Quantifying population dynamics using hidden relatedness

Itsik Pe’er
Columbia University
IGERT, 11/18/2016

Population Genomics 101: Data

ACTTGTTTTTGGGTTGGGTGGGGCATCC…
ATTTGTTTTTGCGTTGGGTGGGGCATCC…

ACTTATTTTTGGGTTGAGTGGGGCATCC…
ACTTATTTTTGGGTTGGGTGGGGCGTCC…

ACTTGTTTTTGCGTTGGGTGGGGCATCC…
ACTTGTTTTTGCGTTGGGTGGGGCGTCC…

ACTTGTTTTTGCGTTGGGTGGGGCATCC…
ACTTGTTTTTGGGTTGGG-GGGGCATCC…

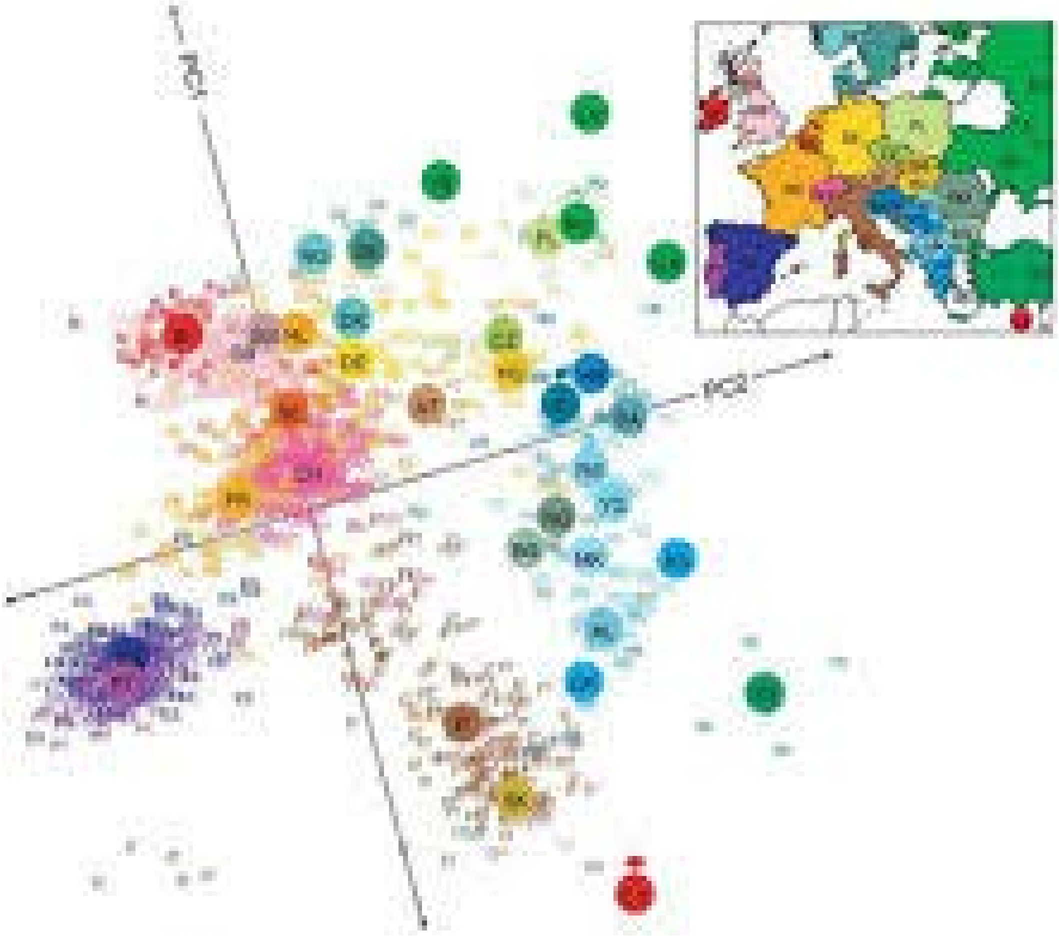
Population Genomics 101: Data

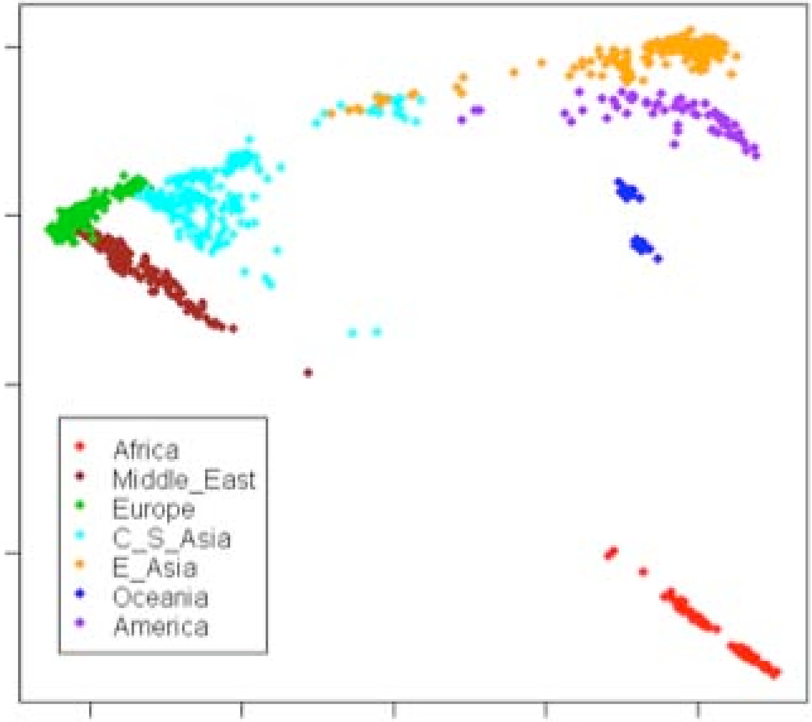
C	G	G	G	T	A	…
T	G	C	G	T	A	…
C	A	G	A	T	A	…
C	A	G	G	T	G	…
C	G	C	G	T	A	…
C	G	C	G	T	G	…
C	G	C	G	T	A	…
C	G	G	G	-	A	…

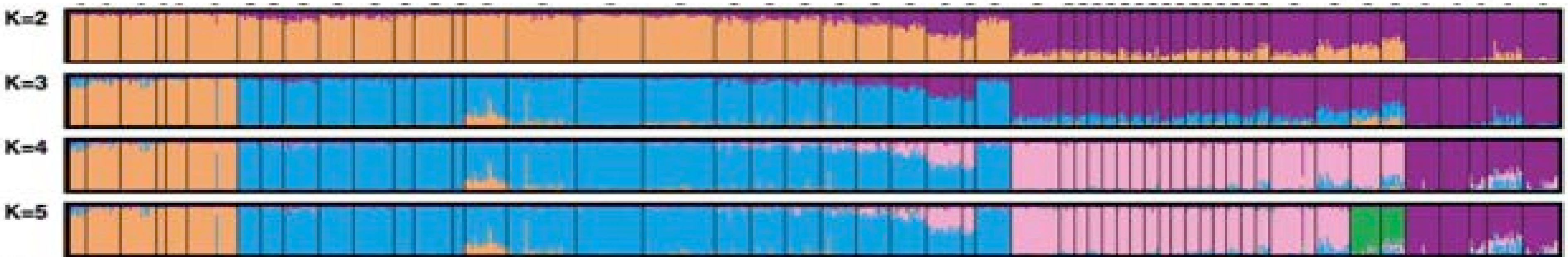
Population Genomics 101: Data

	0	0	0	0	0	0	…
	1	0	1	0	0	0	…
	0	1	0	1	0	0	…
	0	1	0	0	0	1	…
	0	0	1	0	0	0	…
	0	0	1	0	0	1	…
	0	0	1	0	0	0	…
	0	0	0	0	-	0	…

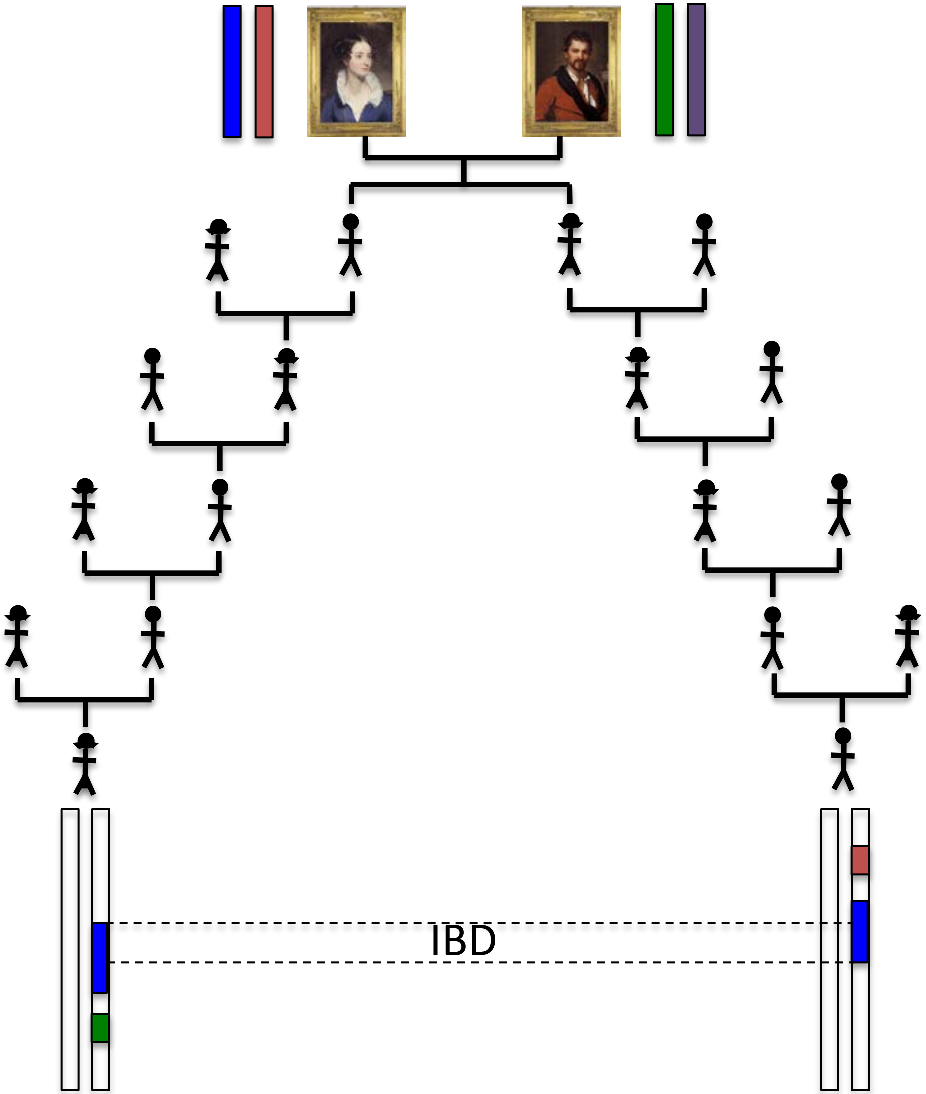
Population Genomics 101: Analysis



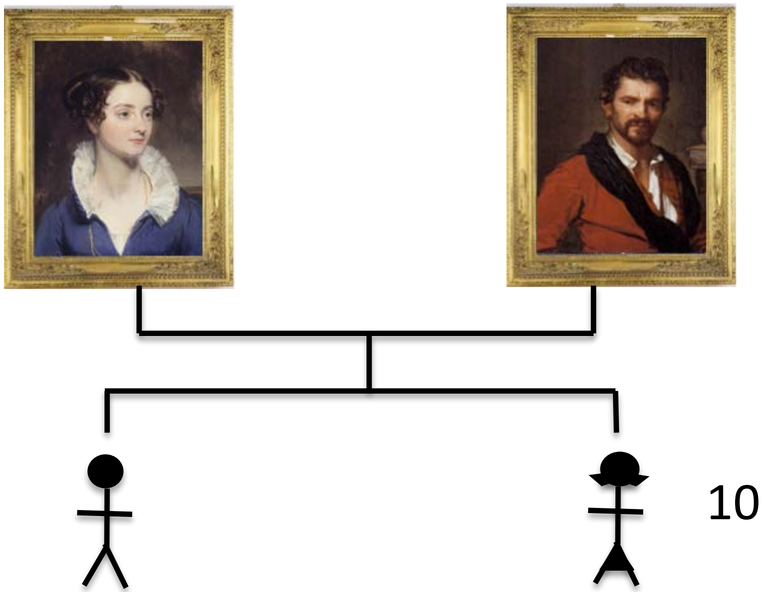




Identity By Descent

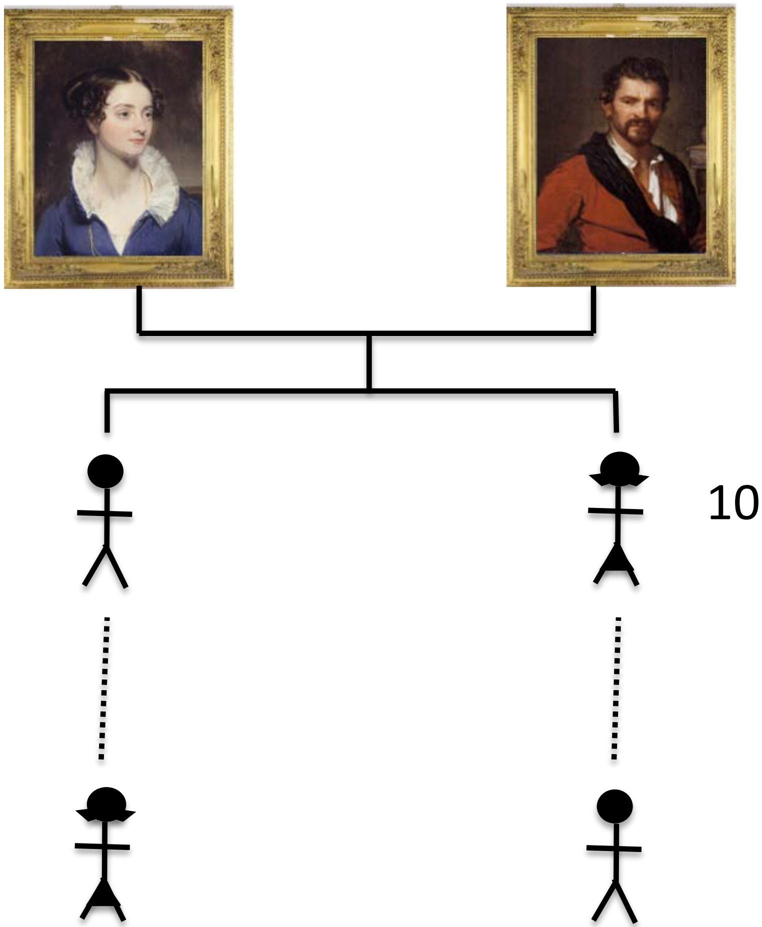


Identity By Descent



Relationship	**Expected Sharing**
Siblings	1/2 of genome 100cM/2 = 50 cM segments

Identity By Descent



Relationship

Expected Sharing

Siblings

1/2 of genome
100cM/2 = 50 cM segments

k-th generation

$2(1/4)^k$ of genome
50/k cM segments
(exponential distance to next recombinstion)

Outline

- Introduction: Identity by Descent

- A model for IBD sharing and demography

- Examples

- IBD and sequence data

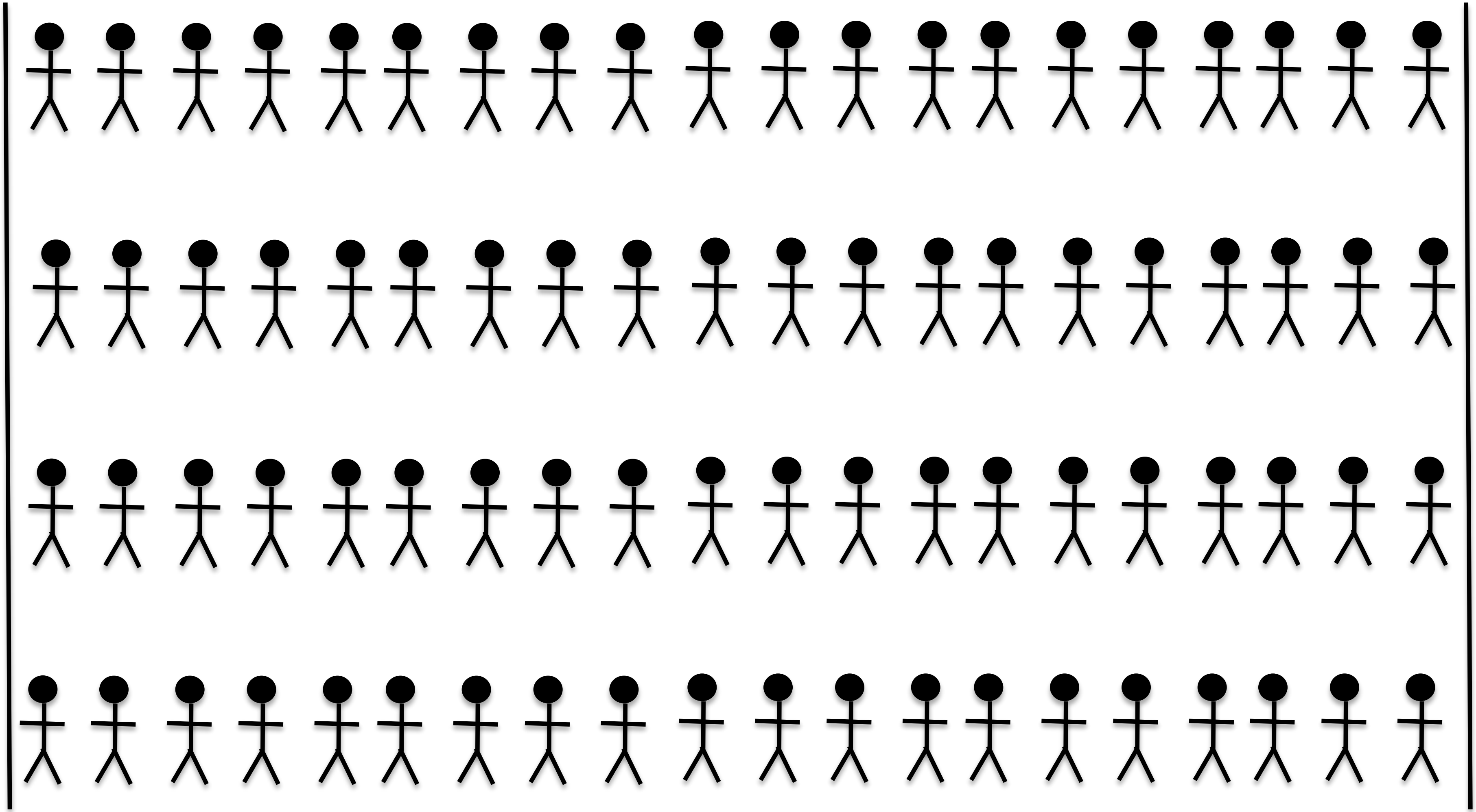
Outline

- Introduction: Identity by Descent

- A model for IBD sharing and demography
  - Migration
- Examples

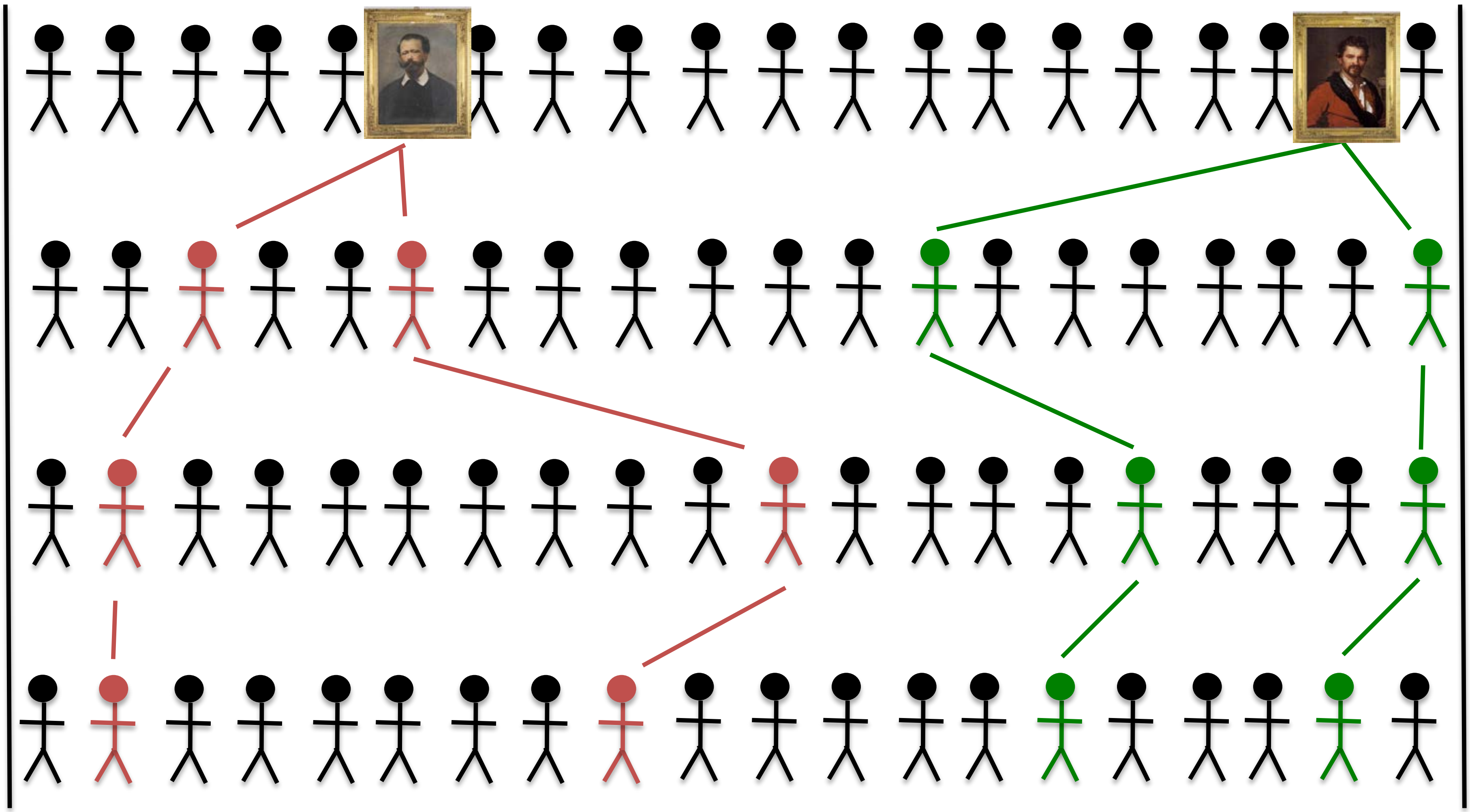
- IBD and sequence data

Common ancestry and demography



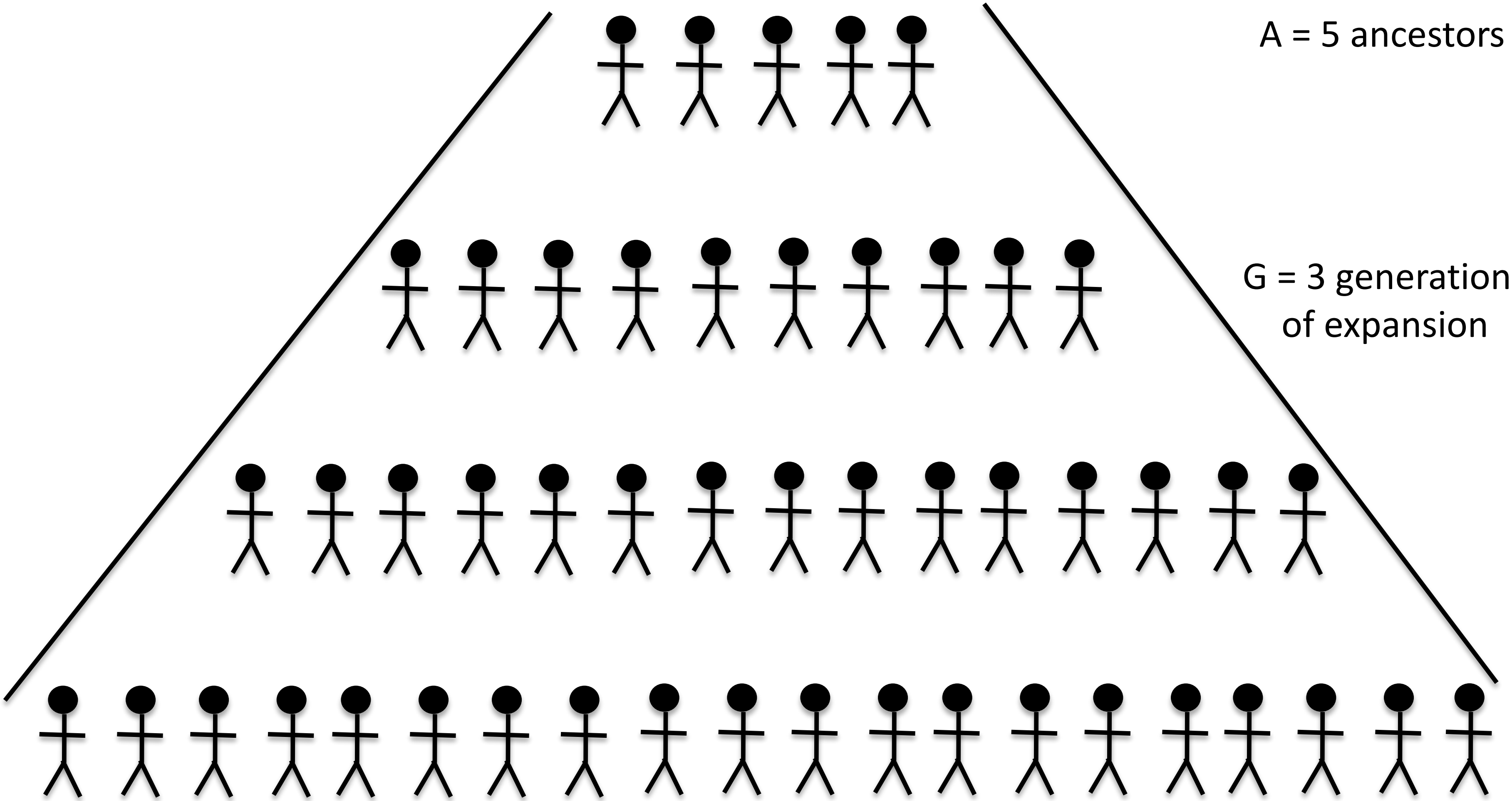
Population of constant size 20

Common ancestry and demography



Population of constant size 20

Common ancestry and demography



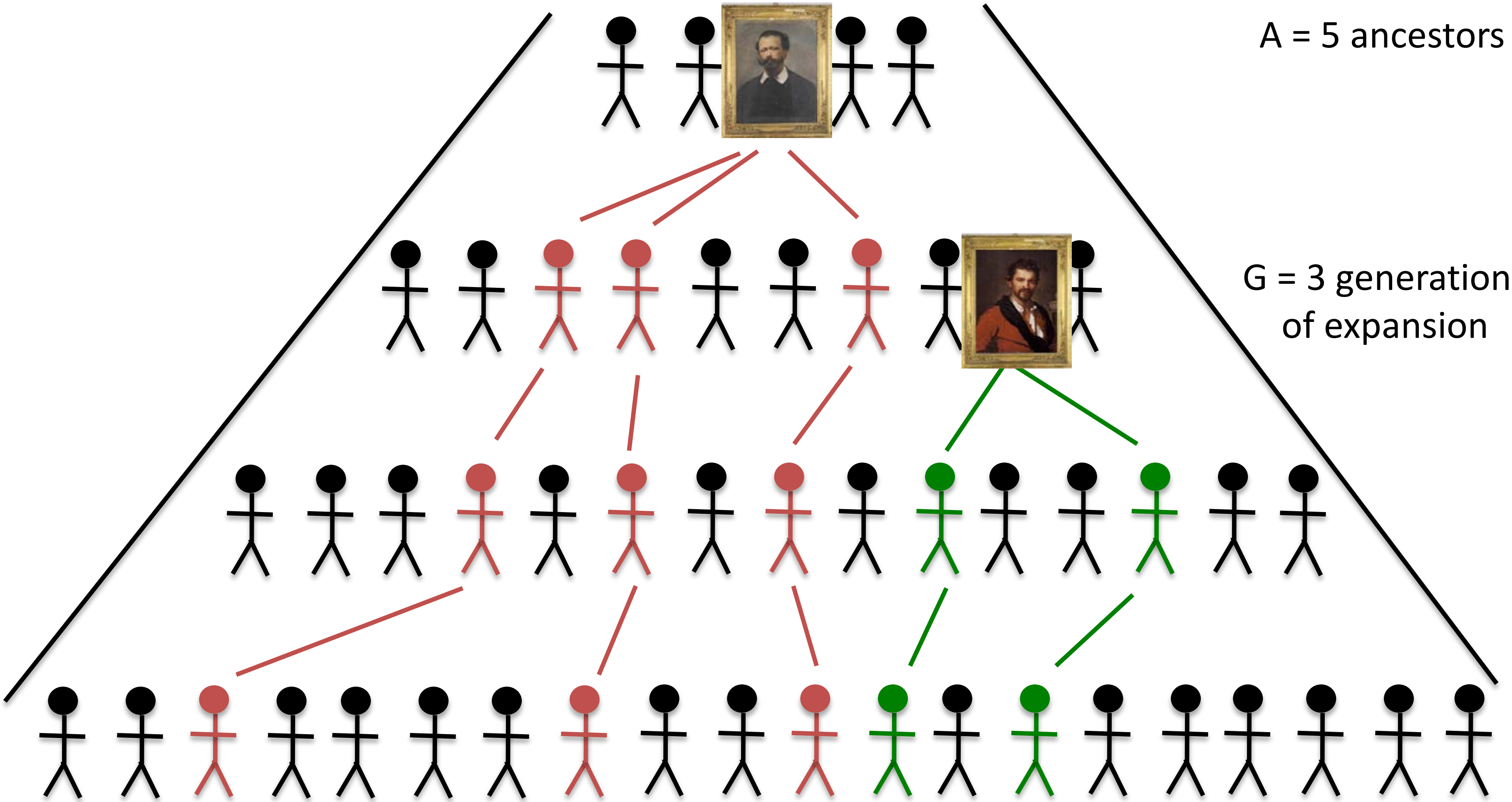
A = 5 ancestors

G = 3 generations
of expansion

C = 20 individuals

Expanding population

Common ancestry and demography



A = 5 ancestors

G = 3 generations
of expansion

Expanding population

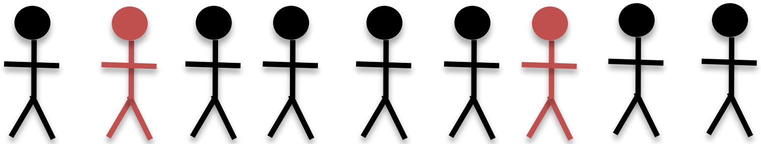
C = 20 individuals

A formal link between demography and IBD

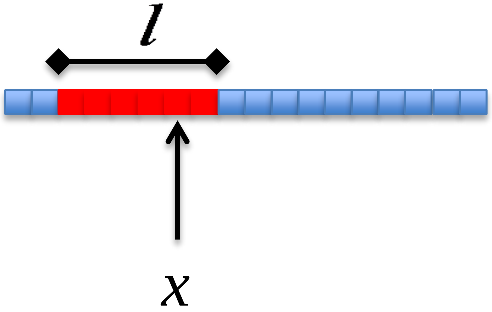
- Random pair in population with demographic history $\boldsymbol{\vartheta}$
  - E.g. Constant population size: $\boldsymbol{\vartheta} = [N]$
  - E.g. Contraction/Expansion: $\boldsymbol{\vartheta} = [A,C,G]$

- For an individual site x, express

  Pr(x spanned by a shared haplotype of length l $[u,v]$)



$$\int_u^v P(l|\boldsymbol{\theta})dl = ?$$



details in Palamara et al. AJHG 2012

A formal link between demography and IBD

- Marginalize generation k of common ancestor

$$\int_u^v P(l|\boldsymbol{\theta})dl = \int_u^v \sum_{k=1}^{\infty} P(l, g = k|\boldsymbol{\theta})\, dl$$

A formal link between demography and IBD

$$\int_u^v P(l|\boldsymbol{\theta})dl = \int_u^v \sum_{k=1}^{\infty} P(l, g = k|\boldsymbol{\theta})\, dl$$

$$\Downarrow$$

$$\ldots = \sum_{k=1}^{\infty} P(g = k|\boldsymbol{\theta}) \int_u^v P(l|g = k)dl$$

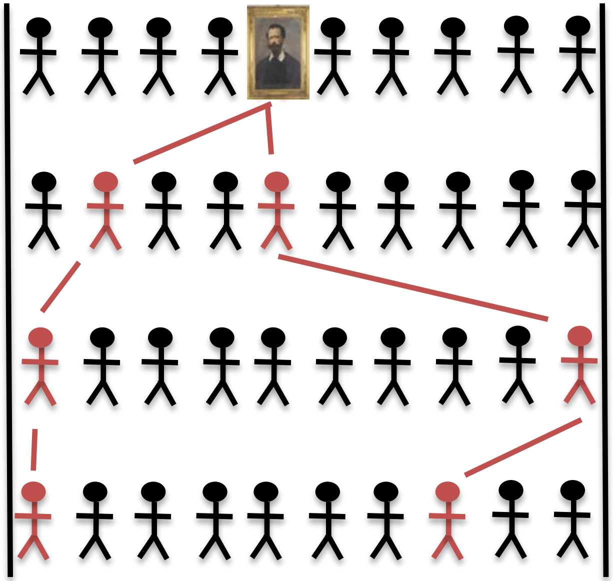
A formal link between demography and IBD

$$\int_u^v P(l|\boldsymbol{\theta})dl = \sum_{k=1}^{\infty} P(g = k|\boldsymbol{\theta}) \int_u^v P(l|g = k)dl$$

A formal link between demography and IBD

$$\int_u^v P(l|\boldsymbol{\theta})dl = \sum_{k=1}^{\infty} \boxed{P(g=k|\boldsymbol{\theta})} \int_u^v P(l|g=k)dl$$

Coalescent distribution,
function of population size



A formal link between demography and IBD

$$\int_u^v P(l|\boldsymbol{\theta})dl = \sum_{k=1}^{\infty} P(g = k|\boldsymbol{\theta}) \boxed{\int_u^v P(l|g = k)dl}$$

Sum of two exponential random variables with same expectation (Erlang-2)

A formal link between demography and IBD

$$\int_u^v P(l|\boldsymbol{\theta})dl = \sum_{k=1}^{\infty} P(g = k|\boldsymbol{\theta}) \int_u^v P(l|g = k)dl$$

If constant population size N:

A formal link between demography and IBD

$$\int_u^v P(l|\boldsymbol{\theta})dl = \sum_{k=1}^{\infty} P(g = k|\boldsymbol{\theta}) \int_u^v P(l|g = k)dl$$

If constant population size N:

$$\int_u^v P(l|\boldsymbol{\theta})dl = \frac{100N^2(v-u)[25(u+v)+uvN]}{(50+uN)^2(50+vN)^2}$$

A formal link between demography and IBD

$$\int_u^v P(l|\boldsymbol{\theta})dl = \sum_{k=1}^{\infty} P(g=k|\boldsymbol{\theta}) \boxed{\int_u^v P(l|g=k)dl} \longrightarrow E[f_{u,v}|k]$$

If constant population size N:

$$\boxed{\int_u^v P(l|\boldsymbol{\theta})dl =} \frac{100N^2(v-u)[25(u+v)+uvN]}{(50+uN)^2(50+vN)^2}$$

$$\Big\downarrow$$

$$E[f_{u,v}|\theta]$$

A formal link between demography and IBD

$$E[f_{u,v}|\boldsymbol{\theta}] = \sum_{k=1}^{\infty} P(g = k|\boldsymbol{\theta})\, E[f_{u,v}|k]$$

If constant population size N:

$$E[f_{u,v}|\boldsymbol{\theta}] = \frac{100N^2(v-u)[25(u+v)+uvN]}{(50+uN)^2(50+vN)^2}$$

A formal link between demography and IBD

$$E[f_{u,v}|\boldsymbol{\theta}] = \sum_{k=1}^{\infty} P(g = k|\boldsymbol{\theta})\, E[f_{u,v}|k]$$

If constant population size N:

$$E[f_{u,v}|\boldsymbol{\theta}] = \frac{100N^2(v-u)[25(u+v)+uvN]}{(50+uN)^2(50+vN)^2}$$

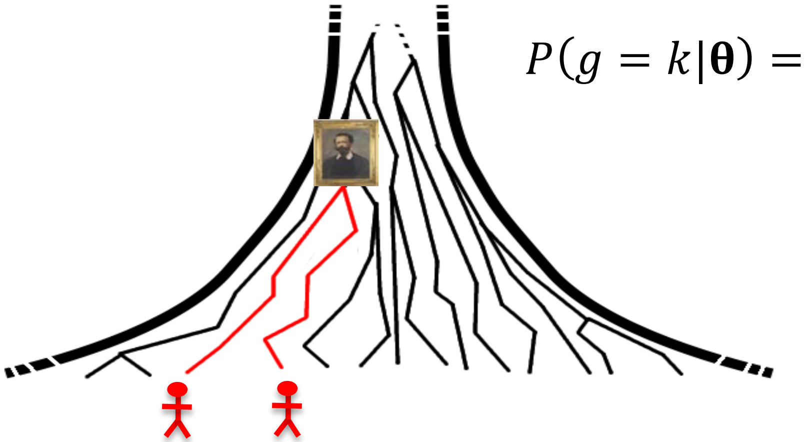
If $v \rightarrow \infty$ and $\hat{f}_u$ is the observed average sharing:

$$\hat{N} = \frac{50(1-\hat{f}_u+\sqrt{1-\hat{f}_u})}{u\hat{f}_u}$$

A formal link between demography and IBD

$$E[f_{u,v}|\boldsymbol{\theta}] = \sum_{k=1}^{\infty} \boxed{P(g=k|\boldsymbol{\theta})}\, E[f_{u,v}|k]$$

If population has arbitrary size $N(k,\boldsymbol{\theta})$



$$P(g=k|\boldsymbol{\theta}) = \frac{1}{N(k,\boldsymbol{\theta})} \prod_{j=1}^{k-1} \left(1 - \frac{1}{N(k,\boldsymbol{\theta})}\right)$$

More IBD quantities

Additional quantities can be similarly derived:

- Distribution of the length s of a random IBD segment

$$\Pr(random\ segment\ is\ of\ length\ l|\theta) =$$

$$\mathrm{P}(\mathrm{s} = \mathrm{l}|\theta) = \frac{P(l|\theta)}{l} \times \frac{1}{\int_0^\infty \frac{\Pr(l|\theta)}{l} dl} = \frac{2 \times 50^2 N_e}{(50 + lN_e)^3}$$

More IBD quantities

Additional quantities can be similarly derived:

- Distribution of the length s of a random IBD segment
- Expectation of the number of IBD segments in range R=[u,v]

$$\lambda_R = \gamma \times \frac{E_R[f|\theta]}{E_R[s|\theta]} = \gamma \times \frac{50N_e^2 uv(100 + N_e(u+v))}{(50+uN_e)^2(50+vN_e)^2}$$

More IBD quantities

Additional quantities can be similarly derived:

- Distribution of the length s of a random IBD segment
- Expectation of the number of IBD segments in range R
- Distribution for the total IBD sharing

Outline

- Introduction: Identity by Descent

- A model for IBD sharing and demography
  - Migration
- Examples

- IBD and sequence data

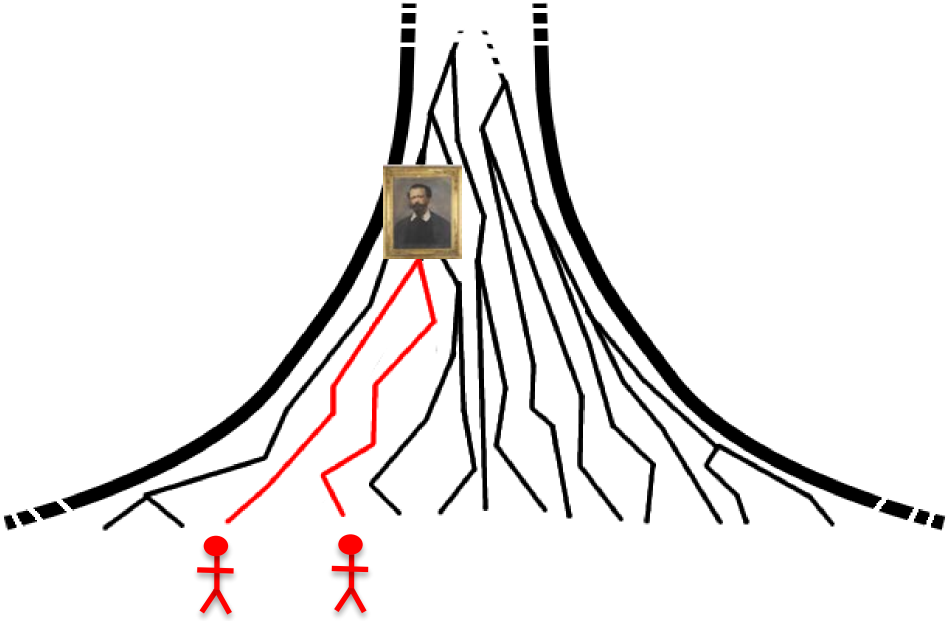
Outline

- Introduction: Identity by Descent
- A model for IBD sharing and demography
  - Migration
- Examples
- IBD and sequence data

IBD and the coalescent

$$E[f_{u,v}|\boldsymbol{\theta}] = \sum_{k=1}^{\infty} \boxed{P(g = k|\boldsymbol{\theta})} E[f_{u,v}|k]$$

IBD in population depends on demographic parameters through the coalescent distribution.

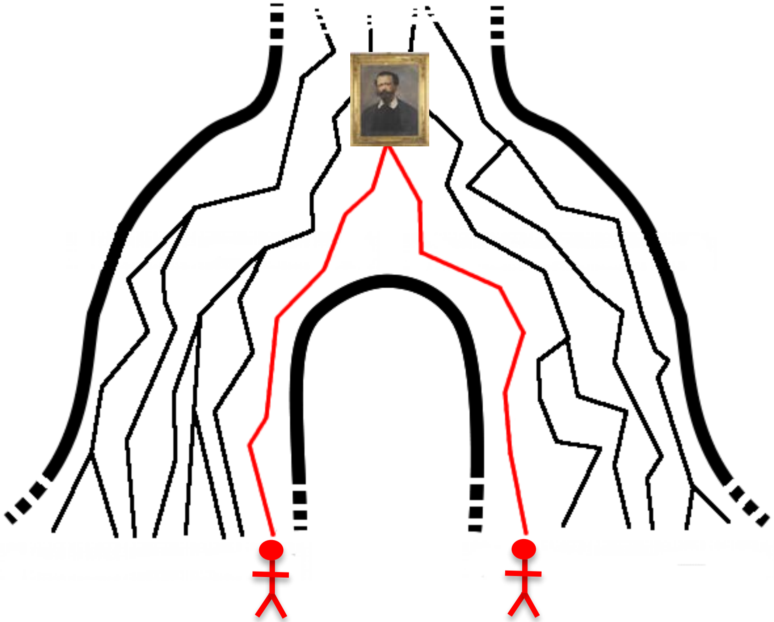


Palamara & Pe’er, Bioinformatics 2013

IBD and the coalescent

$$E[f_{u,v}|\boldsymbol{\theta}] = \sum_{k=1}^{\infty} \boxed{P(g = k|\boldsymbol{\theta})} E[f_{u,v}|k]$$

What if the two individual are sampled from two different populations?

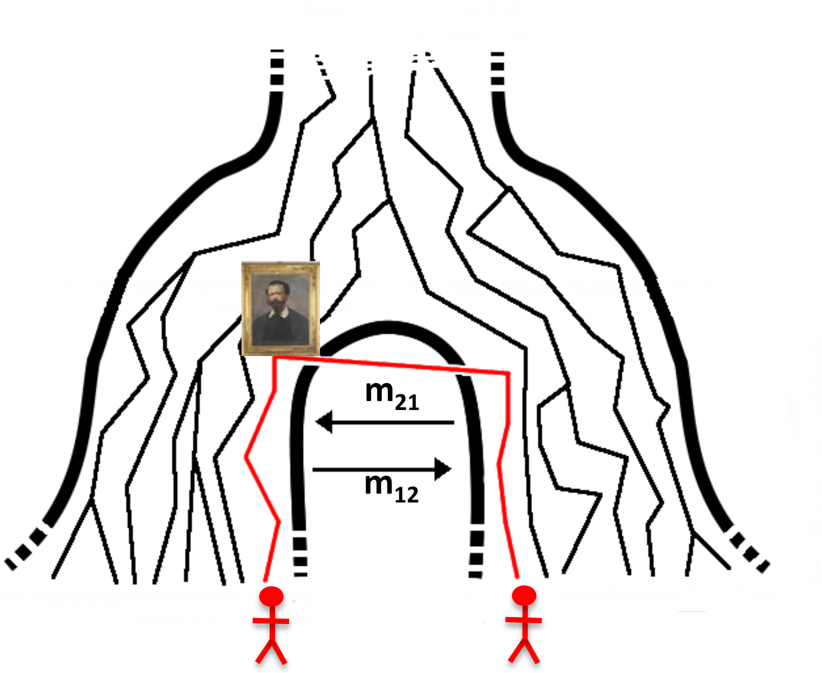


Palamara & Pe’er, Bioinformatics 2013

IBD and the coalescent

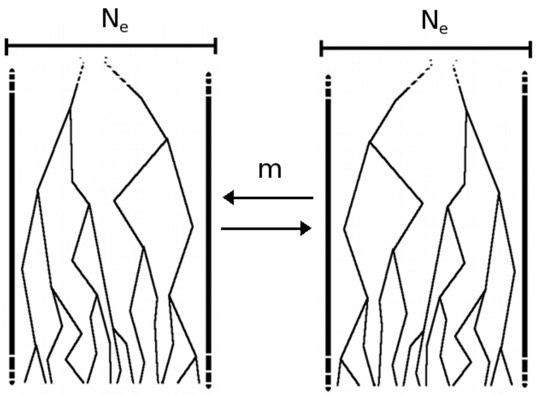
$$E[f_{u,v}|\boldsymbol{\theta}] = \sum_{k=1}^{\infty} \boxed{P(g = k|\boldsymbol{\theta})} E[f_{u,v}|k]$$

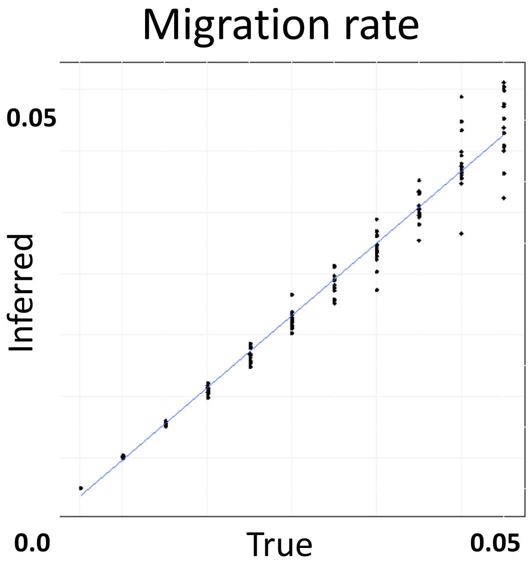
What if the two individual are sampled from two different populations?

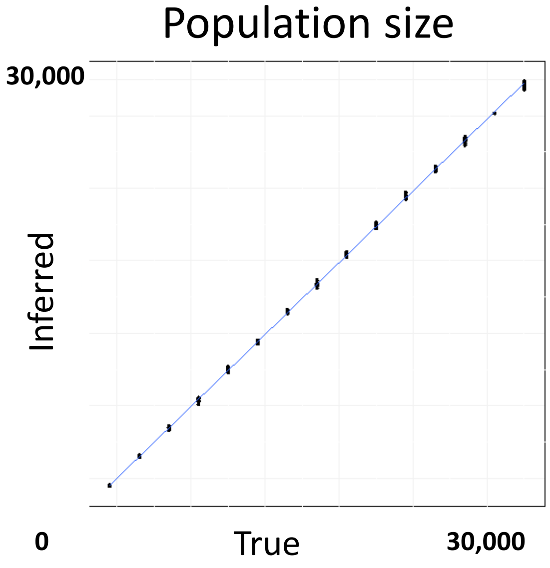


Palamara & Pe'er, Bioinformatics 2013

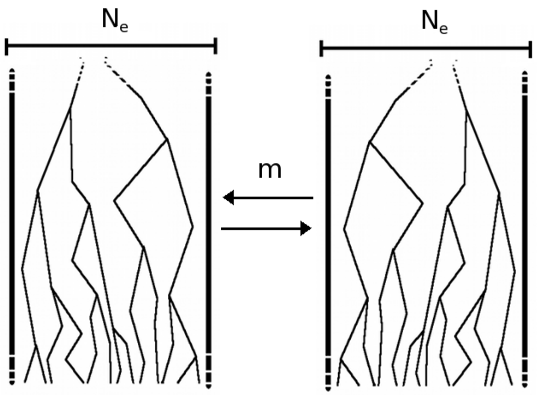
Evaluation - simulations



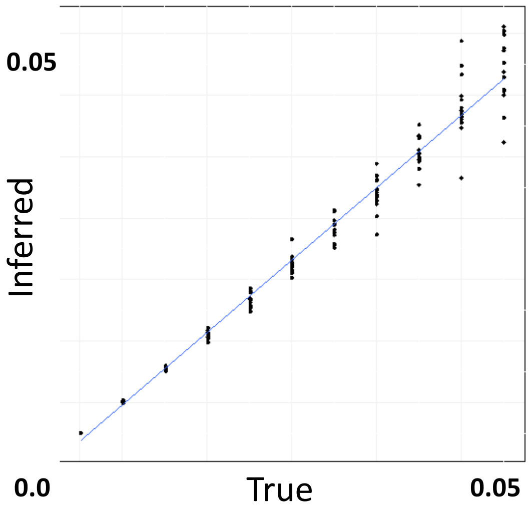




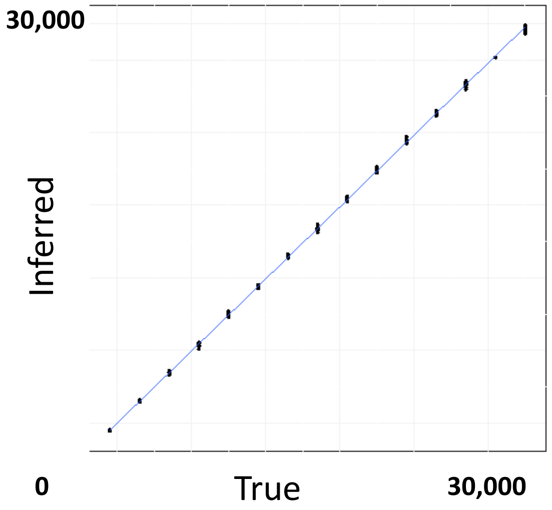
Evaluation - simulations

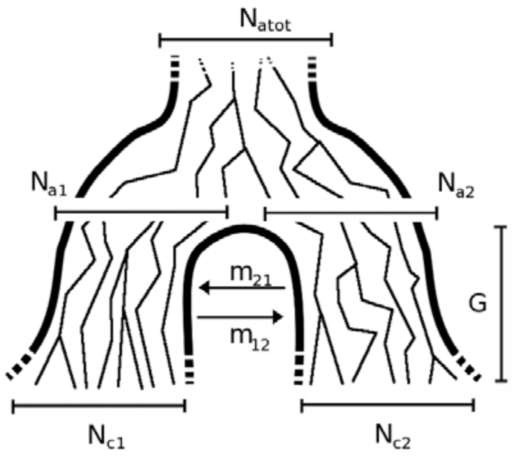


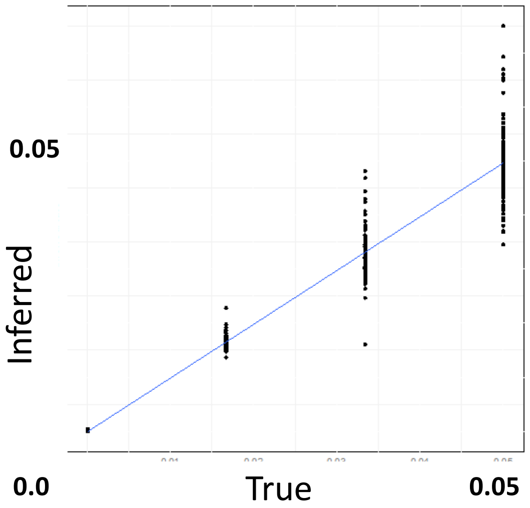
Migration rate

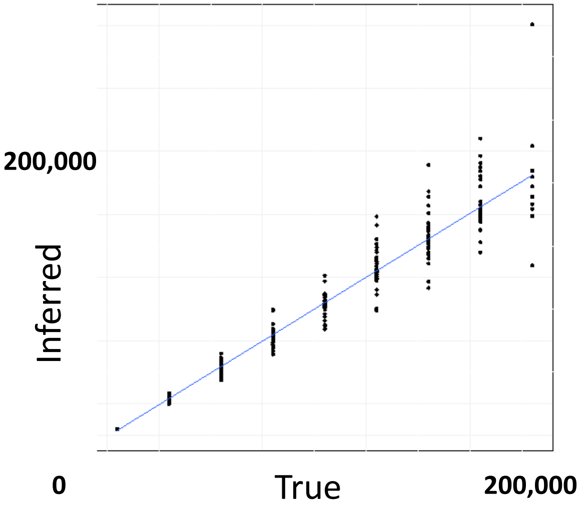


Population size

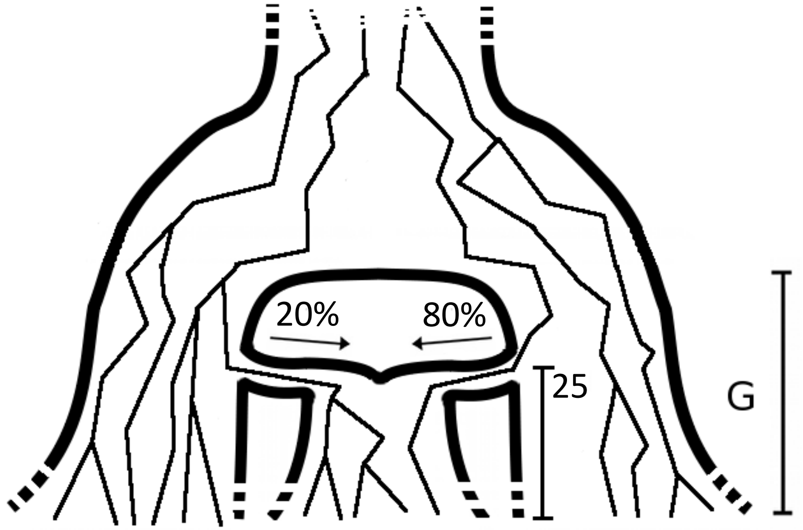


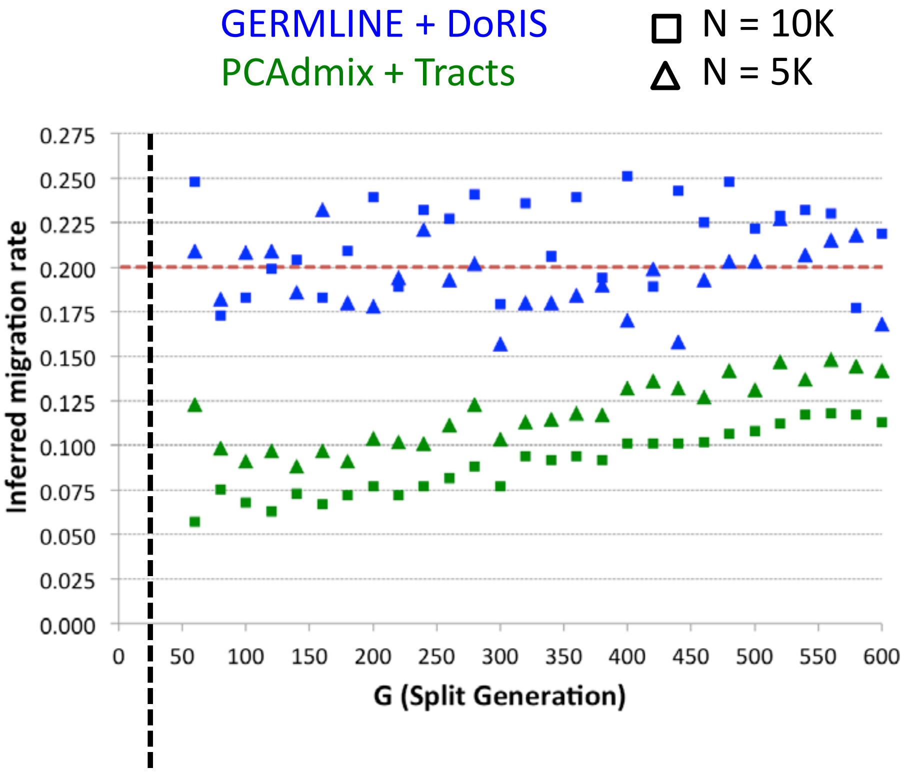






Evaluation – IBD vs ancestry deconvolution

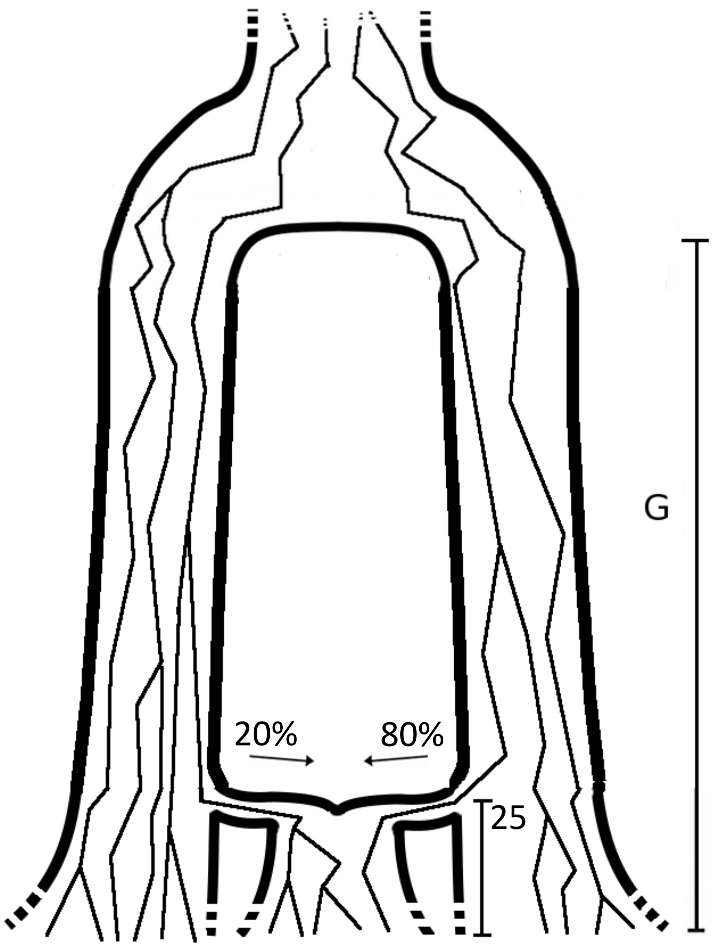




Admixture generation = 25

Palamara & Pe’er, Bioinformatics 2013

Evaluation – IBD vs ancestry deconvolution

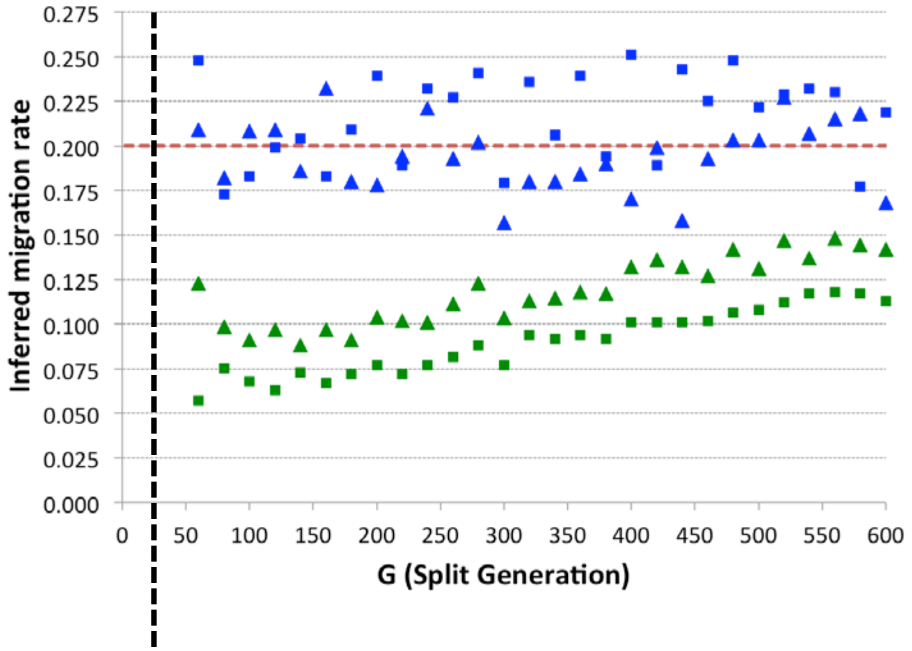


GERMLINE + DoRIS

PCAdmix + Tracts

□ N = 10K

△ N = 5K



Admixture generation = 25

Palamara & Pe’er, Bioinformatics 2013

Outline

- Introduction: Identity by Descent

- A model for IBD sharing and demography
  - Migration
- Examples

- IBD and sequence data

Outline

- Introduction: Identity by Descent

- A model for IBD sharing and demography
  - Migration
- Examples
  - Netherlands
  - Ashkenazi Jews
- IBD and sequence data

The Genome Of Netherlands

SAMPLES: 498 trio parents from 11 provinces
DATA: 14X Illumina, 2.3M SNPs MAF>1%, Trio-phased (*)
IBD COMPUTATION:
- Clean stretches span 2160cM
- Tuned GERMLINE (**) parameters
- Analyzed length distributions

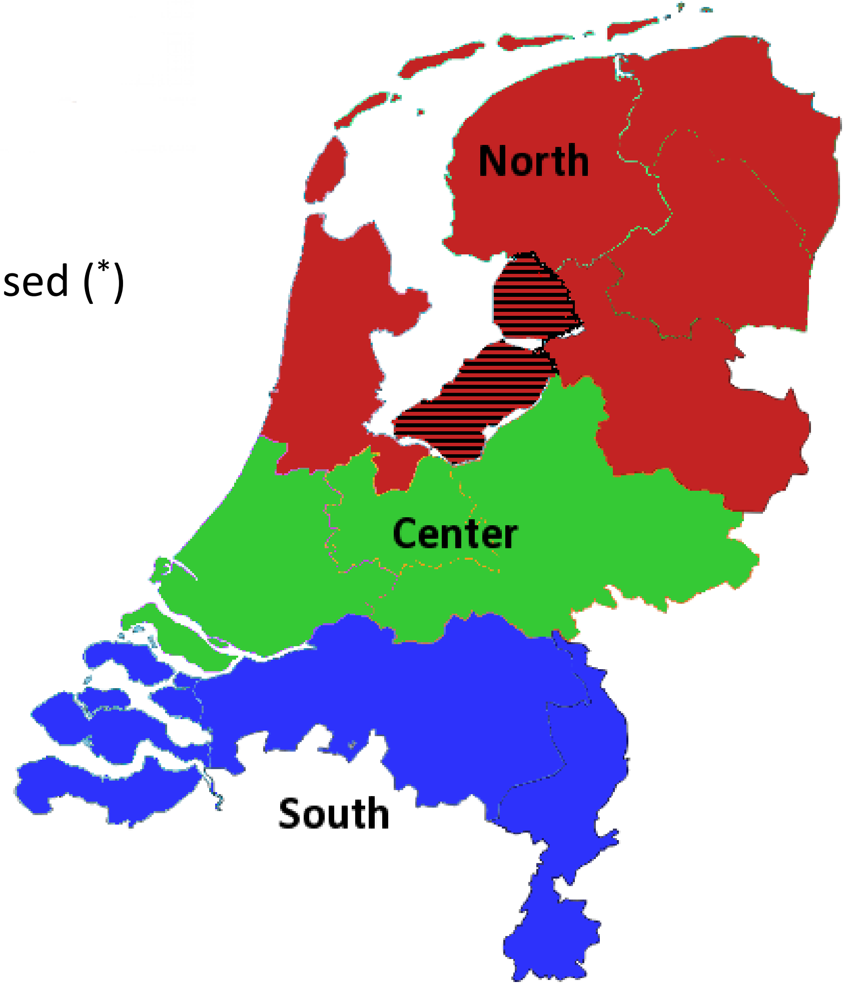
(*) Melanou & Marchini, ‘13
(**) Gusev et al., ‘09

Image adapted from Abdellaoui et al. ‘13

The Genome Of Netherlands

SAMPLES: 498 trio parents from 11 provinces
DATA: 14X Illumina, 2.3M SNPs MAF>1%, Trio-phased (*)
IBD COMPUTATION:

- Clean stretches span 2160cM
- Tuned GERMLINE (**) parameters
- Analyzed length distributions

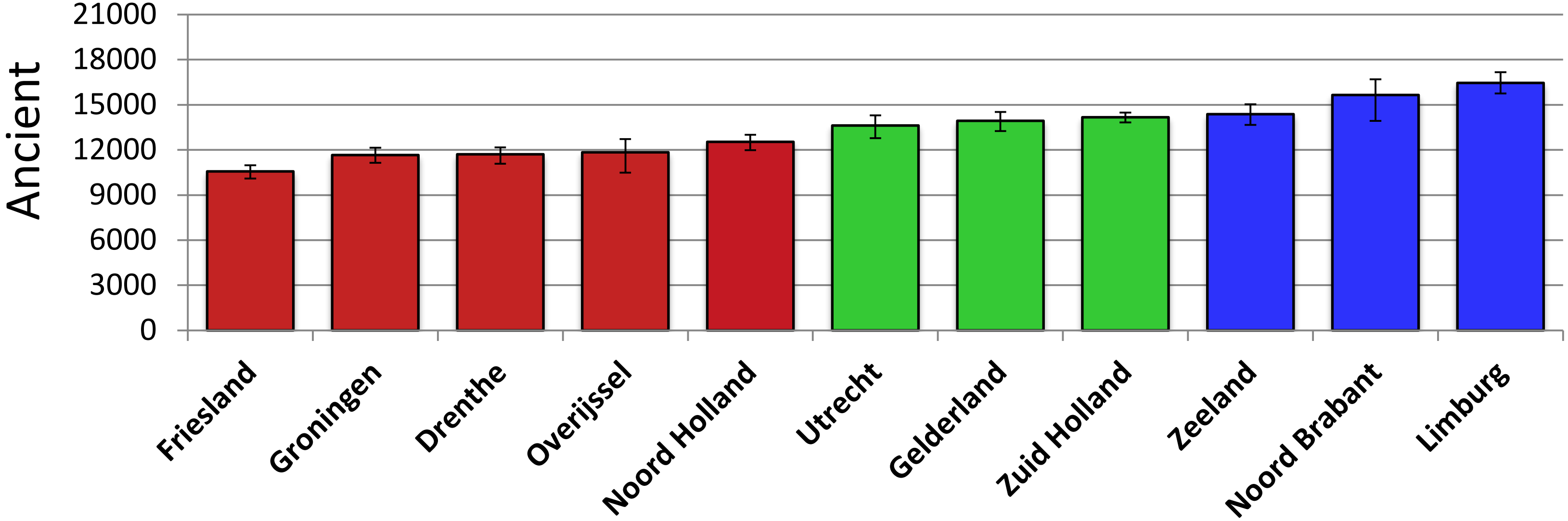


(*) Melanou & Marchini, '13
(**) Gusev et al., '09

Image adapted from Abdellaoui et al. '13

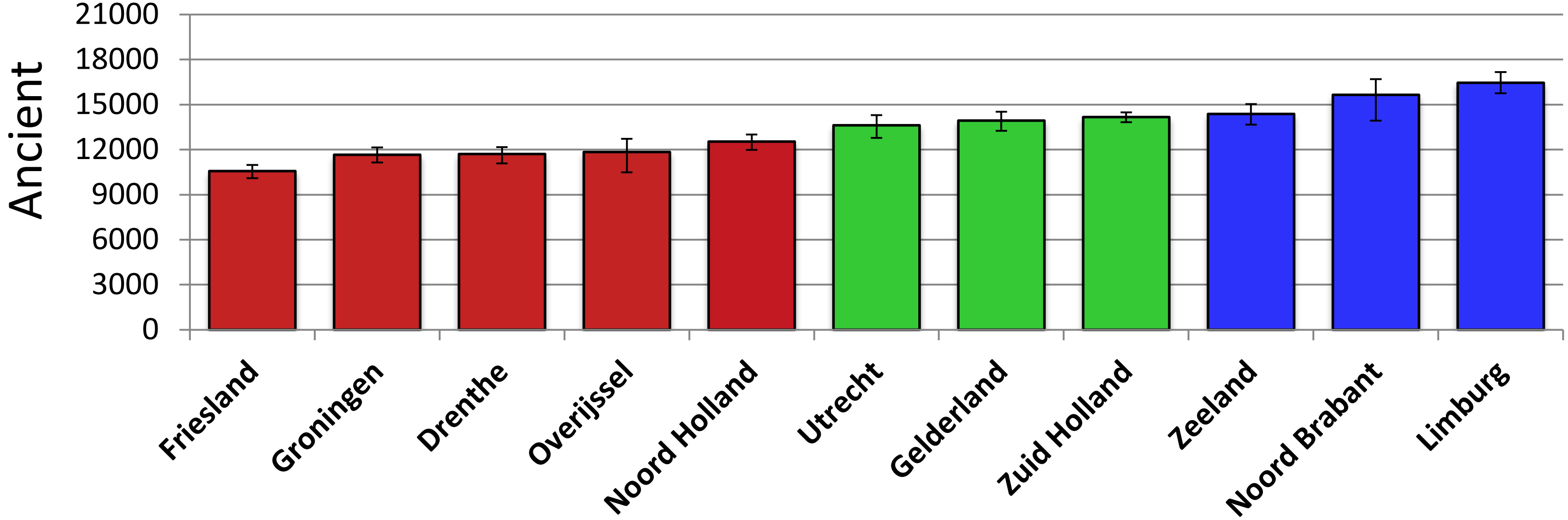
Reconstructed population size in GoNL

1cM ≤ IBD ≤ 2cM

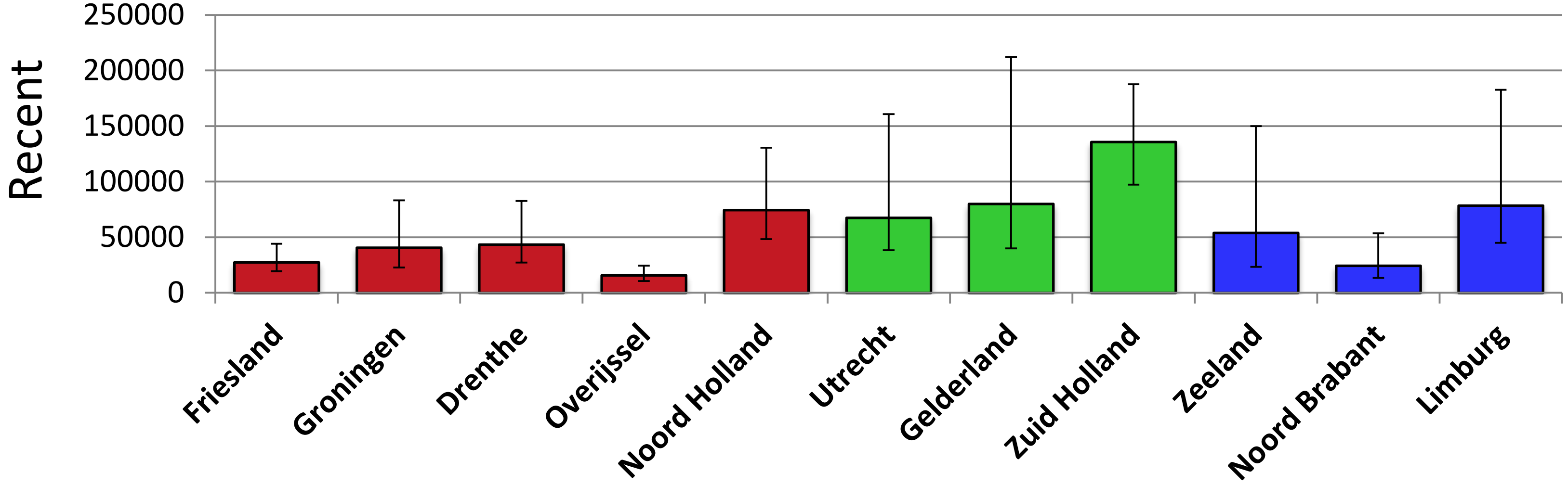


Reconstructed population size in GoNL

1cM ≤ IBD ≤ 2cM



7cM ≤ IBD ≤ 15cM



Inter/intra-province IBD sharing

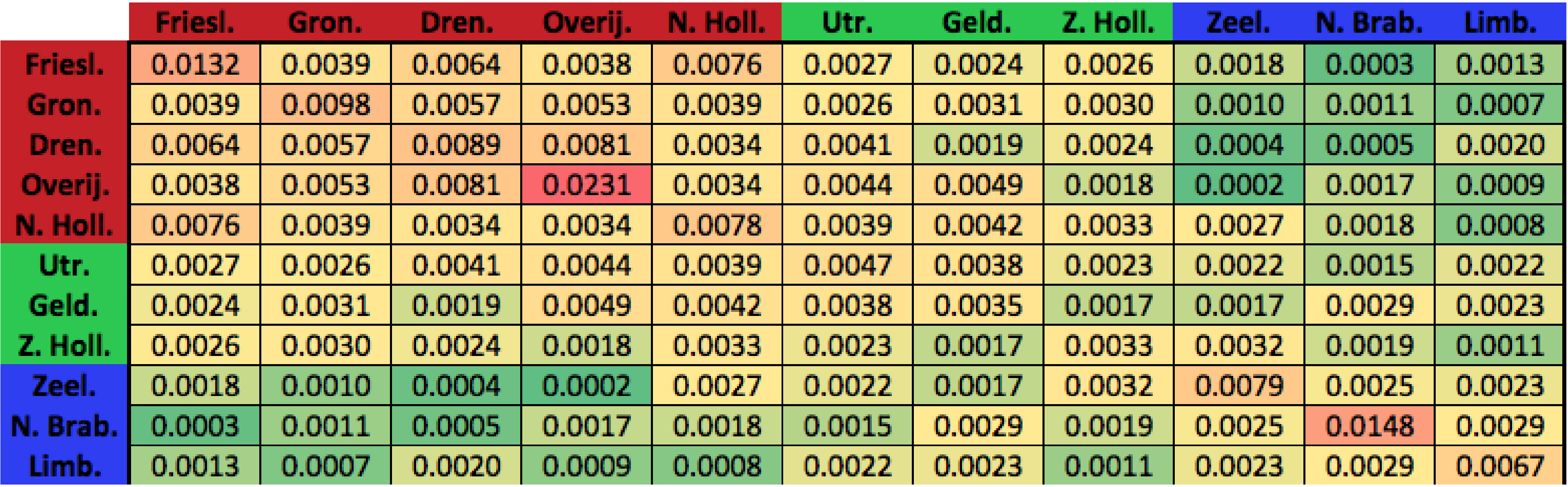
Average number of segments **IBD ≥ 7cM**

	Friesl.	Gron.	Dren.	Overij.	N. Holl.	Utr.	Geld.	Z. Holl.	Zeel.	N. Brab.	Limb.
Friesl.	0.0339	0.0078	0.0095	0.0056	0.0071	0.0031	0.0027	0.0025	0.0018	0.0007	0.0014
Gron.	0.0078	0.0316	0.0103	0.0068	0.0052	0.0016	0.0031	0.0058	0.0010	0.0011	0.0017
Dren.	0.0095	0.0103	0.0237	0.0182	0.0052	0.0035	0.0049	0.0029	0.0010	0.0016	0.0009
Overij.	0.0056	0.0068	0.0182	0.0603	0.0036	0.0038	0.0078	0.0024	0.0012	0.0008	0.0009
N. Holl.	0.0071	0.0052	0.0052	0.0036	0.0133	0.0049	0.0051	0.0039	0.0019	0.0045	0.0012
Utr.	0.0031	0.0016	0.0035	0.0038	0.0049	0.0177	0.0055	0.0053	0.0031	0.0024	0.0030
Geld.	0.0027	0.0031	0.0049	0.0078	0.0051	0.0055	0.0120	0.0031	0.0026	0.0043	0.0030
Z. Holl.	0.0025	0.0058	0.0029	0.0024	0.0039	0.0053	0.0031	0.0073	0.0048	0.0043	0.0019
Zeel.	0.0018	0.0010	0.0010	0.0012	0.0019	0.0031	0.0026	0.0048	0.0206	0.0034	0.0033
N. Brab.	0.0007	0.0011	0.0016	0.0008	0.0045	0.0024	0.0043	0.0043	0.0034	0.0440	0.0035
Limb.	0.0014	0.0017	0.0009	0.0009	0.0012	0.0030	0.0030	0.0019	0.0033	0.0035	0.0130

Expected time of common ancestor transmitting IBD segment

Inter/intra-province IBD sharing

Average number of segments **6cM ≤ IBD ≤ 7cM**



	Friesl.	Gron.	Dren.	Overij.	N. Holl.	Utr.	Geld.	Z. Holl.	Zeel.	N. Brab.	Limb.
Friesl.	0.0132	0.0039	0.0064	0.0038	0.0076	0.0027	0.0024	0.0026	0.0018	0.0003	0.0013
Gron.	0.0039	0.0098	0.0057	0.0053	0.0039	0.0026	0.0031	0.0030	0.0010	0.0011	0.0007
Dren.	0.0064	0.0057	0.0089	0.0081	0.0034	0.0041	0.0019	0.0024	0.0004	0.0005	0.0020
Overij.	0.0038	0.0053	0.0081	0.0231	0.0034	0.0044	0.0049	0.0018	0.0002	0.0017	0.0009
N. Holl.	0.0076	0.0039	0.0034	0.0034	0.0078	0.0039	0.0042	0.0033	0.0027	0.0018	0.0008
Utr.	0.0027	0.0026	0.0041	0.0044	0.0039	0.0047	0.0038	0.0023	0.0022	0.0015	0.0022
Geld.	0.0024	0.0031	0.0019	0.0049	0.0042	0.0038	0.0035	0.0017	0.0017	0.0029	0.0023
Z. Holl.	0.0026	0.0030	0.0024	0.0018	0.0033	0.0023	0.0017	0.0033	0.0032	0.0019	0.0011
Zeel.	0.0018	0.0010	0.0004	0.0002	0.0027	0.0022	0.0017	0.0032	0.0079	0.0025	0.0023
N. Brab.	0.0003	0.0011	0.0005	0.0017	0.0018	0.0015	0.0029	0.0019	0.0025	0.0148	0.0029
Limb.	0.0013	0.0007	0.0020	0.0009	0.0008	0.0022	0.0023	0.0011	0.0023	0.0029	0.0067

Expected time of common ancestor transmitting IBD segment

Inter/intra-province IBD sharing

Average number of segments **5cM ≤ IBD ≤ 6cM**

	Friesl.	**Gron.**	**Dren.**	**Overij.**	**N. Holl.**	**Utr.**	**Geld.**	**Z. Holl.**	**Zeel.**	**N. Brab.**	**Limb.**
Friesl.	0.0307	0.0108	0.0090	0.0083	0.0142	0.0083	0.0061	0.0055	0.0061	0.0029	0.0019
Gron.	0.0108	0.0126	0.0139	0.0115	0.0070	0.0062	0.0038	0.0060	0.0041	0.0024	0.0017
Dren.	0.0090	0.0139	0.0152	0.0141	0.0077	0.0061	0.0078	0.0062	0.0030	0.0030	0.0020
Overij.	0.0083	0.0115	0.0141	0.0309	0.0062	0.0055	0.0082	0.0038	0.0027	0.0020	0.0033
N. Holl.	0.0142	0.0070	0.0077	0.0062	0.0143	0.0067	0.0052	0.0053	0.0046	0.0045	0.0030
Utr.	0.0083	0.0062	0.0061	0.0055	0.0067	0.0089	0.0049	0.0071	0.0051	0.0024	0.0024
Geld.	0.0061	0.0038	0.0078	0.0082	0.0052	0.0049	0.0104	0.0048	0.0022	0.0067	0.0036
Z. Holl.	0.0055	0.0060	0.0062	0.0038	0.0053	0.0071	0.0048	0.0064	0.0064	0.0047	0.0019
Zeel.	0.0061	0.0041	0.0030	0.0027	0.0046	0.0051	0.0022	0.0064	0.0132	0.0074	0.0039
N. Brab.	0.0029	0.0024	0.0030	0.0020	0.0045	0.0024	0.0067	0.0047	0.0074	0.0196	0.0047
Limb.	0.0019	0.0017	0.0020	0.0033	0.0030	0.0024	0.0036	0.0019	0.0039	0.0047	0.0130

Expected time of common ancestor transmitting IBD segment

Inter/intra-province IBD sharing

Average number of segments **4cM ≤ IBD ≤ 5cM**

	Friesl.	Gron.	Dren.	Overij.	N. Holl.	Utr.	Geld.	Z. Holl.	Zeel.	N. Brab.	Limb.
Friesl.	0.0499	0.0238	0.0236	0.0203	0.0302	0.0200	0.0116	0.0140	0.0156	0.0087	0.0078
Gron.	0.0238	0.0295	0.0238	0.0228	0.0190	0.0150	0.0133	0.0120	0.0095	0.0083	0.0085
Dren.	0.0236	0.0238	0.0315	0.0342	0.0190	0.0143	0.0125	0.0112	0.0091	0.0082	0.0064
Overij.	0.0203	0.0228	0.0342	0.0477	0.0168	0.0160	0.0181	0.0123	0.0100	0.0085	0.0066
N. Holl.	0.0302	0.0190	0.0190	0.0168	0.0206	0.0181	0.0124	0.0144	0.0103	0.0084	0.0074
Utr.	0.0200	0.0150	0.0143	0.0160	0.0181	0.0177	0.0157	0.0145	0.0136	0.0096	0.0094
Geld.	0.0116	0.0133	0.0125	0.0181	0.0124	0.0157	0.0146	0.0092	0.0085	0.0107	0.0086
Z. Holl.	0.0140	0.0120	0.0112	0.0123	0.0144	0.0145	0.0092	0.0130	0.0129	0.0116	0.0076
Zeel.	0.0156	0.0095	0.0091	0.0100	0.0103	0.0136	0.0085	0.0129	0.0259	0.0131	0.0074
N. Brab.	0.0087	0.0083	0.0082	0.0085	0.0084	0.0096	0.0107	0.0116	0.0131	0.0286	0.0104
Limb.	0.0078	0.0085	0.0064	0.0066	0.0074	0.0094	0.0086	0.0076	0.0074	0.0104	0.0179

Expected time of common ancestor transmitting IBD segment

Inter/intra-province IBD sharing

Average number of segments **4cM ≤ IBD ≤ 5cM**

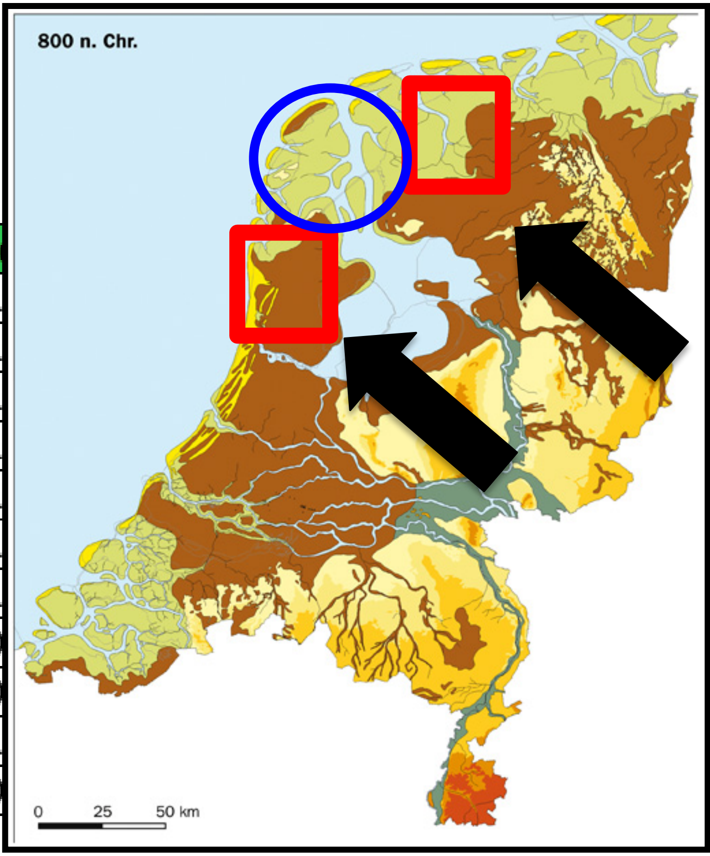
	Friesl.	Gron.	Dren.	Overij.	N. Holl.	Utr.	Geld.	Z. Holl.	Zeel.	N. Brab.	Limb.
Friesl.	0.0499	0.0238	0.0236	0.0203	0.0302	0.0200	0.0116	0.0140	0.0156	0.0087	0.0078
Gron.	0.0238	0.0295	0.0238	0.0228	0.0190	0.0150	0.0133	0.0120	0.0095	0.0083	0.0085
Dren.	0.0236	0.0238	0.0315	0.0342	0.0190	0.0143	0.0125	0.0112	0.0091	0.0082	0.0064
Overij.	0.0203	0.0228	0.0342	0.0477	0.0168	0.0160	0.0181	0.0123	0.0100	0.0085	0.0066
N. Holl.	0.0302	0.0190	0.0190	0.0168	0.0206	0.0181	0.0124	0.0144	0.0103	0.0084	0.0074
Utr.	0.0200	[illegible]150	0.0143	0.0160	0.0181	[illegible]177	0.0157	0.0145	0.0136	0.0096	0.0094
Geld.	0.0116	0.[illegible]3	0.0125	0.0181	0.0124	0.[illegible]7	0.0146	0.0092	0.0085	0.0107	0.0086
Z. Holl.	0.0140	0.012[illegible]	0.0112	0.0123	0.0144	0.014[illegible]	0.0092	0.0130	0.0129	0.0116	0.0076
Zeel.	0.0156	0.0095	0.0091	0.0100	0.0103	0.0136	0.0085	0.0129	0.0259	0.0131	0.0074
N. Brab.	0.0087	0.0083	0.0082	0.0085	0.0084	0.0096	0.0107	0.0116	0.0131	0.0286	0.0104
Limb.	0.0078	0.0085	0.0064	0.0066	0.0074	0.0094	0.0086	0.0076	0.0074	0.0104	0.0179

Expected time of common ancestor transmitting IBD segment

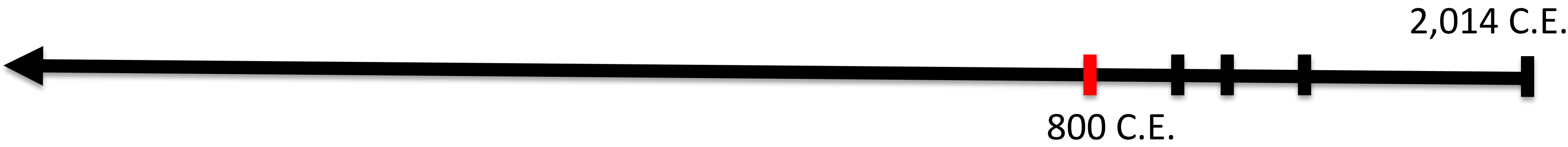
Inter/intra-province IBD sharing

Average number of segments **4cM ≤ IBD ≤ 5cM**

	Friesl.	Gron.	Dren.	Overij.	N. Holl.	Utr.	Gel
Friesl.	0.0499	0.0238	0.0236	0.0203	0.0302	0.0200	0.01
Gron.	0.0238	0.0295	0.0238	0.0228	0.0190	0.0150	0.01
Dren.	0.0236	0.0238	0.0315	0.0342	0.0190	0.0143	0.01
Overij.	0.0203	0.0228	0.0342	0.0477	0.0168	0.0160	0.01
N. Holl.	0.0302	0.0190	0.0190	0.0168	0.0206	0.0181	0.01
Utr.	0.0200	[illegible]	0.0143	0.0160	0.0181	[illegible]	0.01
Geld.	0.0116	[illegible]	0.0125	0.0181	0.0124	[illegible]	0.01
Z. Holl.	0.0140	[illegible]	0.0112	0.0123	0.0144	[illegible]	0.00
Zeel.	0.0156	0.0095	0.0091	0.0100	0.0103	0.0136	0.00
N. Brab.	0.0087	0.0083	0.0082	0.0085	0.0084	0.0096	0.01
Limb.	0.0078	0.0085	0.0064	0.0066	0.0074	0.0094	0.00



Expected time of common ancestor transmitting IBD segment



Map adapted from http://nl.wikipedia.org/wiki/Tijdlijn_van_de_Nederlandse_geschiedenis

Inter/intra-province IBD sharing

Average number of segments **4cM ≤ IBD ≤ 5cM**

	Friesl.	**Gron.**	**Dren.**	**Overij.**	**N. Holl.**	**Utr.**	**Geld.**	**Z. Holl.**	**Zeel.**	**N. Brab.**	**Limb.**
Friesl.	0.0499	0.0238	0.0236	0.0203	0.0302	0.0200	0.0116	0.0140	0.0156	0.0087	0.0078
Gron.	0.0238	0.0295	0.0238	0.0228	0.0190	0.0150	0.0133	0.0120	0.0095	0.0083	0.0085
Dren.	0.0236	0.0238	0.0315	0.0342	0.0190	0.0143	0.0125	0.0112	0.0091	0.0082	0.0064
Overij.	0.0203	0.0228	0.0342	0.0477	0.0168	0.0160	0.0181	0.0123	0.0100	0.0085	0.0066
N. Holl.	0.0302	0.0190	0.0190	0.0168	0.0206	0.0181	0.0124	0.0144	0.0103	0.0084	0.0074
Utr.	0.0200	0.0150	0.0143	0.0160	0.0181	0.0177	0.0157	0.0145	0.0136	0.0096	0.0094
Geld.	0.0116	0.0133	0.0125	0.0181	0.0124	0.0157	0.0146	0.0092	0.0085	0.0107	0.0086
Z. Holl.	0.0140	0.0120	0.0112	0.0123	0.0144	0.0145	0.0092	0.0130	0.0129	0.0116	0.0076
Zeel.	0.0156	0.0095	0.0091	0.0100	0.0103	0.0136	0.0085	0.0129	0.0259	0.0131	0.0074
N. Brab.	0.0087	0.0083	0.0082	0.0085	0.0084	0.0096	0.0107	0.0116	0.0131	0.0286	0.0104
Limb.	0.0078	0.0085	0.0064	0.0066	0.0074	0.0094	0.0086	0.0076	0.0074	0.0104	0.0179

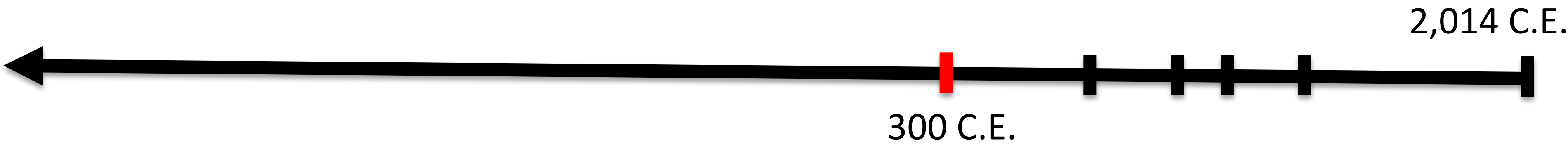
Expected time of common ancestor transmitting IBD segment

Inter/intra-province IBD sharing

Average number of segments **3cM ≤ IBD ≤ 4cM**

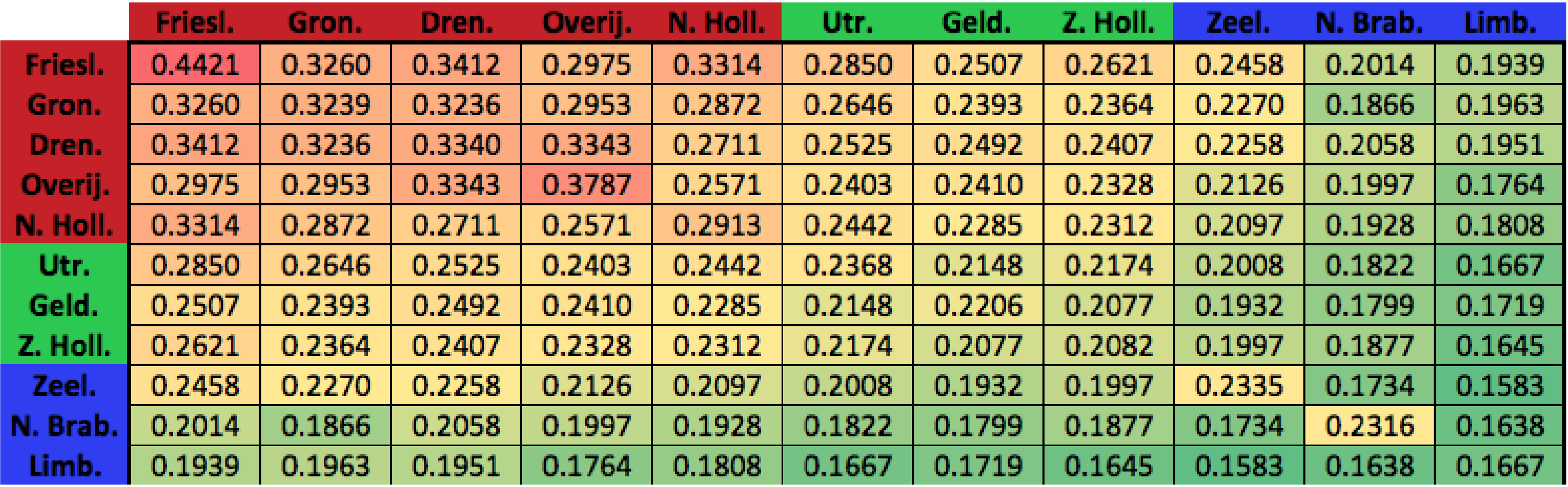
	Friesl.	**Gron.**	**Dren.**	**Overij.**	**N. Holl.**	**Utr.**	**Geld.**	**Z. Holl.**	**Zeel.**	**N. Brab.**	**Limb.**
Friesl.	0.1443	0.0815	0.0787	0.0608	0.0802	0.0512	0.0446	0.0511	0.0485	0.0319	0.0284
Gron.	0.0815	0.0775	0.0709	0.0641	0.0572	0.0453	0.0457	0.0426	0.0385	0.0247	0.0305
Dren.	0.0787	0.0709	0.0892	0.0775	0.0550	0.0473	0.0454	0.0398	0.0330	0.0264	0.0251
Overij.	0.0608	0.0641	0.0775	0.1185	0.0514	0.0465	0.0533	0.0411	0.0345	0.0332	0.0270
N. Holl.	0.0802	0.0572	0.0550	0.0514	0.0700	0.0525	0.0409	0.0480	0.0388	0.0344	0.0278
Utr.	0.0512	0.0453	0.0473	0.0465	0.0525	0.0536	0.0371	0.0413	0.0339	0.0273	0.0217
Geld.	0.0446	0.0457	0.0454	0.0533	0.0409	0.0371	0.0402	0.0335	0.0272	0.0302	0.0281
Z. Holl.	0.0511	0.0426	0.0398	0.0411	0.0480	0.0413	0.0335	0.0392	0.0399	0.0316	0.0250
Zeel.	0.0485	0.0385	0.0330	0.0345	0.0388	0.0339	0.0272	0.0399	0.0544	0.0343	0.0257
N. Brab.	0.0319	0.0247	0.0264	0.0332	0.0344	0.0273	0.0302	0.0316	0.0343	0.0623	0.0290
Limb.	0.0284	0.0305	0.0251	0.0270	0.0278	0.0217	0.0281	0.0250	0.0257	0.0290	0.0330

Expected time of common ancestor transmitting IBD segment



Inter/intra-province IBD sharing

Average number of segments **2cM ≤ IBD ≤ 3cM**



	Friesl.	Gron.	Dren.	Overij.	N. Holl.	Utr.	Geld.	Z. Holl.	Zeel.	N. Brab.	Limb.
Friesl.	0.4421	0.3260	0.3412	0.2975	0.3314	0.2850	0.2507	0.2621	0.2458	0.2014	0.1939
Gron.	0.3260	0.3239	0.3236	0.2953	0.2872	0.2646	0.2393	0.2364	0.2270	0.1866	0.1963
Dren.	0.3412	0.3236	0.3340	0.3343	0.2711	0.2525	0.2492	0.2407	0.2258	0.2058	0.1951
Overij.	0.2975	0.2953	0.3343	0.3787	0.2571	0.2403	0.2410	0.2328	0.2126	0.1997	0.1764
N. Holl.	0.3314	0.2872	0.2711	0.2571	0.2913	0.2442	0.2285	0.2312	0.2097	0.1928	0.1808
Utr.	0.2850	0.2646	0.2525	0.2403	0.2442	0.2368	0.2148	0.2174	0.2008	0.1822	0.1667
Geld.	0.2507	0.2393	0.2492	0.2410	0.2285	0.2148	0.2206	0.2077	0.1932	0.1799	0.1719
Z. Holl.	0.2621	0.2364	0.2407	0.2328	0.2312	0.2174	0.2077	0.2082	0.1997	0.1877	0.1645
Zeel.	0.2458	0.2270	0.2258	0.2126	0.2097	0.2008	0.1932	0.1997	0.2335	0.1734	0.1583
N. Brab.	0.2014	0.1866	0.2058	0.1997	0.1928	0.1822	0.1799	0.1877	0.1734	0.2316	0.1638
Limb.	0.1939	0.1963	0.1951	0.1764	0.1808	0.1667	0.1719	0.1645	0.1583	0.1638	0.1667

Expected time of common ancestor transmitting IBD segment

Inter/intra-province IBD sharing

Average number of segments **1cM ≤ IBD ≤ 2cM**

	Friesl.	Gron.	Dren.	Overij.	N. Holl.	Utr.	Geld.	Z. Holl.	Zeel.	N. Brab.	Limb.
Friesl.	4.0175	3.7754	3.7523	3.6017	3.6568	3.4693	3.3994	3.3494	3.2792	3.1282	3.0038
Gron.	3.7754	3.6667	3.6812	3.5634	3.5052	3.3577	3.3074	3.2891	3.2092	3.0104	2.9785
Dren.	3.7523	3.6812	3.6709	3.6243	3.5120	3.3360	3.3242	3.2605	3.2049	3.0610	2.9707
Overij.	3.6017	3.5634	3.6243	3.6056	3.3688	3.2990	3.2995	3.2169	3.1538	2.9819	2.9594
N. Holl.	3.6568	3.5052	3.5120	3.3688	3.4224	3.2644	3.2037	3.1830	3.1385	2.9654	2.9066
Utr.	3.4693	3.3577	3.3360	3.2990	3.2644	3.1851	3.1290	3.0873	3.0401	2.8945	2.8943
Geld.	3.3994	3.3074	3.3242	3.2995	3.2037	3.1290	3.1073	3.0392	2.9855	2.8928	2.8137
Z. Holl.	3.3494	3.2891	3.2605	3.2169	3.1830	3.0873	3.0392	3.0372	2.9747	2.8539	2.8289
Zeel.	3.2792	3.2092	3.2049	3.1538	3.1385	3.0401	2.9855	2.9747	3.0189	2.8086	2.7525
N. Brab.	3.1282	3.0104	3.0610	2.9819	2.9654	2.8945	2.8928	2.8539	2.8086	2.7693	2.6730
Limb.	3.0038	2.9785	2.9707	2.9594	2.9066	2.8943	2.8137	2.8289	2.7525	2.6730	2.6705

Expected time of common ancestor transmitting IBD segment

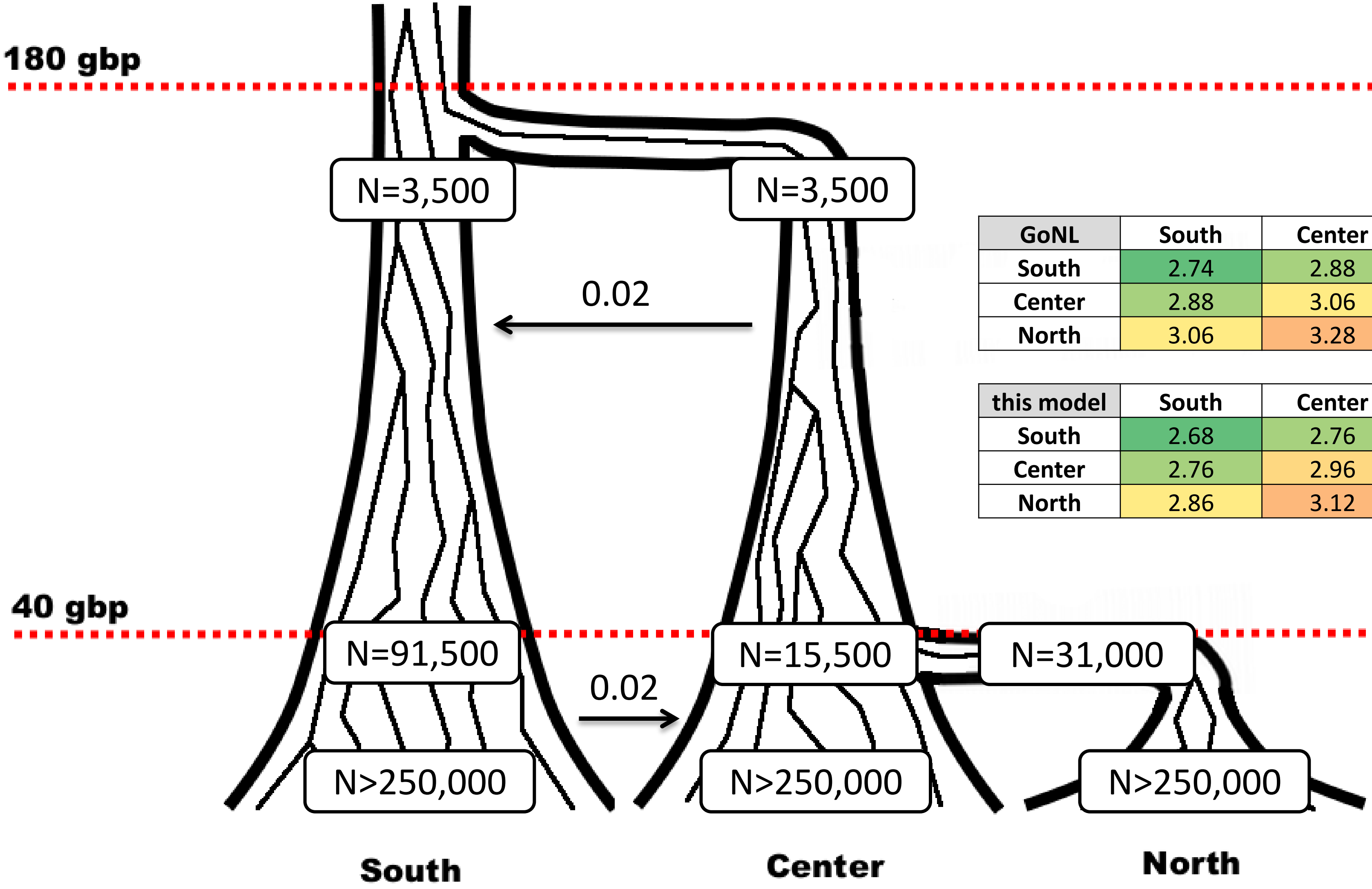
Inter/intra-province IBD sharing

Average number of segments **1cM ≤ IBD ≤ 2cM**

	Friesl.	**Gron.**	**Dren.**	**Overij.**	**N. Holl.**	**Utr.**	**Geld.**	**Z. Holl.**	**Zeel.**	**N. Brab.**	**Limb.**
Friesl.	4.0175	3.7754	3.7523	3.6017	3.6568	3.4693	3.3994	3.3494	3.2792	3.1282	3.0038
Gron.	3.7754	3.6667	3.6812	3.5634	3.5052	3.3577	3.3074	3.2891	3.2092	3.0104	2.9785
Dren.	3.7523	3.6812	3.6709	3.6243	3.5120	3.3360	3.3242	3.2605	3.2049	3.0610	2.9707
Overij.	3.6017	3.5634	3.6243	3.6056	3.3688	3.2990	3.2995	3.2169	3.1538	2.9819	2.9594
N. Holl.	3.6568	3.5052	3.5120	3.3688	3.4224	3.2644	3.2037	3.1830	3.1385	2.9654	2.9066
Utr.	3.4693	3.3577	3.3360	3.2990	3.2644	3.1851	3.1290	3.0873	3.0401	2.8945	2.8943
Geld.	3.3994	3.3074	3.3242	3.2995	3.2037	3.1290	3.1073	3.0392	2.9855	2.8928	2.8137
Z. Holl.	3.3494	3.2891	3.2605	3.2169	3.1830	3.0873	3.0392	3.0372	2.9747	2.8539	2.8289
Zeel.	3.2792	3.2092	3.2049	3.1538	3.1385	3.0401	2.9855	2.9747	3.0189	2.8086	2.7525
N. Brab.	3.1282	3.0104	3.0610	2.9819	2.9654	2.8945	2.8928	2.8539	2.8086	2.7693	2.6730
Limb.	3.0038	2.9785	2.9707	2.9594	2.9066	2.8943	2.8137	2.8289	2.7525	2.6730	2.6705

Expected time of common ancestor transmitting IBD segment

Netherlands: serial founder



180 gbp

N=3,500

N=3,500

0.02

40 gbp

N=91,500

0.02

N=15,500

N=31,000

N>250,000

N>250,000

N>250,000

South

Center

North

GoNL	South	Center	North
South	2.74	2.88	3.06
Center	2.88	3.06	3.28
North	3.06	3.28	3.61

this model	South	Center	North
South	2.68	2.76	2.86
Center	2.76	2.96	3.12
North	2.86	3.12	3.52

Outline

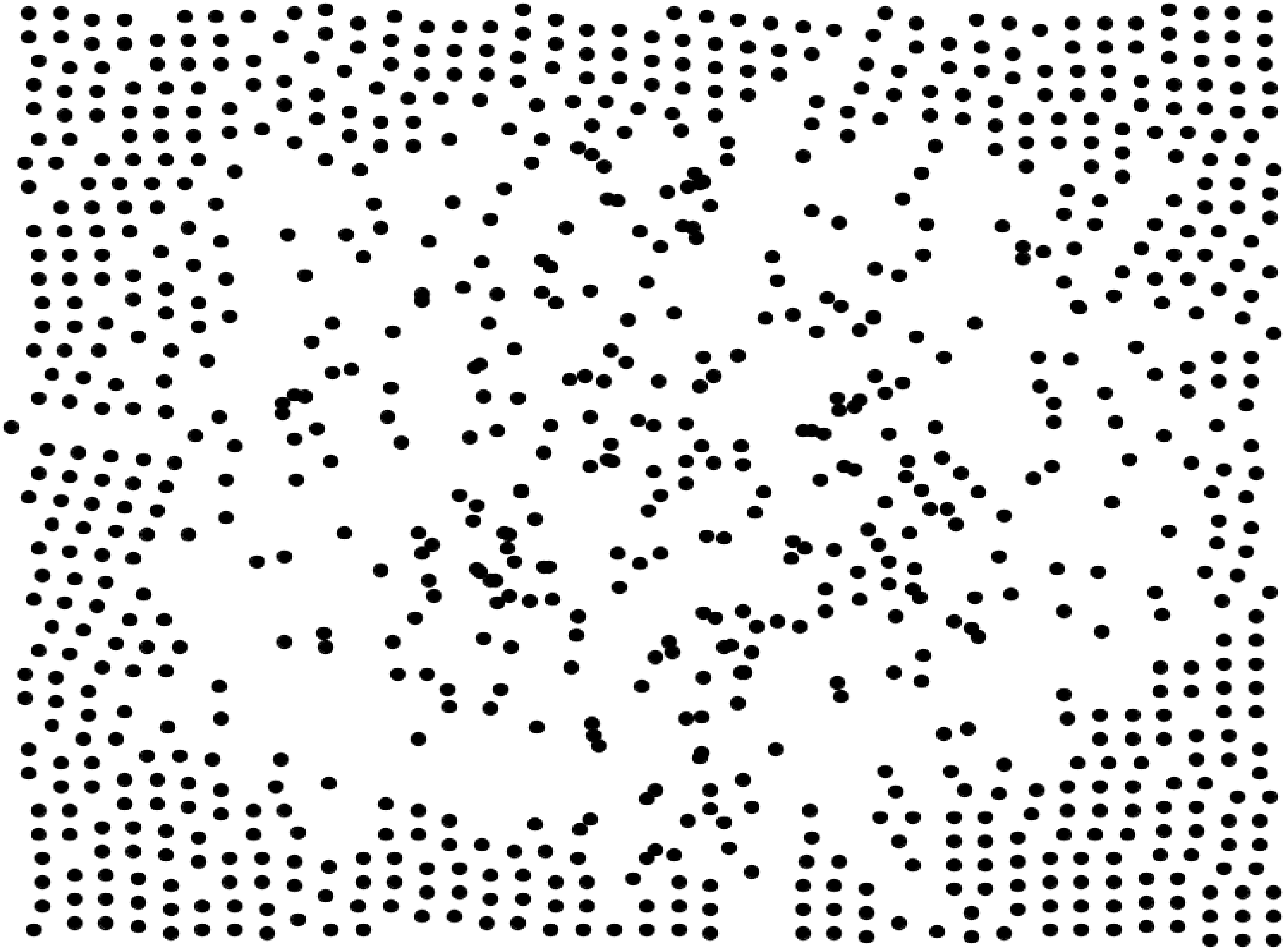
- Introduction: Identity by Descent

- A model for IBD sharing and demography
  - Migration
- Examples
  - Netherlands
  - Ashkenazi Jews
- IBD and sequence data

Outline

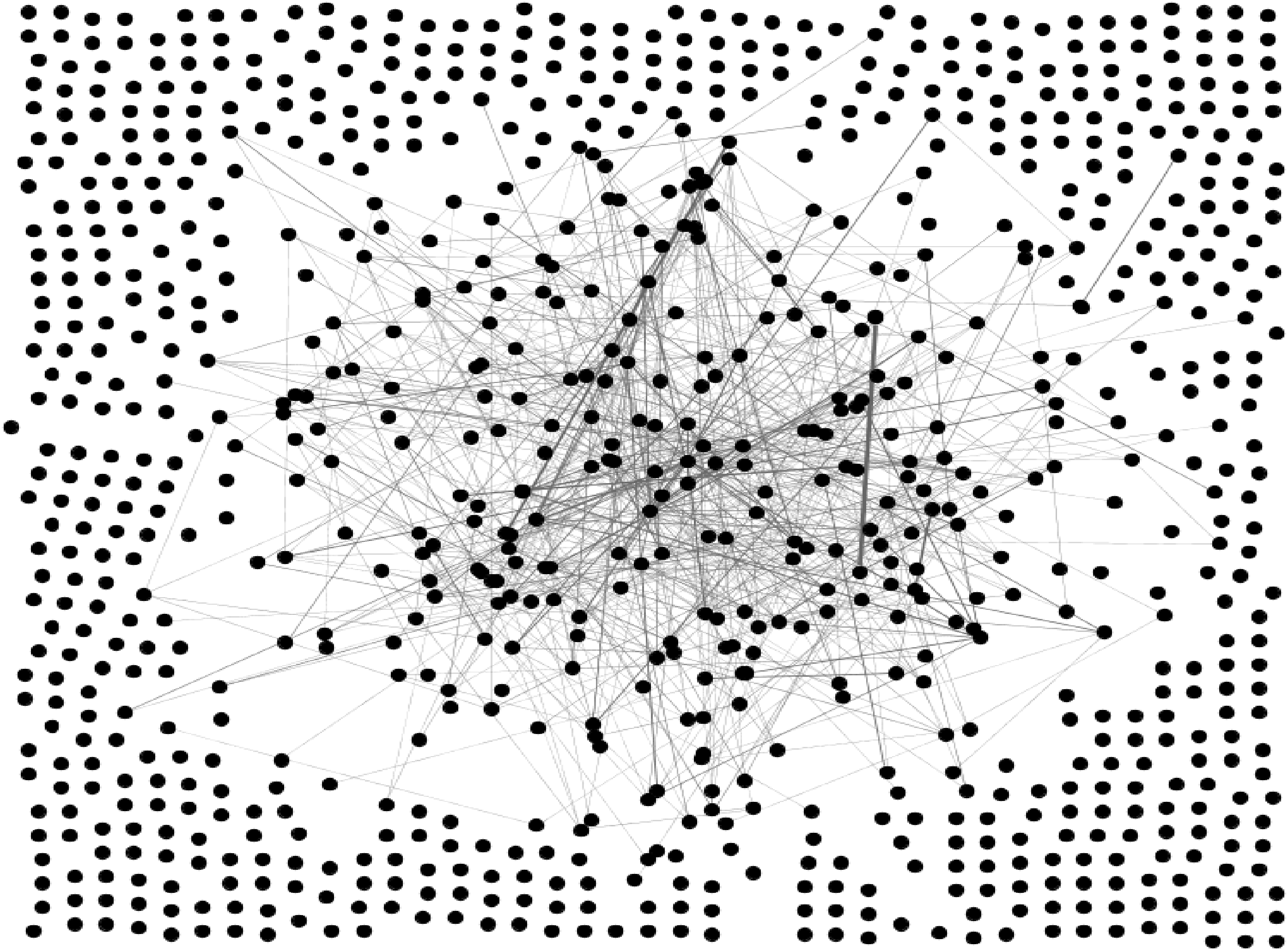
- Introduction: Identity by Descent
- A model for IBD sharing and demography
  - Migration
- Examples
  - Netherlands
  - Ashkenazi Jews
- IBD and sequence data

1000 New Yorkers (EUR)



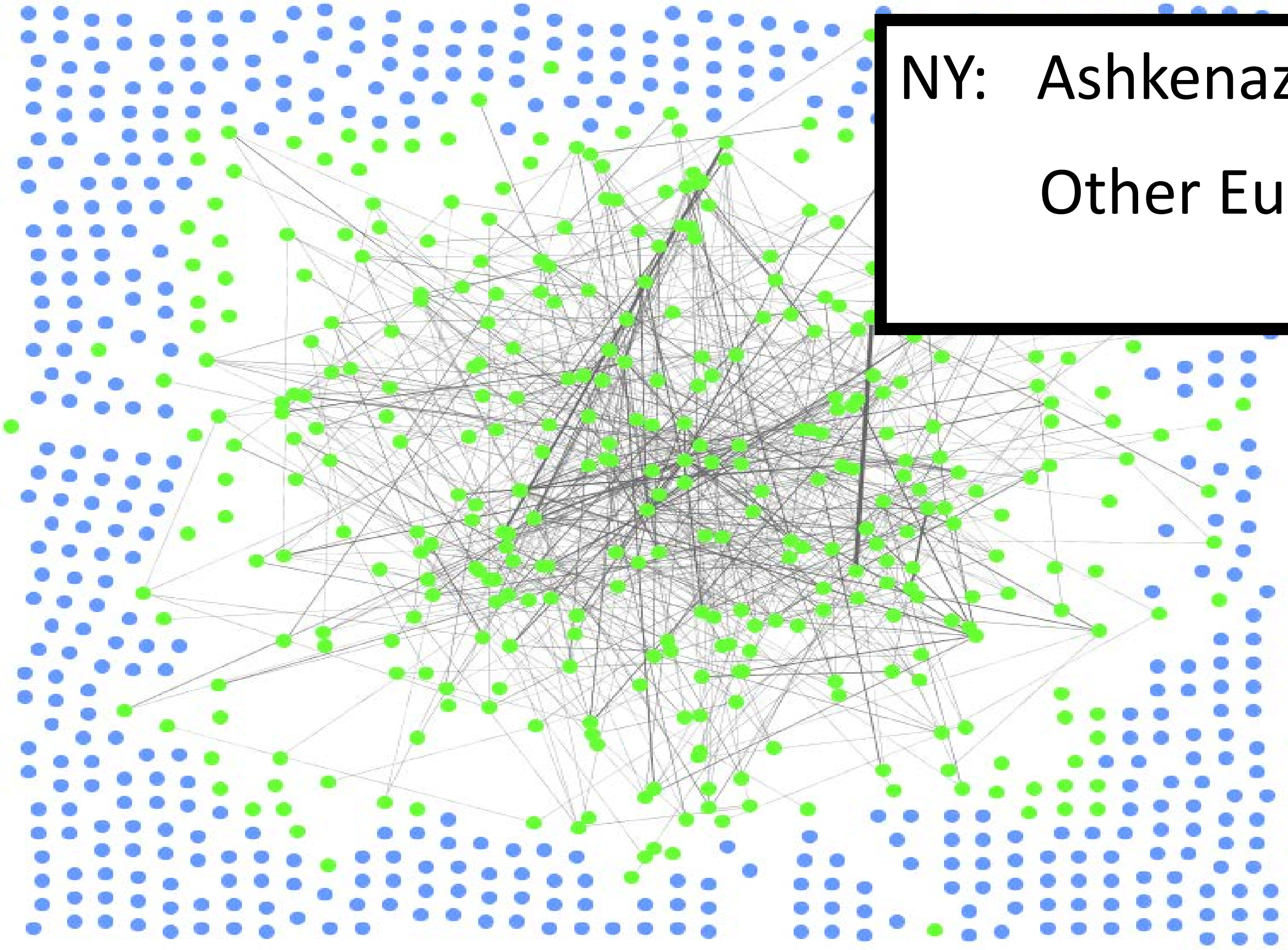
IntragenDB
Gregersen

Hidden Relatedness Detected



IntragenDB
Gregersen

Relatedness vs. Population

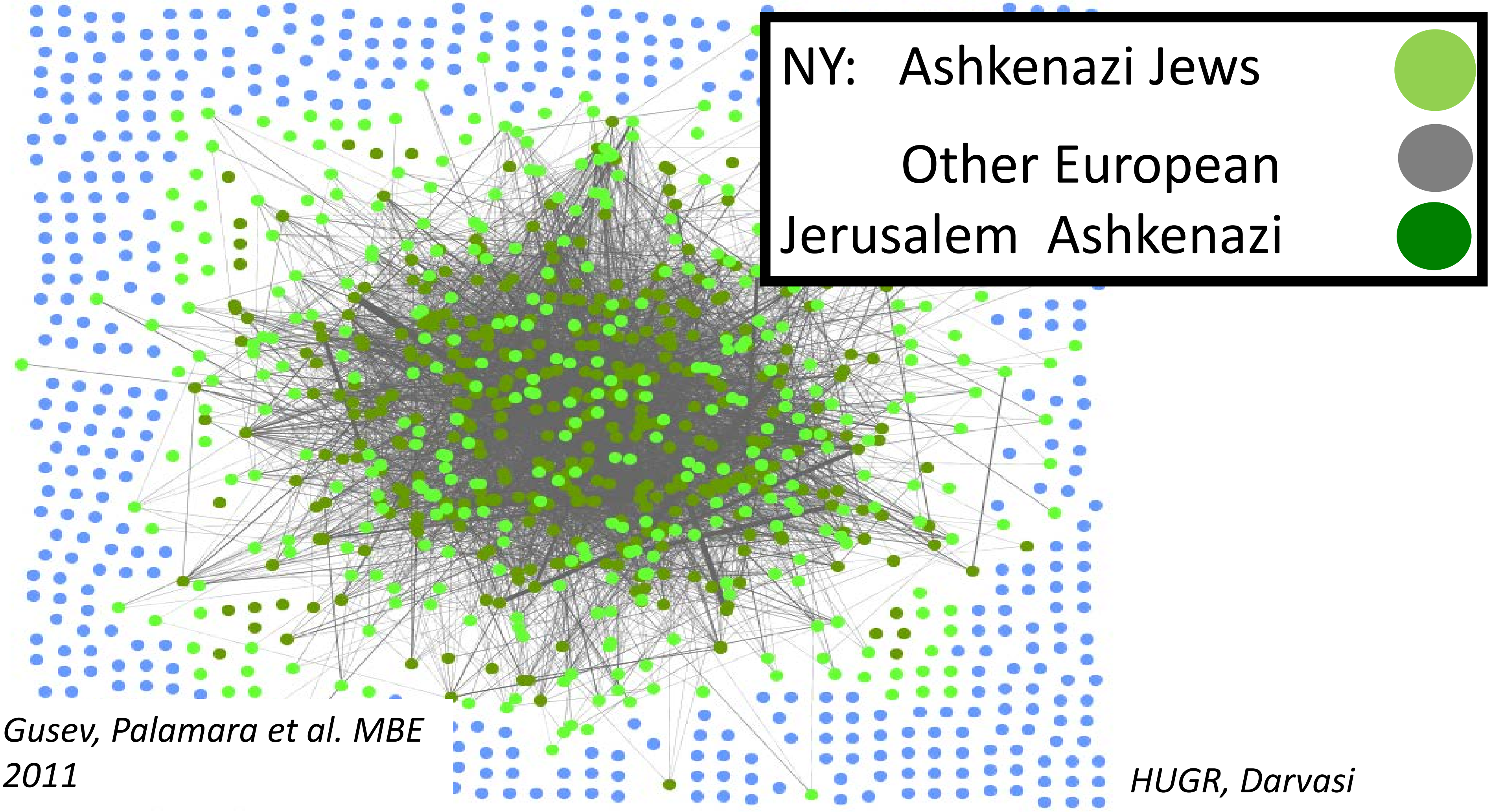


NY: Ashkenazi Jews ●

Other European ●

IntragenDB
Gregersen

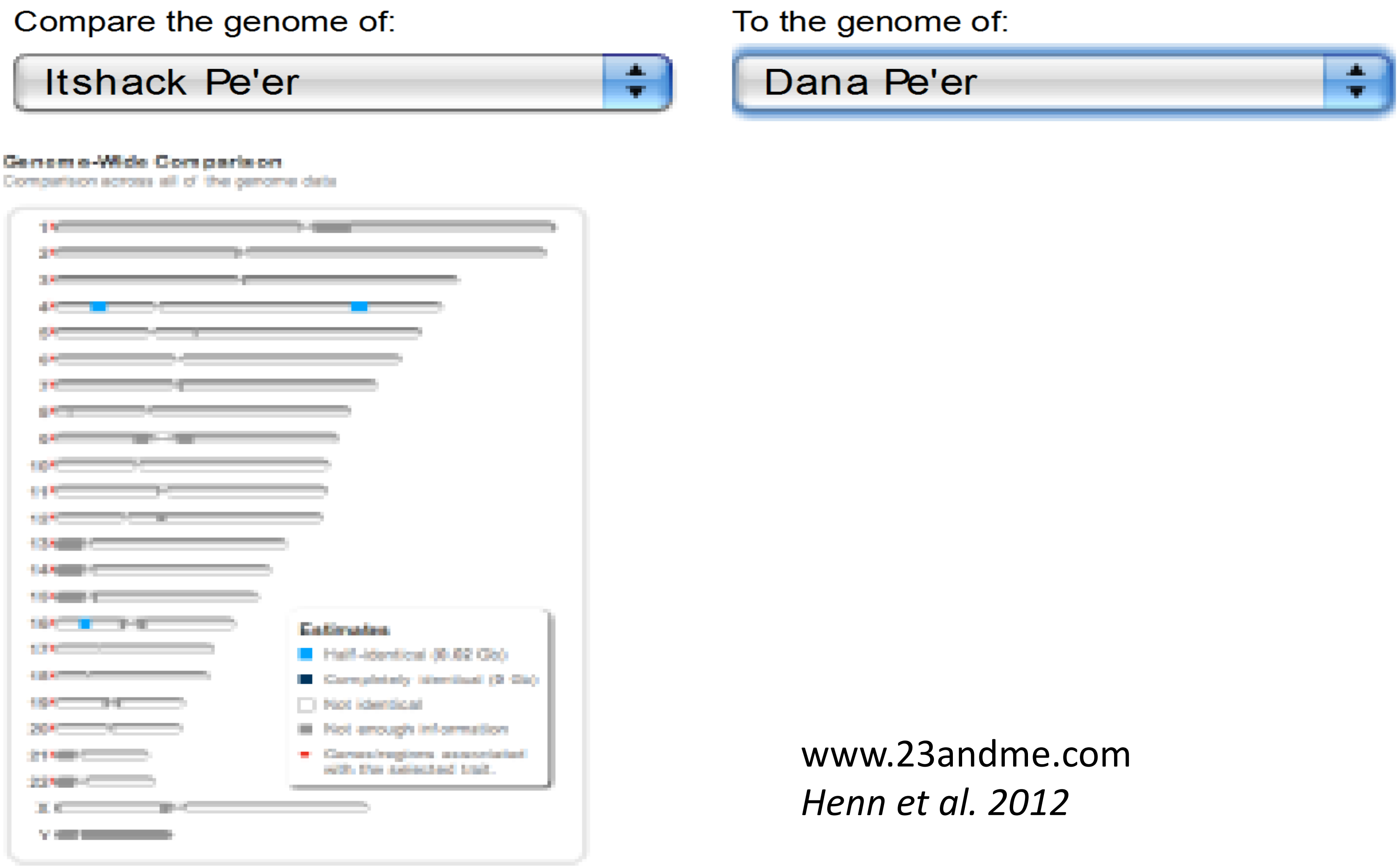
Relatedness vs. Population



Gusev, Palamara et al. MBE 2011

HUGR, Darvasi

Direct To Consumer: Ashkenazi Genetics



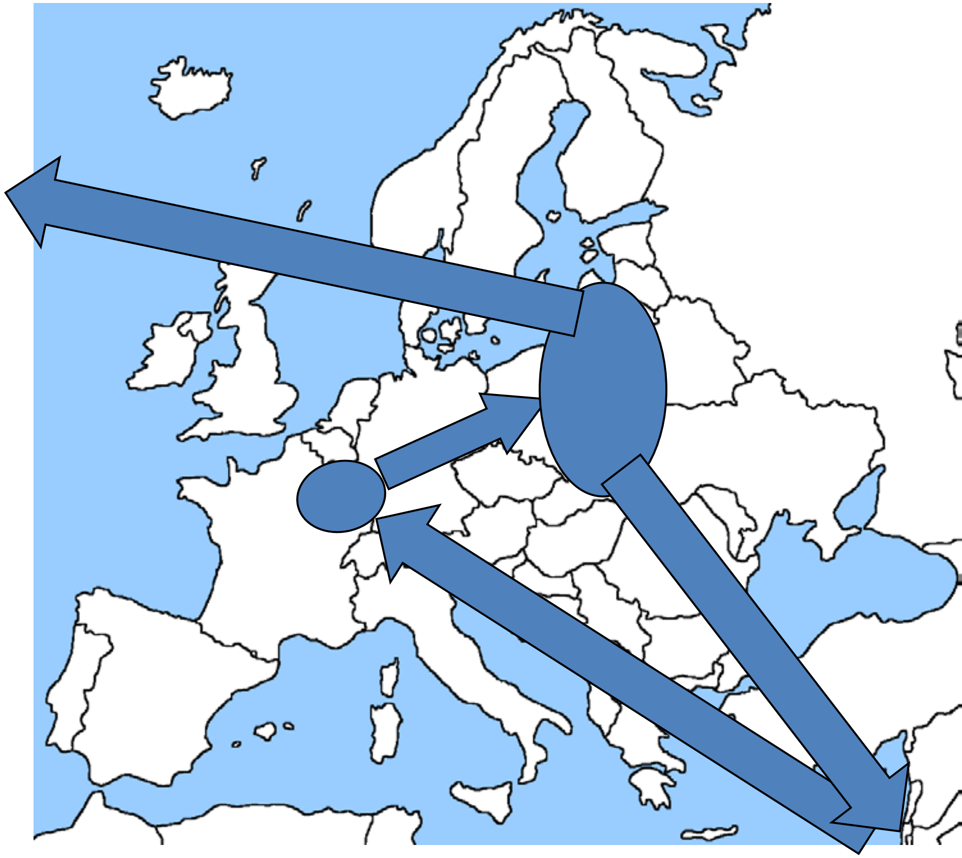
Compare the genome of: Itshack Pe'er

To the genome of: Dana Pe'er

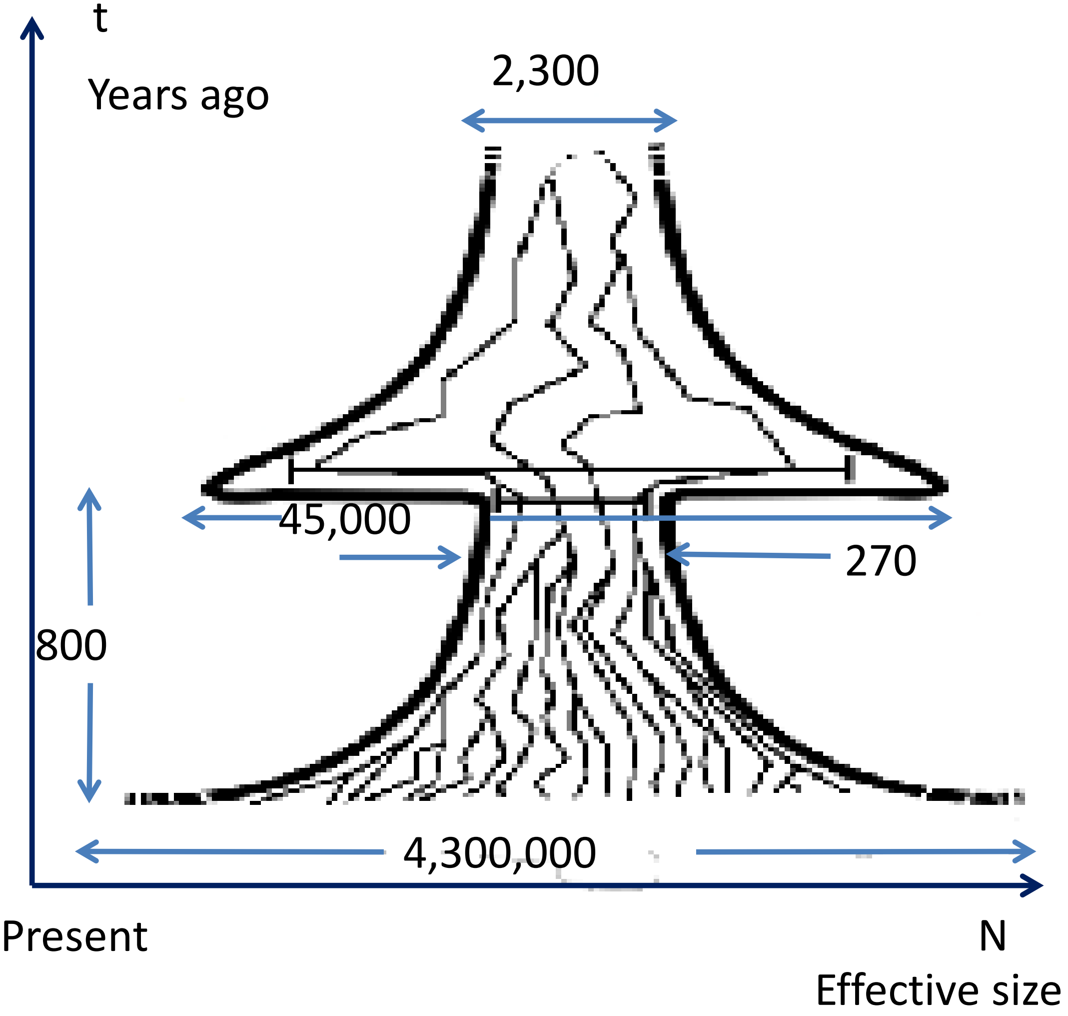
www.23andme.com
Henn et al. 2012

Ashkenazi History

- Mediterranean origin (?)
- 1st millennium:
  Small communities in
  Northern France, Rhineland
- Migration east(?)
- Expansion, relative isolation
- ≈13M pre-war
- Migration to US and Israel
- ≈10M today



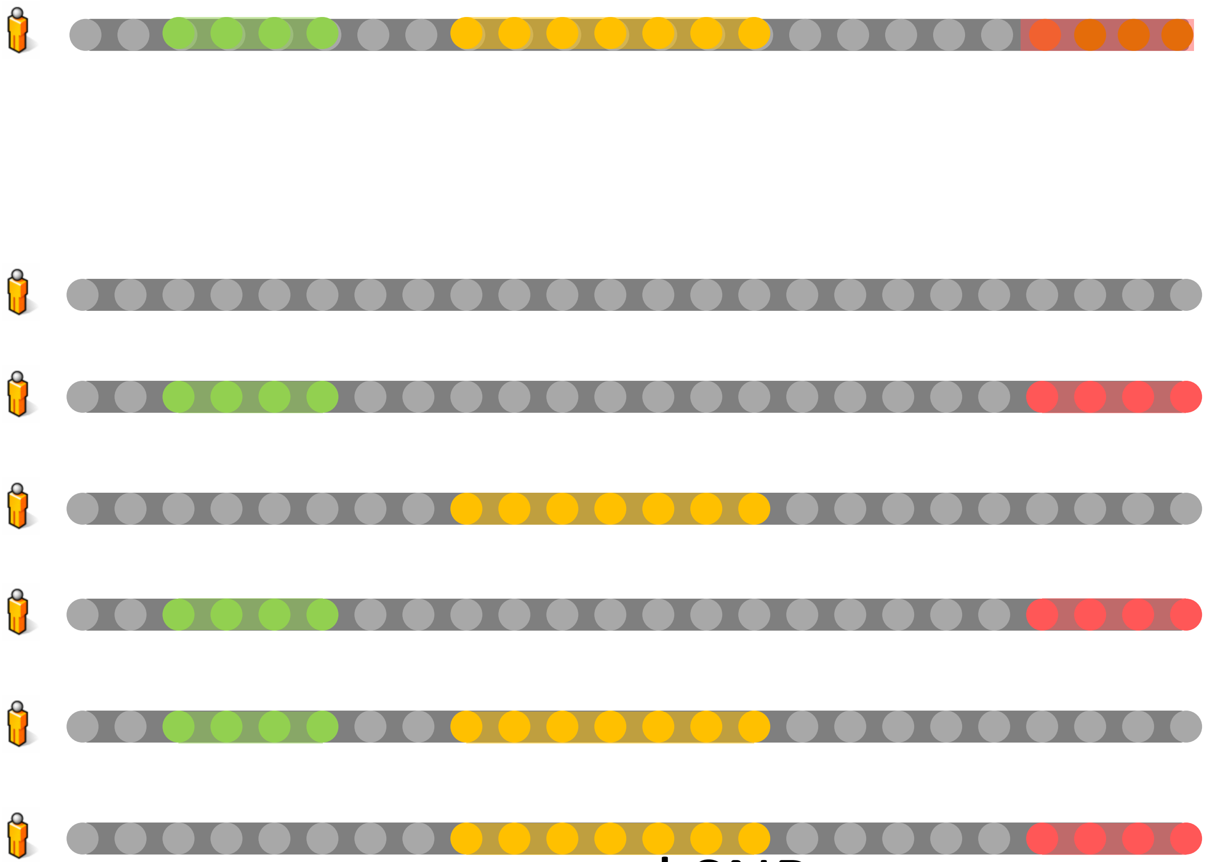
AJ Genetics



Palamara et al., AJHG 2012

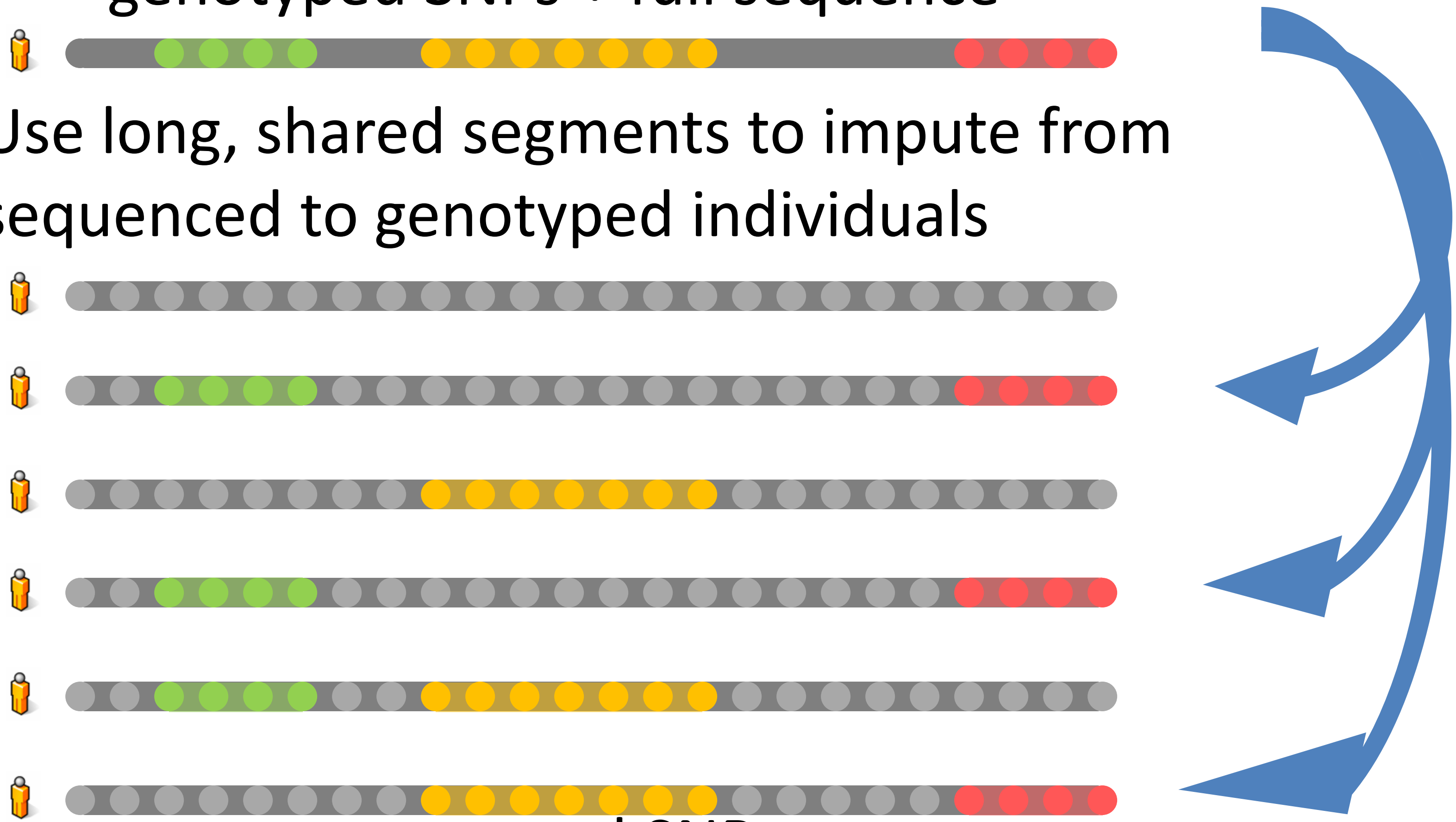
Idea: Imputation by Sharing

Identify shared segments:



genotyped SNPs

Idea: Imputation by Sharing

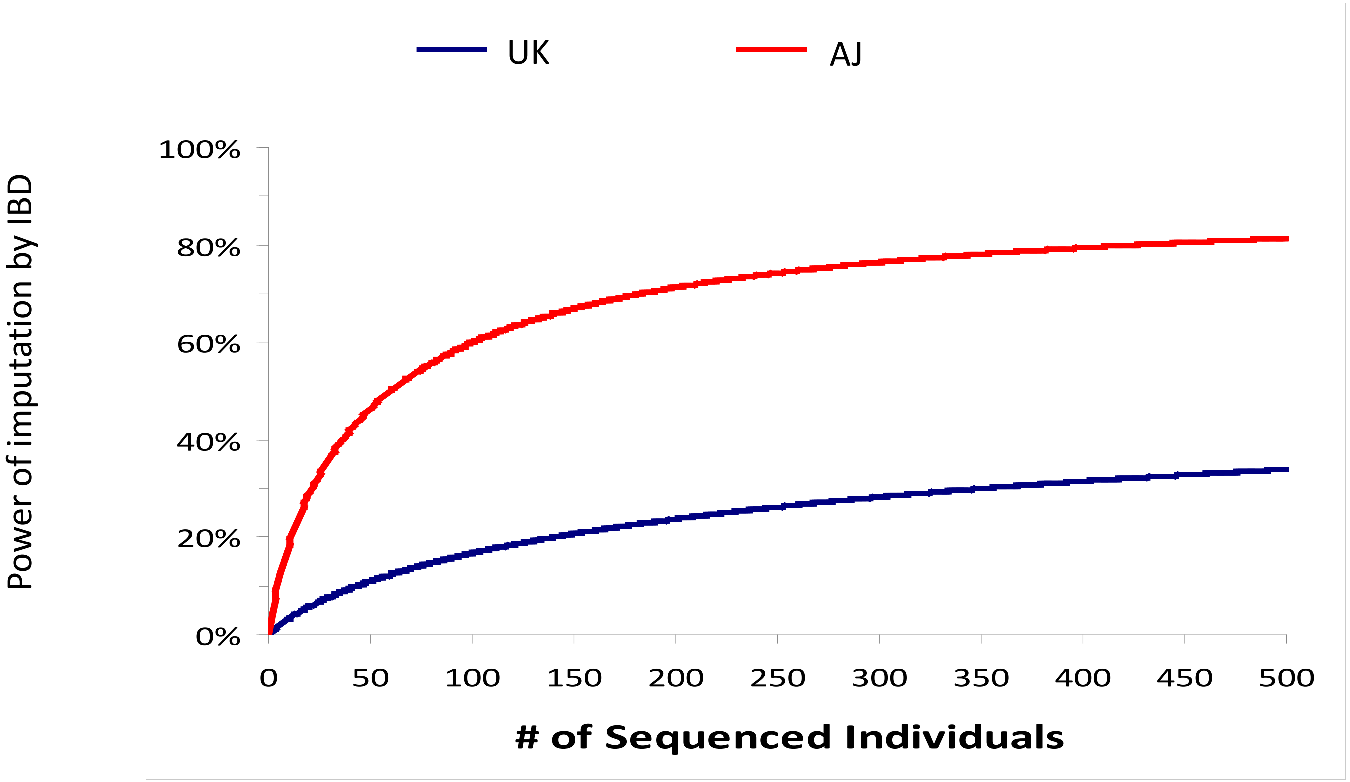


genotyped SNPs + full sequence

Use long, shared segments to impute from sequenced to genotyped individuals

genotyped SNPs

Imputation by IBD



Palamara et al., AJHG 2012

The Ashkenazi Genome Consortium

Phase I:

- 128 healthy Ashkenazi (Carmi et al. ‘14)
- Complete Genomics sequencing

Phase II: +574 Illumina (cases+healthy)

The Hebrew University of Jerusalem

Beth Israel
Continuum Health Partners, Inc

Memorial Sloan-Kettering Cancer Center

Yale University School of Medicine

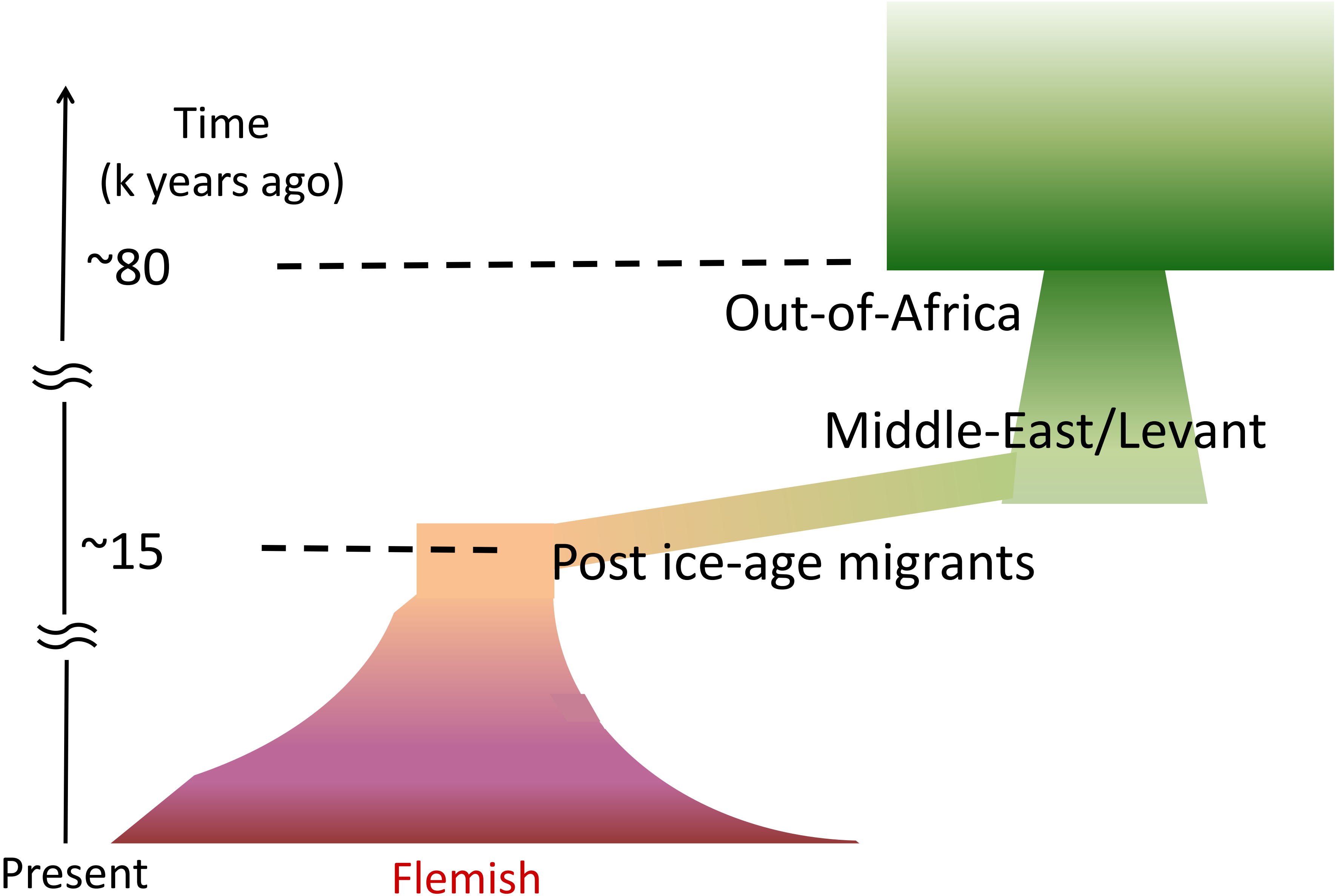
Columbia University Medical Center

The Feinstein Institute for Medical Research
North Shore LIJ
North Shore-Long Island Jewish Health System

EINSTEIN
Albert Einstein College of Medicine
OF YESHIVA UNIVERSITY

Icahn School of Medicine at Mount Sinai

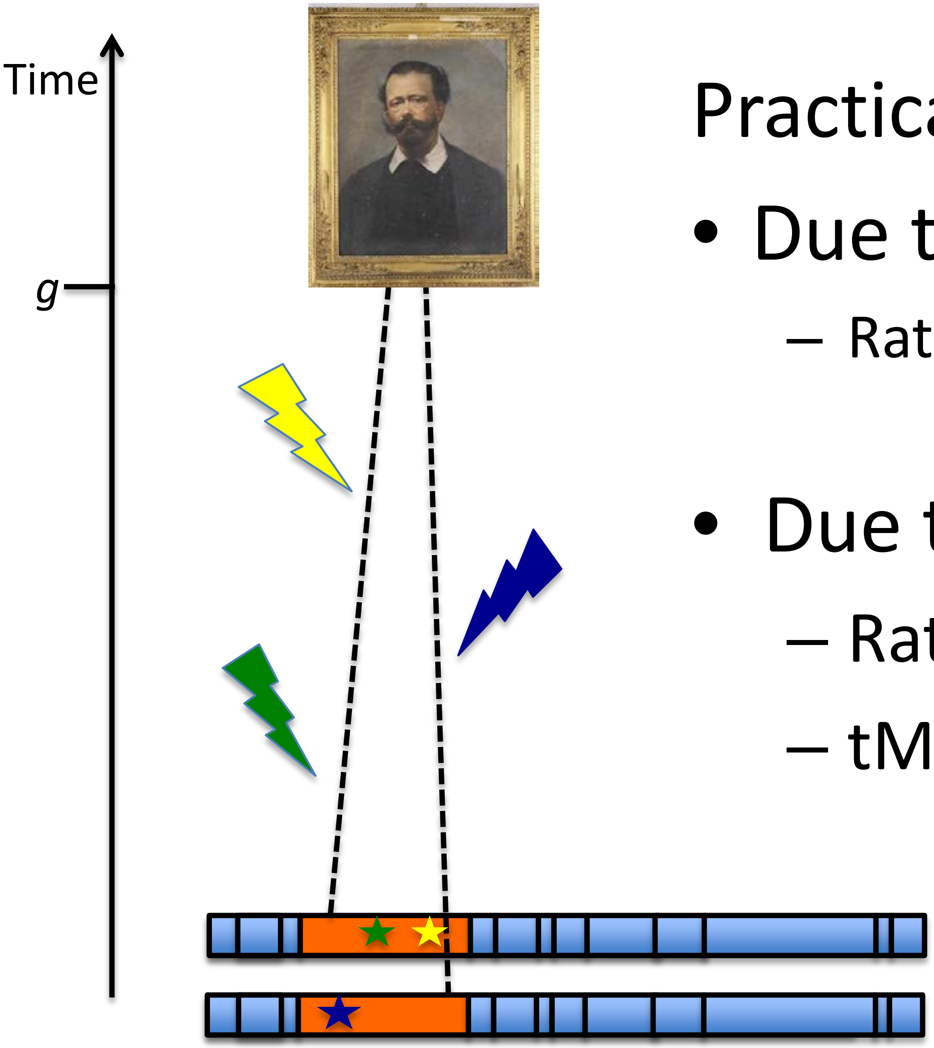
The Inferred Model



Outline

- Introduction: identity by Descent
- A model for IBD sharing and demography
- Examples
- IBD and sequence data

Near Identity-By-Descent (IBD)



Practically IBD sequences differ

- Due to sequencing errors
  - Rate per bp
- Due to true events
  - Rate per bp, per generation
  - tMRCA estimated from length[1]

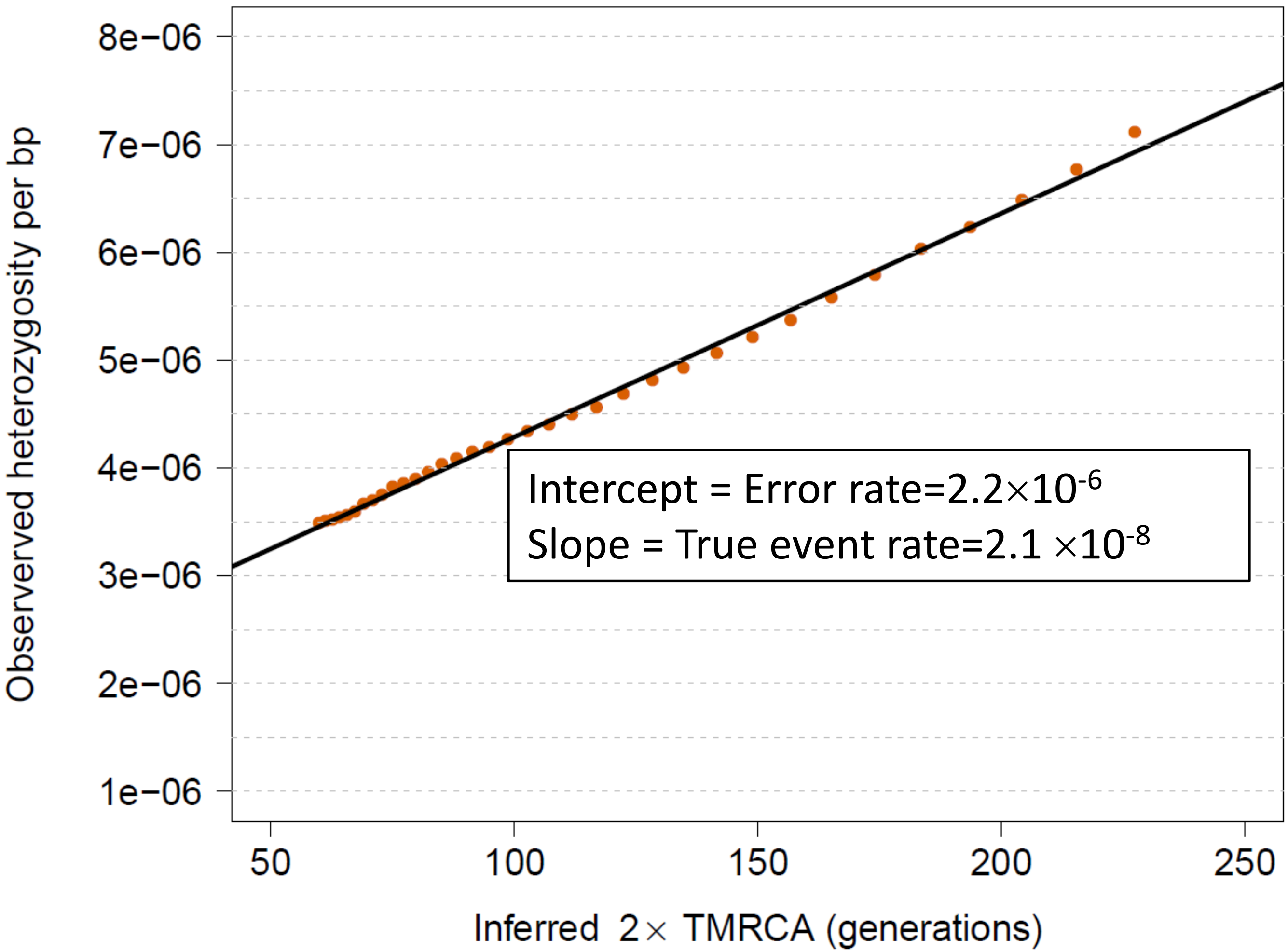
---

[1] Palamara et al. ‘12

Simulated Data

Intercept = Error rate
Slope = True event rate

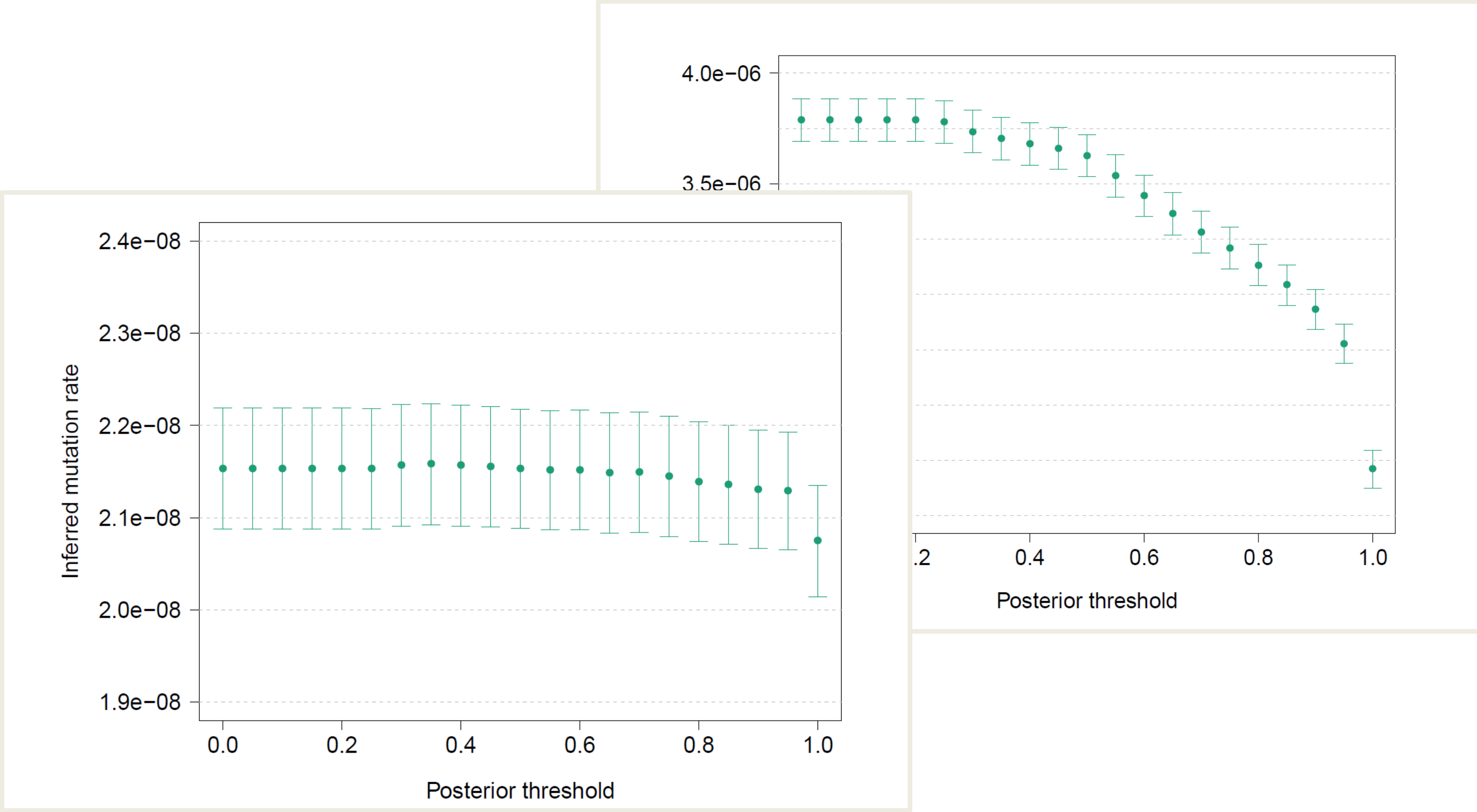
Real Data (GONL)



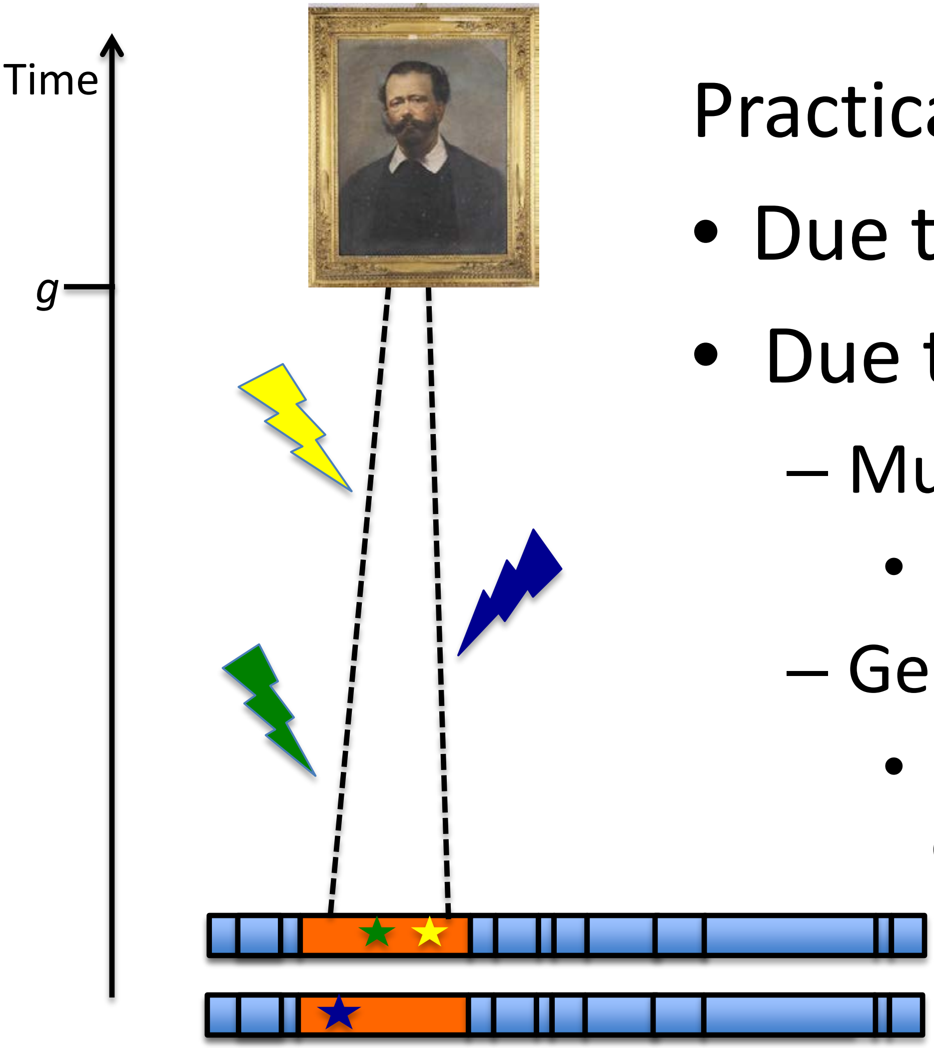
Intercept = Error rate=2.2×10^{-6}

Slope = True event rate=2.1×10^{-8}

Real Data (GONL)



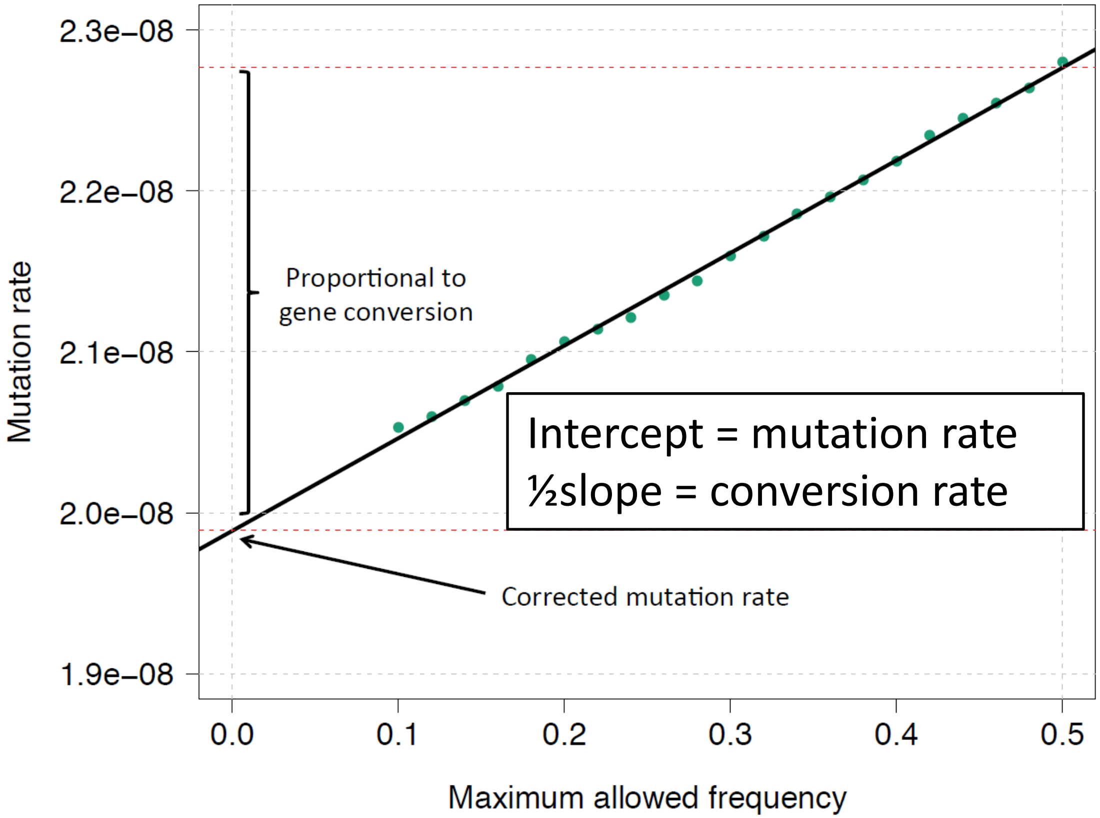
Near Identity-By-Descent (IBD)



Practically IBD sequences differ

- Due to sequencing errors
- Due to true events
  - Mutations
    - New variants
  - Gene conversions
    - Reflect frequency spectrum of single-sample heterozygotes

Simulated Data

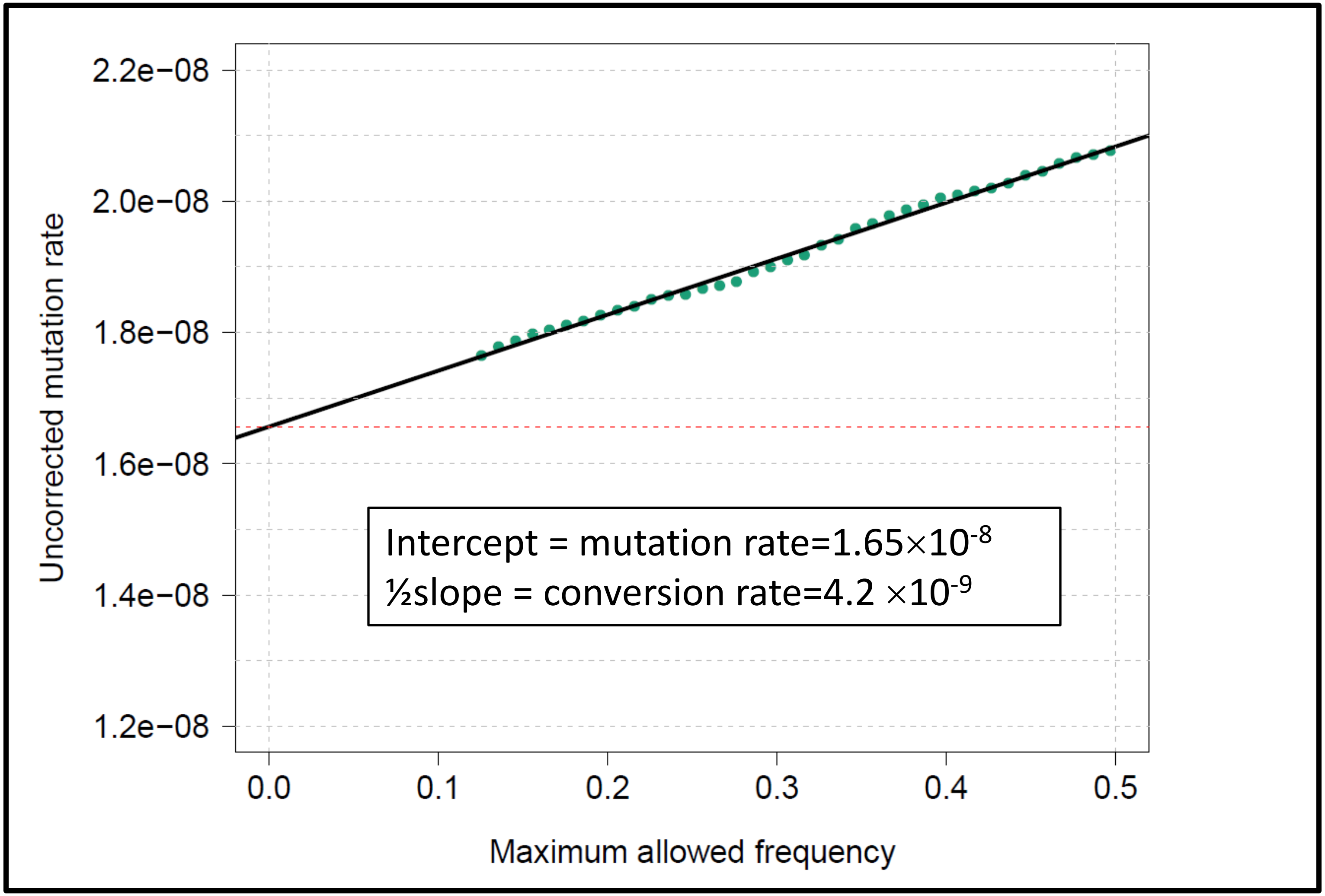


Proportional to gene conversion

Intercept = mutation rate
½slope = conversion rate

Corrected mutation rate

Real data (GONL)



Intercept = mutation rate=1.65×10^{-8}
½slope = conversion rate=4.2×10^{-9}

Thanks

Pier Palamara

GoNL

- Laurent Francioli
- Sarah Pulit
- Clara Elbers
- Androniki Menelaou
- Freerk van Dijk
- Morris Swertz
- Shamil Sunyaev
- Cisca Wijmenga
- Paul deBakker

Thank you!

Columbia University Computer Science:

Shai Carmi

Cameron Palmer

Fillan Grady,

Ethan Kochav,

James Xue

Shlomo Hershkop

Long-Island Jewish Medical Center:
Todd Lencz, Jin Yu, Semanti Mukherjee, Saurav Guha

Albert Einstein College of Medicine:
Gil Atzmon, Harry Ostrer, Nir Barzilai,
Upadhyay, Danny Ben-Avraham

Columbia University Medical Center:
Lorraine Clark, Xinmin Liu

Memorial Sloan Kettering Cancer Center:
Ken Offit, Joseph Vijai

Mount Sinai School of Medicine:
Inga Peter, Laurie Ozelius

Yale School of Medicine:
Judy Cho, Ken Hui, Monica Bowen

Beth Israel:
Susan Bressman

The Hebrew University of Jerusalem:
Ariel Darvasi

VIB, Gent, Belgium
Herwig Van Marck
Stephane Plaisance

Complete Genomics

Omicia

NYGenome

NATIONAL INSTITUTES OF HEALTH

HFSP

NSF