

Diffusion Wave Hypothesis

Alan R. Rogers

November 23, 2009

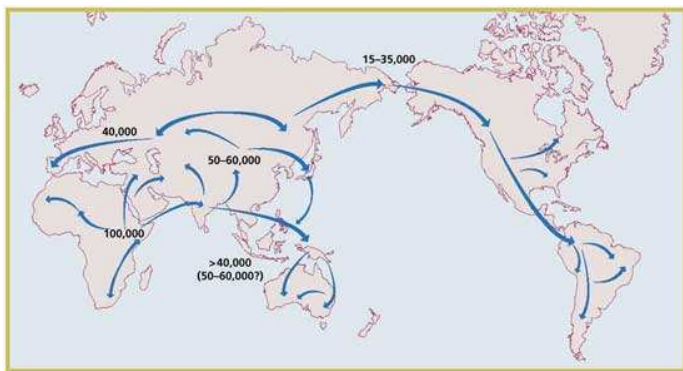
Neandertal & Modern



Navigation icons

Navigation icons

Moderns invade Eurasia



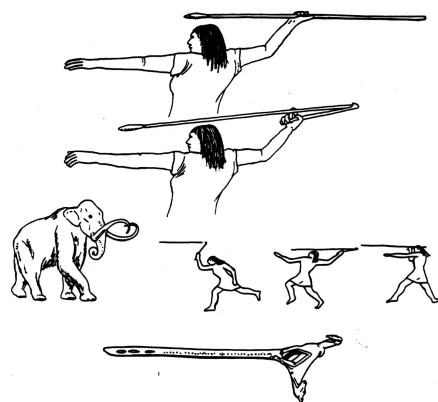
Navigation icons

Moderns made fancy points



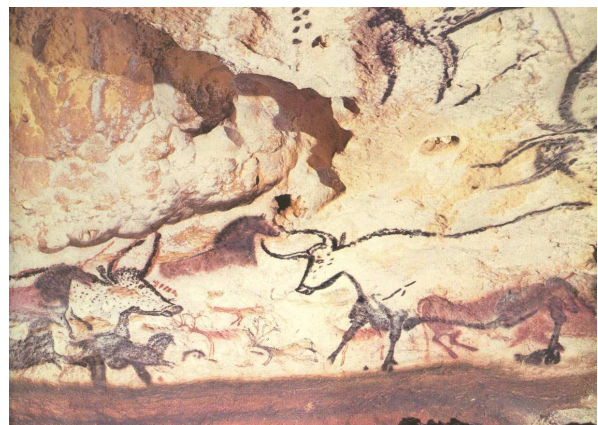
Navigation icons

Spear throwers



Navigation icons

Art



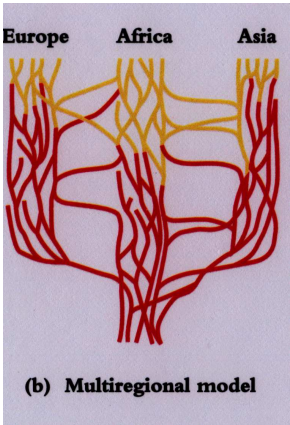
Navigation icons

How did modern humans evolve?

In the 1980s and 90s, there were two main hypotheses

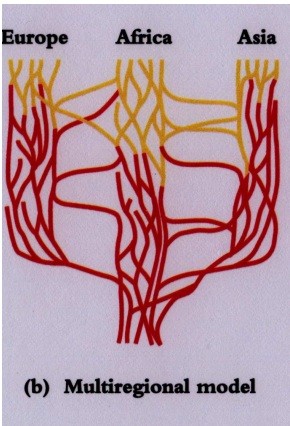
- ▶ Multiregional
- ▶ Replacement

Multiregional hypothesis



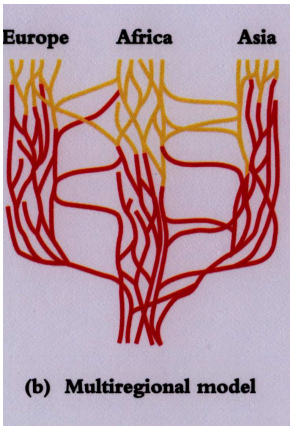
- ▶ *Homo erectus* expands into Eurasia 1.8 mya

Multiregional hypothesis



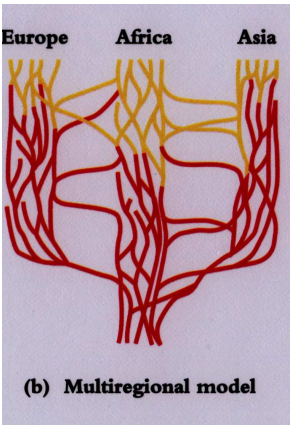
- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Strong gene flow

Multiregional hypothesis



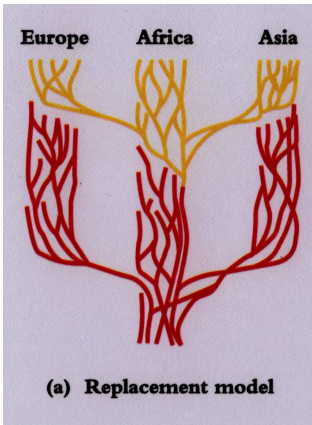
- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Strong gene flow
- ▶ Favorable mutations spread in every direction.

Multiregional hypothesis



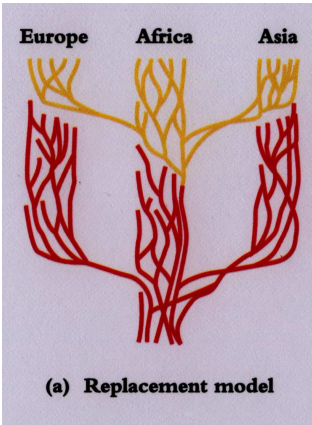
- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Strong gene flow
- ▶ Favorable mutations spread in every direction.
- ▶ Moderns have no geographic origin.

Replacement hypothesis



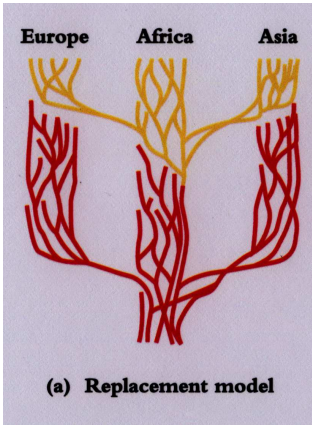
- ▶ *Homo erectus* expands into Eurasia 1.8 mya

Replacement hypothesis



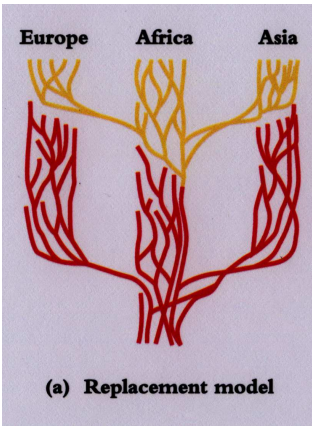
- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Moderns evolve in Africa by 100 kya

Replacement hypothesis



- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Moderns evolve in Africa by 100 kya
- ▶ Expand into Eurasia btw 100 and 30 kya

Replacement hypothesis



- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Moderns evolve in Africa by 100 kya
- ▶ Expand into Eurasia btw 100 and 30 kya
- ▶ No mixing with archaics

Mitochondrial story, circa 1996

This problem has been studied using bones, stones, and genes. Here's how I would have summarized the genetic evidence:

- ▶ Population expansion 50–100 kya

Mitochondrial story, circa 1996

This problem has been studied using bones, stones, and genes. Here's how I would have summarized the genetic evidence:

- ▶ Population expansion 50–100 kya
- ▶ Early population ≤ 5000 females

Mitochondrial story, circa 1996

This problem has been studied using bones, stones, and genes. Here's how I would have summarized the genetic evidence:

- ▶ Population expansion 50–100 kya
- ▶ Early population ≤ 5000 females
- ▶ Late population $\geq 1,000,000$ females

Mitochondrial story, circa 1996

This problem has been studied using bones, stones, and genes. Here's how I would have summarized the genetic evidence:

- ▶ Population expansion 50–100 kya
- ▶ Early population ≤ 5000 females
- ▶ Late population $\geq 1,000,000$ females
- ▶ Consistent with replacement hypothesis

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

Mitochondrial story, circa 1996

This problem has been studied using bones, stones, and genes. Here's how I would have summarized the genetic evidence:

- ▶ Population expansion 50–100 kya
- ▶ Early population ≤ 5000 females
- ▶ Late population $\geq 1,000,000$ females
- ▶ Consistent with replacement hypothesis
- ▶ but not multiregional hypothesis

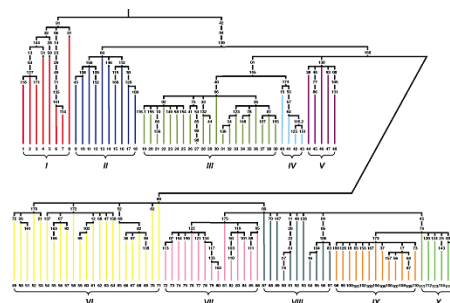
◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

Nuclear genes

Since 1997, we have relied more and more on nuclear genes.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

Some nuclear loci have shallow genealogies

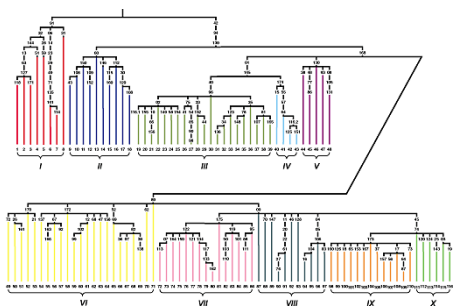


(UNDERHILL & AL 2000)

Y chromosome
49 kyr old

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

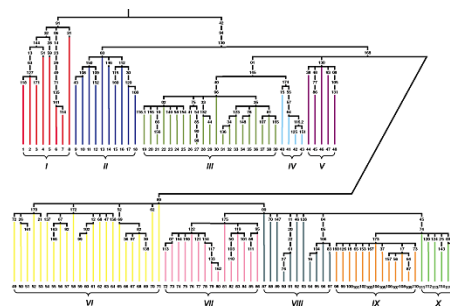
Some nuclear loci have shallow genealogies



(UNDERHILL & AL 2000)

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ≡ ▶ ≡ ▶ ↺ 🔍 ↻

Some nuclear loci have shallow genealogies

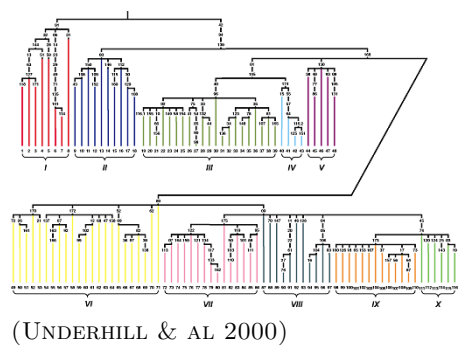


(UNDERHILL & AL 2000)

Y chromosome
49 kyr old
Old branches are
short; suggests
expansion.

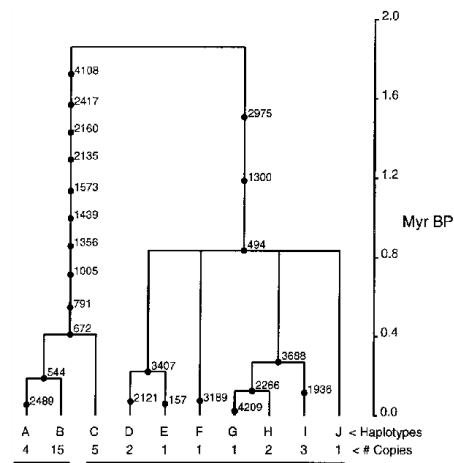
A set of small navigation icons typically found in Beamer presentations, including symbols for back, forward, search, and other slide controls.

Some nuclear loci have shallow genealogies



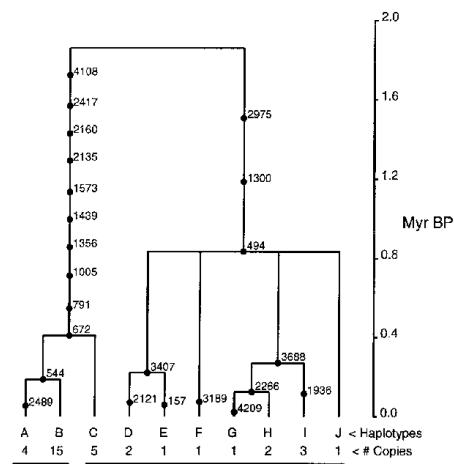
Y chromosome
49 kyr old
Old branches are short; suggests expansion.
Supports replacement hypothesis.

Others have deep genealogies



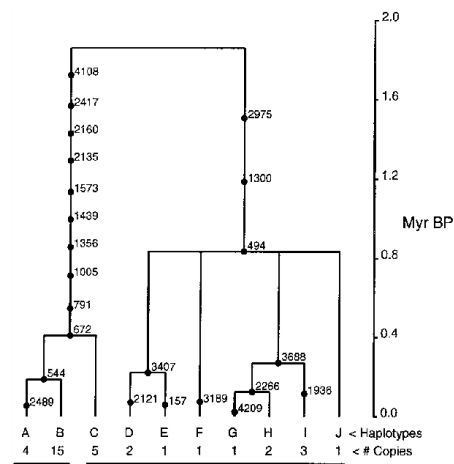
PDHA1 locus
(Harris & Hey 1999)

Others have deep genealogies



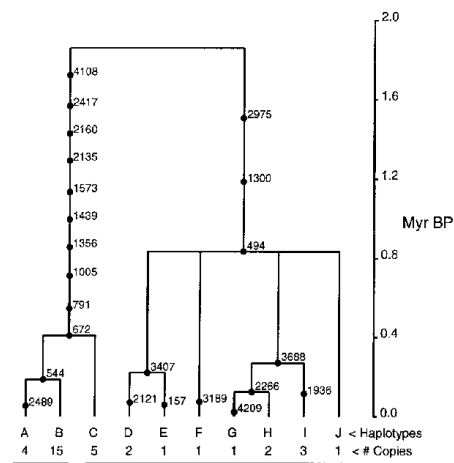
PDHA1 locus
(Harris & Hey 1999)
Nearly 2 myr old.

Others have deep genealogies



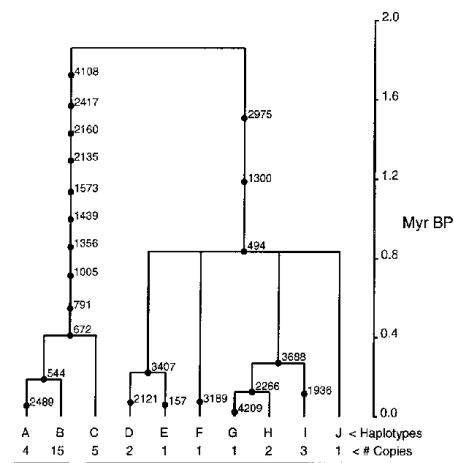
PDHA1 locus
(Harris & Hey 1999)
Nearly 2 myr old.
Old branches long.

Others have deep genealogies



PDHA1 locus
(Harris & Hey 1999)
Nearly 2 myr old.
Old branches long.
Suggests subdivision.

Others have deep genealogies



PDHA1 locus
(Harris & Hey 1999)
Nearly 2 myr old.
Old branches long.
Suggests subdivision.
Supports multiregional hypothesis

The puzzle

Some nuclear loci support one hypothesis;



The puzzle

Some nuclear loci support one hypothesis; some support the other.



The puzzle

Some nuclear loci support one hypothesis; some support the other.

The truth must somehow encompass both.

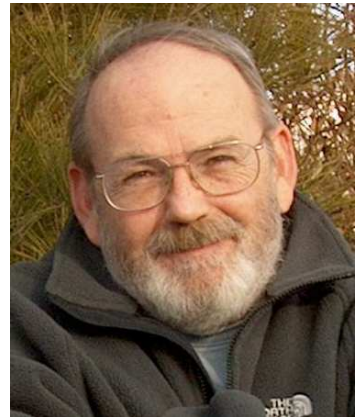


Diffusion wave authors

Vinayak Eswaran



Henry Harpending



Diffusion wave hypothesis

- ▶ *Homo erectus* expands into Eurasia 1.8 mya



Diffusion wave hypothesis

- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Moderns evolve in Africa by 100 kya



Diffusion wave hypothesis

- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Moderns evolve in Africa by 100 kya
- ▶ Expand into Eurasia btw 100 and 30 kya



Diffusion wave hypothesis

- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Moderns evolve in Africa by 100 kya
- ▶ Expand into Eurasia btw 100 and 30 kya
- ▶ Some mixing with archaics



Diffusion wave hypothesis

- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Moderns evolve in Africa by 100 kya
- ▶ Expand into Eurasia btw 100 and 30 kya
- ▶ Some mixing with archaics
- ▶ Co-adapted gene complex reduces fitness of hybrids.



Co-adapted gene complexes

- ▶ Hybrids often have low fitness



Co-adapted gene complexes

- ▶ Hybrids often have low fitness
- ▶ Especially when population is deeply subdivided



Co-adapted gene complexes

- ▶ Hybrids often have low fitness
- ▶ Especially when population is deeply subdivided
- ▶ Probably because each sub-population has evolved a different “co-adapted gene complex.”
- ▶ Co-adapted gene complex: a group of alleles at different loci that work well together, but not apart.
- ▶ For example, the parts in your Toyota work well together, but don’t install them in your Ford.



A detailed line drawing of a larva, likely a copepod, showing its segmented body, antennae, and legs. The larva has a rounded cephalothorax with a single eye visible. It has two long, segmented antennae at the front. The body is divided into several segments, with the first few being larger and more rounded. The posterior end is wider and has several long, thin appendages or legs extending downwards. The drawing is a black and white line illustration.



A detailed line drawing of a larva. The head is at the top with two long, segmented antennae. The body is divided into several segments. The posterior end features a large, rounded cluster of small, circular structures, likely eggs or cells, with several long, thin lines extending downwards from the base of the cluster.

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ≡ ↺ 🔍 ↻

A detailed line drawing of a larva, likely a copepod, showing its segmented body, antennae, and legs. The larva has a rounded cephalon with a single eye, two long antennae, and a segmented thorax. The abdomen is elongated and tapers towards the posterior, where several long, thin appendages are visible. The drawing is oriented vertically.

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ≡ ↺ 🔍 ↻

A detailed line drawing of a larva, likely a copepod, showing its segmented body, antennae, and legs. The larva has a rounded cephalon with a single eye, followed by a thorax and an abdomen. It has two long, segmented antennae at the front and several pairs of legs. The posterior end of the abdomen is fan-shaped and bears several long, thin setae.

- ◀ ◻ ▶ ◀ 📄 ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ 🔍 ↺

A detailed line drawing of a larva, likely a copepod, showing its segmented body, antennae, and legs. The larva has a rounded cephalon with a single eye, two long antennae, and a segmented thorax. The abdomen is elongated and tapers towards the posterior, where several long, thin appendages are visible. The drawing is oriented vertically with the head at the top.

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ≡ ↺ 🔍 ↻

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ≡ ≡ ≡ ↺ 🔍 ↻

Modelling a co-adapted gene complex

- ▶ Several unlinked loci
- ▶ Each locus has a “modern” and an “archaic” allele.
- ▶ Modern individuals have 2 copies of the modern allele at each locus. All others are archaic.



Modelling a co-adapted gene complex

- ▶ Several unlinked loci
- ▶ Each locus has a “modern” and an “archaic” allele.
- ▶ Modern individuals have 2 copies of the modern allele at each locus. All others are archaic.
- ▶ Moderns have a selective advantage.



Modelling a co-adapted gene complex

- ▶ Several unlinked loci
- ▶ Each locus has a “modern” and an “archaic” allele.
- ▶ Modern individuals have 2 copies of the modern allele at each locus. All others are archaic.
- ▶ Moderns have a selective advantage.
- ▶ Advantage disappears if you have even 1 copy of an archaic allele



Neutral loci

- ▶ There are 40 unlinked neutral loci, in addition to the co-adapted gene complex.

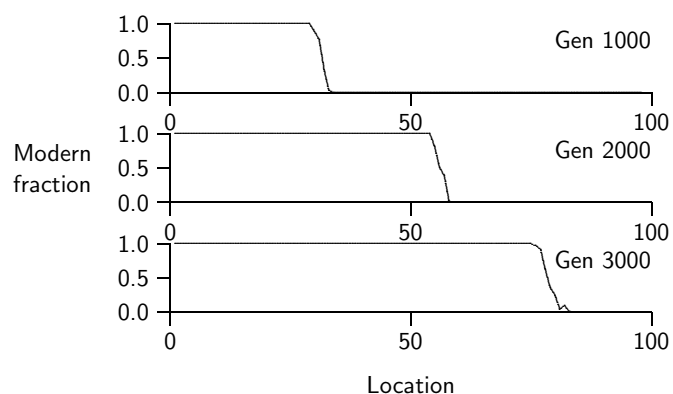


Neutral loci

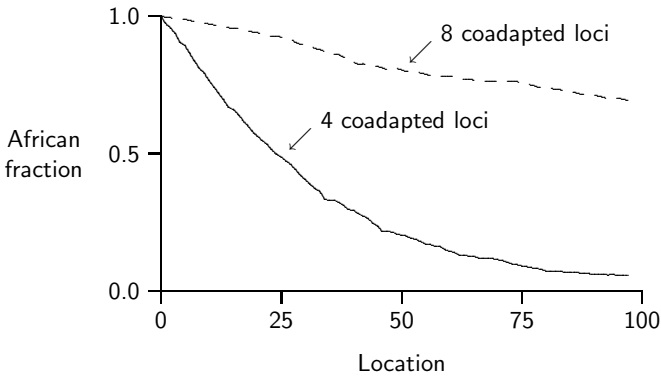
- ▶ There are 40 unlinked neutral loci, in addition to the co-adapted gene complex.
- ▶ The statistics we report are based on these neutral loci, not on the co-adapted gene complex.



Progress of diffusion wave



Fraction of neutral loci that are African, after diffusion wave



Diffusion wave predicts archaic admixture at some (but not all) loci.

Diffusion wave predicts archaic admixture at some (but not all) loci.

- ▶ At few loci if adaptation is complex.

Diffusion wave predicts archaic admixture at some (but not all) loci.

- ▶ At few loci if adaptation is complex.
- ▶ At many loci if adaptation is simple.

Diffusion wave predicts archaic admixture at some (but not all) loci.

- ▶ At few loci if adaptation is complex.
- ▶ At many loci if adaptation is simple.

Just the sort of confusion that we see in real data.

Diffusion wave predicts archaic admixture at some (but not all) loci.

- ▶ At few loci if adaptation is complex.
- ▶ At many loci if adaptation is simple.

Just the sort of confusion that we see in real data.

Next... the evidence

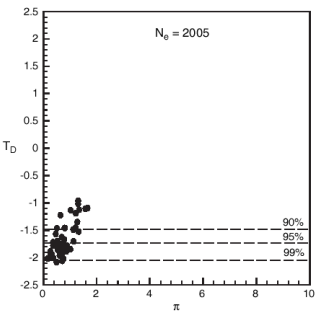
Tajima's D

- ▶ a statistic that is sensitive both to selection and population growth.

Tajima's D

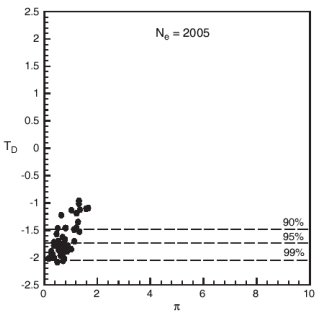
- ▶ a statistic that is sensitive both to selection and population growth.
- ▶ behaves differently under the replacement, multiregional, and diffusion wave hypotheses.

Replacement hypothesis



Simulated Tajima's D

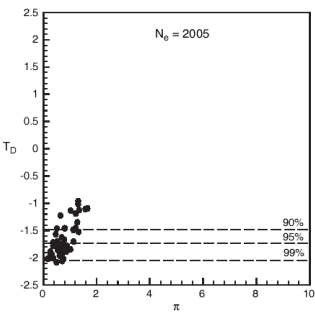
Replacement hypothesis



Simulated Tajima's D

- ▶ Many near -2.

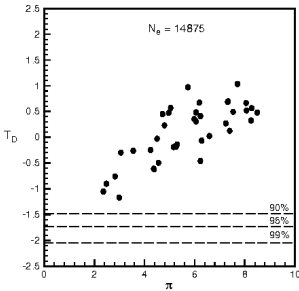
Replacement hypothesis



Simulated Tajima's D

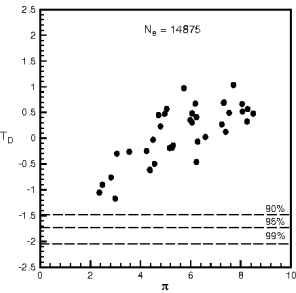
- ▶ Many near -2.
- ▶ None much above -1.

Multiregional hypothesis



Simulated Tajima's D

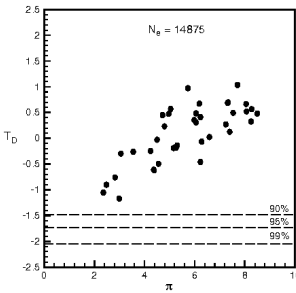
Multiregional hypothesis



Simulated Tajima's D

- Symmetric about zero

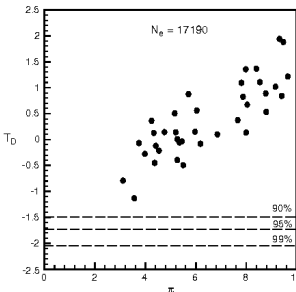
Multiregional hypothesis



Simulated Tajima's D

- Symmetric about zero
- Would be > 0 with less migration

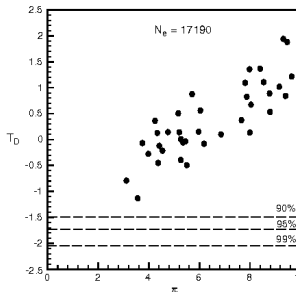
Multiregional hypothesis again



Simulated Tajima's D

- With more population structure

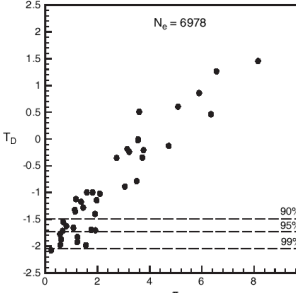
Multiregional hypothesis again



Simulated Tajima's D

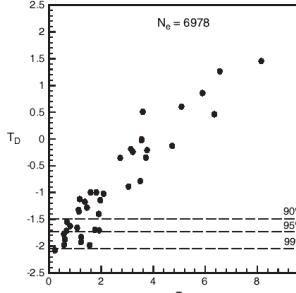
- With more population structure
- Average above zero.

Diffusion Wave Hypothesis



Simulated Tajima's D

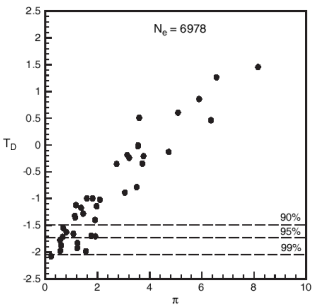
Diffusion Wave Hypothesis



Simulated Tajima's D

- From -2 to $+1.5$.

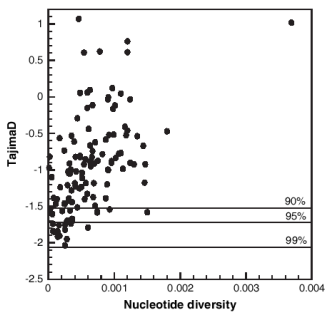
Diffusion Wave Hypothesis



Simulated Tajima's D

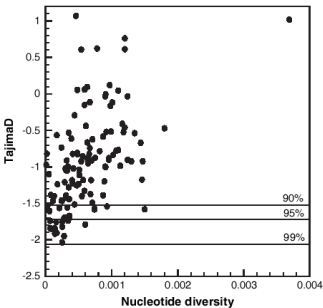
- ▶ From -2 to +1.5.
- ▶ Most are negative.

Reality



Observed Tajima's D (Stephens et al 2003)

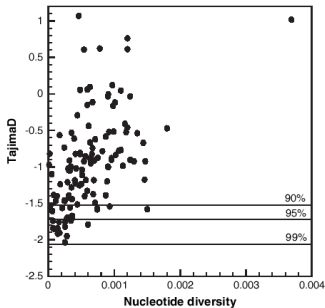
Reality



Observed Tajima's D (Stephens et al 2003)

- ▶ Too negative for multiregional hypothesis.

Reality



Observed Tajima's D (Stephens et al 2003)

- ▶ Too negative for multiregional hypothesis.
- ▶ Not negative enough for replacement hypothesis.

Summary of Tajima's D

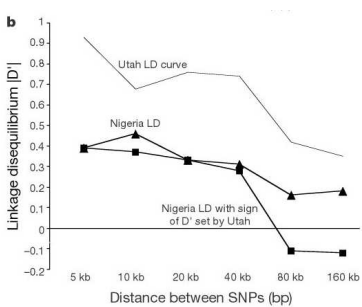
Explains the pattern in the data better than any other hypothesis.

Summary of Tajima's D

Explains the pattern in the data better than any other hypothesis.

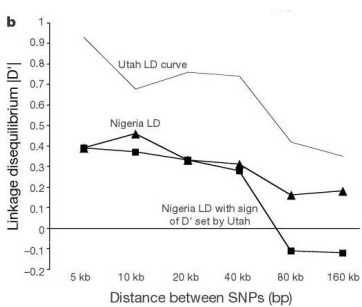
If this hypothesis is correct, many human loci should have archaic admixture. How can we tell?

More disequilibrium in Europe than Africa



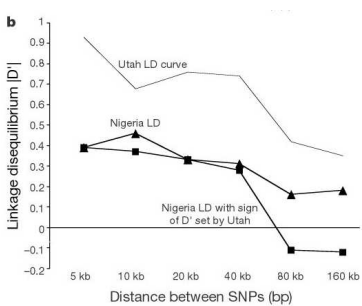
- ▶ More LD in Europe than Africa

More disequilibrium in Europe than Africa



- ▶ More LD in Europe than Africa
- ▶ Could reflect archaic admixture in Europe

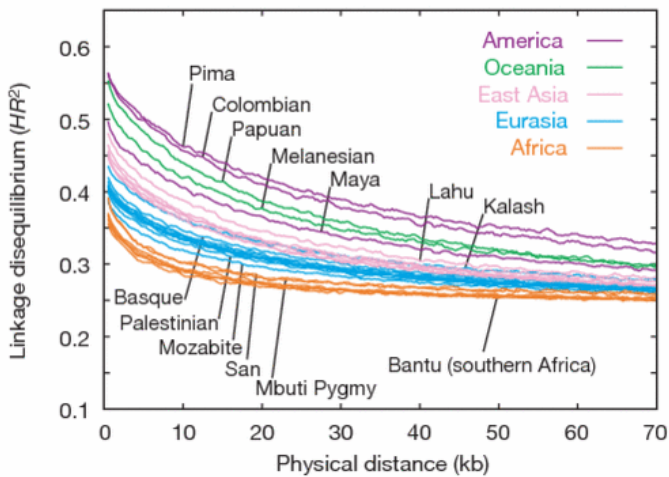
More disequilibrium in Europe than Africa



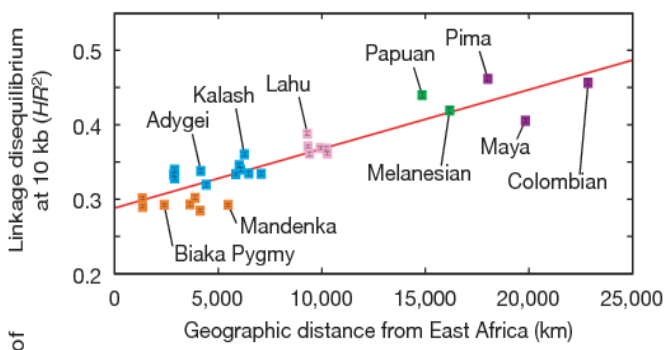
- ▶ More LD in Europe than Africa
- ▶ Could reflect archaic admixture in Europe
- ▶ or a European bottleneck.

(REICH ET AL 2001)

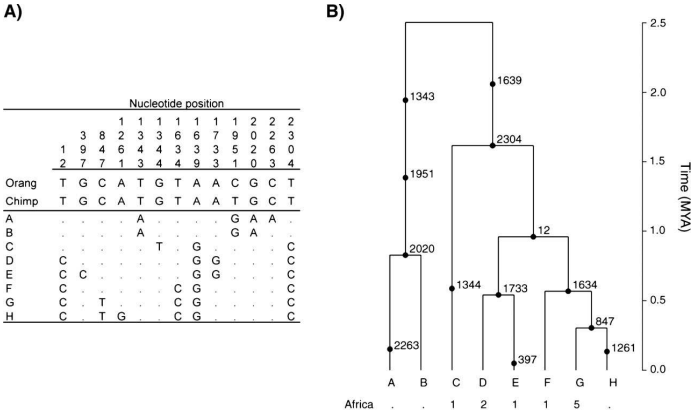
Decay of LD with distance along chromosome



LD increases with distance from Africa

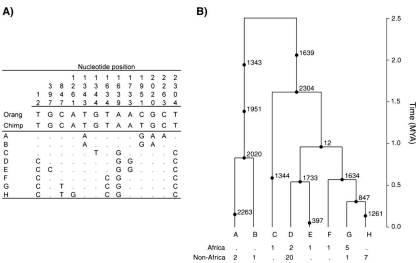


Disequilibrium at locus RRM2P4



(GARRIGAN ET AL 2004)

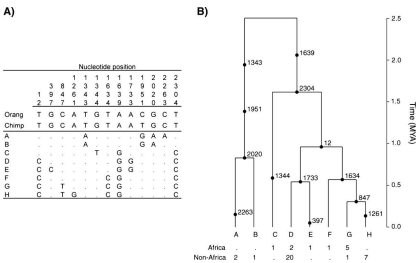
RRM2P4 continued



(GARRIGAN ET AL 2004)

- ▶ H'types A,B in strong disequilibrium
- ▶ No surprise: sequence is short (2.4 kb)
- ▶ Locus is very old: 2.5 myr
- ▶ Isolated populations or balancing sel'n

RRM2P4 continued



(GARRIGAN ET AL 2004)

- ▶ H'types A,B in strong disequilibrium
- ▶ No surprise: sequence is short (2.4 kb)
- ▶ Locus is very old: 2.5 myr
- ▶ Isolated populations or balancing sel'n

Deep disequilibrium at locus Xp21.1

	111	8899999	1111111111
	1247023	4523446	5555666677
	5133691	9071093	4456046633
	5708906	1246458	5646823446
			2380065185
a	..t..a	c.a...	..a...a..
b	.c.t...	.a...	caa...a..
C	..G.TC.	.G...G	...T...TC
d	..t...	.a...	..a...a..
e	..t...	c.a...	..a...atc
f	..t...	.a...	..a...atc
g	c..t...	..a..g	..a...gatc
h	c..t...	.a.tg.	..a...gatc
i	c..t...	.a.g.	..a...gatc
j	c..t...	.a...	..a...gatc
k	c..t...	.a.g.	..a...atc
l	c..t...	.a.g.	..aag..gatc
m	..t..a	c.a...	..a..C.a..
n	..t...	c.ag...	..a...a..
o	..t...	.ag...	..a...a..
p	..t...	c.a...	..a...a..c
q	c..t...	..ag.g.	..a...gatc

- ▶ Sequence is 17.5 kb long

Deep disequilibrium at locus Xp21.1

	111	8899999	1111111111
	1247023	4523446	5555666677
	5133691	9071093	4456046633
	5708906	1246458	5646823446
			2380065185
a	..t..a	c.a...	..a...a..
b	.c.t...	.a...	caa...a..
C	..G.TC.	.G...G	...T...TC
d	..t...	.a...	..a...a..
e	..t...	c.a...	..a...atc
f	..t...	.a...	..a...atc
g	c..t...	..a..g	..a...gatc
h	c..t...	.a.tg.	..a...gatc
i	c..t...	.a.g.	..a...gatc
j	c..t...	.a...	..a...gatc
k	c..t...	.a.g.	..a...atc
l	c..t...	.a.g.	..aag..gatc
m	..t..a	c.a...	..a..C.a..
n	..t...	c.ag...	..a...a..
o	..t...	.ag...	..a...a..
p	..t...	c.a...	..a...a..c
q	c..t...	..ag.g.	..a...gatc

- ▶ Sequence is 17.5 kb long
- ▶ MRCA is 1.9 myr old

Deep disequilibrium at locus Xp21.1

	111	8899999	1111111111
	1247023	4523446	5555666677
	5133691	9071093	4456046633
	5708906	1246458	5646823446
			2380065185
a	..t..a	c.a...	..a...a..
b	.c.t...	.a...	caa...a..
C	..G.TC.	.G...G	...T...TC
d	..t...	.a...	..a...a..
e	..t...	c.a...	..a...atc
f	..t...	.a...	..a...atc
g	c..t...	..a..g	..a...gatc
h	c..t...	.a.tg.	..a...gatc
i	c..t...	.a.g.	..a...gatc
j	c..t...	.a...	..a...gatc
k	c..t...	.a.g.	..a...atc
l	c..t...	.a.g.	..aag..gatc
m	..t..a	c.a...	..a..C.a..
n	..t...	c.ag...	..a...a..
o	..t...	.ag...	..a...a..
p	..t...	c.a...	..a...a..c
q	c..t...	..ag.g.	..a...gatc

- ▶ Sequence is 17.5 kb long
- ▶ MRCA is 1.9 myr old
- ▶ H'type C in strong disequilibrium

Deep disequilibrium at locus Xp21.1

	111	8899999	1111111111
	1247023	4523446	5555666677
	5133691	9071093	4456046633
	5708906	1246458	5646823446
			2380065185
a	..t..a	c.a...	..a...a..
b	.c.t...	.a...	caa...a..
C	..G.TC.	.G...G	...T...TC
d	..t...	.a...	..a...a..
e	..t...	c.a...	..a...atc
f	..t...	.a...	..a...atc
g	c..t...	..a..g	..a...gatc
h	c..t...	.a.tg.	..a...gatc
i	c..t...	.a.g.	..a...gatc
j	c..t...	.a...	..a...gatc
k	c..t...	.a.g.	..a...atc
l	c..t...	.a.g.	..aag..gatc
m	..t..a	c.a...	..a..C.a..
n	..t...	c.ag...	..a...a..
o	..t...	.ag...	..a...a..
p	..t...	c.a...	..a...a..c
q	c..t...	..ag.g.	..a...gatc

- ▶ Sequence is 17.5 kb long
- ▶ MRCA is 1.9 myr old
- ▶ H'type C in strong disequilibrium
- ▶ Not balancing selection: too much disequilibrium

Deep disequilibrium at locus Xp21.1

	111	8899999	111111111
	1247023	4523446	5555666677
	5133691	9071093	4456046633
	5708906	1246458	5646823446
			2380065185
a	. . . t . . a	c . a aa a . .
b	. c . t a	caa a . .
C	. . G . TC .	. G . . . G T . . TC
d	. . . t a aa a . .
e	. . . t . . .	c . a aa a t c
f	. . . t a aa a t c
g	c . . t a . . . g	. aa . . . gat c
h	c . . t a . tg .	. aa . . . gat c
i	c . . t a . . g .	. aa . . . gat c
j	c . . t a aa . . . gat c
k	c . . t a . . g .	. aa . . . a t c
l	c . . t a . . g .	. aag . . gat c
m	. . . t . . a	c . a aa . . c . a . .
n	. . . t . . .	c . ag aa . . . a . .
o	. . . t ag aa . . . a . .
p	. . . t . . .	c . a aa . . . a . c
q	c . . t ag . g .	. aa . . . gat c

- ▶ Sequence is 17.5 kb long
- ▶ MRCA is 1.9 myr old
- ▶ H'type C in strong disequilibrium
- ▶ Not balancing selection: too much disequilibrium
- ▶ Requires isolated populations (GARRIGAN ET AL 2005)

Microcephalin (Evans et al 2005)

- ▶ Sequenced a large (23.4 kb) region.

Microcephalin (Evans et al 2005)

- ▶ Sequenced a large (23.4 kb) region.
- ▶ One haplotype has been increasing for 37 kyr.

Microcephalin (Evans et al 2005)

- ▶ Sequenced a large (23.4 kb) region.
- ▶ One haplotype has been increasing for 37 kyr.
- ▶ In “near-complete LD across the entire region.”

Microcephalin (Evans et al 2005)

- ▶ Sequenced a large (23.4 kb) region.
- ▶ One haplotype has been increasing for 37 kyr.
- ▶ In “near-complete LD across the entire region.”
- ▶ MRCA 1.7 mya

Microcephalin (Evans et al 2005)

- ▶ Sequenced a large (23.4 kb) region.
- ▶ One haplotype has been increasing for 37 kyr.
- ▶ In “near-complete LD across the entire region.”
- ▶ MRCA 1.7 mya
- ▶ Can't be balancing selection (too much LD).

Microcephalin (Evans et al 2005)

- ▶ Sequenced a large (23.4 kb) region.
- ▶ One haplotype has been increasing for 37 kyr.
- ▶ In “near-complete LD across the entire region.”
- ▶ MRCA 1.7 mya
- ▶ Can't be balancing selection (too much LD).
- ▶ Strongly suggests archaic admixture.

Summary

- ▶ Hypothesis: our ancestors expanded out of Africa 50 kya, and mated with surrounding archaic populations.

Summary

- ▶ Hypothesis: our ancestors expanded out of Africa 50 kya, and mated with surrounding archaic populations.
- ▶ Yet only a few archaic genes entered the modern population.

Summary

- ▶ Hypothesis: our ancestors expanded out of Africa 50 kya, and mated with surrounding archaic populations.
- ▶ Yet only a few archaic genes entered the modern population.
- ▶ Explains the contradictory signals given by different genetic loci.

Summary

- ▶ Hypothesis: our ancestors expanded out of Africa 50 kya, and mated with surrounding archaic populations.
- ▶ Yet only a few archaic genes entered the modern population.
- ▶ Explains the contradictory signals given by different genetic loci.
- ▶ It also predicts that some nuclear genes should show signs of archaic admixture: deep gene trees and high LD.

Summary

- ▶ Hypothesis: our ancestors expanded out of Africa 50 kya, and mated with surrounding archaic populations.
- ▶ Yet only a few archaic genes entered the modern population.
- ▶ Explains the contradictory signals given by different genetic loci.
- ▶ It also predicts that some nuclear genes should show signs of archaic admixture: deep gene trees and high LD.
- ▶ We find this pattern at two loci: Xp21.1 and Microcephalin.

Summary

- ▶ Hypothesis: our ancestors expanded out of Africa 50 kya, and mated with surrounding archaic populations.
- ▶ Yet only a few archaic genes entered the modern population.
- ▶ Explains the contradictory signals given by different genetic loci.
- ▶ It also predicts that some nuclear genes should show signs of archaic admixture: deep gene trees and high LD.
- ▶ We find this pattern at two loci: Xp21.1 and Microcephalin.
- ▶ Are we part Neandertal?