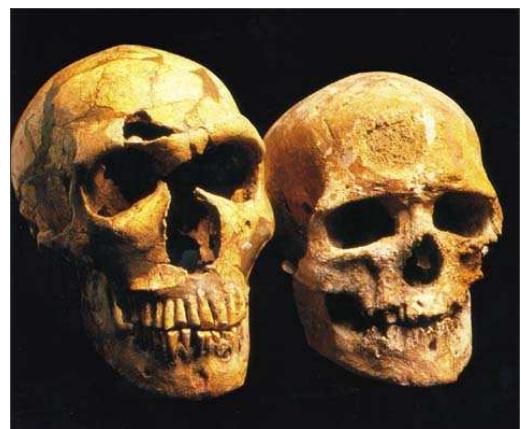


Diffusion Wave Hypothesis

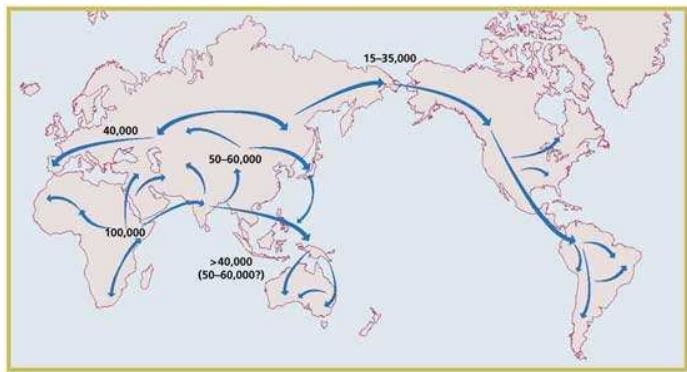
Alan R. Rogers

November 23, 2009

Neandertal & Modern



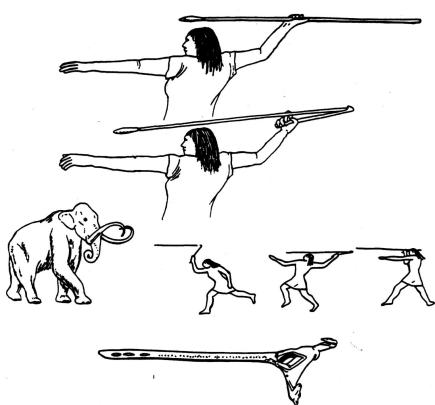
Moderns invade Eurasia



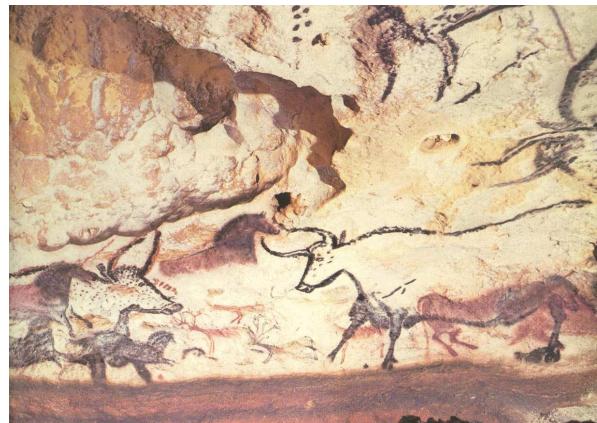
Moderns made fancy points



Spear throwers



Art

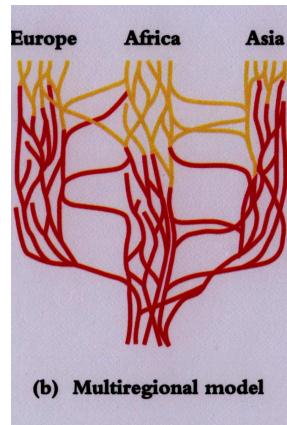


How did modern humans evolve?

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

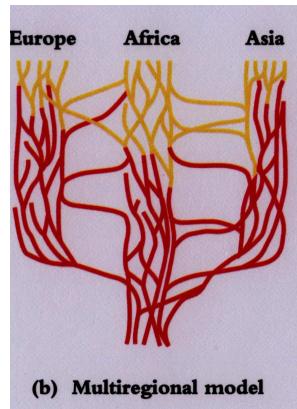
- ▶ Multiregional
- ▶ Replacement

Multiregional hypothesis



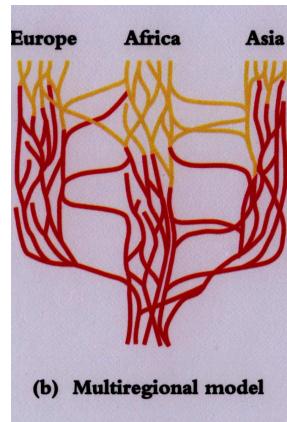
- ▶ *Homo erectus* expands into Eurasia 1.8 mya

Multiregional hypothesis



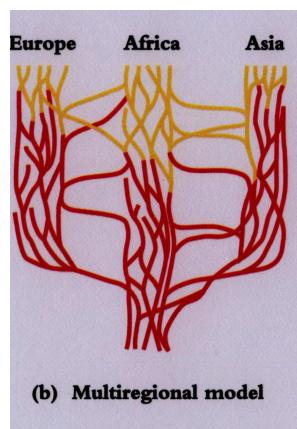
- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Strong gene flow

Multiregional hypothesis



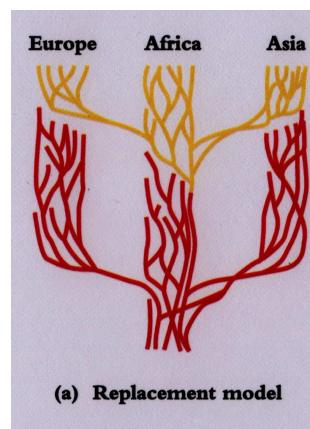
- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Strong gene flow
- ▶ Favorable mutations spread in every direction.

Multiregional hypothesis



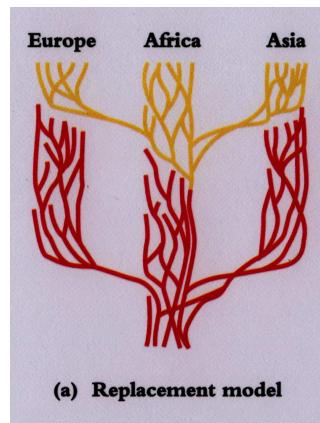
- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Strong gene flow
- ▶ Favorable mutations spread in every direction.
- ▶ Moderns have no geographic origin.

Replacement hypothesis



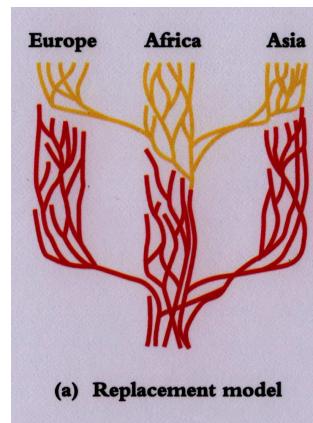
- ▶ *Homo erectus* expands into Eurasia 1.8 mya

Replacement hypothesis



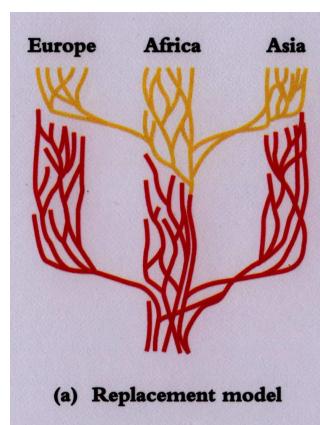
- ▶ *Homo erectus* expands into Eurasia 1.8 mya
- ▶ Moderns evolve in Africa by 100 kya

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 \leq 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 \leq 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 \leq 5000 females
- ▶ Late population \geq 1,000,000 females
- ▶ Consistent with replacement hypothesis

Mitochondrial story, circa 1996

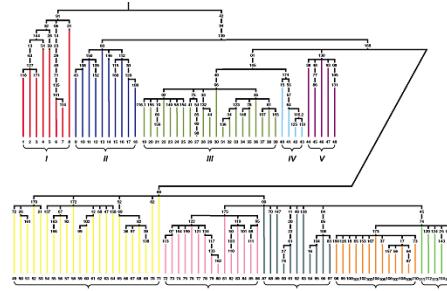
This problem has been studied using bones, stones, and genes. Here's how I would have summarized the genetic evidence:

- ▶ Population expansion 50–100 kya
- ▶ Early population \leq 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.

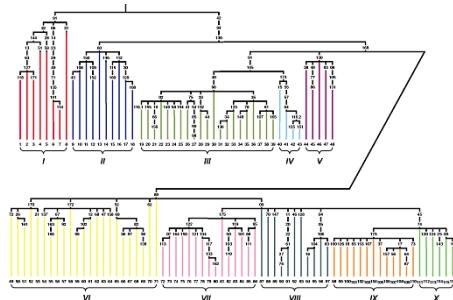
Some nuclear loci have shallow genealogies



Y chromosome
49 kyr old

(UNDERHILL & AL 2000)

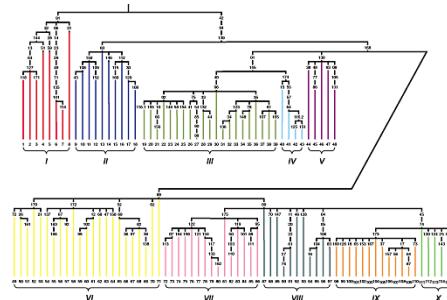
Some nuclear loci have shallow genealogies



(UNDERHILL & AL 2000)

Y chromosome
49 kyr old
Old branches are
short;

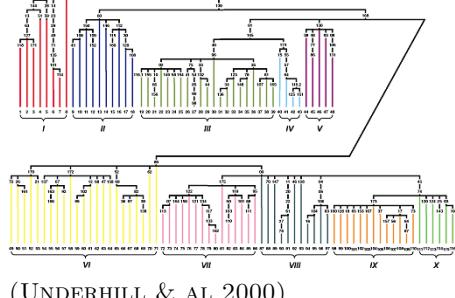
Some nuclear loci have shallow genealogies



(UNDERHILL & AL 2000)

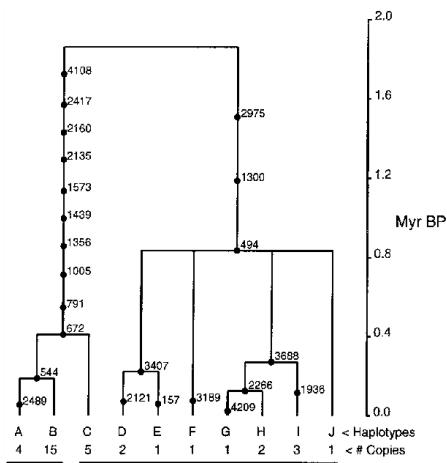
Y chromosome
49 kyr old
Old branches are
short; suggests
expansion.

Some nuclear loci have shallow genealogies



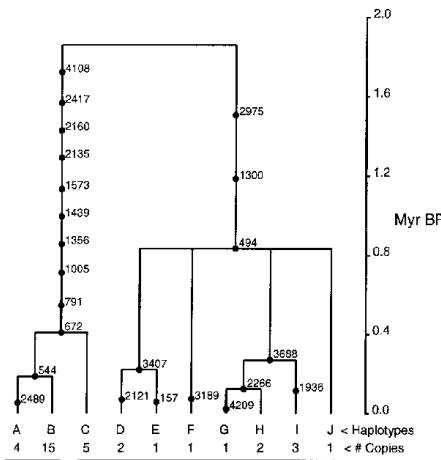
Y chromosome
49 kyr old
Old branches are short; suggests expansion.
Supports replacement hypothesis.

Others have deep genealogies



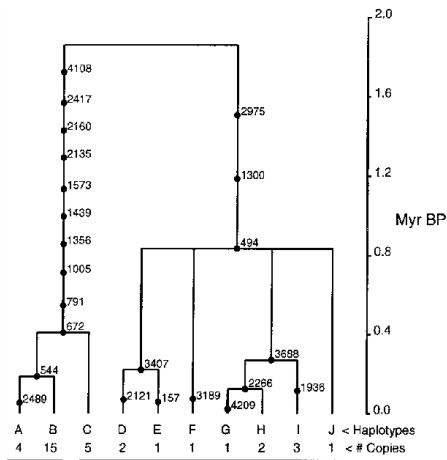
PDHA1 locus
(Harris & Hey 1999)

Others have deep genealogies



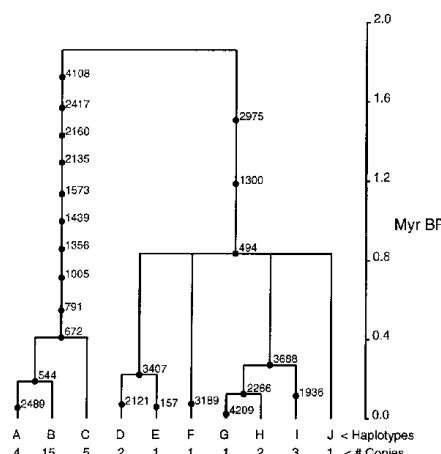
PDHA1 locus
(Harris & Hey 1999)
Nearly 2 myr old.

Others have deep genealogies



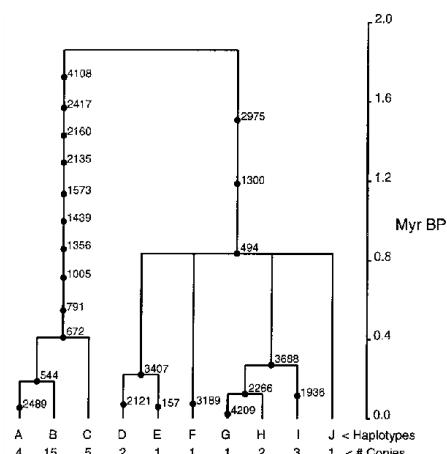
PDHA1 locus
(Harris & Hey 1999)
Nearly 2 myr old.
Old branches long.

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.

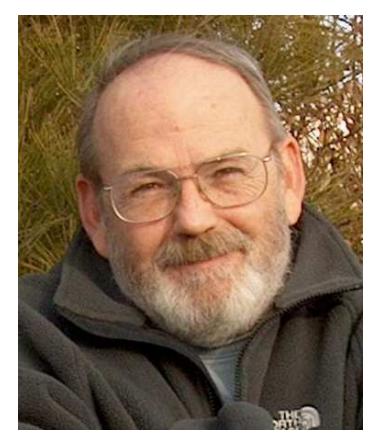


Diffusion wave authors

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

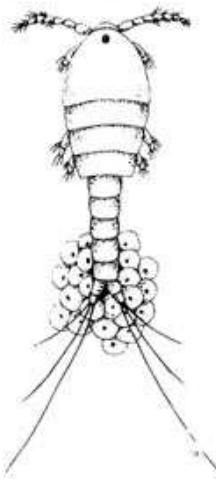
Co-adapted gene complexes

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

Co-adapted gene complexes

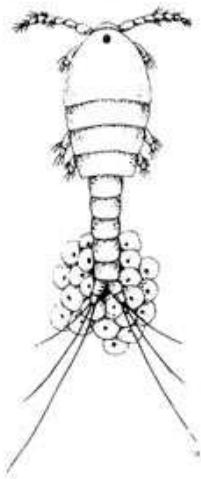
- ▶ Hybrids often have low fitness
- ▶ Especially when population is deeply subdivided
- ▶ Probably because each sub-population has evolved a different “co-adapted gene complex.”
- ▶ Co-adapted gene complex: a group of alleles at different loci that work well together, but not apart.
- ▶ For example,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

Copepod *Tigriopus californicus*



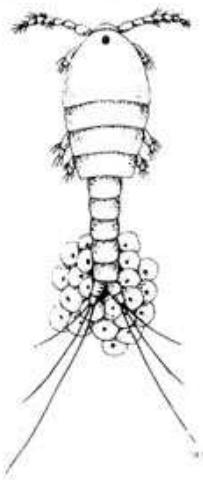
- A set of small, semi-transparent navigation icons located at the bottom of the slide. From left to right, they include: a left arrow, a square, a right arrow, a double left arrow, a double square, a double right arrow, a double left arrow, a double square, a double right arrow, a list icon, a refresh/circular arrow icon, a magnifying glass icon, and a circular arrow icon.

Copepod *Tigriopus californicus*



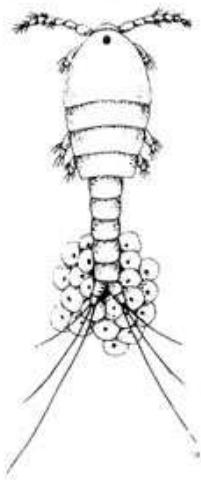
- ▶ Live in tide pools

Copepod *Tigriopus californicus*



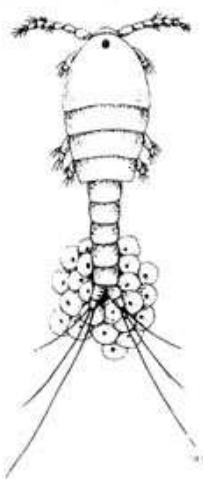
- ▶ Live in tide pools
 - ▶ Isolated populations

Copepod *Tigriopus californicus*



- ▶ Live in tide pools
 - ▶ Isolated populations
 - ▶ Hybrids have low fitness

Copepod *Tigriopus californicus*



- ▶ Live in tide pools
 - ▶ Isolated populations
 - ▶ Hybrids have low fitness
 - ▶ Different complexes of coadapted alleles

Modelling a co-adapted gene complex

- ▶ Several unlinked loci

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

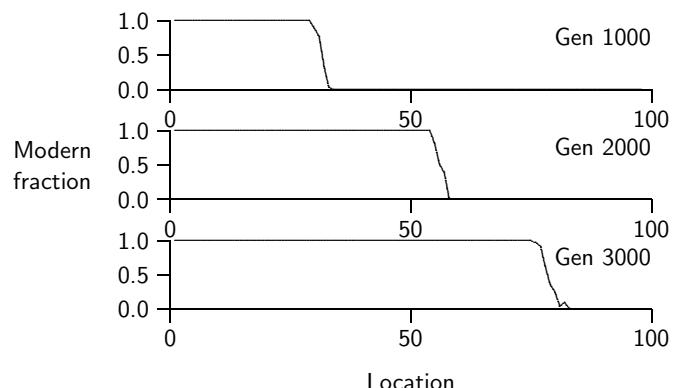
Neutral loci

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

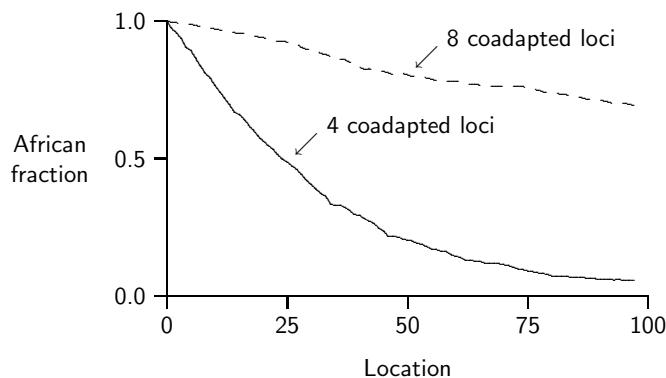
Neutral loci

- ▶ 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

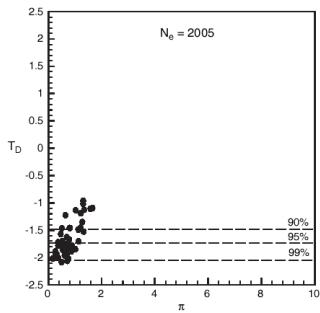
Tajima's D

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

Tajima's D

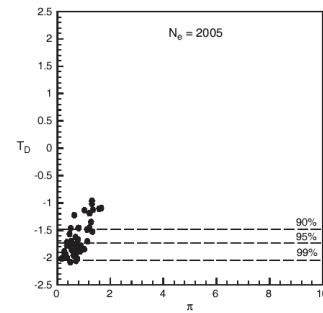
- ▶ a statistic that is sensitive both to selection and population growth.
- ▶ behaves differently under the replacement, multiregional, and diffusion wave hypotheses.

Replacement hypothesis



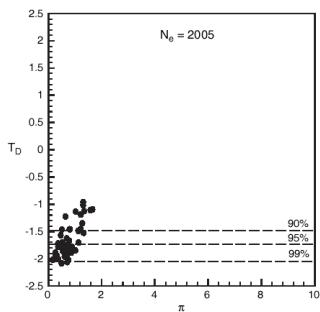
Simulated Tajima's D

Replacement hypothesis



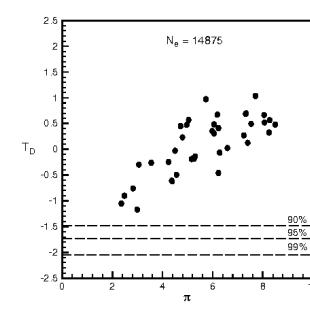
Simulated Tajima's D
▶ Many near -2 .

Replacement hypothesis



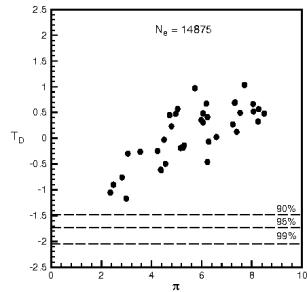
Simulated Tajima's D
▶ Many near -2 .
▶ None much above -1 .

Multiregional hypothesis



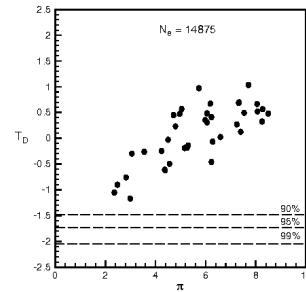
Simulated Tajima's D

Multiregional hypothesis



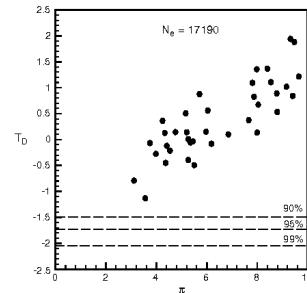
Simulated Tajima's D
► Symmetric about zero

Multiregional hypothesis



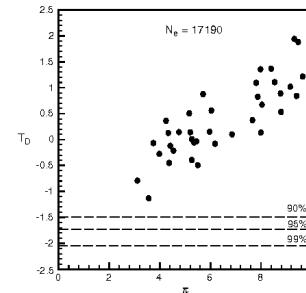
Simulated Tajima's D
► Symmetric about zero
► Would be > 0 with less migration

Multiregional hypothesis again



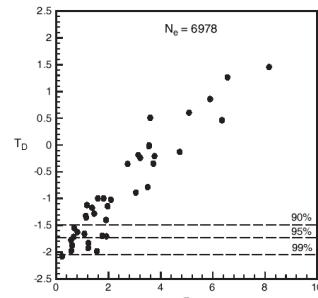
Simulated Tajima's D
► With more population structure

Multiregional hypothesis again



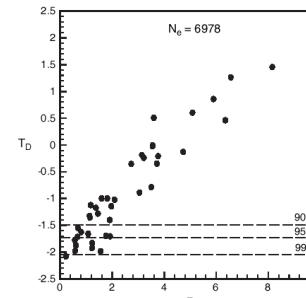
Simulated Tajima's D
► With more population structure
► Average above zero.

Diffusion Wave Hypothesis



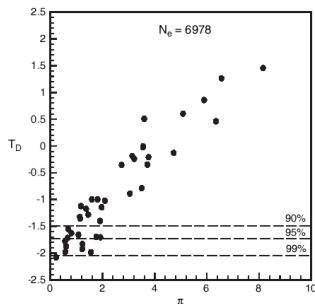
Simulated Tajima's D

Diffusion Wave Hypothesis



Simulated Tajima's D
► From -2 to $+1.5$.

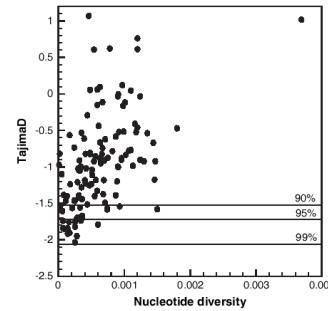
Diffusion Wave Hypothesis



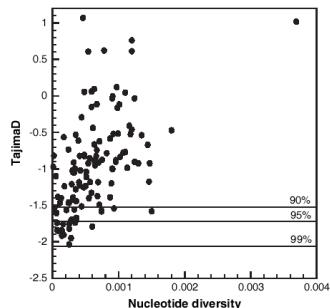
Simulated Tajima's D

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

Reality

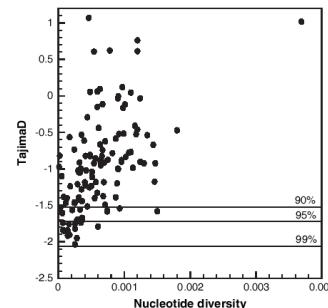


Reality



- ▶ Too negative for multiregional hypothesis.

Reality



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

Summary of Tajima's D

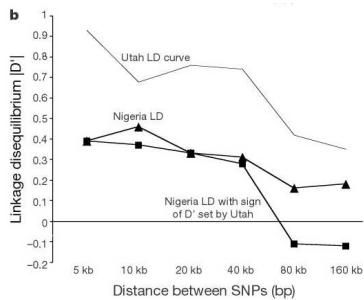
Explains the pattern in the data better than any other hypothesis.

Summary of Tajima's D

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

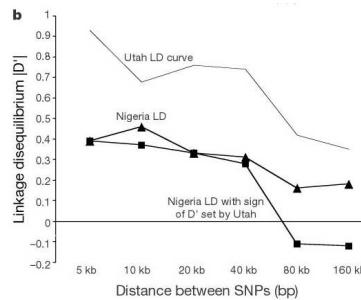
If this hypothesis is correct, many human loci should have archaic admixture. How can we tell?

More disequilibrium in Europe than Africa



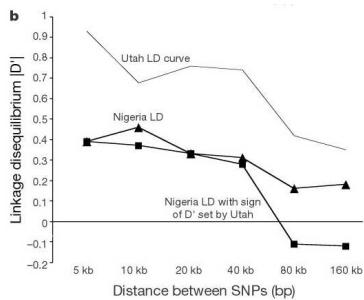
- More LD in Europe than Africa

More disequilibrium in Europe than Africa



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

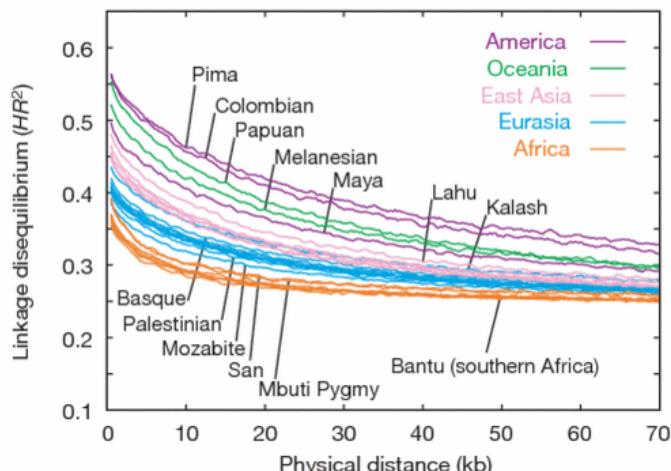
More disequilibrium in Europe than Africa



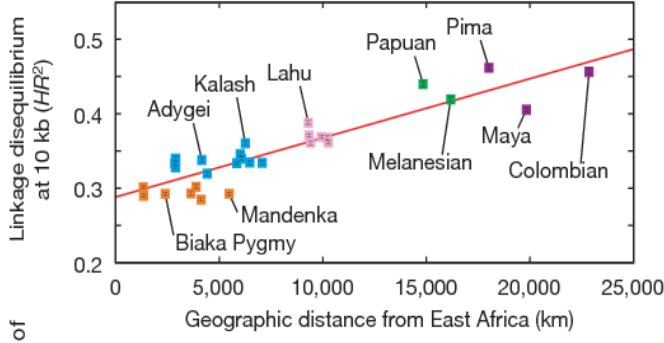
- More LD in Europe than Africa
- Could reflect archaic admixture in Europe
- or a European bottleneck.

(REICH ET AL 2001)

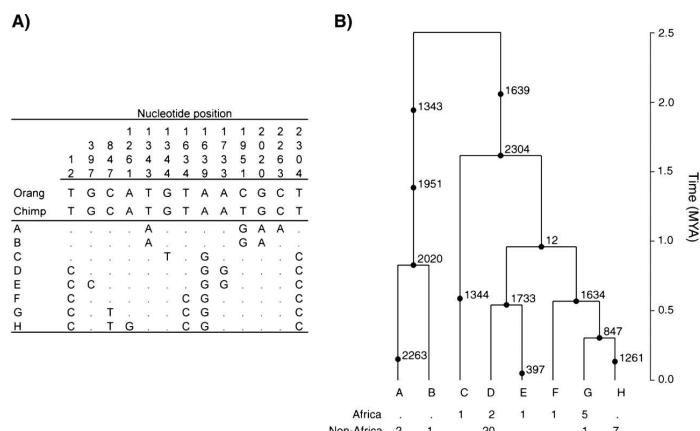
Decay of LD with distance along chromosome



LD increases with distance from Africa

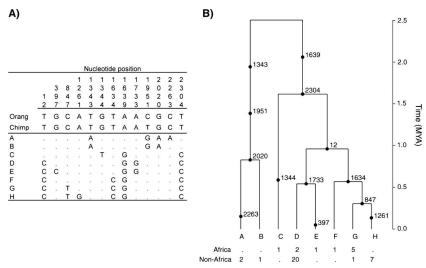


Disequilibrium at locus RRM2P4



(GARRIGAN ET AL 2004)

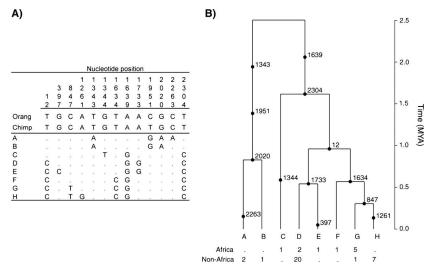
RRM2P4 continued



(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 | 5555666677 |
|---------|---------|------------|------------|
| 1247023 | 4523446 | 4456046633 | |
| 5133691 | 9071093 | 5646823446 | |
| 5708906 | 1246458 | 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....atc |
| f | ...t.. | ..a.... | .aa....atc |
| g | ...t.. | ..a....g | .aa....gatc |
| h | ...t.. | ..a.tg. | .aa....gatc |
| i | ...t.. | ..a....g | .aa....gatc |
| j | ...t.. | ..a.... | .aa....gatc |
| k | ...t.. | ..a....g | .aa....atc |
| l | ...t.. | ..a....g | .aag...gatc |
| m | ...t.. | a | 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....gatc |

- ▶ Sequence is 17.5 kb long

Deep disequilibrium at locus Xp21.1

| | 111 | 8899999 | 5555666677 |
|---------|---------|------------|------------|
| 1247023 | 4523446 | 4456046633 | |
| 5133691 | 9071093 | 5646823446 | |
| 5708906 | 1246458 | 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....atc |
| f | ...t.. | ..a.... | .aa....atc |
| g | ...t.. | ..a....g | .aa....gatc |
| h | ...t.. | ..a.tg. | .aa....gatc |
| i | ...t.. | ..a....g | .aa....gatc |
| j | ...t.. | ..a.... | .aa....gatc |
| k | ...t.. | ..a....g | .aa....atc |
| l | ...t.. | ..a....g | .aag...gatc |
| m | ...t.. | a | 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....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 | 5555666677 |
|---------|---------|------------|------------|
| 1247023 | 4523446 | 4456046633 | |
| 5133691 | 9071093 | 5646823446 | |
| 5708906 | 1246458 | 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....atc |
| f | ...t.. | ..a.... | .aa....atc |
| g | ...t.. | ..a....g | .aa....gatc |
| h | ...t.. | ..a.tg. | .aa....gatc |
| i | ...t.. | ..a....g | .aa....gatc |
| j | ...t.. | ..a.... | .aa....gatc |
| k | ...t.. | ..a....g | .aa....atc |
| l | ...t.. | ..a....g | .aag...gatc |
| m | ...t.. | a | 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....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 | 1111111111 |
| 1247023 | 4523446 | 5555666677 |
| 5133691 | 9071093 | 4456046633 |
| 5708906 | 1246458 | 5646823446 |
| 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....atc |
| f |t... | ..a....aa....atc |
| g |t... | ..a....aa....gatc |
| h |t... | ..a.tg..aa....gatc |
| i |t... | ..a..g..aa....gatc |
| j |t... | ..a....aa....gatc |
| k |t... | ..a..g..aa....atc |
| l |t... | ..a..g..aag..gatc |
| 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 |t... | ..ag.g..aa....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
- ▶ 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?