Human Paleogenetics

Laurent Excoffier
Mathias Currat
Nicolas Ray

CMPG
Computational and Molecular Population Genetics Lab

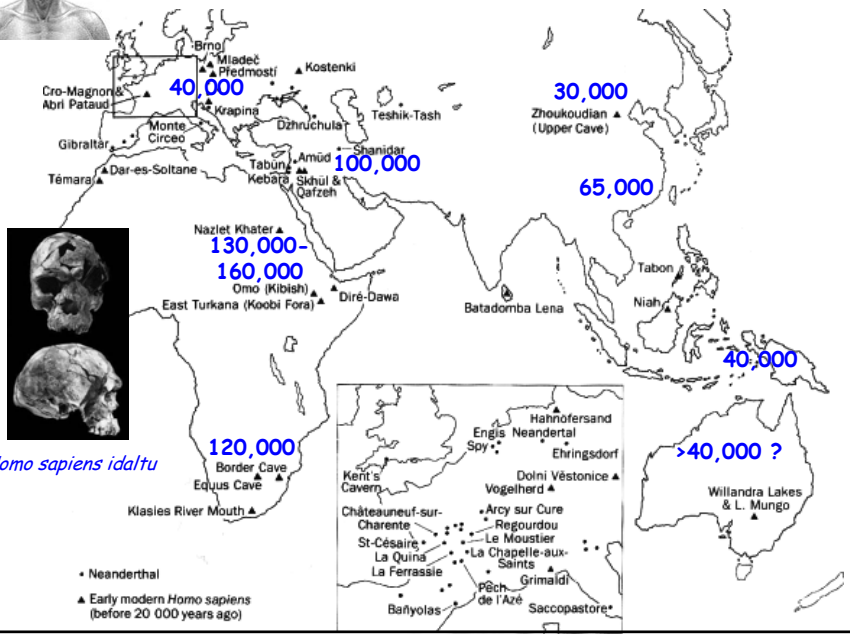
Zoological Institute
University of Bern, Switzerland

Paris, June 2005

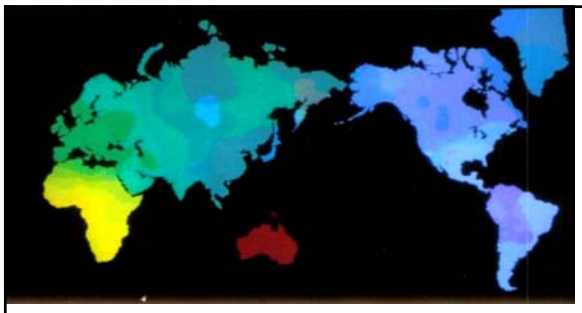
Outline

- Some facts about human genetic diversity
- Different scenarios of human evolution
- The transition between Neanderthals and modern humans
- Ability to distinguish between different scenarios of human evolution, and finding the geographic origin of modern humans

Early modern humans sites



Homo sapiens idaltu

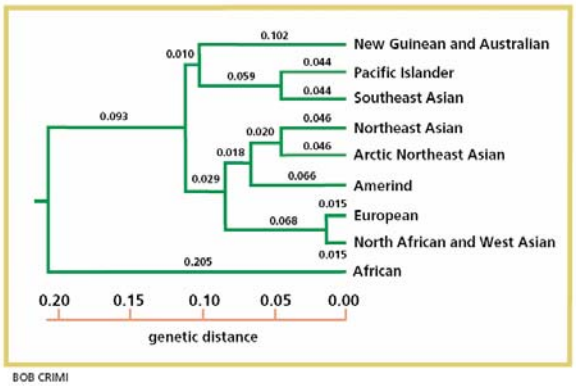


Non-molecular diversity

Cavalli-Sforza et al. 1994

Cavalli-Sforza and Feldman 2003

120 alleles
(blood groups, immunological and protein markers)



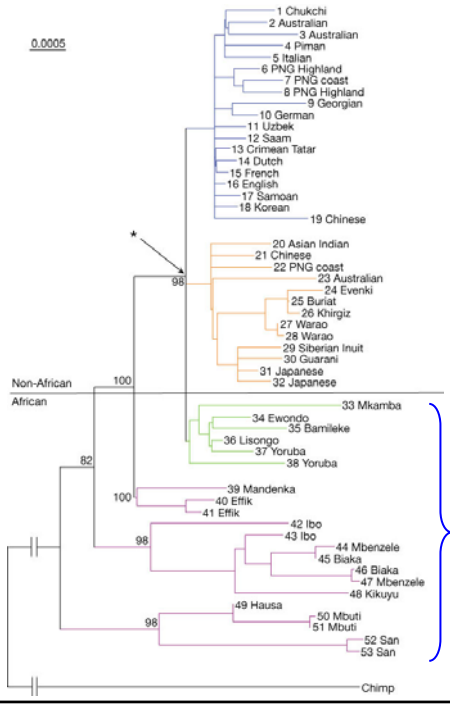
BOB CRIMI

Extent of molecular diversity on different continents

Locus	Marker system	Africa (*N*)	Europe (*N*)	Asia (*N*)	Highest genetic diversity*
MtDNA	HVSI	72	120	63	Africa
MtDNA	HVSII	72	120	63	Africa
NRY	6 STRPs	72	120	63	Africa
NRY	43 biallelic	360	507	1415	Africa‡
Autosome	60 STRPs	72	120	63	Africa
Autosome	45 STRPs	216	246	302	Africa
Autosome	30 RFLPs	72	120	63	Europe
Autosome	13 Alus	72	120	63	Africa
Autosome	8 Alus	176	334	359	Africa
PLAT	STRP1/Alu	924	352	386	Africa
PLAT	STRP2/Alu	1030	410	422	Africa

Tishkoff and Williams 2003

Analysis of 53 complete mtDNA sequences



Global T_{MRCA} = 171.5 KY +-50 KY

T_{MRCA} * = 52 KY +-27.5 KY

Africa

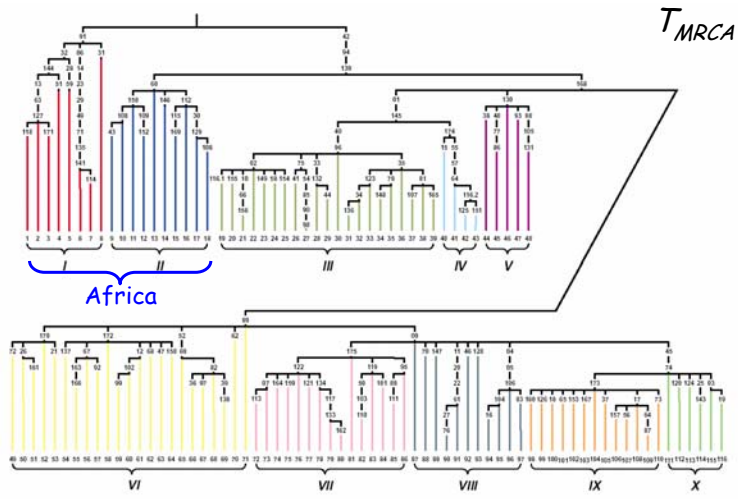
Ingman et al. 2000

Y chromosome worldwide diversity

1009 men

166 SNPs +YAP

T_{MRCA} = 59 Ky



Underhill et al. 2000

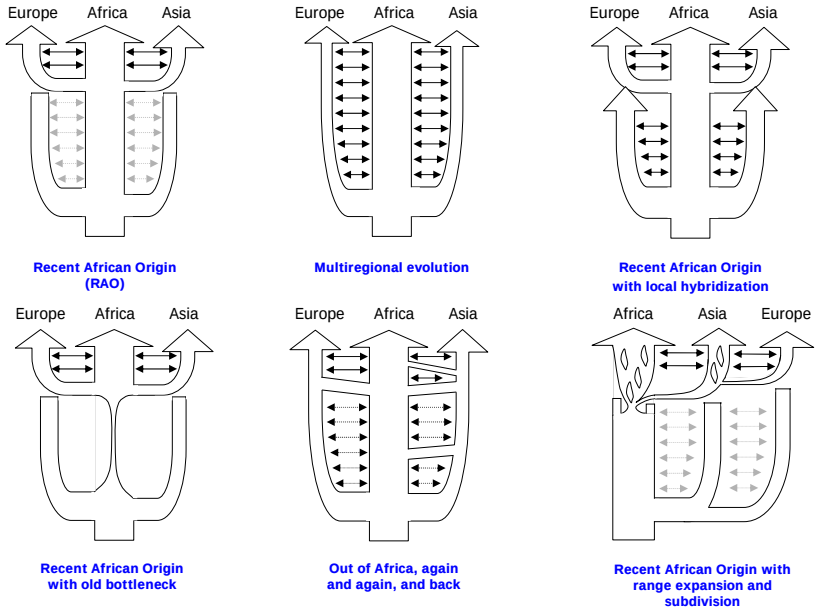
Age and location of ancestral sequences

Summary of human sequence studies.

Location	Gene	Length (Kb)	d*	n†	T_{MRCA}‡ KY	Earliest branching lineage in Africa only
mtDNA	*Whole genome*	16.5	–	53	170	Yes
mtDNA	*HVR1* non-coding	0.4	62	2778	–	Yes
mtDNA	Whole coding genome	15.5	5	179	240	–
Yp11.3	ZFY	0.7	15	205		–
NRY	*SMCY*	40	5	53	41-68	Yes
NRY	*DBY*	9	5	70	39-100	Yes
NRY	*DFFRY*	15	5	70	40-65	Yes
NRY	Not known	69	5	108	89	–
Xp11.4-3	*MAO-A* (five segments)	18.8	8	56		No
Xp21.3	*Gk* intron 1	1.9	8	10	410	No
Xp21.3	*ZFX* intron	1.1	15	336	1090	Yes
Xp22.2-1	*Pdha1* introns 9 and 10	1.7	8	10	1050	Yes
Xp22.2-1	*Pdha1* exons 7 to 10	4.2	8	35	1780	Yes
Xq13.3	Non-coding region	10	–	69	540	Yes
Xq21	*Dmd* intron 7	2.3	4	41	210	–
Xq21	*Dmd* intron 44	1.4	8	10	1350	Yes
Xq21	*Dmd* intron 44	3	4	41	1560	Yes
Xq21-22	*Plp* intron 5	0.7	8	10	1280	Yes
Xq26.1	*Hprt* introns 2 and 8	2.7	8	10	530	Yes
Xq27.2-1	*FIX* intron 4	3.7	11	36	282	No
1q24	Mostly introns	10	3	61	1376	Yes
1q21	*psGBA*	5.4	12	100	91–196	Yes
8p22	*LPL*	10	3	71	–	–
11p15.5	*β-globin*	3	9	326**	800	Yes
14q24	*EDN*	1.2	4	67	1150	–
14q31	*ECP*	1.2	4	54	1090	–
16q24.3	*MC1R*	0.95	16	672**	1000	Yes
16q24.3	*MC1R* promoter	6.7	3	54	1520	Yes
16p13.3	5′ region *MS205* intron	11.7	5	50	1040	Yes
17q23	*ACE*	24	2	11	1113	–
19q13.2	*APOE*	5.5	4	96	311	–
22q11.2	Non-coding	9.9	16	64	1288	…

Excoffier 2002

Alternative models of human evolution

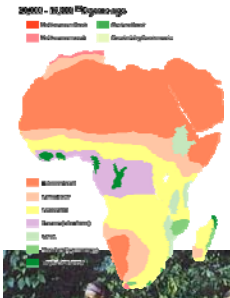


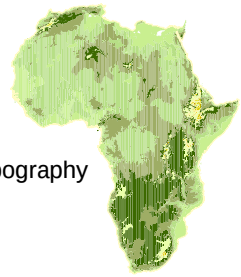
Use more realistic models of human evolution

- Spatially explicit models
- Take environmental information into account
- Model interactions and potential competition between populations

Environmental variables affect migrations and demography

Current and past vegetation



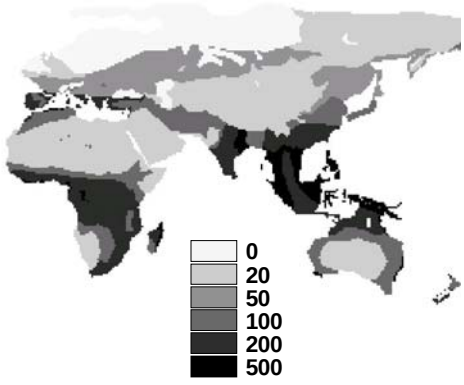


Topography

Hydrography and coastlines

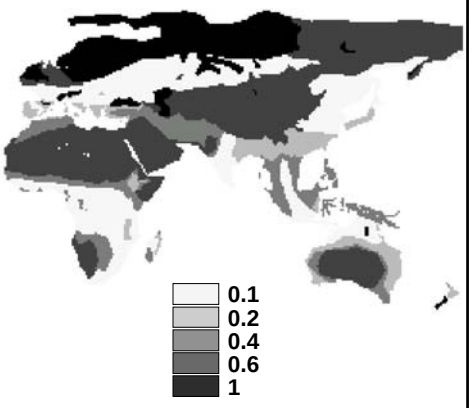
Environmental information can be translated into:

Carrying capacity

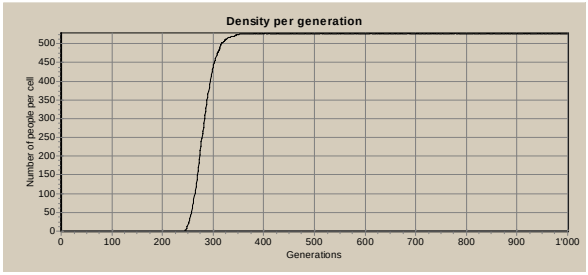


individuals / 10,000 km²

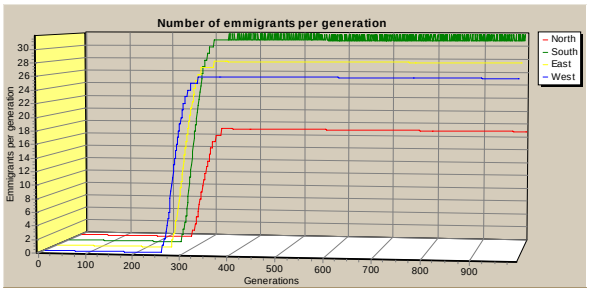
Relative Friction



Demographic simulations



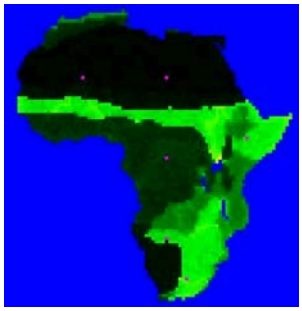
Demographic data base

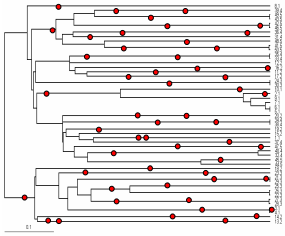


Simulating genetic data

Demographic data base

Migration rates population densities





Backward in time

Observed genetic data at the same sample location

Comparison through summary statistics

Simulated genetic data

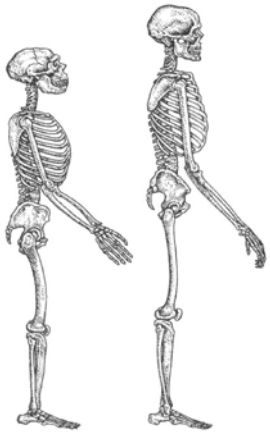
Inference, parameter estimation

Neanderthal replacement in Europe

- Successors of *H. erectus*
- Evolution over more than 400,000 years
- Final morphology around 120,000 BP

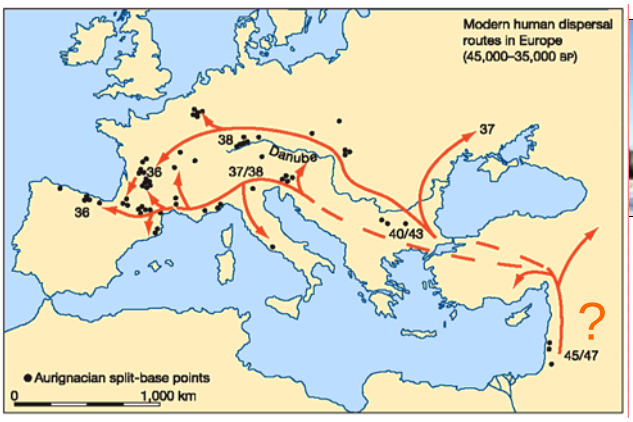
Neanderthal Modern

Klein, 2003



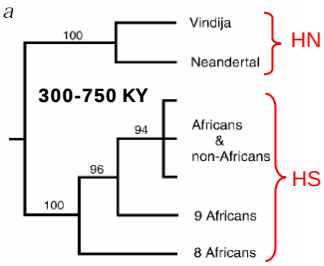
Expansion of modern humans into Europe

- Arrival in Europe around 45-30 Ky BP
- East to West Colonization
- Originated from Near-East?
- Simultaneous retreat et disappearance of Neanderthals

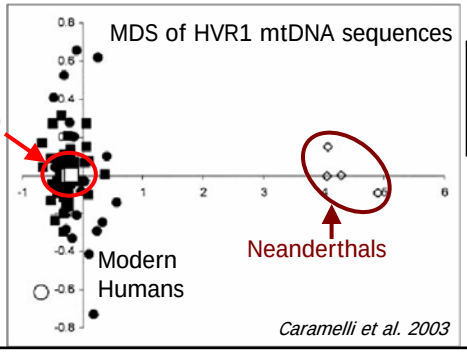


Mellars, 2004

Neanderthal genetic remains



Krings et al. 2000



Caramelli et al. 2003

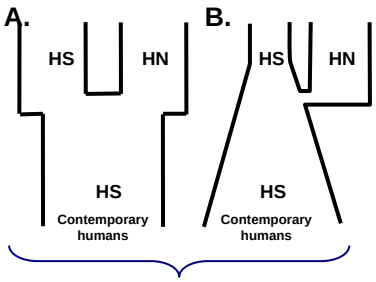
Neanderthal mtDNA is very different from modern mtDNA

Confirmed by Serre et al. (2004) with the addition of 4 Neanderthals & 5 "Cro-Magnon" sequences

Estimations of hybridization rate between Neanderthals and modern humans

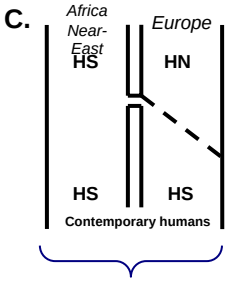
Observation: Total absence of Neanderthal sequences in modern humans

Compatible with up to 25% Neanderthal initial introgression into modern gene pool under a simple demographic scenario



Previous models :
- Instantaneous admixture
- Unsubdivided populations

Nordborg, 1998 ; Serre et al. 2004

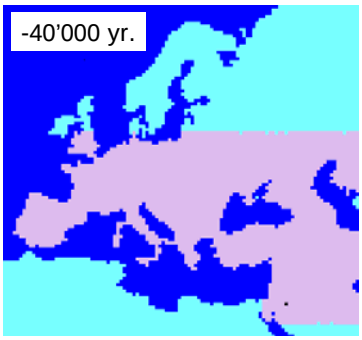


Our model :
- Spatial expansion
- Progressive hybridization
- Subdivided population

Currat & Excoffier 2004

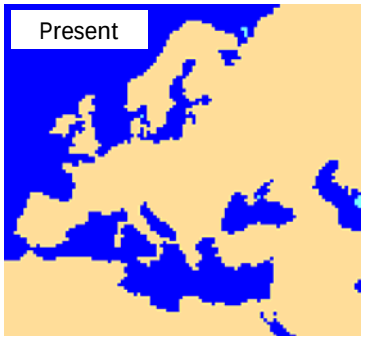
Simulation conditions

Neanderthal range



Uniform Environment

Human range

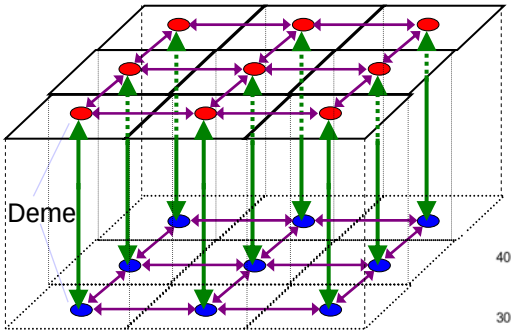


3,500 demes
K=10-25 females
Density=0.015-0.03 ind./km^2

7,250 demes
K=40 females
Density=0.06 ind./km^2

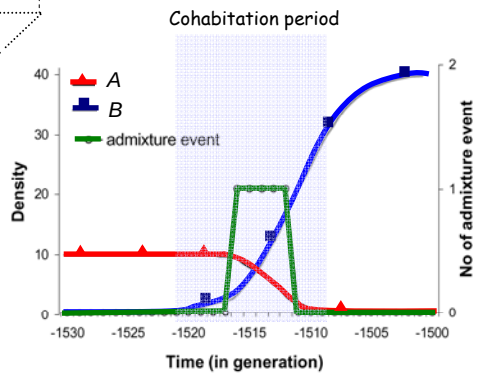
Simulating colonization and interaction

Population *A*



Population *B*

Each generation:
1. Hybridization (admixture)
2. Logistic regulation (including density-dependent competition)
3. Migrations

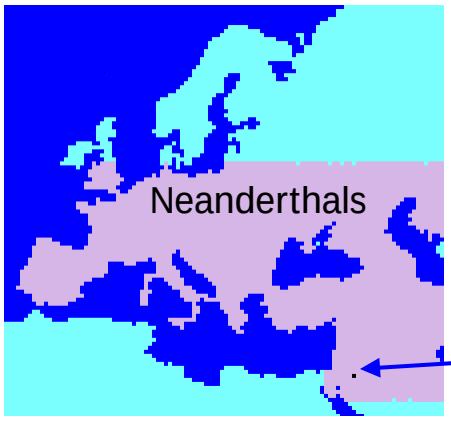


1- Demographic simulation

1'600 generations ago (~ 40'000 years)

Neanderthals

Modern H.

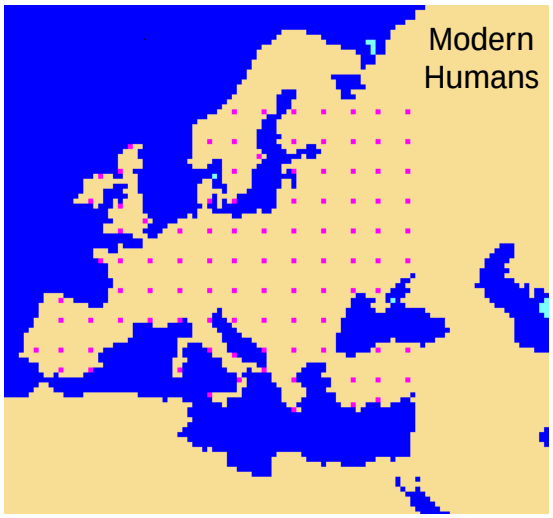


Unoccupied

Cohabitation

2 - Genetic simulation

Present: 4,000 mtDNA sequences in 100 demes (100 samples of 40 genes)



Local admixture rate: 0.01

Unoccupied

Neanderthals

Modern HS

Cohabitation

Modern genes

Neanderthal genes

Genetic simulations used to estimate admixture proportions

Data

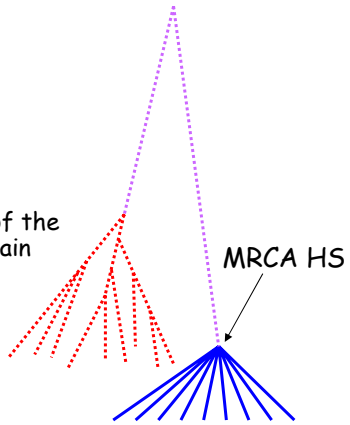
>4000 monophyletic modern human sequences

Simulation

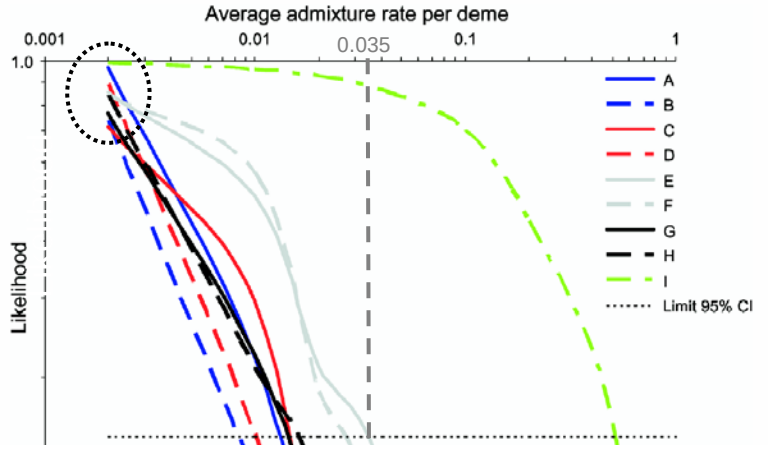
Genealogy of 100 samples of 40 genes distributed uniformly over Europe, for different demographic scenarios and different admixture proportions

Estimation of the likelihood

Proportion of simulations for which the genealogy of the 4000 sampled genes is monophyletic (does not contain Neanderthal genes)



Upper limit of Neanderthal contribution



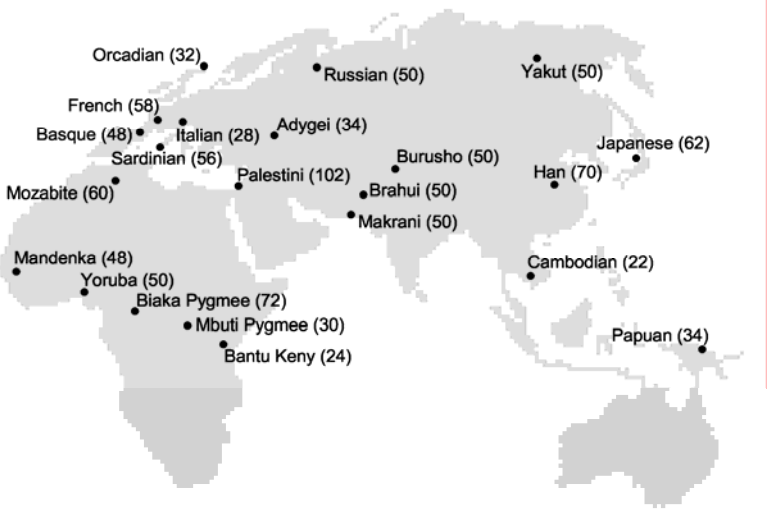
Total number of hybridization:	120	1863
Neanderthal initial contribution to modern gene pool :	0.09%	1.33%
	400x smaller	20x smaller

First conclusions

- Absence or very low levels of hybridization between Neanderthal females and modern men
- Implies sterility or lower fertility of hybrids if mtDNA is neutral
- Support for the Recent African Origin model
- Same phenomenon would be expected for interaction between *H. erectus* and *H. sapiens* in Asia.
- Does not completely exclude the possibility of gene flow through male Neanderthals

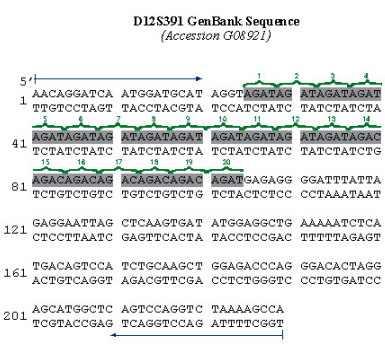
⇒ Need to look at nuclear markers

STR data set
377 STR loci in 22 populations

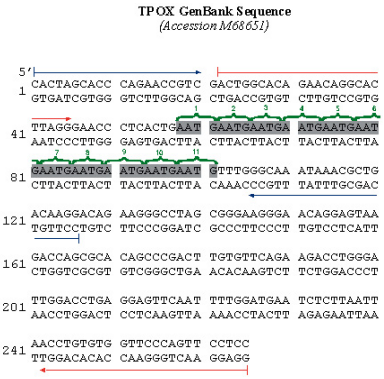


Rosenberg et al. 2002

Short Tandem Repeats (STRs)

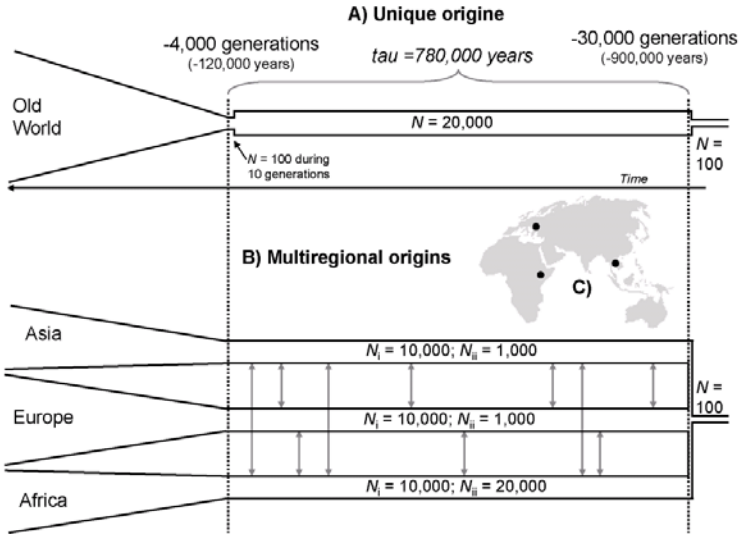


$(AGAT)_n$

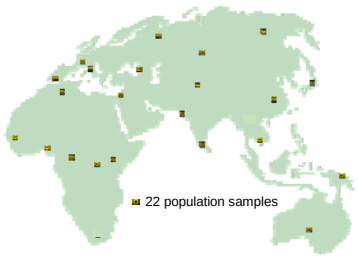


$(AATG)_n$

Tested scenarios of human evolution



25 potential geographic origins



25 simulated origins

Multiregional scenarios

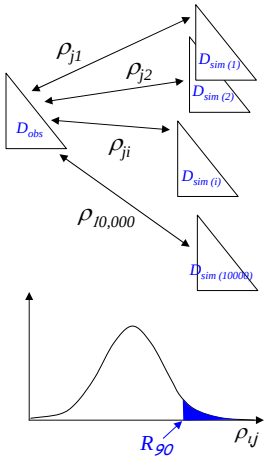
Table 1: Description of the nine ME scenarios (numbered 26 to 34).

		N			*m* ($\times 10^4$)		
Scenario#		*Africa*	*Europe*	*Asia*	*Africa*	*Europe/Asia*	
26	Nm = 0.1 (all regions)	7,000	7,000	7,000	0.143	0.143	Equal continental sizes, equal migration rates
27	Nm = 1 (all regions)	7,000	7,000	7,000	1.43	1.43	
28	Nm = 10 (all regions)	7,000	7,000	7,000	14.3	14.3	
29	Nm = 0.1 (all regions)	20,000	1,000	1,000	0.05	1	Equal migration rates between continents
30	Nm = 1 (all regions)	20,000	1,000	1,000	0.5	10	
31	Nm = 10 (all regions)	20,000	1,000	1,000	5	100	
32	Nm = 0.1 (in Africa)	20,000	1,000	1,000	0.05	0.05	Africa send more migrants than it receives
33	Nm = 1 (in Africa)	20,000	1,000	1,000	0.5	0.5	
34	Nm = 10 (in Africa)	20,000	1,000	1,000	5	5	

N is the effective number of genes in a continent.
m is the probability of emigration per gene per generation.
Nm is the number of emigrant genes per generation per continent.

Assignment scores for different scenarios

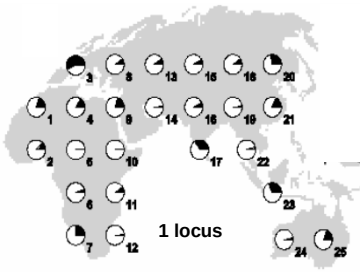
1. Compute observed genetic distances (F_{ST}) between all pairs of populations → D_{obs}
2. For geographic origin j, j = 1 ... 25
   1. Simulate 10,000 genetic data sets (1, 20, or 377 STR loci) and D_{sim} (i)
   2. Compute the correlations between the observed and the simulated genetic distance matrices ρ_{ji} = $corr(D_{obs}, D_{sim}\ (i))$
   3. From the distribution of ρ_{ji} , take the 90% quantile value (R_{90}) as the assignment score for the j-th origin
3. Select the evolutionary scenario with the largest assignment score (largest R_{90}), thus giving the best fit between observed and simulated data.

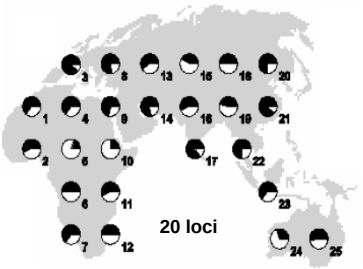


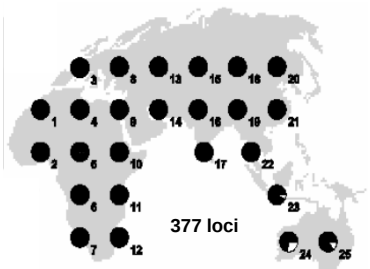
Ability to recover a scenario

Unique origin vs. multiregional

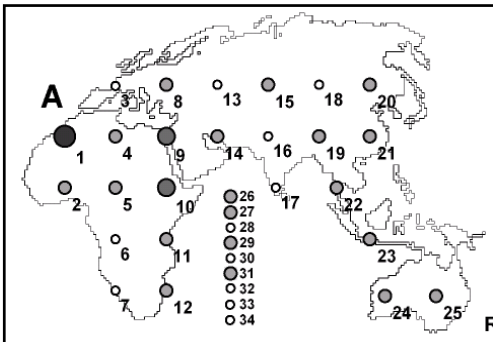
	Frequencies of correct assignment		
	Single origin	Multiregional	Total
1 locus	0.935 (0.438)	0.388 (0.420)	0.790 (0.433)
20 loci	0.995 (0.884)	0.992 (0.877)	0.994 (0.882)
377 loci	1.000 (0.991)	1.000 (0.991)	1.000 (0.991)



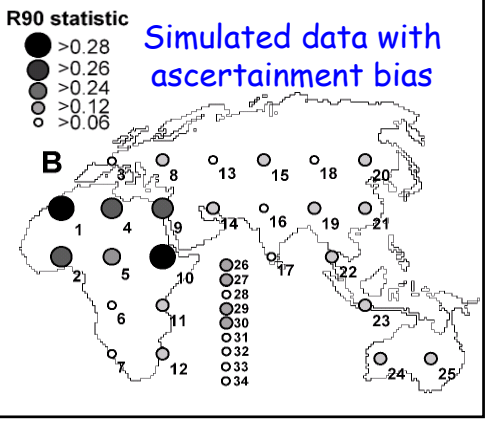




Scores of different scenarios



Possibility of STR ascertainment bias: STR loci chosen for their high diversity in Europe



Simulated data with ascertainment bias

Conclusions and perspectives

- Multiregional scenarios are clearly rejected
- Best fit with a unique and East-African origin
- But…
  - Relatively low correlation
  - Even more complex scenario required!
    - Competition
    - Dynamic environment
    - Culture
    - Neolithic transition
    - Selection
- Need to integrate simulations into a an inference framework

  ⇒ Approximate Bayesian Computation : ABC

  (Beaumont et al. 2002)

Acknowledgements

Pierre Berthier

Swiss NSF