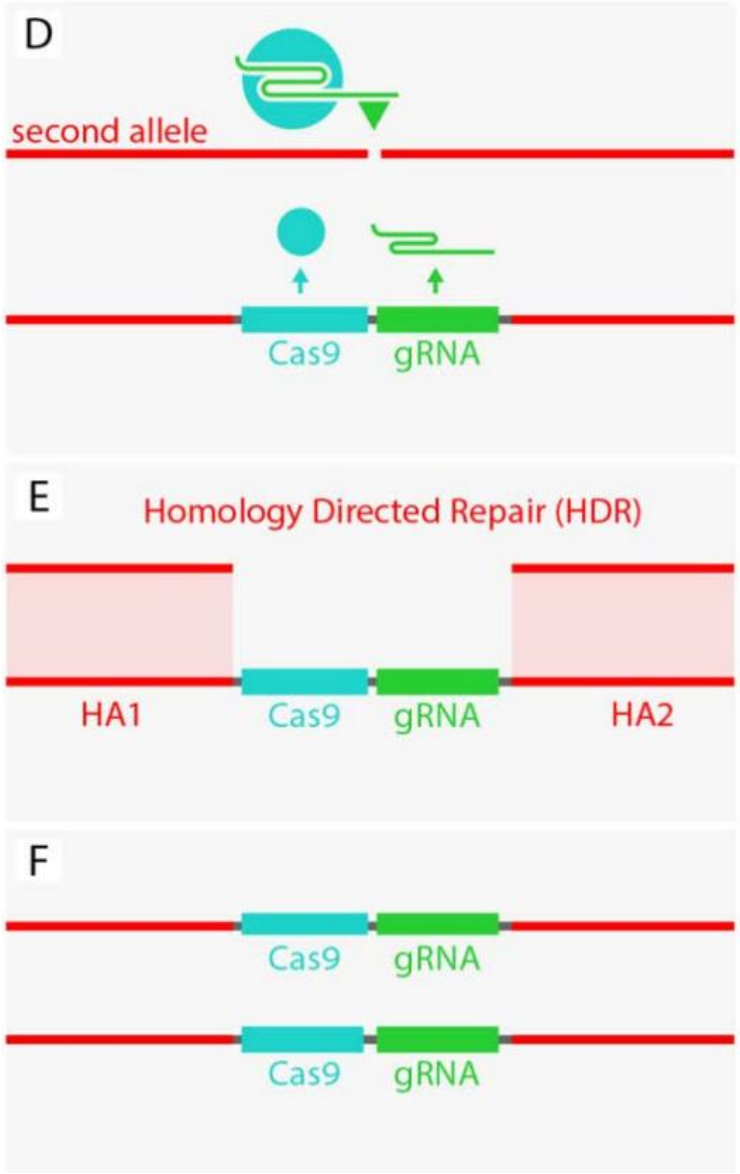
Gantz and Bier 2015 Science 348:442-444



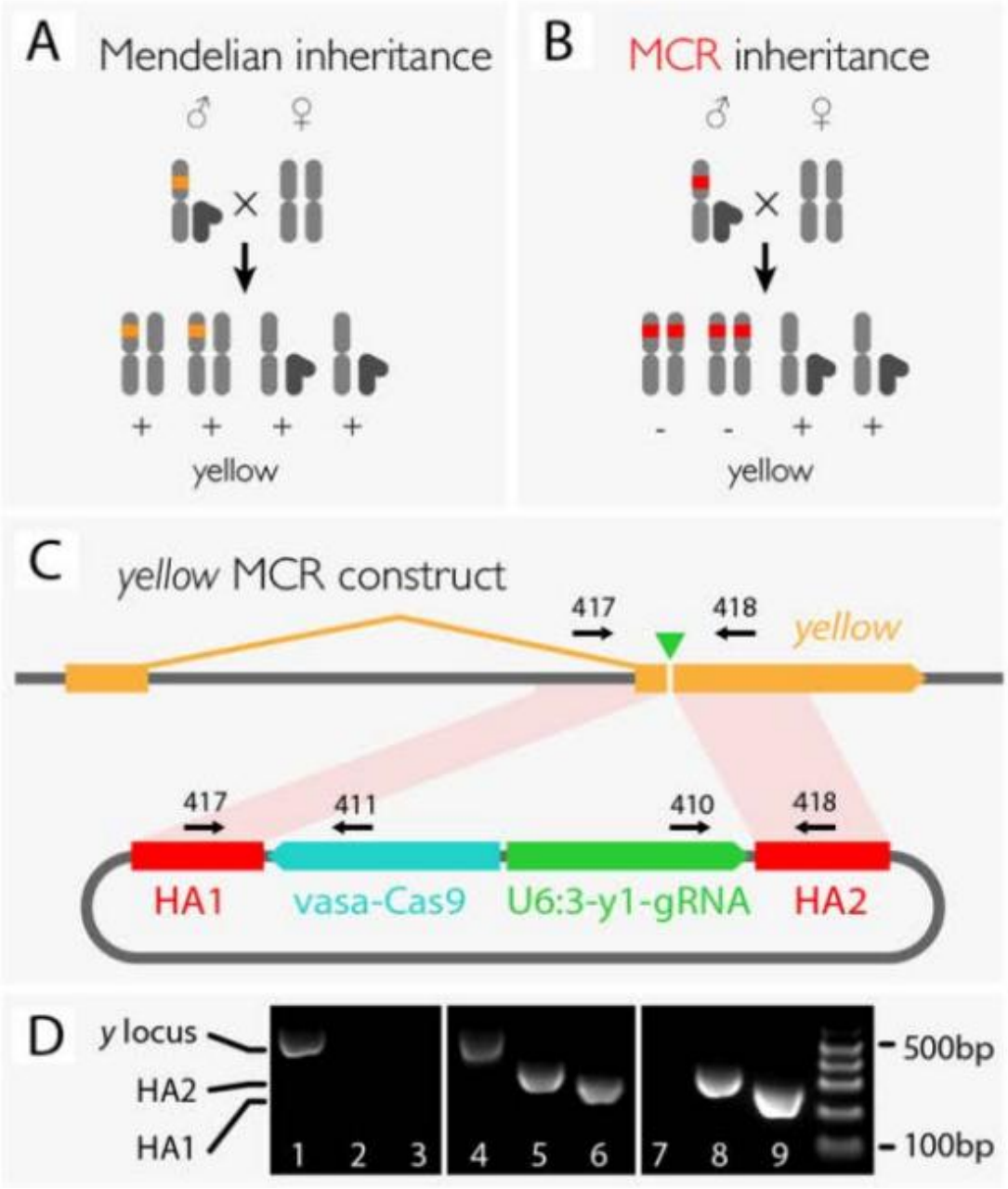
Gantz and Bier 2015 Science 348:442-444



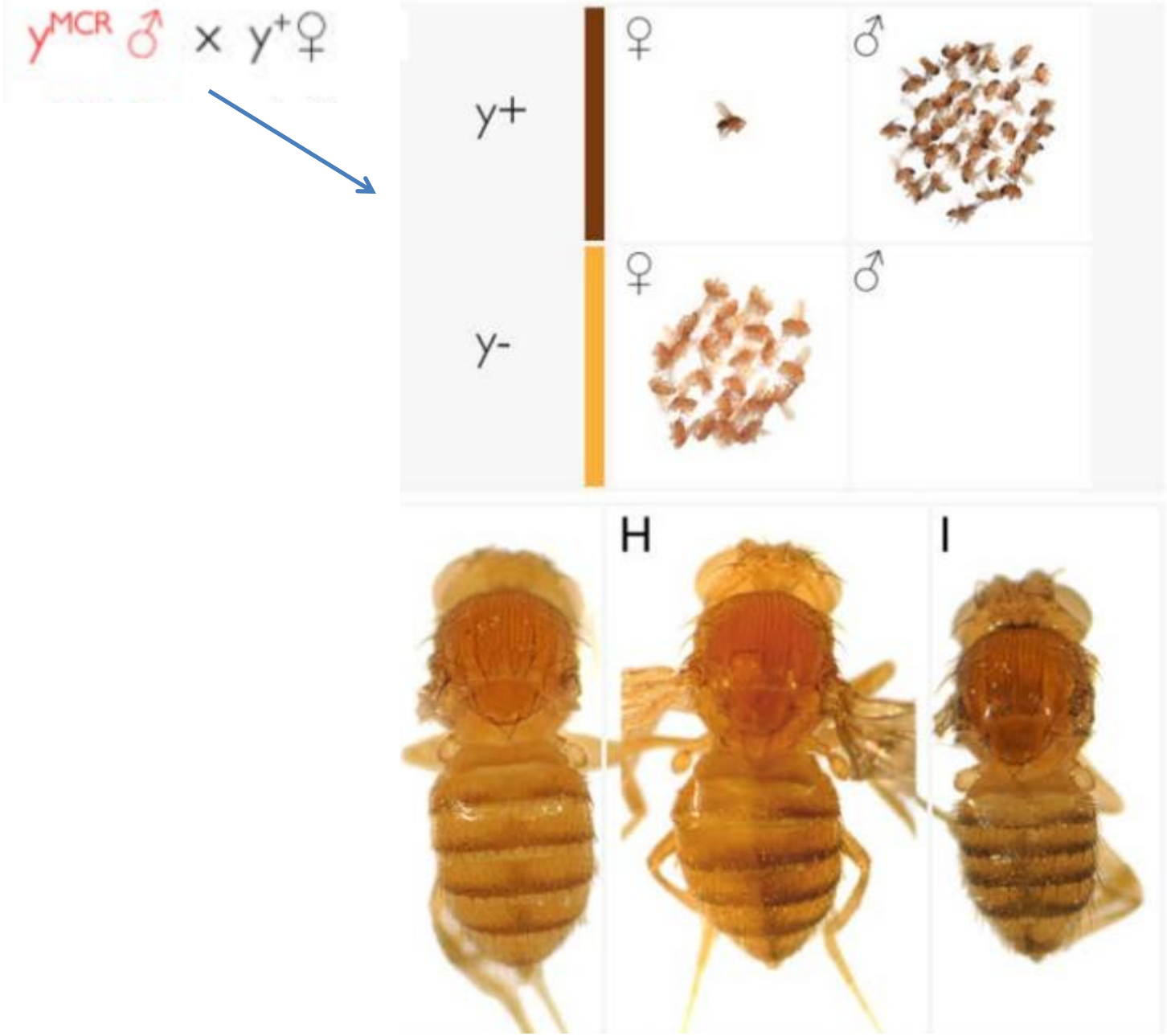
Gantz and Bier 2015 Science 348:442-444



Gantz and Bier 2015 Science 348:442-444



Gantz and Bier 2015 Science 348:442-444



Gantz and Bier 2015 Science 348:442-444

Questions

- What are the factors impacting MCR dynamics?
- What determines fixation time?
- Can MCR produce imperfect copies?
- What are the consequences of this?
- What can we expect if MCR is introduced into the genome of a complex, evolving species?

CRISPR/Cas9 editing is not flawless

- **HR** – homologous recombination repair of Cas9 nick results in perfect allele replacement
- **NHEJ** – non-homologous end joining of Cas9 DSB produces "Dead-on-Arrival" alleles

- Let δ be the proportion HR and 1-δ for DoA.

Recursions allowing for NHEJ

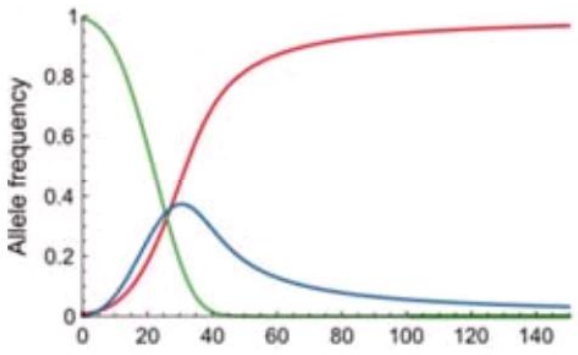
DoA $p' = (1-s)(qrc(1-\delta)+p^2)+pq(1-s)+pr$

MCR $q' = (1-s)(qrc(1-\delta)+2qrc\delta+q^2)+qr(1-c)+pq(1-s)$

Wild $r' = qr(1-c)+pr+r^2$

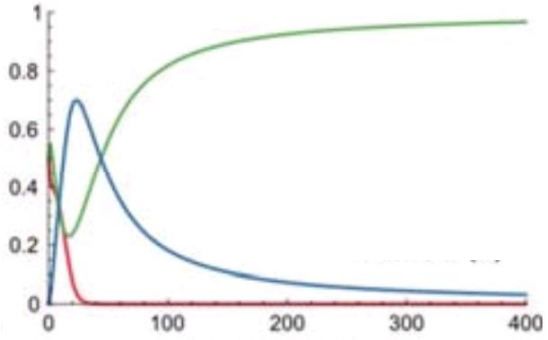
Three possible outcomes
with NHEJ making **DoA** alleles

MCR fixes



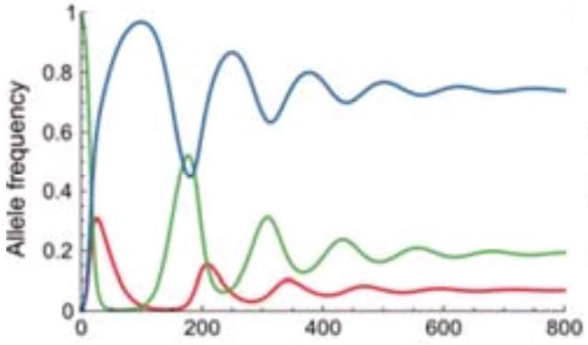
δ, c sufficiently high
s sufficiently low

Wildtype fixes



δ, c sufficiently low
s sufficiently high

Stable polymorphism



δ, c intermediate
s sufficiently low

Modeling natural polymorphism for resistance to gene drive

- Consider 4 alleles
  - Wildtype
  - MCR
  - Two classes of resistant alleles
    - One that knocks out the target and the other does not

- Evolution of resistant alleles is virtually guaranteed.

Messer et al. 2017 PLoS Genetics

Outline

- Demographic inference
- Population structure and history
- Admixture/ Introgression
- Random genetic drift
- Natural selection
- Recombination
- Mutation spectrum
- Cryptic relatives
- Disease association
- Genome function
- Gene drive